ANALYSIS OF GENES UNDERLYING MATE SELECTIVITY IN ARABIDOPSIS: REGULATION OF THE S-LOCUS RECEPTOR KINASE AND IDENTIFICATION OF NOVEL CANDIDATE RAPIDLY-EVOLVING REPRODUCTIVE GENES

by Susan Rebecca Strickler

ANALYSIS OF GENES UNDERLYING MATE SELECTIVITY IN *ARABIDOPSIS*: REGULATION OF THE *S*-LOCUS RECEPTOR KINASE AND IDENTIFICATION OF NOVEL CANDIDATE RAPIDLY EVOLVING REPRODUCTIVE GENES

A Dissertation
Presented to the Faculty of the Graduate School
of Cornell University
In Partial Fulfillment of the Requirements for the Degree of
Doctor of Philosophy

by
Susan Rebecca Strickler
August 2010

ANALYSIS OF GENES UNDERLYING MATE SELECTIVITY IN *ARABIDOPSIS*: REGULATION OF THE *S*-LOCUS RECEPTOR KINASE AND IDENTIFICATION OF NOVEL CANDIDATE RAPIDLY-EVOLVING REPRODUCTIVE GENES

Susan Rebecca Strickler, Ph. D.
Cornell University 2010

Mate selectivity in plants, as in other eukaryotes, is often based on the activity of highly polymorphic and rapidly evolving reproductive proteins. In flowering plants, selectivity in intraspecific and interspecific matings is observed not only in interactions between female and male gametes, but also in interactions between the male gametophyte with cells of the pistil that line the path of pollen tubes as they make their way from the stigma through the transmitting tract and into the ovary. In the following studies, the model plant *Arabidopsis thaliana* will provide a useful platform for analysis of intra and interspecific mate selectivity.

Intraspecific mate selectivity is often enforced by self-incompatibility (SI), a barrier to self-pollination that inhibits productive pollen-pistil interactions. Consistent with their role in recognition, some genes involved in SI are highly polymorphic and exhibit signals of positive selection. *A. thaliana* is self fertile, but can be made to express SI by transformation with two functional SI genes isolated from its close self-incompatible relative, *A. lyrata.* In Chapter 2, an induced mutation that suppresses the SI phenotype of these

transformants is analyzed to investigate the regulation of SI. Map-based cloning determined that the mutation disrupted NRPD1a, a plant-specific polymerase required for DNA methylation and production of some types of silencing RNAs. Subsequent analyses showed that NRPD1a, along with the RNA-dependent RNA polymerase RDR2, is required for SI in some *A. thaliana* accessions.

In Chapter 3, to assess interspecific mate selectivity, a screen was conducted to find rapidly evolving genes in *Arabidopsis spp*. A full genome sequence and the availability of extensive polymorphism data and genetic resources make *A. thaliana* ideal for identifying rapidly evolving reproductive genes. Previously generated *A. thaliana* data sets containing genes expressed in cells of the pistil, pollen, or pollen tubes were used to search for orthologous genes in the recently sequenced genome of *A. lyrata*. Within- and between-species variation in candidate rapidly evolving genes was investigated and statistical analyses of positive selection were performed. Several candidate genes were identified which may play an important role in species-specific reproductive function.

BIOGRAPHICAL SKETCH

Susan R. Strickler was born March 26, 1979 at Latrobe Area Hospital. She is the only child of Pamela G. Strickler and Roger D. Strickler. After graduating from Mt. Pleasant Area High School in 1997, she decided to pursue her life-long passion for horticulture at The Pennsylvania State University. While working on her Bachelor of Science degree, she conducted undergraduate research with Majid Foolad to identify tomato germplasm resistant to abiotic stresses for use in breeding. This work, along with the revolution occurring in genome sequencing at that time, sparked an interest in genetics, which would eventually lead her on to graduate school at Cornell University.

To my parents and all those who gave me the strength to persevere.

ACKNOWLEDGMENTS

I thank June Nasrallah for her help and guidance with my research and writing and also for enabling me to do these projects. I also thank my two other committee members, Charles Aquadro and Jeff Doyle. Chip allowed me to be a teaching assistant for his population genetics class, which was an enormous asset while working on many of the problems in Chapter 3. He also introduced me to my collaborator, Dr. Nathan Clark, and provided a number of helpful suggestions to improve my project. I appreciate that Jeff Doyle was able to join my committee and provide valuable input on short notice.

Dr. Clark was essential in the conduction of the high-throughput analysis of datasets in Chapter 3. I have enjoyed many engaging discussions with him concerning this project and his comments have been very constructive.

All of the members of the Nasrallah Lab, past and present, deserve my gratitude. In particular, I would like to acknowledge Krithika Yogeeswaran, Chih Wei Tung, and Sue Sherman-Broyles for their assistance and patience when I was starting in the lab. Titima Tantikanjana was critical in the development of Chapter 2 and Suganthi Kandasamy contributed to candidate gene analysis in Chapter 3. Tiffany Crispell assisted with experiments in both chapters and maintained plant material for research use. Additionally, I have enjoyed a wonderful friendship with Nathan Boggs and many fun and engaging discussions with Anne Rea.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

CHAPTER 1:

INTRODUCTION

Mankind's first basic understanding of flowering plant reproduction may have been the beginnings of one of the most influential discoveries since the dawn of humanity. As a result, seed could be saved from previous generations of plants, cultivation of crops in specific locations became possibly reducing the need for foraging, and eventually plants could be specifically bred to produce offspring of desirable phenotypes. Without this early knowledge of plant reproduction, agriculture and most likely civilization would not exist in its current state. It is therefore quite apparent that our interest in the reproduction of plants has had a major impact on our lives and will continue to be important in the future as population sizes increase and arable land diminishes. In the following, our current understanding of this important area of research as it applies to the processes of pollination, fertilization, and the plant structures involved, will be reviewed. Recognition and discrimination processes that occur both within and between species at the pollen/pistil level will be examined. Reproductive mechanisms in the Brassicaceae will be highlighted, as representatives from this family are among the best-characterized organisms for plant reproduction research. Additionally, since some genes involved in reproduction are thought to be rapidly evolving and have a potential role in speciation, the evolution of reproductive genes will be discussed.

I. Overview of Reproductive Structures in Angiosperms

There is great diversity among the floral structures of angiosperm species. Despite this variety, the basic arrangement of the various floral organs directly involved in reproduction and their general purpose in the fertilization processes tend to be repeated across flowering plant families. In the following section, common tissues involved in plant reproduction and the function of each constituent will be reviewed.

A. Pollen

Pollen is the male gametophytic tissue. It is supported nutritionally within the pollen sac by the tapetum and carried within the floral structure by the anther. Pollen is formed from four microspores that develop in the pollen sac, or microsporangia, during microsporogenesis. These microspores then develop into pollen grains during a process known as microgametogenesis. During microgametogenesis, an asymmetric division of the microspore results in two cells of differing size. The larger of these two cells, the tube cell, engulfs the smaller cell known as the generative cell. The generative cell subsequently divides again and the timing of this second division can have a major impact on the resulting pollen. In some pollen the generative cell divides immediately. This type of pollen is known as trinucleate and tends to have shortened viability. Binucleate pollen typically only divides after it has been released from the anther. This type of pollen tends to be more robust. In any case, the division of the generative cell results in the formation of two sperm cells, both of

which are used in fertilization of the female gametophyte[1].

Despite its seemingly quiescent state of existence, the pollen is transcriptionally active although less so than the sporophyte from which it originates. A recent microarray study found 992 genes in the *Arabidopsis* pollen transcriptome and almost 40% of these genes were expressed uniquely in the pollen grain[2]. Over-represented mRNAs included those coding for proteins involved in cell wall metabolism, cytoskeleton, and signaling. Under-represented mRNAs tended to be involved in transcription and protein synthesis.

The attributes of pollen vary widely across species. Size, shape, and amounts of pollen released into the environment by a plant can differ greatly. Some plants, such as lily, have extremely desiccated pollen[3] whereas others are more hydrated. Despite the great diversity found in pollen, the basic building blocks of the pollen grain are conserved across species. A pollen grain consists of the tube cell containing two haploid sperm cells, surrounded by a protective matrix of proteins, lipids, and other components. We will now look at each of these components in greater detail.

1. Pollen coat

The main purpose of the pollen coat is to protect the sperm cells from pathogen attack, extreme environmental conditions, to aid in transport of the pollen to a suitable stigma by wind, animal vectors, and other means, and to harbor molecules required for successful interaction with the stigmatic surface. The pollen coat, which is largely derived from the tapetum, is composed of lipids, proteins,

pigments, and aromatic compounds that fill in the cavities of the exine, a resistant interior layer[4]. A study designed to look specifically at proteins found in the *A. thaliana* pollen coat revealed sequences for two putative lipases and five oleosins known as glycine-rich proteins (GRPs)[5]. These lipases and oleosins accounted for more than 90% of the pollen coat proteins detected in this study. In addition, the lipids of the pollen coat may be important in hydration of the pollen when it reaches a suitable stigma[6, 7].

2. Exine

The exine is the multi-layered outer wall of the pollen and is made up of a very resistant substance known as sporopollenin. This layer of the pollen is derived from the tapetum and the microspores. Typically, the exine is elaborately sculptured in a way that is characteristic of the plant species from which it originated. Openings or apertures in the exine often exist that, in some species, may allow for hydration of the pollen grain once it lands on a suitable stigma and the emergence of the pollen tube from the interior of the pollen grain.

3. Intine

The intine is the innermost wall of the mature pollen grain and is synthesized by the actual pollen grain itself, rather than by sporophytic tissue. The intine may help to protect the pollen at the apertures of the exine, since the intine typically is thicker at the gaps of the exine[4]. The intine is normally composed of mostly cellulose and pectin that may help polarize and strengthen the pollen tube during

and after its emergence from the pollen grain. As the pollen tube emerges from the pollen grain, the intine becomes continuous with the pollen tube cell wall and may be involved in pistil tissue invasion and self-incompatibility reactions in some plant species[4].

4. Pollen Tube

After hydration of the pollen grain at the stigma surface, the pollen tube emerges from the outer pollen layers through apertures in the exine, although this is not always the case. The sperm cells are then carried within the pollen tube through the female tissues of the pistil until they reach the ovary.

B. Pistil

The pistil houses the female gametophytic tissue. It is made up of several carpels that usually become fused during floral development. The interior of the pistil may be completely filled in by the transmitting tract tissue or, in some instances, it may remain hollow with only a thin lining of transmitting tract tissue. The pistil is made up of several different tissues including the stigma, the style, and the ovary, each of which plays an essential role in pollination and fertilization.

1. Stigma

The stigma is found at the top of the pistil and is responsible for the reception of pollen grains. In *Brassica*, it is covered by a proteinous layer known as the pellicle, which is found on top of the

cuticle. Some stigmas, such as those found in tobacco and lily, are wet due to the presence of exudates produced by the cells of the stigma. In other species, such as members of the Brassicaceae, the stigma is dry. In these species the stigma epidermis consists of many large papillar cells that catch and interact with the pollen directly. Species with dry stigmas rely on a cycle of events that must occur before fluids are released from the stigma and the pollen becomes hydrated.

2. Style/Transmitting Tract

The transmitting tract is found at the interior of the pistil within a structure known as the style, which connects the stigma to the ovary, and in the ovary itself. The transmitting tract is not an immediately evident structure, but can be identified as the tissues in which the pollen tubes travel within the style and ovary. Transmitting tract cells are metabolically active secretory cells that provide the appropriate mechanical and nutritive environment for normal pollen tube growth. Some species, such as *Arabidopsis thaliana* have a solid style that is filled with the tissues of the transmitting tract. Species of *Lilium* have a hollow style that has a thin lining of transmitting tract tissue through which the pollen tubes travel.

3. Ovary

The ovary is the expanded area found at the base of the pistil. It houses the ovules, which contain the female gametophytic tissue called the embryo sac. The development of the embryo sac is a rather

curious phenomenon. The megaspore mother cell is the progenitor of the embryo sac and undergoes meiosis to form four haploid megaspores. For unknown reasons, three of these megaspores dissolve, leaving only one megaspore to develop into the embryo sac. The nucleus of the megaspore divides three times to form one large cell with eight haploid nuclei. Cytokinesis results in an embryo sac containing seven cells: the egg cell flanked by two synergid cells, a central cell containing the two polar nuclei, and three antipodal cells. The egg cell and synergids are located near the micropyle, which is an opening in the covering of the ovule known as the integument. The antipodal cells are located at the opposite end from the egg cell and synergids. When a growing pollen tube reaches the ovary, it travels down a structure known as the funiculus. This structure is basically a stalk that connects the ovule to the tissues of the ovary. The pollen tube then moves into the embryo sac through the micropyle. Upon arrival at the embryo sac, the pollen tube releases two haploid sperm cells that are employed in a unique process known as double fertilization. One of these cells fuses with the egg cell resulting in the creation of the embryo, while the other fuses with the two polar nuclei to form the endosperm tissue. The endosperm is a specialized storage structure that supports the growing embryo by providing nutrients. After fertilization, the outer covering of the ovule, known as the integument, will harden to form the seed coat, which protects the developing embryo from unfavorable environmental conditions. Another protective layer, the fruit, develops around the seed coat from tissues of the ovary.

II. Recognition Processes Occurring within a Species

A. Compatible Reproductive Interactions

When pollination and fertilization occur successfully and a viable embryo is produced, a compatible interaction has occurred. Usually, compatible interactions occur between members of the same species although viable offspring can be produced as the result of interspecific hybridization in plants. The initial site of contact between male and female tissues is the stigma surface, where the pollen grain tightly adheres in a compatible pollination. Next, as the pollen tube passes through the stigma epidermal cell wall and starts its journey through the interior of the pistil, it encounters cells of another important tissue in plant reproduction, the transmitting tract. The last tissue believed to play a role in successful fertilization is the female gametophytic tissue, which is involved in pollen tube attraction and guidance. Each of these important steps in compatible reproduction will now be discussed in greater detail with particular emphasis on specific interactions that take place between male and female participants.

1. Stigma and Pollen

The stigma epidermal cells are responsible for initial reception of the pollen grain. The stigma must be at the proper developmental stage for the pollen to successfully adhere and penetrate the stigma surface. In some wet stigma species, a layer of cuticle exists on the surface that prohibits pollen germination until the stigma is mature[3]. As the stigma reaches maturity, various factors, such as insects,

enzymes, and turgidity of the stigma may assist in the breakdown of the cuticle. Eventually the stigma deteriorates and while it may in some instances still support pollen grain hydration and germination, it is not longer conducive to pollen tube growth[8].

When a pollen grain lands on a receptive stigma a series of events follows. The speed at which these events occur depends on the ability of the pollen grains to adhere and hydrate. As described previously, the succulence of the stigma surface varies between species and this trait is an important factor that can affect pollen adherence and hydration. In species with wet stigmas, such as lily, most pollen that lands on the stigma is able to stick to the stigma surface and hydrate, but in species such as *Brassica* that have a dry stigma, a compatible interaction must occur between the stigma and pollen grain in order for pollen adhesion and hydration to occur. Pollen tends to be extremely desiccated before it is sent loose into the environment, so the degree of dehydration can influence the time it takes for it to become properly rehydrated. In the Liliaceae family, pollen is very dehydrated and dormant when released, so germination can take up to one hour after the pollen reaches a suitable stigma[3]. The pollen coat and exine also play an important role in adhesion and hydration, especially in species with dry stigmas.

Wet and dry stigmas may still have some parallel processes in adherence and hydration, despite their obvious differences (to be described in greater detail below). Aquaporins are proteins found to be involved in water transport across biological membranes[9]. Aquaporin expression has been observed in the pistil of wet-stigma Solanaceae

species [10, 11] and dry-stigma Brassicaceae species [12, 13] which leads to the enticing possibility that they play a role in transporting water for pollen hydration. Since only one aquaporin, MIP-MOD of *Brassica*, has been found to be expressed in the stigma epidermal cells to date[13]these proteins must use a mechanism that allows indirect contact with the pollen grain if these genes are indeed involved in pollen hydration. Alternatively, the aquaporins could be important in supporting pollen tube elongation after it exits the stigma epidermis. Unfortunately, this is a large gene family making genetic analysis difficult. In addition, when a class of PIP2 aquaporins in tobacco was knocked-down using RNA interference, no effect on pollen hydration and tube growth was observed[14].

Various interacting proteins may be necessary to assist in hydration and germination of the pollen grain, regardless of the succulence of the stigma surface. One example in a wet stigma species is LAT52 and LePRK2. LAT52 interacts with a pollen coat protein known as *Lycopersicon esculentum* Protein-specific Receptor-like Kinase (LePRK2)[15] and this communication may be important in pollen tube growth. This information is supported by antisense experiments that knockdown *LAT52* in tomato. LAT52 is a protein found on the male side and is believed to be involved in pollen hydration and germination. In antisense–*LAT52* transgenic plants, pollen is unable to germinate *in vitro*, although *in vivo* experiments showed a less severe phenotype suggesting factors on the stigma may compensate for the lack of *LAT52*[16]. *LePRK2* is a member of a gene family that contains other pollen-specific genes that may also be

involved in reproduction[17]. Homologs of the *LePRKs* have been found in maize and in *Arabidopsis*[17]. The finding of *LePRK* homologs in *Arabidopsis* is of particular interest, since it may suggest that this gene also has a role in fertilization in dry-stigma species. A pistil factor has also been found to interact with LePRK2. LeSTIG1 was found to displace LAT52 during *in vitro* binding and also to stimulate the growth of cultured pollen tubes [18]. A model has been proposed in which LePRK2 dissociates from LAT52 on the stigma surface and binds LeSTIG1. LeSTIG1 may then help direct the pollen tube into the style[18].

Lipids are also believed to play a role in pollen hydration, but their location may vary depending on the type of stigma involved. In wet-stigma species, lipids important in pollination are believed to be found mainly in the stigma exudates, whereas in dry-stigma species such lipids reside in the pollen coat[19].

In *Arabidopsis*, mutations that affect lipid synthesis and pollen coat formation often cause a defect in pollen hydration[7, 20]. Mutations in *Eceriferum* (*CER*) genes can result in pollen that is unable to hydrate on the stigma surface, but is able to hydrate *in vitro*. These genes are thought to encode lipids important in some biophysical property of the pollen coat and may play a role in pollen-stigma recognition events that are necessary for the release of water from the stigma[20]. One of several oleosins found in the pollen coat of *A. thaliana*, GRP17, is also thought to interact with the dry stigma in a manner that allows for the hydration and germination of the pollen grain[21]. On a similar note, mutations that eliminate the secretions found on the surface of a wet

stigma cause female sterility, which can be alleviated by the application of exogenous lipids[6, 22]. Collectively, these results show that lipids play a role in permeability of the stigma cuticle to water in both wet- and dry-stigma species[23, 24]. Lipids also may be important in preventing pollen tube growth into tissues of the plant unrelated to reproductive processes. In *fiddlehead* mutants, which have a defect in lipid synthesis, the epidermis becomes more permeable and pollen is able to hydrate and germinate on non-reproductive surfaces[24]. In addition, exogenous lipid application allows pollen to germinate on *Arabidopsis* leaves[6]. These experiments suggest that lipids mediate the flow of water through membrane surfaces. Under normal circumstances, impermeability is retained to prevent dehydration. When a compatible interactant lands on these surfaces, perhaps a change in the lipid constituents somehow allows the passage of water. This suggests a need for interactions between male and female molecules involved in hydration which may in part prevent pollen from hydrating through the apertures when it lands on a wet surface other than the stigma.

After the pollen grain is properly hydrated, the pollen tube proceeds to pass through the outer layers of the pollen grain and penetrate the stigma surface by growing through the cell walls of the papillar cells. This process acts to anchor the pollen grain onto the stigma surface. Water and lipids may play an important role in this process by polarizing the pollen grain and guiding the tube to the stigma surface[6, 25]. Proof comes from pollen-coat deficient mutants that are able to hydrate, but unable to locate the stigma[19]. The pollen

may exit the pollen grain either through exine apertures or by breaking directly through the exine[4]. Enzymes and turgor force may further assist the pollen tube in leaving the barriers of the exine[4]. Cell wall-degrading enzymes are also necessary in allowing the pollen tube to penetrate the stigma surface in species with closed styles. These include cutinase[26], polygalacturonase[27], pectin esterase[28], glucanase[29, 30], and endoxylanase[31].

In the following, our current understanding of interactions found specifically on wet or dry stigma surfaces will be discussed in greater depth.

a. Wet stigma surfaces

Many cells are found on the surface of a wet stigma that release exudates containing proteins, lipids, polysaccharides, and pigments[4]. These exudates help to capture and anchor incoming pollen grains. Since wet stigmas have the ability to hydrate most pollen grains that land on their surfaces and the extracellular matrices (ECM) cover the stigma surface, prevention of pollen tube growth usually occurs at the style or some point after the pollen tube attempts to penetrate the stigma surface[32]. This is especially apparent in the hollow style of lily where no penetration of stigma tissues is necessary for the pollen tube to gain entry to the style.

It has been shown in tobacco that lipids in the ECM provide a gradient of water that may play a role in directing the growing pollen tube into the tissues of the stigmas[6]. Other species may rely on a chemotropic chemical for pollen tube guidance. In lily, stigma/stylar

cysteine-rich adhesin (SCA) protein[33] and chemocyanin[34] have been found to attract pollen tubes *in vitro*. SCA shows pH-dependent binding and can exist in a bound form, in which it is bound to pectin, and an unbound form. It has been suggested that a gradient of SCA may play a role in directing pollen tube growth[35]. A central canal is found on the stigma of lily that leads to the transmitting tract located at the interior of the hollow style. SCA may be important in guiding the pollen tube along the stigma surface and into the opening of the central canal.

b. Dry stigma surfaces

Since only pollen from the same species is usually able to adhere and germinate on a dry stigma surface, this type of stigma epidermis provides a good mechanism to avoid incompatible pollinations. In order to maintain this discriminating surface, dry stigmas must rely on a series of specific interactions between the outer surfaces of the pollen and the stigma epidermis for proper adhesion and hydration of the pollen grain. Adhesion of the pollen grain usually occurs in three main steps in these species. First, a quick initial interaction occurs between the pollen exine and stigma papillar cells to catch the pollen grain. The next step in adhesion involves interactions between pollen coat proteins and stigma-expressed proteins. Lastly, as the pollen tube starts to penetrate the stigma surface, the pollen grain is firmly tethered onto the stigma surface.

A study by Zinkl *et al.* showed that the exine may be important in the rapid initial stages of pollen adhesion in *A. thaliana*, a dry-

stigma species[36]. In this study, *cer6-2,* a mutant lacking a pollen coat, and pollen treated with a pollen coat-removing chemical were compared in their ability to bond to the stigma. Both types of pollen, as well as purified exine, were able to adhere to the stigma surface, suggesting the pollen coat is not necessary in this preliminary interaction. Undulations in the cell walls of the stigma were found within five minutes after pollination, suggesting that the stigma also changes quickly in response to the pollen contact, possibly to provide a greater surface for pollen adhesion. It has also been suggested that in *Brassica*, the self-incompatibility system has no influence over this preliminary stage of adhesion[37].

After the pollen is firmly anchored to the stigma surface, hydration of the grain is initiated. Water is believed to travel from the stigma through the pollen foot to hydrate the pollen[38, 39]. It is known that the absence of the pollen coat interferes with proper pollen hydration and germination[7]. As described previously, the pollen coat contains many lipids that are important in hydrating the pollen when it reaches the dry stigma surface[6, 7]. In dry-stigma plants, the stigma cells may have a special permeability that allows water to travel through to reach the pollen grain at the correct time.

In the later events of pollen germination and growth, it seems that factors specific to the stigma epidermis are essential. Along this line, *Pis63* is a gene known to be expressed abundantly in the stigma epidermal cells[40]. Transgenic *Brassica napus* plants in which the level of *Pis63* transcripts were reduced had decreased pollen germination on the stigma surface but not adhesion[41]. It seems that this gene plays a

role in events that occur after adhesion of the pollen grain, but the details of this task are not yet known[41]. It is also known that the papillar cells themselves play an essential role in pollen tube growth in *Brassica*[42], although strangely, ablation of the papillar cells in *A. thaliana* has little effect on wild type pollen tube growth[43]. This suggests that there is some fundamental difference in the recognition systems of these two members of the Brassicaceae, perhaps something related to the loss of another recognition system in *A. thaliana*: the self-incompatibility system.

2. Transmitting Tract and Pollen Tube

The transmitting tract is found in the interior of the style and is important in promoting and guiding pollen tubes towards the ovules[44-47]. As a demonstration of this fact, pollen tubes grown *in vitro* never reach the lengths of those grown *in vivo*[35]. It has also been shown that a depletion in stylar reserves occurs during pollen tube growth[48]. This may be due to the uptake of molecules present in the pistil by the growing pollen tube. These molecules, which may include sugars, ions, and proteins, are probably used in supporting the nutritional needs of the pollen tube and directing it toward the ovary.

Various factors seem to be involved in directing the pollen tube to the ovary. It has been suggested that some molecules, such as calcium and □-aminobutyric acid (GABA), may form a chemotropic gradient that helps guide the pollen tube to the proper place. In another experiment involving minute latex beads, results suggested that a constituent of the style may physically direct movement of

pollen tubes through a predefined track. The inert particles were able to travel through the transmitting tract of *Hemerocallis flava*, *Raphanus raphanistrum*, and *Vicia faba* at the same rate as pollen tubes[49]. It is not known whether pollen tubes move through a predefined track or if chemotropic gradients are the determining force in pollen tube travel. Perhaps both of these mechanisms work together in pollen tube guidance or pollen tubes may be guided by different methods in different species.

The configuration of the transmitting tract tissue in the style can vary by species. Species in *Lilium* have a hollow style and the transmitting tract lines the exterior of this cavity. Many species, such as *Nicotiana* and *Brassica spp.* have a solid style and the transmitting tract tissue completely fills the inside of the pistil. The details of pollen tube growth and guidance in each of these style types follow below.

a. Hollow Styles

In hollow-styled species, the growing pollen tubes must adhere to the transmitting tract tissue lining the inner cavity of the pistil. In lily, SCA and pectin form a matrix along the transmitting tract that is thought to assist in the adherence of pollen tubes in the hollow stigma[33, 49]. Recently, it has been found that SCA can induce pollen tube adhesion in the absence of the stylar pectin suggesting that SCA is interacting with a component of the pollen tube surface[50]. It is not clear how SCA guides pollen tubes through the style, but it is unlikely that a chemotropic gradient is involved since pollen will grow towards

both the stigma and ovary when placed on an open style[51]. Since many pollen tubes typically grow through the lily style, pollen tubes that germinate later are forced to adhere to earlier germinating pollen tubes that cover the transmitting tract matrix. SCA also seems to play a role in the binding of pollen tubes to each other when the transmitting tract tissues cannot be directly contacted[50].

b. Solid Styles

In solid-styled species the pollen tube gains access to the transmitting tract by penetrating through the stigma tissue. Upon arrival at the interior tissues of the style, various factors may be necessary to guide the pollen tube through the transmitting tract and into the ovary. The pollen-expressed LAT52 and LePRK2, and the stigma factor, LeSTIG1, may be important for this guidance in tomato. In plants expressing an antisense-LAT52 construct, the pollen tube is unable to travel normally through the transmitting tract and arrests in the style, although these results were less severe than *in vitro* studies, suggesting that a molecule in the style may compensate for LAT52[16]. LeSTIG1 is expressed in the stigma and style[18]. Exogenous applications of LeSTIG1 abolishes the interaction of LAT52 and LePRK2 in pollen and may interact with the pollen tube in a way that helps to direct its growth through the transmitting tract tissue[18]. It is possible that once the pollen tube reaches the ovary, another binding partner replaces LeSTIG1, since this protein does not seem to be present in ovary tissues.

In *Arabidopsis, VANGUARD1* (*VDG1*) is a pollen and pollen tube-

expressed gene that enhances pollen tube growth in the transmitting tract[52]. A mutation in this gene resulted in greatly retarded pollen tube growth and tubes that were prone to bursting. Since *vdg1* mutants were able to grow normally on the stigma surface, it was suggested that VDG1 may be important in interactions that occur between the pollen tube and transmitting tract tissue.

Various chemotropic chemicals may be present in the transmitting tract to direct the pollen tube towards the ovary. A gradient of TTS (transmitting tract specific glycoprotein) glycosylation occurs in the style with higher levels of carbohydrates found near the ovary and may be important in directing the pollen tube to the ovary[53]. TTS proteins are known to be deglycosylated during pollen tube growth and the resulting sugars may be incorporated into the pollen tube wall[35]. In tobacco and *Nicotiana alata*, TTS stimulates pollen tube growth *in vitro*[54, 55]. Transgenic plants with reduced expression of TTS have a reduced rate of pollen tube growth. An experiment to find the localization of SCA used labeling with quantum dots to show that SCA localizes at the pollen tube tip and is taken up by the tube cell[56]. It has been proposed that when SCA binds its putative receptor, it may activate an intracellular signal transduction cascade in the pollen tube that is important for pollen tube binding[57]. It may then be transported into the pollen tube by endocytosis[57]. Calcium also plays a role in directing pollen tube growth and has been found to exist in a chemotropic gradient in *Antirrhinum*[58]. In *Arabidopsis*, a chemotropic gradient of □-aminobutyric acid (GABA) was found in the pistil[59]. Low levels of GABA are found on the stigma and the concentration

increases throughout the style, with the highest concentration found in the ovary. Other molecules are most likely involved in guiding the pollen tube since pollen tubes do not respond to an in vitro GABA gradient[14].

3. Ovule and Pollen Tube

As described previously, angiosperms undergo a unique reproductive process known as double fertilization. To recapitulate, major players in fertilization on the pollen side include the tube cell and the two sperm cells that it contains. The embryo sac is the important gametophytic tissue on the female side and includes the egg cell surrounded by two synergid cells, two polar nuclei, and three antipodal cells. One sperm cell fuses with the egg cell to form the zygote and another sperm cell fuses with the two polar nuclei to result in the triploid endosperm. As discussed in the following, the female gametophytic tissue is believed to be involved in signaling and guiding the pollen tube to the ovule.

An experiment using an *A. thaliana* strain that was heterozygous for a chromosomal translocation resulted in the arrest of approximately half of the ovules[60]. In this experiment, ovules with a defective embryo sac were unable to attract pollen tubes. Another experiment in *A. thaliana* identified *magatama* (*maa*) mutants, which have delayed development of the female gametophyte. These mutants are unable to direct pollen tubes into the micropyle[61]. The tubes grow normally within the style and ovary, but upon arrival at the micropyle they grow in random directions. The specific gametophytic cells

involved in this loss of pollen tube orientation are not known, but synergid cells have been identified as important signaling factors in *Torenia*. In this study, laser ablation was used to destroy all cells of the embryo sac[62]. Only when the synergids were destroyed was a defect in pollen tube guidance observed. It is now known that defensin-like polypeptides named LUREs are important pollen tube-guidance molecules expressed mainly in the synergid cells of *Torenia*[63]. A receptor-like kinase, *FERONIA*, is also expressed in the synergid membrane and is important in pollen tube reception. In *feronia* mutants, the pollen tube fails to rupture and properly deliver the sperm cells to the female gametophyte[64]. In maize, *ZmEA1* (*Zea mays EGG APPARATUS1*) was found to be expressed in the ovule and synergid cells[63]. Down-regulation of this gene resulted in loss of pollen tube guidance into the micropyle. Further work is needed to determine whether the egg cells are specifically involved in this response. The central cell of *A. thaliana* also may play a role in pollen tube attraction. In central cell guidance (*ccg*) mutants, pollen tubes have difficulty finding their way into the micropyle[64].

In addition to chemotropic compounds, many species have small bump-like structures that face towards the ovule. These structures, called obturators, are believed to physically direct the pollen tube into the ovule entrance. In some species, such as peach[65] and kiwi[66], the obturators enter a secretory phase that assists in the growth of the pollen tube along the obturator surface.

Pollen tubes must also have some method of interacting with tissues in the female reproductive tract, so that they can effectively

receive guidance and respond to guidance cues. Two *FERONIA* homologs, *ANXUR1* and *ANXUR2*, have been found necessary for pollen tube growth. Both genes encode receptor-like serine threonine kinases that are targeted to the plasma membrane at the pollen tube tip and are thought to promote and protect tip growth until it reaches the female gametophyte[67]. The sperm cells also appear to play a role in targeting the tube to its final destination. *HAP2* encodes a membrane-bound protein found in the haploid sperm cells. Mutants of this gene have inefficient targeting to the ovules and are often unable to complete fertilization[68].

B. Incompatible Reproductive Interactions

Approximately 75% of angiosperm species give rise to hermaphroditic floral structures, meaning both male and female gametophytes are located in the same flower[69]. Considering this fact, it seems that the majority of flowering plant species would take advantage of this close proximity and rely on self-pollination and inbreeding. However, many angiosperms are out-crossing to some degree and some species are obligate out-crossers. How can this be so? Plants have developed a number of ways to enforce out-crossing, including floral polymorphisms that hinder selfing and genetic barriers to pollination such as self-incompatibility. As a general rule, inbreeding can lead to an accumulation of recessive deleterious mutations and loss of genetic diversity. Therefore, it is in the best interest of some species to maintain out-crossing so as to increase genetic diversity and allow the creation of new gene combinations. In

the following, various methods by which plants enforce out-crossing will be discussed in greater detail.

1. Floral Polymorphisms

Various floral polymorphisms can enforce outcrossing. For example, spatial separation of the pollen and stigma, known as herkogamy, can reduce self-pollination. Temporal separation, known as dichogamy, in which one of the plant gametes matures before the other gamete is prepared for fertilization, results in plants that must outcross. This occurs often in monoecious species that have both male and female reproductive structures on the same individual. Another fairly common floral polymorphism that enforces outcrossing is dioecy, where the male and female gametes reside on entirely different plants. In many plant species that exhibit floral polymorphism, selfing can still occur. For example, in *Narcissus longispathus*, a herkogamic species, no correlation between herkogamy and outcrossing was found, possibly due to the presence of pollinators that deposit the pollen to the stigma surface[70]. Since many factors such as animal vectors, wind, rain, and insects can overcome structural barriers to pollination, using floral polymorphisms alone may not be the most effective means of avoiding outcrossing.

2. Genetic Barriers to self-pollination

Stricter enforcement of outcrossing is obtained in plant species that have genetic barriers to self-pollination, known as self-incompatibility (SI). There are two classes of SI, heteromorphic and

homomorphic. Hetermorphic self-incompatible plants have structural differences within the floral organ that are associated with SI genotype. Pollination and fertilization are only successful in crosses between plants having different floral morphologies, but not between plants having the same floral structure. This mechanism differs from that of floral polymorphism pollination barriers because in the former, differences are typically controlled by a single genetic locus. A classic example of a plant exhibiting heteromorphic self-incompatibility is *Primula*. *Primula* exhibits a condition known as distyly, where there are two different flower types [or floral morphs]. One morph, known as pin, has a long style and short stamens, while the other morph, known as thrum, has a short style and long stamens. Only pollinations between pin and thrum flowers are compatible. This heteromorphic SI system is genetically controlled by a locus consisting of seven linked genes [71].

Homomorophic SI is not associated with floral polymorphism. Flowers of incompatible plants are similar, but their genotype in a region of the genome known as the *S* locus differs. The *S* locus is a highly polymorphic region of closely linked genes involved in the SI response. Since these genes are closely linked and recombination is minimal, the combination of alleles found in this region is referred to as a haplotype. Classically, homomorphic self-incompatible plants were grouped into two classes. One class exhibits gametophytic self-incompatibility (GSI), in which the success of fertilization between two plants is dependent upon the haploid genotype of the sperm cells carried within the pollen grain. Another class exhibits sporophytic

self-incompatibility (SSI), in which fertilization can occur only if the diploid genotype of the anther from which the pollen grain is derived is dissimilar from the genotype of the female parent. In either case, when the pollen source plant and the plant housing the stigma share matching alleles, pollination is inhibited, either at the stigma surface (as in the Brassicaceae and grasses) or within the style (as in the Solanaceae, Rosaceae, and Plantaginaceae).

a. Gametophytic Self-incompatibility (GSI)

GSI is the most prevalent form of genetic SI. The Solanaceae and the Papaveraceae are the two main model systems for GSI, but the specific mechanisms used to maintain GSI in these two taxa differ. In the GSI system of the Solanaceae, S-RNases in the style interact with factors in the pollen tube to result in degradation of the growing pollen tube (see McClure 2009 for a review)[72]. The Papaveraceae system involves an interaction between a receptor found in the pollen tube plasma membrane and its glycoprotein ligand found on the stigma[73].

In the Solanaceae as well as the Rosaceae and Plantaginaceae, pollen tube rejection occurs in the upper portion of the style, after the pollen tube has already penetrated the stigma surface. The two genes known to be involved in the specificity of this reaction are the *S-RNase* gene, which is the female determinant of SI, and the *SLF* (*S-locus F-box*) or *SFB* (*S-haplotype-specific F-box*), which is the male determinant of SI. The *S-RNase* gene was first identified in *Nicotiana alata*[74] and further gain- and loss-of-function experiments revealed that this gene was indeed responsible for the *S*-haplotype specificity of the pistil[75, 76].

Pollen rejection results from the activity of the S-RNase, which degrades the cytoplasmic RNA of incompatible pollen tubes[77],[78]. The male determinant of GSI has been identified in *Nicotiana, Petunia,* and *Prunus*, a member of the Rosaceae, as an F-box gene[79]. *SLF* is expressed exclusively in the anther and pollen grains. Proof that *SLF* was indeed the male determinant in GSI was obtained by transformation of *Petunia inflata* with a particular *SLF* allele, and then relying on competitive interaction to breakdown SI in plants pollinated with pollen carrying two different *SLF* alleles[80]. In addition to *S-RNase* and *SLF*, other loci appear to play a role in the GSI response. One such locus is *HT-B:* plants in which this gene was suppressed by the inhibitory RNA (RNAi) approach were found to show a significant decrease in pollen rejection when pollinated with plants that should generate the SI response[81]. In addition to *N. alata*, homologs of HT-B have been found in two other genera of the Solanaceae, *Lycopersicon*, and *Solanum,* suggesting that this gene may have a conserved role in this family[81-83].

b. Sporophytic Self-incompatibility (SSI)

Although less common than GSI, SSI has been identified in the Brassicaceae, Convolvulaceae, Asteraceae, Geraniaceae, and the Caprifoliaceae. In SSI, pollen tubes are usually unable to hydrate, germinate, and grow through the stigma surface in an incompatible interaction.

In the Brassicaceae, which has been the model for analysis of SSI, SI is controlled by alleles of the *S* locus, and self-pollination is

prevented when pollen and stigma are derived from plants that express the same *S*-locus variant. Two *S*-locus genes are known to be essential for specificity in the SI response of the Brassicaceae: the *S*-locus Receptor Kinase (*SRK*) and the S-locus Cysteine-Rich protein (*SCR*) genes. SRK is a receptor kinase expressed in the stigma epidermis and its SCR ligand is found in the pollen coat. An SCR variant can bind and activate only the SRK variant encoded in the same *S*-locus haplotype. This "self" interaction is thought to trigger a poorly understood signal transduction cascade within the epidermal cell, which prevents pollen hydration and penetration of the pollen tube into the stigma epidermal cell wall.

The first *S*-locus gene to be identified using biochemical and molecular techniques was the *S*-locus glycoprotein (SLG) of *Brassica*[84]. The discovery of SLG was a key step in the eventual identification of the true female determinant of SI, SRK[85]. SLG and the extracellular domain of SRK share extensive sequence similarity, and this similarity was critical for the identification of SRK. Unlike SLG, SRK has a transmembrane domain and an intracellular serine/threonine kinase domain. The extracellular domain of the SRK receptor contains several highly polymorphic sites with an average sequence similarity among alleles of 78.2% within *B. oleracea*[85, 86]. *SRK* is expressed in the papillar epidermal cells, with maximal expression observed just before anthesis of the flower. Experiments with transgenic *B. rapa* showed that *SRK* was indeed the stigma determinant of SI specificity, while *SLG* enhanced, but was not necessary for, SI specificity[87]. Further proof was obtained by transforming self-compatible *B. napus*

with *SRK*, which resulted in plants that showed the SI response when pollinated with "self" pollen[88]. The role of *SLG* in the SI response is unclear, since some lines of SI *Brassica* and also *A. lyrata*, another self-incompatible member of the Brassicaceae, do not have an *SLG* gene[89-91].

SCR was first identified by the cloning of the *S*-locus region, which allowed for the identification *SCR* as a sequence unusually rich in cysteine residues[92]. When *B. oleracea* was transformed with a specific allele of *SCR*, the pollen of transformants acquired the specificity of the transgenic *SCR* allele, demonstrating that *SCR* is the male determinant in SSI[92]. *SCR* is known to be highly polymorphic and *SCR* variants can exhibit as little as 30% sequence similarity within species [84, 93], [100]. SCR is a small globular protein that is highly polymorphic except for eight conserved cysteine residues that maintain its overall structure[94].

Conclusive evidence for both the stigma and the pollen determinants of SI was obtained by transforming *A. thaliana* ecotype Col with *SRKb* and *SCRb* from *A. lyrata*[95]. Transformants of the normally selfing species were able to elicit the SI response when pollinated with "self" pollen. These results suggest that *A. thaliana* may have remnants of a once functional SI system. Additionally, a recent study found that some accessions of *A. thaliana* may still have a functional *SRK* and the *SCR* is recoverable[96].

Several genes unlinked to the *S* locus have been proposed to be required for SI, either as components of the downstream signal transduction cascade that occurs after SRK activation or as regulators

of *SRK* expression or SRK activity. One such gene was identified by analysis of a self-compatible mutant strain of *B. rapa* carrying a recessive mutation at the *M* or *MOD* locus. This mutation was originally thought to disrupt an aquaporin-like gene expressed in the stigma epidermis[13]. However, a later study reported that a nearby gene, *MLPK*, was actually responsible for the self-compatible phenotype[97]. This conclusion was based on the finding that transient expression of the *MLPK* allele isolated from a self-incompatible strain in self-compatible mutants could restore the SI response. To date, however, there have been no reports of stable transformation experiments to confirm this finding. In any case, it is hypothesized that MLPK, which is tethered to the plasma membrane via an N-terminal myristoylation motif[98], is recruited when SRK becomes activated by incompatible pollen.

Another candidate effector of the SI response is the armadillo repeat-containing 1 protein (ARC1), which was identified in a yeast 2-hybrid screen that used the cytoplasmic kinase domain of SRK as bait[99, 100]. In transgenic plants expressing an anti-sense *ARC1* construct, a partial breakdown of SI was observed[101]. ARC1 has ubquitin-ligase activity and is thought to target substrates for ubiquitination[102]. Recently, a candidate target substrate for ubiquitination by ARC1 has been suggested. The protein, Exo70A1, is necessary for pollen tube growth in compatible pollinations[103]. Loss of this gene in the stigmas of both Brassica and A. thaliana prevents compatible pollen growth. Overexpression of *Exo70A1* in *Brassica* results in a partial breakdown of the SI response. Exo70A1 is

predicted to be part of the exocyst complex and involved in vesicle docking at the plasma membrane. It is possible that it is involved in the transport of resources to the stigma surface that are necessary for pollen growth. Ubiquitination and subsequent degradation of this protein may deny pollen the necessary criteria for growth on the dry stigma surface found in the Brassicaceae. Regulation of Exo70A1 though ubiquitination by ARC1, seems a plausible method for controlling SI.

Two other proteins, the thioredoxin h-like proteins THL1 and THL2, were also identified as proteins that interact with the kinase domain of SRK in the yeast 2-hybrid system[104]. It has been shown *in vitro* that autophosphorylation of SRK is prevented by THL1 in the absence of an incompatible SCR[105]. In addition, antisense suppression of thioredoxin in self-compatible *B. napus* led to a low level pollen rejection[106]. THL1 and THL2 are suggested to maintain SRK in an inactive state in the absence of an incompatible SCR.

Finally, studies of cryptic variation among *A. thaliana* accessions for their ability to express a robust and stable SI response upon transformation with *A. lyrata SRK* and *SCR* gene pairs led to the identification of an *S*-linked modifier of SI, *PLANT U-BOX8 PUB8*[107]. PUB8 is involved in regulating *SRK* expression and a hypomorphic *PUB8* allele is suggested to be responsible for the transient SI or pseudo-self-compatibility phenotype of certain *A. thaliana* accessions.

Epigenetics is now finding a role in the SI response of the Brassicaceae. In interspecific hybrids between *A. lyrata* and *A. thaliana*, SI is found to breakdown in the F1 generation, most likely

due to aberrant processing of the full-length *SRK* transcript[108]. Hybrids between *C. rubella* and *C. grandiflora* show a breakdown of SI probably due to silencing of *SCR* in the pollen[108]. In addition, a component of the RNA silencing pathway, RNA-dependent RNA polymerase (*RDR6*), has been recently implicated in SI. *A. thaliana SRK-SCR* transformants of the Col-0 accession exhibit transient SI, but in the presence of the *rdr6* mutation, these plants exhibit an enhanced SI response, as well as stigma exsertion[109]. It is possible that silencing may have played a role in the loss of SI from *A. thaliana*. In support of this hypothesis, *SRK,* the nonfunctional ortholog of *SRK* from *A. lyrata*, has a plethora of small RNAs generated from its 3' end[110]. It will be interesting to see if silencing plays a role in the regulation of SI in a natively SI species such as *A. lyrata*.

III. Recognition Processes Occurring between Species

A. Mechanisms of Discrimination and Isolation between Species

In order to remain reproductively isolated and genetically distinct, a species must be able to discriminate between pollen from a member of its species group and pollen from an "outsider". There are several methods by which this discrimination can occur.

1. Pre-zygotic Isolation

Isolation of species often results from pre-fertilization barriers that prevent productive egg-sperm interactions from post-zygotic aberrations in embryo development. In angiosperms, a major barrier to interspecific hybridization operates pre-zygotically during the

interactions between the pollen and the pistil. Interestingly, there are similarities between the inhibition of intraspecific pollen in SI and the arrest of pollen in interspecific pollinations, especially in the site within the pistil where inhibition occurs. Thus, in Solanaceae, interspecific pollen is generally arrested within the transmitting tract of the style. In *Brassicaceae,* incompatible interspecific pollinations result in the failure of pollen to hydrate and germinate at the surface of the dry stigma[20]. However, some Brassicaceae species have more permissive stigmas. For example, the *A. thaliana* stigma allows the hydration and germination of pollen grains from other species within the Brassicaceae,[20] although the strength of adhesion of pollen grains to the stigma diminishes in relation to its evolutionary distance from *A. thaliana*[36]. In this species, pollen from closely-related species will germinate and pollen tubes will grow into the transmitting tract of the style, but lose their way in the ovary[46], suggesting that transmitting tract genes might function in species recognition[7, 20, 111]. Another interesting connection between SI and interspecific incompatibility is the fact that reciprocal pollinations between self-incompatible and self-compatible species often have different outcomes. These pollinations fail when the self-incompatible species is used as female parent but succeed when the self-compatible species is used as female parent. This observation has led to the hypothesis that the *S*-locus genes themselves function in discrimination against pollen from other species. However, experiments in *Brassica* do not support this hypothesis, since *Brassica* stigmas that do not express a functional SRK are still able to reject *A. thaliana* pollen[46].

2. Post-zygotic Isolation

Post-zygotic isolation results from incompatibilities that occur after the fertilization process. Some general causes can be differing chromosomal numbers and unfavorable interactions between loci, which can lead to hybrid sterility and necrosis. Often, this phenomenon can be explained by the Bateson-Dobzhansky-Muller model[112], where divergence between the parental species has led to detrimental synergistic interactions between genes in the offspring. A good example of this type of interaction occurs between *Drosophila melanogaster* and *Drosophila simulans*. Lethality in hybrids of these two species is partly due to incompatibilities between two genes, *Hybrid male rescue* (*Hmr*) and *Lethal hybrid rescue* (*Lhr*). Population genetic analysis of *Hmr* has shown that the gene has diverged under positive selection in many lineages of *Drosophila*[114,115]. A similar interaction occurs between isolates of *Caenorhabditis elegans*. This effect can be attributed to some degree to two tightly-linked genes, zygotic *epistatic embryonic lethal-1* (*zeel-1*) and *paternal effect epistatic embryonic lethal-1* (*peel-1*)[113]. Embryos carrying incompatible alleles of these genes arrest as embryos. In *A. thaliana*, hybrid incompatibility was identified by segregation distortion between the Col-0 and Cvi-0 ecotypes[114]. These ecotypes maintain alternate paralogous copies of a gene essential for histidine biosynthesis, with one copy being either silenced or lacking. When the two nonfunctional copies combine in the F2 offspring of a cross, the seeds abort.

B. Rapidly-Evolving Reproductive Genes and Their Role in Recognition and Species Isolation.

Several recent studies have found that genes involved in sexual reproduction tend to be more divergent than genes expressed in other tissues[115, 116]. Rapidly-evolving reproductive genes have been found in many organisms (see Swanson and Vacquier 2002a for a review)[117] including primates[118, 119], *Chlamydomonas reinhardtii*[120], marine invertebrates[121], *Drosophila*[122, 123], and plants [124]. Rapid evolution has also been found to be associated with processes such as pathogen response[125, 126] and environmental adaptation[127].

A rapidly-evolving gene is defined as a gene that encodes a protein with a higher-than-average percentage of amino-acid substitutions between species[117]. Statistically, rapid evolution can be identified by an $\omega > 1$, where ω is the number of nonsynonymous substitutions per nonsynonymous site (dN) divided by the number of synonymous substitutions per synonymous site (dS)[128]. When an amino-acid substitution confers a selective advantage, it will generally become fixed in the population and ω will be greater than one. This is referred to as positive or diversifying selection. When amino-acid substitutions occur at the same rate as synonymous substitutions, the sequence is evolving neutrally and ω will be approximately 1. This occurrence is characteristic of a pseudogene. A protein under purifying selection will typically have an ω value that is less than 1, meaning that the sequence is under strong constraint.

Several hypotheses have been proposed to explain the rapid evolution of reproductive genes, including sperm competition, sexual selection, and sexual conflict[129]. Sperm competition refers to the selection pressure exerted on individual sperm cells during their race to be the first to reach the egg. Some individual sperm cells may be better adapted than others for the process of swimming towards the egg and penetrating the egg envelope. The female has also been found to play an important role in determining the success of individual sperm, so it is conceivable that female genes may also be changing in response to sperm competition[130]. Cryptic female choice is a form of sexual selection that refers to the phenomenon of an egg cell binding preferentially to a particular type of sperm cell. This form of selection is known to occur in *Echinometra* eggs[131]. Sexual conflict describes a situation in which characteristics that increase the fitness of one sex decrease the fitness of the other sex[132]. For example, polyspermy is unfavorable to the egg and as a consequence, the female has evolved mechanisms to prevent fertilization by more than one sperm. The sperm cells may then evolve in a way that allows them to overcome the polyspermy barrier. This coevolution can continue between the two sexes and may lead to reproductive isolation between different populations due to differences in the direction and rate of evolution[133]. Other factors that may drive rapid diversification of a reproductive gene include relaxed constraint on gene repeats, reinforcement, gene duplication events, pathogens, and within-population variation that allows for assortative mating (see Swanson and Vacquier 2002b for a review)[129].

The fruit fly *Drosophila* has received much attention in the identification of rapidly evolving genes. In particular, the male accessory-gland proteins of *Drosophila* have been the focus of such studies. It is estimated that these proteins are twice as divergent between *Drosophila* species as nonreproductive proteins[116]. The most rapidly-evolving accessory-gland protein found to date, Acp26Aa, has a dN/dS ratio of 1.6 when sequences from *D. melanogaster* and *D. yakuba* are compared[123]. Two other accessory-gland proteins, Acp36DE and Acp29AB, have also been found to be subject to positive selection[134, 135]. Some accessory-gland proteins have species-specific functions, since the correct species form of the proteins must be present after mating to ensure viability of interspecific hybrids[136].

A few female components of the reproductive processes have also been found to be highly divergent. Three mammalian reproductive proteins expressed in the female reproductive tract, zona pellucida 2 (ZP2) and zona pellucida 3 (ZP3), both egg-coat proteins, and oviductal glycoprotein (OGP), have been found to be undergoing positive selection[137]. Of particular interest, a region of ZP3 shown to be undergoing positive selection is known to be involved with sperm interaction[137].

The rapid evolution of some reproductive proteins and their resulting diversification can result in reproductive barriers that lead to speciation. Genes that cause ecological, sexual, or post-mating isolation between neospecies are called speciation genes (see Wu and Ting 2004 for a review[138]). It is conceivable that if a reproductive gene is rapidly evolving, it may change in such a way as to prohibit the

organism from successfully reproducing with members outside of its population, eventually leading to the formation of a new species. In *Drosophila*, the *Odysseus* (*OdsH*) gene is an example of a rapidly evolving reproductive gene involved in speciation. *OdsH* is expressed in the testis and is known to cause male sterility when introgressed from *D. mauritiana* into *D. simulans*[139]. This sterility is believed to be due to divergent regulation of the gene between *Drosophila* species. Two additional rapidly-evolving genes with roles in speciation are the hybrid male rescue (*Hmr*) gene, which causes lethality and female sterility in *Drosophila* hybrids[140], and *Nup96*, a gene that encodes a nucleoporin component of the nuclear pore complex, which causes inviability in hybrids between species[141]. It is not yet known if these two genes are involved in reproductive processes.

In plants, tissues directly involved in pollination and fertilization activities (reviewed in prior sections) are likely to express a set of genes that underlie reproductive success or contribute to specificity or discrimination in the pollination process within and between species. In particular, the pollen grain interacts closely with the stigma and relies on it for proper germination and growth. Once the pollen tube emerges and initiates its journey through the style, it must depend on support from the papillar cells, transmitting tract, and ovary. Upon arrival at the female gametophyte, a mechanism is in place to prevent the entrance of more than one pollen tube into the micropyle. Changes in any of the key players necessary for successful fertilization on either the male or female side could disrupt these finely tuned interactions and potentially lead to prezygotic isolation.

In agreement with the aforementioned hypothesis, a few rapidly evolving reproductive genes have been identified in plants. Plant *S* loci are known to be rapidly evolving, consistent with their role in intraspecific mate recognition. The extracellular domain of the SRK receptor contains several highly polymorphic sites with an average sequence similarity among alleles of 78.2% within *Brassica oleracea*[85, 86]. SCR variants exhibit as little as 30% sequence similarity within species[92, 94]. *S* locus genes are thought to be involved mainly in intraspecific recognition and as discussed previously, they most likely do not play a role in interspecific discrimination.

Another class of rapidly evolving reproductive genes is a group of glycine-rich proteins (GRPs) of the pollen coat of Brassicaceae. A recent study showed that the amino-acid sequences of the most conserved oleosin domain in these GRPs are only 40-63% identical between *Brassica oleracea* and *Arabidopsis thaliana*[5]. The members of this gene family were found to be clustered into closely linked groups on chromosome 5. In view of the above-average polymorphism and divergence in the GRP gene cluster region, haplotypes of this gene cassette may play a role in recognition and possibly discrimination between species. An interspecific analysis of GRPs observed a relaxation of selective constraints on GRP loci in comparison to flanking conserved genes[142]. Another study found the GRPs to be present in plants with different breeding systems, which may suggest that they play a role in species recognition[143]. A closer look showed that there were many indels in the exons of GRPs between species, but there were no frameshift mutations, suggesting there is some

functional constraint on the protein[143]. Both studies have found GRPs to be divergent between species, suggesting that these proteins might function in interspecific reproductive isolation. Interestingly, this characteristic seems to be unique to GRPs as other pollen coat lipids tend to be rather conserved across the Brassicaceae[144].

Since the *A. thaliana* stigma allows the hydration and germination of pollen grains from other species within the Brassicaceae, it is possible that transmitting tract genes and female gametophyte would be relied upon for between-species recognition[7, 111]. In support of this hypothesis, the embryo sac has several components known to be involved in pollen tube attraction and regulation as discussed in a previous section of this chapter. In particular, two of these aforementioned genes have been found to show evidence of divergence in sequence and function between species and seem likely candidate speciation genes. LUREs, implicated in attracting pollen tubes to the female gametophyte of Torenia[63] are known to diverge rapidly between species[63,145]. In addition, pollen from various species of *Torenia* preferentially grows towards synergid cells from its own species[155]. This is most likely due to the presence of the appropriate LURE in the synergid cells. *FERONIA*, which encodes a receptor most likely acting in the filiform apparatus, may be involved in one of the final steps of fertilization, the release of the sperm cells from the pollen tube[64,150]. Interspecific hybrids between *A. thaliana* and *A. lyrata* showed that *FERONIA* might play a role in fertilization barriers between these two species. Sequence and analysis of the FERONIA receptor showed that there is higher divergence in the

putative extracellular binding domain of this protein, possibly contributing to species isolation between *A. thaliana* and *A. lyrata*[64]. To date, no pollen tube-expressed genes have been identified as possible speciation genes, although it is likely that further work in the field will identify some candidates.

With this background in mind, we now turn to the work to be covered in this dissertation. In the following chapters, the regulation of intraspecific discrimination will be addressed. Key players in the self-incompatibility response will be studied and new roles in SI for genes known to be involved in other processes will be considered. Interspecies interactions will also be assessed with the goal of identifying novel genes that are necessary for reproductive success within a species while enforcing species boundaries. The goal of these studies is to shed light on discriminatory processes occurring both within and between species and to gain a better understanding of the critical components of a successful pollination and fertilization.

REFERENCES

1. Yang, W. and Sundaresan, V. Genetics of gametophyte biogenesis in *Arabidopsis*. Growth and Development 3, 53-57 (2000).
2. Honys, D. and Twell, D. Comparative analysis of the *Arabidopsis* pollen transcriptome. Plant Physiology 132, 640-652 (2003).
3. Heslop-Harrison, Y. Control gates and micro-ecology: The pollen stigma interaction in perspective. Annals of Botany 85, 5-13 (2000).
4. Edlund, A. F., Swanson, R. and Preuss, D. Pollen and stigma structure and function: the role of diversity in pollination. Plant Cell 16 Suppl, S84-97 (2004).
5. Mayfield, J. A., Fiebig, A., Johnstone, S. E. and Preuss, D. Gene families from the *Arabidopsis thaliana* pollen coat proteome. Science 292, 2482-5 (2001).
6. Wolters-Arts, M., Lush, W. M., and Mariani, C. Lipids are required for directional pollen tube growth. Nature 392, 818-821 (1998).
7. Preuss, D., Lemieux, B., Yen, G. and Davis, R. W. A conditional sterile mutation eliminates surface components from *Arabidopsis* pollen and disrupts cell signaling during fertilization. Genes and Development 7, 974-985 (1993).
8. Sanzol, J., Rallo, P., and Herrero, M. Ashynchronous development of stigmatic receptivity in the pear (*Pyrus communis*, Rosa) flower. American Journal of Botany 90, 78-84 (2003).
9. Johanson, U., Karisson, M., Johansson, I., Gustavsson, S., Sjovall, S., Fraysse, L., Weig, A. R., and Kjellbom, P. The complete set of genes encoding major intrinsic proteins in *Arabidopsis* provides a framework for a new nomenclature for major intrinsic proteins in plants. Plant Physiology 126, 1358-1369 (2001).

10. O'Brien, M., Bertrand, C., and Matton, D. P. Characterization of a fertilization-induced and developmentally regulated plasma-membrane aquaporin expressed in reproductive tissues, in the wild potato *Solanum chacoense*. Bitt Planta 215, 485-493 (2002).

11. Bots, M., Feron, R., Uehlein, N., Weterings, K., Kaldenhoff, R., and Mariani, T. *PIP1* and *PIP2* aquaporins are differentially expressed during tobacco anther and stigma development. Journal of Experimental Botany 56, 113-121 (2005).

12. Marin-Oliver, M., Chevalier, T., Fobis-Loisy, I., Dumas, C., and Gaude, T. Aquaporin *PIP* genes are not expressed in the stigma papillae in *Brassica oleracea*. Plant Journal 24, 231-240 (2000).

13. Dixit, R., Rizzo, C., Nasrallah, M., and Nasrallah, J. The *Brassica MIP-MOD* gene encodes a functional water channel that is expressed in the stigma epidermis. Plant Molecular Biology 45, 51-62 (2001).

14. Sanchez, A. M., Bosch, M., Bots, M., Nieuwland, J., Feron, R., and Mariani, C. Pistil Factors Controlling Pollination. Plant Cell 16, S98-S106 (2004).

15. Muschietti, J., Eyal, Y., and McCormick, S. Pollen tube localization implies a role in pollen-pistil interactions for the tomato Receptor-like Protein Kinases *LePRK1* and *LePRK2*. The Plant Cell 10, 319-330 (1998).

16. Muschietti, J., Dircks, L., Vancanneyt, G., and McCormick, S. LAT52 protein is essential for tomato pollen development: pollen expressing antisense *LAT52* RNA hydrates and germinates abnormally and cannot achieve fertilization. The Plant Journal 6, 321-338 (1994).

17. Kim, H. U., Cotter, R., Johnson, S., Senda, M., Dodds, P., Kulikauska, R., Tang, W., Ezcura, I., Herzmark, P. and McCormick, S. New pollen-specific receptor kinases identified in tomato, maize and *Arabidopsis*: the tomato kinases show overlapping but distinct localization patterns on pollen tubes. Plant Mol Biol 50, 1-16 (2002).

18. Tang, W., Kelley, D., Ezcurra, I., Cotter, R. and McCormick, S. LeSTIG1, an extracellular binding partner for the pollen receptor kinases LePRK1 and LePRK2, promotes pollen tube growth *in vitro*. Plant J 39, 343-53 (2004).

19. Dickinson, H. Dry stigmas, water and self-incompatibility in *Brassica*. Sexual Plant Reproduction 8, 1-10 (1995).

20. Hulskamp, M., Kopczak, S. D., Horejsi, T. F., Kihl, B. K. and Pruitt, R. E. Identification of genes required for pollen-stigma recognition in *Arabidopsis thaliana*. Plant J 8, 703-14 (1995).

21. Mayfield, J. A. and Preuss, D. Rapid initiation of *Arabidopsis* pollination requires the oleosin-domain protein GRP17. Nat Cell Biol 2, 128-30 (2000).

22. Goldman, M. H., Goldberg, R. B., and Mariani, C. Female sterile tobacco plants are produced by stigma-specific cell ablation. European Molecular Biology Organization Journal 13, 2976-2984 (1994).

23. Lolle, S. J., Hsu, W., and Pruitt, R. E. Genetic analysis of organ fusion in *Arabidopsis thaliana*. Genetics 149, 607-619 (1998).

24. Pruit, R. E., Viell-Calzada, J. P., Ploense, S. E., Grossniklaus, U., and Lolle, S. J. *FIDDLEHEAD*, a gene required to suppress epidermal cell interactions in *Arabidopsis*, encodes a putative lipid biosynthetic enzyme. Proc Natl Acad Sci U S A 97, 1311-1316 (2000).

25. Lush, W. M., Grieser, F., and Wolters-Arts, M. Directional guidance of *Nicotiana alata* pollen tubes *in vitro* and on the stigma. Plant Physiology 118, 733-741 (1998).

26. Hiscock, S. J., Dewey, F. M., Coleman, J. O. K., and Dickinson, H. G. Identification and localization of an active cutinase in the pollen of *Brassica napus* L. Planta 193, 377-384 (1994).

27. Dearnaley, J. D. W., and Daggard, G. A. Expression of a polygalacturonase enzyme in germinating pollen of *Brassica napus*. Sexual Plant Reproduction 13, 265-271 (2001).

28. Mu, J. H., Stains, J. P., and Kao, T. H. Characterization of a pollen-expressed gene encoding a putative pectin esterase of *Petunia inflata*. Plant Mol Biol 25, 539-544 (1994).

29. Kotake, T., Li, Y-Q, Takahashi, M., and Sakurai, N. Characterization and function of wall-bound exo-beta-glucanases of *Lilium longiflorum* pollen tubes. Sexual Plant Reproduction 13, 1-9 (2000).

30. Doblin, M. S., De Melis, L., Newbigin, E., Bacic, A., and Read, S. M. Pollen tubes of *Nicotiana alata* express two genes from different beta-glucan synthase families. Plant Physiol 125, 2040-2052 (2001).

31. Bih, F. Y., Wu, S. S. H., Ratnayake, C., Walling, L. L., Nothnagel, E. A., and Huang, A. H. C. The predominant protein on the surface of maize pollen is an endozylanase synthesized by a tapetum mRNA with a long 5' leader. Journal of Biological Chemistry 274, 22884-22894 (1999).

32. Wheeler, M. J., Franklin-Tong, V. E., and Franklin, F. C. H. The molecular and genetic basis of pollen-pistil interactions. New Phytologist 151, 565-584 (2001).

33. Park, S. Y., Jauh, G. Y., Mollet, J. C., Eckard, K. J., Nothnagel, E. A., Walling, L. L. and Lord, E. M. A lipid transfer-like protein is necessary for lily pollen tube adhesion to an in vitro stylar matrix. Plant Cell 12, 151-64 (2000).

34. Kim, S., Mollet, J. C., Dong, J., Zhang, K., Park, S. Y. and Lord, E. M. Chemocyanin, a small basic protein from the lily stigma, induces pollen tube chemotropism. Proc Natl Acad Sci U S A 100, 16125-30 (2003).

35. Lord, E. M. Adhesion and guidance in compatible pollination. J Exp Bot 54, 47-54 (2003).

36. Zinkl, G. M., Zwiebel, B. I., Grier, D. G., and Preuss, D. Pollen-stigma adhesion in *Arabidopsis*: a species-specific interaction mediated by lipophilic molecules in the pollen exine. Development 126, 5431-5440 (1999).

37. Luu, D. T., Heizmann, P., and Dumas, C. Pollen-stigma adhesion in kale is not dependent on the self-(in)compatibility genotype. Plant Physiology 115, 1221-1230 (1997).

38. Elleman, C. J. and Dickinson, H. G. Pollen-stigma interactions in *Brassica*. IV. Structural reorganization in the pollen grains during hydration. Journal of Cell Science 80, 141-157 (1986).

39. Elleman, C. J., Franklin-Tong. V., and Dickinson, H. G. Pollination in species with dry stigmas: The nature of the early stigmatic response and the pathway taken by pollen tubes. New Phytologist 121, 413-424 (1992).

40. Robert, L., Allard, S., Gerster, J. L., Cass, L., and Simmonds, J. Molecular analysis of two *Brassica napus* genes expressed in the stigma. Plant Molecular Biology 26, 1217-1222 (1994).

41. Kang, Y. R., and Nasrallah, J. B. Use of genetically ablated stigmas for the isolation of genes expressed specifically in the stigma epidermis. Sexual Plant Reproduction 14, 85-94 (2001).

42. Kandasamy, M. K., Thorsness, M. K., Rundle, S. J., Goldberg, M. L., Nasrallah, J. B., and Nasrallah, M. E. Ablation of Papillar Cell Function in *Brassica* Flowers Results in the Loss of Stigma Receptivity to Pollination. The Plant Cell 5, 263-275 (1993).

43. Thorsness, M. K., Kandasamy, M. K., Nasrallah, M. E., and Nasrallah, J. B. Genetic Ablation of Floral Cells in *Arabidopsis*. The Plant Cell 5, 253-261 (1993).

44. Dickinson, H. G., Doughty, J., Hiscock, S. J., Elleman, C. J. and Stephenson, A. G. Pollen-stigma interactions in *Brassica*. Symp Soc Exp Biol 51, 51-7 (1998).

45. Hill, J. P. and Lord, E. M. Dynamics of pollen tube growth in the wild radish *Raphanus raphanistrum* (*Brassica*ceae). II. Morphology, cytochemistry and ultrastructure of transmitting tissues, and path of pollen tube growth. American Journal of Botany 74, 988-997 (1987).

46. Kandasamy, M. K., Nasrallah, J. B., and Nasrallah, M. E. Pollen-pistil interactions and developmental regulation of pollen tube growth in *Arabidopsis*. Development 120, 3405-3418 (1994).

47. Lennon, K. A., Roy, S., Hepler, P. K., and Lord, E. M. The structure of the transmitting tissue of *Arabidopsis thaliana* (L.) and the path of pollen tube growth. Sexual Plant Reproduction Volume 11, 49 - 59 (1998).

48. Herrero, M., and Hrmaza, J. I. Pistil strategies controlling pollen tube growth. Sexual Plant Reproduction 9, 343-347 (1996).

49. Sander, L. C. and Lord, E. M. Directed movement of latex particles in the gynoecia of three species of flowering plants. Science 243, 1606-1608 (1989).

50. Zhao, J., Mollet, J-C, and Lord, E. M. Lily (*Lilium longiflorum* L.) pollen protoplast adhesion is increased in the presence of the peptide SCA. Sexual Plant Reproduction 16, 227-233 (2004).

51. Iwanami, Y. Physiological studies of pollen. Journal of the Yokohama Municpal University 116, 1-137 (1959).

52. Jiang, L., Yang, S., Xie, L., Puah, C. S., Zhang, Z., Yang, W., Sunderasan, V., and Ye, D. *VANGUARD1* encodes a pectin methylesterase that enhances pollen tube growth in the *Arabidopsis* style and transmitting tract. Plant Cell 17, 584-596 (2005).

53. Wu, H. M., Wang, H., and Cheung, A. Y. A pollen tube growth stimulatory glycoprotein is deglycosylated by pollen tubes and displays a glycosylation gradient in the flower. Cell 82, 395-403 (1995).

54. Cheung, A. Y., Wang, H., and Wu, H. A floral transmitting tissue-specific glycoprotein attracts pollen tubes and stimulates their growth. Cell 82, 383-393 (1995).

55. Wu, H. M., Wong, E., Ogdahl, J. and Cheung, A. Y. A pollen tube growth-promoting arabinogalactan protein from *Nicotiana alata* is similar to the tobacco TTS protein. Plant J 22, 165-76 (2000).

56. Ravindran, S., Kim, S., Martin, R., Lord, E. M., and Ozkan, C. S. Quantum dots as bio-labels for the localization of a small plant adhesion protein. Nanotechnology 16, 1-4 (2005).

57. Kim, S. T., Zhang, K., Dong, J. and Lord, E. M. Exogenous free ubiquitin enhances lily pollen tube adhesion to an *in vitro* stylar matrix and may facilitate endocytosis of SCA. Plant Physiol. 142, 1397-1411 (2006).

58. Mascarenhas, J. P., and Machlis, J. P. Chemotropic response of *Antirrhinum majus* pollen to calcium. Nature 196, 292-293 (1962).

59. Palanivelu, R., Brass, L., Edlund, A. F. and Preuss, D. Pollen tube growth and guidance is regulated by *POP2*, an *Arabidopsis* gene that controls GABA levels. Cell 114, 47-59 (2003).

60. Ray, S. M., Park, S. S. and Ray, A. Pollen tube guidance by the female gametophyte. Development 124, 2489-98 (1997).

61. Shimizu, K. K., and Okada, K. Attractive and repulsive interactions between female and male gametophytes in *Arabidopsis* pollen tube guidance. Development 127, 4511-4518 (2000).

62. Higashiyama, T., Yabe, S., Sasaki, NH., Nishimura, Y., Miyagishima, S., Kuroiwa, H., and Kuroiwa, T. Pollen tube attraction by the synergid cell. Science 293, 1480-1483 (2001).

63. Marton, M. L., Cordts, S., Broadhvest, J., and Dresselhaus, T. Micropylar pollen tube guidance by egg apparatus 1 of maize. Science 307, 573-576 (2005).

64. Chen, Y.-H., Li, H.-J., Shi, D.-Q., Yuan, L., Liu, J., Sreenivasan, R., Baskar, R., Grossniklaus, U. and Yang, W.-C. The central cell plays a critical role in pollen tube guidance in *Arabidopsis*. Plant Cell 19, 3563-3577 (2007).

65. Arbeloa, A., and Herrero, M. The significance of the obturator in the control of pollen tube entry into the ovary in peach (*Prunus persica*). Annals of Botany 60, 681-685 (1987).

66. Gonzalez, M. V., Coque, M., and Herrero, M. Pollen-pistil interaction in kiwifruit (*Actinidia deliciosa*: Actinidiaceae). Am J Bot 83, 148-154 (1996).

67. Boisson-Dernier, A., Roy, S., Kritsas, K., Grobei, M. A., Jaciubek, M., Schroeder, J. I. and Grossniklaus, U. Disruption of the pollen-expressed *FERONIA* homologs *ANXUR1* and *ANXUR2* triggers pollen tube discharge. Development 136, 3279-3288 (2009).

68. von Besser, K., Frank, A. C., Johnson, M. A. and Preuss, D. *Arabidopsis HAP2* (*GCS1*) is a sperm-specific gene required for pollen tube guidance and fertilization. Development 133, 4761-4769 (2006).

69. Fobis-Loisy, I., Miege, C., and Gaude, T. Molecular Evolution of the *S* Locus Controlling Mating in the *Brassica*ceae. Plant Biology 6, 109-118 (2004).

70. Medrano, M., Herrera, C. M., and Barrett, S. C. Herkogamy and mating patterns in the self-compatible Daffodil *Narcissus longispathus*. Annals of Botany Preview (2005).

71. Kurian, V. and Richards, A. J. A new recombinant in the heteromorphy '*S*' supergene in *Primula*. Heredity 78, 383-390 (1996).

72. McClure, B. Darwin's foundation for investigating self-incompatibility and the progress toward a physiological model for S-RNase-based SI. J. Exp. Bot. 60, 1069-1081 (2009).

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

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

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

76. Murfett, J., Atherton, T. L., Mou, B., Gasser, S. S., and McClure, B. A. S-RNase expressed in transgenic *Nicotiana* causes *S*-allele-specific pollen rejection. Nature 367, 563-566 (1996).

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

78. McClure, B. A., Gray, J. E., Anderson, M. A., Clarke, A. E. Self-incompatibility in *Nicotiana alata* involves degradation of pollen rRNA. Nature 341, 757-560 (1990).

79. Entani, T., Iwano, M., Shiba, H., Che, F-S, Isogai, A., and Takayama, S. Comparative analysis of the self-incompatibility (S-) locus region of *Prunus mume*: identification of a pollen-expressed F-box gene with allelic diversity. Genes Cells 8, 203-213 (2003).

80. Sijacic, P., Wang, X., Skirpan, A. L., Wang, Y., and Dowd, P. E. Identification of the pollen determinant of S-RNase-mediated self-incompatibility. Nature 429, 302-305 (2004).

81. O'Brien, M., Kapfer, C., Major, G., Laurin, M., and Bertrand, C. Molecular analysis of the stylar-expressed *Solanum chacoense* small asparagine-rich protein family related to the HT modifier of gametophytic self-incompatibility. Plant Journal 32, 985-996 (2002).

82. Kondo, K., Yamamoto, M., Matton, D. P., Sato, T., and Hirai, M. Cultivated tomato has defects in both *S-RNase* and *HT* genes required for stylar function of self-incompatibility. Plant Journal 29, 627-636 (2002).

83. Kondo, K., Yamamoto, M., Itahashi, R., Sato, T., and Egashira, H. Insights into the evolution of self-compatibility in *Lycopersicon* from a study of stylar factors. Plant Journal 30, 143-153 (2002).

84. Nasrallah, J. B., Kao, T-h, Chen, C-H, Goldberg, M. L., Nasrallah, M. E. Amino-acid sequence of glycoproteins encoded by three alleles of the *S* locus of *Brassica oleracea*. Nature 326 (1987).

85. Stein, J. C., Howlett, B., Boyes, D. C., Nasrallah, M. E. and Nasrallah, J. B. 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, 8816-20 (1991).

86. Nishio, T. and Kusaba, M. Sequence Diversity of SLG and SRK in *Brassica oleracea* L. Annals of Botany 85, 141-146 (2000).

87. Takasaki, T., Hatakeyama, K., Suzuki, G., Watanabe, M., Isogai, A., and Hinata, K. The *S* receptor kinase determines self-incompatibility in *Brassica* stigma. Nature 913-916 (2000).

88. Silva, N. F., Stone, S. L., Christie, L. N., Sulaman, W., and Nazarian, K. A. P. Expression of the *S* receptor kinase in self-compatible *Brassica napus* cv. Weststar leads to the allele-specific rejection of self-incompatible *Brassica napus* pollen. Molecular Genetics and Genomics 265, 552-559 (2001).

89. Kusaba, M., Dwyer, K., Hendershot, J., Vrebalov, J., Nasrallah, J. B., and Nasrallah, M. E. Self-Incompatibility in the genus *Arabidopsis*: Characterization of the *S* Locus in the outcrossing *A. lyrata* and its autogamous relative *A. thaliana*. The Plant Cell 13, 627-643 (2001).

90. Suzuki, G., Kusaba, M., Matsushita, M., Okazaki, K., and Nishio,T. Characterization of *Brassica S*-haplotypes lacking S-locus glycoprotein. FEBS Lett 482, 102-108 (2000).

91. Suzuki, G., Kakizaki, T., Takada, Y., Shiba, H., and Takayama, S. The *S* haplotypes lacking SLG in the genome of *Brassica rapa*. Plant Cell Reports 21, 911-915 (2003).

92. Schopfer, C. R., Nasrallah, M. E. and Nasrallah, J. B. The male determinant of self-incompatibility in *Brassica*. Science 286, 1697-700 (1999).

93. Schierup, M. H., Mable, B. K., Awadalla, P., and Charlesworth, D. Identification and Characterization of a Polymorphic Receptor Kinase Gene Linked to the Self-Incompatibility Locus of *Arabidopsis lyrata*. Genetics 158, 387-399 (2001).

94. Watanabe, M., Ito, A., Takada, Y., Ninomiya, C., Kakizaki, T., Takahata, Y., Hatakeyama, K., Hinata, K., Suzuki, G., Takasaki, T., Satta, Y., Shiba, H., Takayama, S. and Isogai, A. Highly divergent sequences of the pollen self-incompatibility (S) gene in class-I S haplotypes of *Brassica campestris* (syn. *rapa*) L. FEBS Lett 473, 139-44 (2000).

95. Nasrallah, M. E., Liu, P. and Nasrallah, J. B. Generation of Self-Incompatible *Arabidopsis thaliana* by Transfer of Two S Locus Genes from A. *lyrata*. Science 297, 247-249 (2002).

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

97. Murase, K., Shiba, H., Iwano, M., Che, F-S., and Watanabe, M. A membrane-anchored protein kinase involved in *Brassica* self-incompatibility signaling. Science 303, 1516-1519 (2004).

98. Kakita, M., Murase, K., Iwano, M., Matsumoto, T., Watanabe, M., Shiba, H., Isogai, A. and Takayama, S. Two Distinct Forms of M-Locus Protein Kinase Localize to the Plasma Membrane and Interact Directly with S-Locus Receptor Kinase to Transduce Self-Incompatibility Signaling in *Brassica rapa*. Plant Cell 19, 3961-3973 (2007).

99. Gu, T., Mazzurco, M., Sulaman, W., Matias, D. D., and Goring, D. R. Binding of an arm repeat protein to the kinase domain of the *S*-locus receptor kinase. Proc Natl Acad Sci U S A 95, 382-387 (1998).

100. Muzzurco, M., Sulaman, W., Elina, H., Cock, J. M., and Goring, D. R. Further analysis of the interactions between the *Brassica* S receptor kinase and three interacting proteins (ARC1, THL1 and THL2) in the yeast two-hybrid system. Plant Mol Biol 45, 365-375 (2001).

101. Stone, S. L., Arnoldo, M., and Goring, D. R. A breakdown of *Brassica* self-incompatibility in *ARC1* antisense transgenic plants. Science 286, 1729-1731 (1999).

102. Stone, S. L., Anderson, E. M., Mullen, R. T. and Goring, D. R. 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 (2003).

103. Samuel, M. A., Chong, Y. T., Haasen, K. E., Aldea-Brydges, M. G., Stone, S. L. and Goring, D. R. 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 (2009).

104. Bower, M. S., Matias, D. D., Fernandes-Carvalho, E., Muzzurco, M., Gu, T., Rothstein, S. J., and Goring, D. R. Two Members of the Thioredoxin-h Family Interact with the Kinase Domain of a *Brassica* S Locus Receptor Kinase. The Plant Cell 8, 1641-1650 (1996).

105. Cabrillac, D., Cock, J. M., Dumas, C., and Gaude, T. The *S*-locus receptor kinase is inhibited by thioredoxins and activated by pollen coat proteins. Nature 410, 220-223 (2001).

106. Haffani, Y. Z., Gaude, T., Cock, M., and Goring, D. R. Antisense suppression of thioredoxin h mRNA in *Brassica napus* cv. Westar pistils causes a low level constitutive pollen rejection response. Plant Mol Biol 55, 619-630 (2004).

107. Liu, P., Sherman-Broyles, S., Nasrallah, M. E. and Nasrallah, J. B. A cryptic modifier causing transient self-incompatibility in *Arabidopsis thaliana*. 17, 734 (2007).

108. Nasrallah, J. B., Liu, P., Sherman-Broyles, S., Schmidt, R. and Nasrallah, M. E. Epigenetic mechanisms for breakdown of self-incompatibility in interspecific hybrids. Genetics 175, 1965-1973 (2007).

109. Tantikanjana, T., Rizvi, N., Nasrallah, M. E. and Nasrallah, J. B. A dual role for the S-Locus Receptor Kinase in self-incompatibility and pistil cevelopment revealed by an *Arabidopsis rdr6* Mutation. Plant Cell 21, 2642-2654 (2009).

110. Lu, C., Kulkarni, K., Souret, F. F., MuthuValliappan, R., Tej, S. S., Poethig, R. S., Henderson, I. R., Jacobsen, S. E., Wang, W., Green, P. J. and Meyers, B. C. MicroRNAs and other small RNAs enriched in the *Arabidopsis RNA-dependent RNA polymerase-2* mutant. Genome Research 16, 1276-1288 (2006).

111. Hiscock, S. J. a. Dickinson, H. G. Unilateral incompatibility within the *Brassica*ceae: further evidence for the involvement of the self-incompatibility (*S*)-locus. Theoretical Applied Genetics 86, 744-753 (1993).

112. Coyne, J. A. and Orr, H. A. Speciation. (2004).

113. Seidel, H. S., Rockman, M. V. and Kruglyak, L. Widespread genetic incompatibility in *C. elegans* maintained by balancing selection. Science 319, 589-594 (2008).

114. Bikard, D., Patel, D., Le Mette, C., Giorgi, V., Camilleri, C., Bennett, M. J. and Loudet, O. Divergent evolution of duplicate genes leads to genetic incompatibilities within *A. thaliana*. Science 323, 623-626 (2009).

115. Vacquier, V. D. Evolution of gamete recognition proteins. Science 281, 1995-8 (1998).

116. Singh, R. S. and Kulathinal, R. J. Sex gene pool evolution and speciation: A new paradigm. Genes and Genetic Systems Vol. 75, pp.119-130 (2000).

117. Swanson, W. J. and Vacquier, V. D. The rapid evolution of reproductive proteins. Nature Reviews Genetics 3, 137 -144 (2002).

118. Makalowski, W. and Boguski, M. S. Evolutionary parameters of the transcribed mammalian genome: an analysis of 2,820 orthologous rodent and human sequences. Proc Natl Acad Sci U S A 95, 9407-12 (1998).

119. Wyckoff, G. J., Wang, W. and Wu, C. I. Rapid evolution of male reproductive genes in the descent of man. Nature 403, 304-9 (2000).

120. Ferris, P. J., Pavlovic, C., Fabry, S. and Goodenough, U. W. Rapid evolution of sex-related genes in *Chlamydomonas*. Proc Natl Acad Sci U S A 94, 8634-9 (1997).

121. Biermann, C. H. The molecular evolution of sperm bindin in six species of sea urchins (Echinoida: Strongylocentrotidae). Mol Biol Evol 15, 1761-71 (1998).

122. Tsaur, S. C. and Wu, C. I. Positive selection and the molecular evolution of a gene of male reproduction, *Acp26Aa* of *Drosophila*. Mol Biol Evol 14, 544-549 (1997).

123. Tsaur, S. C., Ting, C. T. and Wu, C. I. Positive selection driving the evolution of a gene of male reproduction, *Acp26Aa*, of *Drosophila*: II. Divergence versus polymorphism. Mol Biol Evol 15, 1040-1046 (1998).

124. Sainudiin, R., Shuk, W., Wong, W., Yogeeswaran, K., Nasrallah, J. B., Yang, Z., and Nielsen, R. Detecting site-specific physiochemical slective pressures. Journal of Molecular Evolution 60, 315-326 (2005).

125. Maxwell, A. I., Morrison, G. M. and Dorin, J. R. Rapid sequence divergence in mammalian beta-defensins by adaptive evolution. Mol Immunol 40, 413-21 (2003).

126. Watanabe, Y., Maruoka, T., Walter, L. and Kasahara, M. Comparative genomics of the Mill family: a rapidly evolving MHC class I gene family. Eur J Immunol 34, 1597-607 (2004).

127. Yang, L. and Gui, J. F. Positive selection on multiple antique allelic lineages of transferrin in the polyploid *Carassius auratus*. Mol Biol Evol 21, 1264-77 (2004).

128. Miyata, T. and Yasunaga, T. Molecular evolution of mRNA: a method for estimating evolutionary rates of synonymous and amino acid substitutions from homologous nucleotide sequences and its application. J Mol Evol 16, 23-36 (1980).

129. Swanson, W. J. and Vacquier, V. D. Reproductive protein evolution. Annual Review of Ecology and Systematics 33, 161-179 (2002).

130. Clark, A. G., Begun, D. J. and Prout, T. Female x male interactions in *Drosophila* sperm competition. Science 283, 217-20 (1999).

131. Palumbi, S. R. All males are not created equal: fertility differences depend on gamete recognition polymorphisms in sea urchins. Proc Natl Acad Sci U S A 96, 12632-7 (1999).

132. Stockley, P. Sexual conflict resulting from adaptations to sperm competition. Trends in Ecology and Evolution 12, 154-159 (1997).

133. Gavrilets, S. Rapid evolution of reproductive barriers driven by sexual conflict. Nature 403, 886-9 (2000).

134. Begun, D. J., Whitley, P., Todd, B. L., Waldrip-Dail, H. M. and Clark, A. G. Molecular population genetics of male accessory gland proteins in *Drosophila*. Genetics 156, 1879-88 (2000).

135. Aguade, M. Positive selection drives the evolution of the Acp29AB accessory gland protein in *Drosophila*. Genetics 152, 543-51 (1999).

136. Fuyama, Y. Species-specificity of paragonial substances as an isolating mechanism in *Drosophila*. Experientia 39, 190-192 (1983).

137. Swanson, W. J., Yang, Z., Wolfner, M. F. and Aquadro, C. F. Positive Darwinian selection drives the evolution of several female reproductive proteins in mammals. Proceedings of the National Academy of Sciences of the United States of America 98, 2509-2514 (2001).

138. Wu, C. and Ting, C. Genes and Speciation. Nature Reviews Genetics 5, 114 - 122 (2004).

139. Sun, S. Functional analysis of the hybrid male sterility gene *Odysseus* in *Drosophila*. University of Chicago Press, Chicago (2003).

140. Barbash, D. A., Siino, D. F., Tarone, A. M. and Roote, J. A rapidly evolving MYB-related protein causes species isolation in *Drosophila*. Proc Natl Acad Sci U S A 100, 5302-7 (2003).

141. Presgraves, D. C., Balagopalan, L., Abmayr, S. M. and Orr, H. A. Adaptive evolution drives divergence of a hybrid inviability gene between two species of *Drosophila*. Nature 423, 715-9 (2003).

142. Fiebig, A., Kimport, R. and Preuss, D. Comparisons of pollen coat genes across *Brassica*ceae species reveal rapid evolution by repeat expansion and diversification. Proc Natl Acad Sci U S A 101, 3286-91 (2004).

143. Schein, M., Yang, Z., Mitchell-Olds, T. and Schmid, K. J. Rapid Evolution of a Pollen-Specific Oleosin-Like Gene Family from *Arabidopsis thaliana* and Closely Related Species. Mol Biol Evol 21, 659-669 (2004).

144. Piffanelli, P., Ross, J. H. E., and Murphy, D. J. Biogenesis and function of the lipidic structures of pollen grains. Sexual Plant Reproduction 11, 65-80 (1998).

145. Higashiyama, T., Inatsugi, R., Sakamoto, S., Sasaki, N., Mori, T., Kuroiwa, H., Nakada, T., Nozaki, H., Kuroiwa, T. and Nakano, A. Species preferentiality of the pollen tube attractant derived from the synergid cell of *Torenia fournieri*. Plant Physiol. 142, 481-491 (2006).

CHAPTER 2:

IDENTIFICATION AND CHARACTERIZATION OF *NRPD1A* AS A REGULATOR OF SELF-INCOMPATIBILITY IN *ARABIDOPSIS THALIANA*

INTRODUCTION

In this chapter, I will focus on intraspecific mate selectivity in *Arabidopsis thaliana* by two separate means: outcrossing and self-fertilization (see Chapter 1 for a review). Outcrossing species often utilize self-incompatibility as a method of preventing pollination by pollen from genetically related plants and allowing pollination by pollen of dissimilar genotype. In the Brassicaceae, self-incompatibility (SI) is controlled by haplotypes of the *S* locus, and self-pollination is prevented when pollen and stigma are derived from plants that express the same *S*-locus variant. Two *S*-locus genes are known to be essential for specific recognition of "self" pollen in the SI response in this family: the S-locus Receptor Kinase (*SRK*) and the S-locus Cysteine-Rich protein (*SCR*) genes. SRK is a receptor kinase localized on the surface of stigma epidermal cells and its SCR ligand is located in the pollen coat. An SCR variant can bind and activate only the SRK variant encoded in the same *S*-locus haplotype. This "self' interaction is thought to trigger a poorly-understood signal transduction cascade within the stigma epidermal cell that prevents pollen hydration and germination, as well as growth of the pollen tube into the stigma epidermal cell wall.

The self-fertile species *A. thaliana* contains non-functional alleles of the *SRK* and *SCR* genes[1], but some accessions can be made to express SI by transformation with functional *SRK* and *SCR* genes isolated from its close self-incompatible relative *A. lyrata*[2, 3](see Chapter 1 for a more detailed explanation). These *A. thaliana SRK-SCR* transformants have been used in screens for mutants disrupted in the SI response in an effort to identify genes required for SI[4]. Here we report on a recessive mutation that suppresses the SI phenotype of *SRK-SCR* transformants of the Columbia (Col-0) accession. The gene disrupted by the mutation was identified as *NRPD1a* by a map-based cloning strategy. Subsequent characterization of *nrpd1a* mutants showed that NRPD1a plays a varied and complex role in the regulation of self-incompatibility.

MATERIALS AND METHODS

Plant growth conditions

The mutation analyzed in this study, designated *self compatible 1* (*sc1),* was identified in a previous screen of an EMS-mutagenized population of *A. thaliana* plants of the Col-0 accession that had been transformed with *SRKb* and *SCRb,* which are the *SRK* and *SCR* alleles isolated from the *Sb* haplotype of *A. lyrata.*[4]. Seed was plated on MS media under sterile conditions and vernalized for 3 days at 4°C. Kanamycin at 25 ug/ul was used for selection of plants harboring the *SRKb* and *SCRb* transgenes. Seedlings were transplanted to soil once they had four true leaves. Plants were grown under long-day

conditions (16 hour days/8 hour nights) at a temperature of 20°C in a controlled-environment chamber.

Pollination assays

Pollination tests were performed on stigmas of floral buds at stage 13 of flower development (staging according to Smyth *et al*, 1990[5]; hereafter referred to as the -1 bud stage[4]). To identify plants exhibiting a breakdown of self-incompatibility, pollination tests were typically performed by pollinating three -1 bud-stage stigmas per individual plant with pollen from a plant known to express functional *SCRb* (hereafter *SCRb* pollen). Pollinated buds were incubated at room temperature for 2 hours and subsequently processed for visualization with a UV-fluorescence microscope as described previously[6]. Under these conditions, very few (<5) pollen tubes are observed on the stigmas of wild type *SRKb* plants. Plants whose stigmas exhibited more than 20 pollen tubes penetrating the epidermal cell wall were considered to exhibit a breakdown of the self-incompatibility response and were scored as self-compatible (SC). These plants were retested on a separate day to confirm the results, and the SC phenotype was again confirmed in the F3 plants progenies of these SC plants.

Plant genomic DNA preparation and DNA gel blot analysis

Genomic DNA was extracted using the CTAB method[7] and digested overnight with EcoRI (New England Biolabs, Ipswich, MA). Capillary transfers were performed under alkaline conditions overnight onto Hybond N+ membrane (GE Healthcare, Little Chalfont, UK).

Membranes were probed with a ^{32}P-labeled 1.6-kb probe derived from the 3' UTR of the *SRKb* gene. Blots were exposed to phosphor screens (GE Healthcare, Little Chalfont, UK) and scanned with a STORM 860 PhosphorImager (GE Healthcare, Little Chalfont, UK).

Map-based cloning and DNA sequencing

A genetic mapping population was generated by crossing a plant homozygous for the *SRKb-SCRb* transgenes and for the recessive *sc1* mutation as the female parent, to a wild type plant of the Landsberg *erecta* (Ler-0) accession. Plants carrying the *SRKb* transgene were identified by PCR amplification of the *SRKb* transgene with gene-specific primers (Table A.1). Mapping was performed using simple sequence length repeats (SSLP) and cleaved amplified polymorphism (CAPS) markers listed on The Arabidopsis Information Resource (TAIR) website (Table A.1). Additional SSLP and single nucleotide polymorphism (SNP) markers were generated using the Landsberg erecta random sequence database (www.tigr.org)(Table A.1). A population of 1600 F2 plants was used to map the mutation, and linkage to a 160-kb region on chromosome 1 was found. To identify the gene disrupted by the mutation within the mapping interval, gene-specific primers were designed for all 39 genic regions found in the interval. PCR products were sequenced at the Cornell Bioresource Center (Ithaca, NY) using an Applied Biosystems 3730xl DNA analyzer using big dye terminator chemistry and AmpliTaq-FS DNA polymerase.

Two Col-0 SALK[8] lines, SALK_143437 and SALK_128428, containing a T-DNA insertion in the gene *NRPD1a* were obtained from

ABRC (Columbus, Ohio). The locations of the T-DNA insertions were in exon 14 and exon 9, respectively[8]. These lines were screened for homozygosity using gene-specific primers (Table A.2). Plants homozygous for each of the T-DNA insertions were crossed to a *SRKb-SCRb sc1* homozygous mutant.

Additionally, Col-0 strains carrying T-DNA insertion in several other genes were obtained from the ABRC. The following list gives the strain designations with the genes disrupted by the T-DNA insertion shown in parentheses: SALK_005512 (*dcl3*), SALK_059661 (*rdr2*), SALK_071772 (*ago4*), SALK_076129 (*nrpd1b*), SALK_095689 (*nrpd2a*), SALK-021316 (*drm1*), SALK_150863 (*drm2*), SALK_132061 (*drd1*), and SALK_057148 (*cmt3*). The *nrpd1a ros1-1* double mutant was obtained courtesy of Craig Pikaard's Laboratory (Indiana University].

Table A.2 and Table A.3 shows a list of primers used for screening and analysis of T-DNA insertion lines.

Methylation analysis by chop PCR

SRKb-SCRb sc1 plants and plants homozygous for mutations that disrupt genes involved in DNA methylation were tested for a loss of methylation as follows. Genomic DNA was digested overnight with the methylation-sensitive enzyme, HaeIII. PCR primers for the methylated *AtSN1* repeat and for the non-methylated At2g12990 were designed according to a previous study[9].

RNA gel blot analysis

For RNA gel blot analysis, the following tissues were collected

from 100 -1-stage flower buds: stigmas, pistils without the stigmas, petals, anthers, and sepals. PolyA RNA was extracted using the FastTrak RNA isolation kit (Invitrogen, Carlsbad, CA) and subjected to gel blot analysis as described previously[1]. The blots were hybridized with a ^{32}P-labeled probe generated from the first exon of *SRKb* and subsequently with a ^{32}P-labeled *actin* probe. Visualization of hybridization signals was as described above for DNA gel blot analysis. Signal intensities were quantified using the ImageQuant software package and normalized using *actin* hybridization signals.

Real-time RT-PCR

RNA was isolated from 50 stigmas using Trizol reagent (Invitrogen, Carlsbad, CA). Three replicate RNA samples for each genotype were prepared. For each sample, the RNA was treated with DNase I and 1 ug was used for cDNA synthesis using the First Strand cDNA Synthesis Kit for Real-time PCR (United States Biochemical (USB), Cleveland, OH) and oligodT primers. Real-time PCR was performed using iQ SYBR Green Supermix (Bio-Rad, Hercules, CA) on an ABI Prism 7900HT sequence detection system. The primers used for real-time RT-PCR were as follows: for *SRKb* (rt-*SRKb*4), 5'-CTAAGCCTTGATTCTCATCTCTTTACA-3' and 5'-GAAGTCCCCGAGCAATACCAT-3'; and for ubiquitin, gene-specific primers were developed in a previous study[10]. In order to confirm that no genomic DNA remained in the RNA samples, both ubiquitin and *SRKb* primers were designed to span an intron. Results were analyzed using the SDS software package (Applied Biosystems, Foster City, CA).

The relative amounts of transcripts were calculated using the comparative ct method and normalized to ubiquitin. The mean values were calculated from 3 replicate samples collected on different days.

Analysis of *SRKb* promoter activity using the GUS reporter

The chimeric *Ats1pr*::*uidA*::*nos* gene was described previously[1,12]. An *SRKbpr*::*uidA*::*nos* gene was constructed by Titima Tantikanjana as follows: a fragment corresponding to a 950 base pairs region upstream of the *SRKb* initiating methionine codon was amplified by PCR from an *SRKb*-containing plasmid using the forward primer: 5'- ATTGTTAGTTCTTTCATCAGTTCG-3' and the reverse primer: 5'- CACTTTCCCATGGCTCTCCTTC-3'. Each of the two reporter genes was introduced into plants of the Col-0 and C24 accessions (*SRKbpr*::*uidA*::*nos* was introduced by Titima Tantikanjana). In addition, plants of the Cvi-0, Hodja, Kas, Ler-1, Rld, and Sha accessions were transformed with *SRKbpr*::*uidA*::*nos*. All transformations were performed using the *Agrobacterium*-mediated floral dip method[11]. Transformants were selected on MS medium containing 40 ug/ul hygromycin for *SRKbpr*::*uidA* and 25 ug/ul for *Ats1pr*::*uidA*::*nos*. Homozygous plants carrying a single-transgene integration were identified by DNA gel blot analysis and screening on selective media. Histochemical assays for GUS activity were performed as described previously.[1] With the exception of the *AtS1pr::uidA::nos* transformation, all other transformations and selection of transformants were performed by Tiffany Crispell who also performed the histochemical GUS assays on the majority of transformants.

Stigmas and transmitting tracts were harvested separately for 30 buds. Total RNA was prepared and treated with RNase as above. Gene-specific primers were designed for GUS: 5'-TCCTACCGTACCTCGCATTACC-3' and 5'-GACAGCAGCAGTTTCATCAATCAC-3'. Actin primers were constructed previously[10]. RT-PCR was performed using the Superscript III One Step RT-PCR System (Invitrogen, Carlsbad, CA).

RESULTS

Preliminary mutant analysis

Col-0 plants transformed with the *SRK* and *SCR* alleles derived from the *Sb* haplotype of *A. lyrata* (hereafter *SRKb* and *SCRb*) exhibit transient SI, i.e. their stigmas are self-incompatible at the -1-bud stage but they lose the ability to inhibit self pollen in older flowers.[2] In a previous study, Col-0 plants carrying the *SRKb* and *SCRb* genes integrated at unlinked positions in the genome were mutagenized with ethyl methanesulphonate (EMS), and M2 plants derived from these mutagenized plants were screened by manual self-pollination of stigmas at the -1-bud stage[4]. This screen identified a mutant, *sc1,* which exhibited breakdown of self-incompatibility at the -1-bud stage of stigma development. Crosses between the *sc1* mutant and wild type Col-0 *SRKb* restored self-incompatibility in -1-stage stigmas, demonstrating that the *sc1* mutation is recessive. Additionally, when the *sc1* mutant was crossed to WT Col-0, the F1 plants became self-incompatible, showing that the mutant phenotype is not the result of a mutation in the *SRKb* transgene. DNA gel blot analysis of *sc1* mutant

plants showed that they harbored a single integration of the *SRKb* transgene.

Identification of *NRPD1a* as a modifier of self-incompatibility

A mapping population was generated by crossing a Col-0 *SRKb sc1* homozygote with a wild type plant of the Ler accession. A population of 1600 F2 plants derived from this cross was analyzed by pollinating -1-stage stigmas with *SCRb*-expressing pollen. Plants showing a consistent loss of self-incompatibility in these stigmas were genotyped using a combination of SSLP, CAPS, and SNP markers (Table A.1). A total of 125 individuals informative for mapping were found. In addition, 28 F3 plants were screened both for mapping purposes and to confirm F2 phenotypes. Based on this strategy, the mutation was mapped to a 165.5-kilobase (kb) region between markers F16P17b and F9N12c on chromosome 1 (Table A.1). Sequencing of the 39 genes within this region revealed a single amino-acid substitution at position 3188 in the genomic sequence of the *NRPD1a* gene. The cytosine-to-thymine transition at this position is a missense mutation resulting in a glutamine to asparagine substitution in the eighth exon of the gene. Because seven mutations (*nrpd1a*-1 to *nrpd1a*-7) had already been identified in *NRPD1a,* the *sc1* mutation was named *nrpd1a-8*.

NRPD1a is a plant-specific RNA-polymerase[12]. It is known to be involved in various silencing phenomena including generation of small RNAs (sRNAs)[12], methylation of DNA[13], and the spread and reception of the silencing signal[16,17]. Alignments of the predicted amino-acid

sequences of NRPD1a and RNA polymerase II (PolII) from yeast have identified the putative structural domains of NRPD1a[12]. Based on these alignments, the *nrpd1a-8* mutation is located in the region homologous to the binding site of transcription factor IIb (TFIIb) in yeast PolII (see Herr *et al*, 2005). Specifically, the mutation is located at amino acid 614 in the putative dock domain[12].

nrpd1a-8 mutants exhibit a loss of methylation

Previously-identified *nrpd1a* mutants have been shown to exhibit a loss of cytosine methylation[13]. To determine if the *nrpd1a-8* mutation affected methylation, *nrpd1a-8* homozygotes were assayed by examining the methylation status of the highly-methylated *AtSN1* retroelement (Figure 2.1)[13]. In this assay, genomic DNA is digested with the methylation-sensitive enzyme HaeIII prior to PCR amplification with *AtSN1*-specific primers. An amplification product is obtained if the DNA is methylated and therefore insensitive to HaeIII digestion. The unmethylated gene, At2g12990, contains no methylation sites and is used as a positive control. The *nrpd1a-8* mutants, exhibited loss of methylation manifested by the failure to amplify *AtSN1* fragments from HaeIII-digested, as illustrated in Figure 2.1.

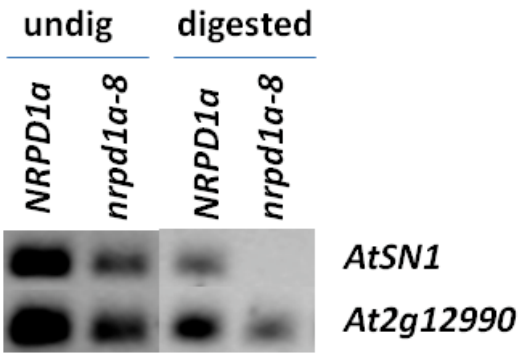

Figure 2.1. *nrpd1a-8* mutants exhibit a loss of methylation at the *AtSN1* repeat.

***nrpd1a-8* mutants exhibit a decrease in the levels of *SRKb* full-length transcripts in stigmas**

Sub-optimal levels of *SRKb* are known to lead to loss of SI[10]. To determine if the self-compatible phenotype of -1-stage stigmas in *nrpd1a-8* is due to reduced *SRKb* transcript, RNA gel blots were performed. The *SRKb* full-length ~3-kb transcript was approximately 2.2 fold more abundant in wild type transgenic plants than in *nrpd1a-8* mutants (Figure 2.2). In contrast, the ~1.6-kb *SRK* transcript, designated *eSRK*, which corresponds to the first exon of the gene, was approximately 2-fold more abundant in *nrpd1a* mutants than in wild type. Additionally, both *SRKb* and *eSRK* transcripts were elevated in the transmitting tract, petal, and stamens of *nrpd1a-8* mutants. No expression was observed in sepals in either mutant or wild type plants.

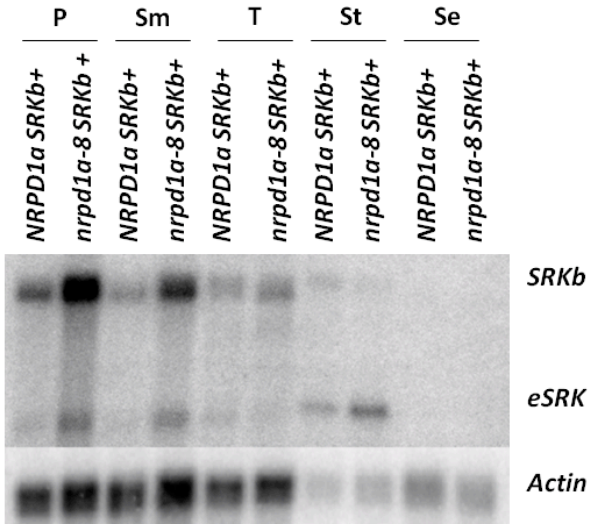

Figure 2.2. Variation in expression of *SRKb* and *eSRK* in *nrpd1a* mutants in different floral tissues. P: petal; Sm: stamen; T: transmitting tract; St: stamen; Se: sepal

RDR2 is required for self-incompatibility in Col-0 *SRKb-SCRb* plants

In addition to DNA methylation, *NRPD1a* is also known to play a role in production of natural-antisense (NAT) RNA. Other mutants in the NAT pathway had been analyzed previously and found to have no breakdown of SI, so they were not further analyzed here[4]. Instead, mutants in the methylation silencing pathway were analyzed to determine if they also caused breakdown of self-incompatibility in *SRKb-SCRb* plants. Strains containing T-DNA insertion mutations in *rdr2, ago4, dcl3, nrpd2a, drm1, drm2, drd1, cmt3,* and *nrpd1b* were each crossed to a Col-0 plant carrying one insert of a linked *SRKb-SCRb* gene pair. For two insertion lines, *nrpd2a* and *dcl3*, genotyping produced ambiguous results, and these lines may not have had the T-DNA insertion in the expected genomic location. For all lines, F2 progenies that contained the *SRKb-SCRb* transgenes and were homozygous for the insertional mutation were analyzed by pollination assays of stigmas at the -1 bud stage. Of the mutants tested, only *rdr2* mutants had a phenotype similar to that of *nrpd1a-8* mutants. All other mutants were strongly SI at the -1-bud stage.

Methylation mutants exhibit reduced *SRKb* expression levels in stigmas

The different phenotypes produced by *rdr2* and other mutations in the NAT and DNA methylation pathways might be due to differences in their effect on *SRKb* transcript levels. To address this issue, real-

time RT-PCR was performed to compare *SRKb* expression levels in the self-compatible T-DNA insertion lines and a subset of self-incompatible T-DNA insertion lines. Additionally, to ensure comparable *SRKb* expression, real-time RT-PCR was also performed on two lines carrying independent *SRKb* transgene integrations: the original line used for mutagenesis, which had a single *SRKb* integration, and the line used in crosses to the methylation mutants, which contained a single integration of linked *SRKb* and *SCRb* transgenes.

cDNA was synthesized from stigma and transmitting tract RNA isolated from F3 plants that were homozygous for both *SRKb-SCRb* and the T-DNA insertional mutation in silencing-pathway genes, specifically, the self-compatible *nrpd1a SRKb-SCRb* and *rdr2 SRKb-SCRb* plants and the self-incompatible *nrpd1b SRKb-SCRb* and *ago4 SRKb-SCRb* plants. Real-time RT-PCR was performed using intron-spanning primers for *SRKb* and a ubiquitin control gene. All methylation mutants tested were found to have a significant decrease in expression of full-length *SRKb* transcripts in the stigma, and there was no significant expression difference between the two *NRPD1a* transgenic lines analyzed (Figure 2.3). The self-incompatible -1-bud stigmas of *nrpd1b SRKb-SCRb* and *ago4 SRKb-SCRb* plants tended to have slightly higher expression than the self-compatible -1-bud stigmas of *nrpd1a SRKb-SCRb* and *rdr2 SRKb-SCRb* plants. However,

this difference was not significant, at least not in comparisons of *ago4 SRKb-SCRb* with the two self-compatible lines, as shown by the overlapping error bars (Figure 2.3).

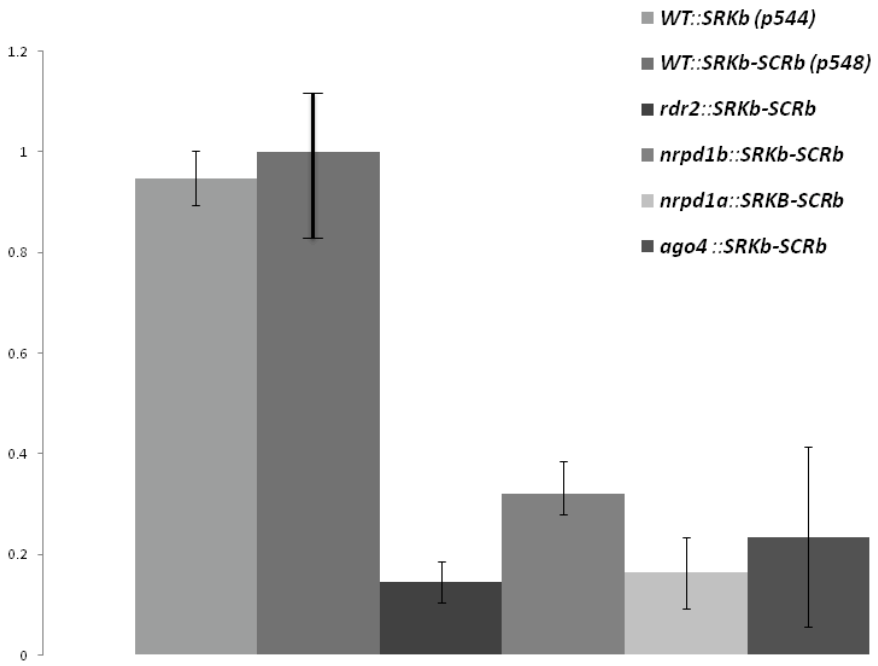

Figure 2.3. Average-fold change of *SRKb* transcripts in the stigmas of methylation mutants homozygous for *SRKb-SCRb* relative to WT *SRKb-SCRb* and WT *SRKb*.

Loss of *NRPD1a* function is not sufficient for the breakdown of SI in inter-ecotypic hybrids.

Ecotypic variation in expression of SI is known to occur between various strains of *A. thaliana*. Unlike Col-0 *SRKb-SCRb* plants, which exhibit transient SI, *SRKb-SCRb* transformants of a few accessions (e.g. C24 and Sha) exhibit strong and developmentally-stable SI[2, 14]. To determine if *nrpd1a-8* could cause a breakdown of SI in these accessions, an *nrpd1a-8* mutant was crossed to C24 *SRKb-SCRb* and to Sha *SRKb-SCRb* plants. The F1 progeny of the *nrpd1a-8* x Sha *SRKb-SCRb* cross were self-incompatible at the -1 bud stage but had near full seed set, indicating that SI broke down at later stages of stigma development. In contrast, the F1 progeny of the *nrpd1a-8* x C24 *SRKb-SCRb* cross were fully self-incompatible at all stages of stigma development. In 4 F2 families analyzed for each cross, F2 plants were recovered that were self-compatible despite harboring the *SRKb-SCRb* transgenes (3/36 in the Sha cross and 5/45 in the C24 cross). All of these plants were homozygous for the *nrpd1a-8* mutation, and their F3 progenies were uniformly self-compatible. However, some F2 *nrpd1a-8* homozygotes were self-incompatible (12/36 for the Sha cross and 12/45 in the C24 cross). The phenotypic ratio of *nrpd1a* plants is 4 SI:1 SC for Sha and 2.4 SI:1 SC for C24. Since the C24 *SRKb-SCRb* parent had a single integration at a different genomic location than *nrpd1a-8* and the Sha *SRKb-SCRb* parent had 5 integrations, copy number of the *SRKb-SCRb* transgene was determined by DNA gel blot analysis. There was no correlation between copy number of *SRKb* (as determined by DNA gel blot

analysis) and SI phenotype.

Modifier loci are known to contribute to differences in the SI phenotype of *SRKb-SCRb* transformants of different accessions[3, 10]. Markers for two modifiers whose genomic locations are known, one on chromosome 3[14] and another on chromosome 4[10], were used to determine if they might be responsible for the variable effect of *nrpd1a-8* on SI. No association was found between the phenotype and either of these markers. Additional markers were tested throughout the genome, and loose linkage of a putative modifier was detected to marker NGA139 on chromosome 5 (Table 2.1).

Table 2.1. Preliminary mapping of Col-0 *nrpd1a*-8 x C24 *SRKb-SCRb* F2 individuals. All plants listed were *nrpd1a*-8 homozygotes.

F2 Family	**pollination test 1**	**pollination test 2**	***SRKb***	**NGA139**	**CIW9**
1-7	+++	+++	+	H	H
1-8	+++	+++	+	Col	Col
3-12	+	20	+	Col	H
6-3	+	++	+	Col	Col
1-1	0	1	+	H	H
2-3	0	3	+	C24	-
2-6	0	0	+	H	H
3-11	3	0	+	H	Col

NRPD1a is not necessary for SI in a stably self-incompatible accession.

An *nrpd1a* mutant in the C24 background was obtained and crossed to C24 carrying the *SRKb* and *SCRb* transgenes. This particular version of *SRKb* is N-terminally tagged with YFP and lacks introns rendering it unable to produce *eSRK* transcripts. Three separate F2 families, totaling 88 plants, were genotyped and phenotyped. None of the 6 *nrpd1a* homozygotes that harbored the *SRKb-SCRb* transgenes exhibited a breakdown of self-incompatibility, and they set a low amount of seed similar to *SRKb-SCRb* plants lacking the *nrpd1a* mutation.

Since the *SRKb* transgene used in the above cross produces no *eSRK*, it was of interest to ensure that the absence of *eSRK* was not a factor in the unexpected phenotype. The C24 *nrpd1a* mutant was crossed to a C24 *SRKb-SCRb* plant that carried an unmodified *SRKb* gene capable of producing *eSRK* transcripts. All 16 F2 plants that had the *SRKb-SCRb* transgenes, including 3 *nrpd1a* homozygotes, were self-incompatible as demonstrated by little to no seed set.

***SRKb* promoter activity varies among *A. thaliana* accessions**

The analysis of the *nrpd1a* and other silencing-pathway mutations indicated that the regulation of *SRKb* gene expression is complex. Furthermore, a previous study had shown that the levels of *SRKb* transcripts are differentially affected by modifiers in different *A. thaliana* accessions[3, 10]. To determine if regulation of *SRKb* at the promoter level is responsible for the variation in expression of self-

incompatibility observed in different accessions of *A. thaliana*, an *SRKb*::*uidA* construct was introduced into plants belonging to several accessions (Table 2.2), and primary transformants were assayed for GUS staining in floral tissues (Figure 2.4). All accessions that expressed stable SI exhibited GUS activity in papillar cells and the majority of those also exhibited GUS activity in the transmitting tract (Table 2.2), similar to the activity previously reported for other *SRK* promoters[1]. In contrast, the Col-0, Ler, and Rld accessions, all of which exhibit transient SI, showed little or no staining in papillar cells and transmitting tract (Table 2.2), and in some cases showed aberrant GUS staining in other floral tissues. For example, Col-0 exhibited no staining in the stigma or transmitting tract, but often had staining in stamens. The absence of GUS staining in papillar cells of Col-0 plants was not due to general misregulation of promoters expressed in papillar cells, but was specific to the *SRKb* promoter. As shown in Figure 2.4 (panel D), transformation of Col-0 with an *AtS1 pr::uidA::nos* construct produced 6 transformants, all of which exhibited strong GUS staining in papillar cells, as expected for the papillar cell-specific *AtS1* promoter[15].

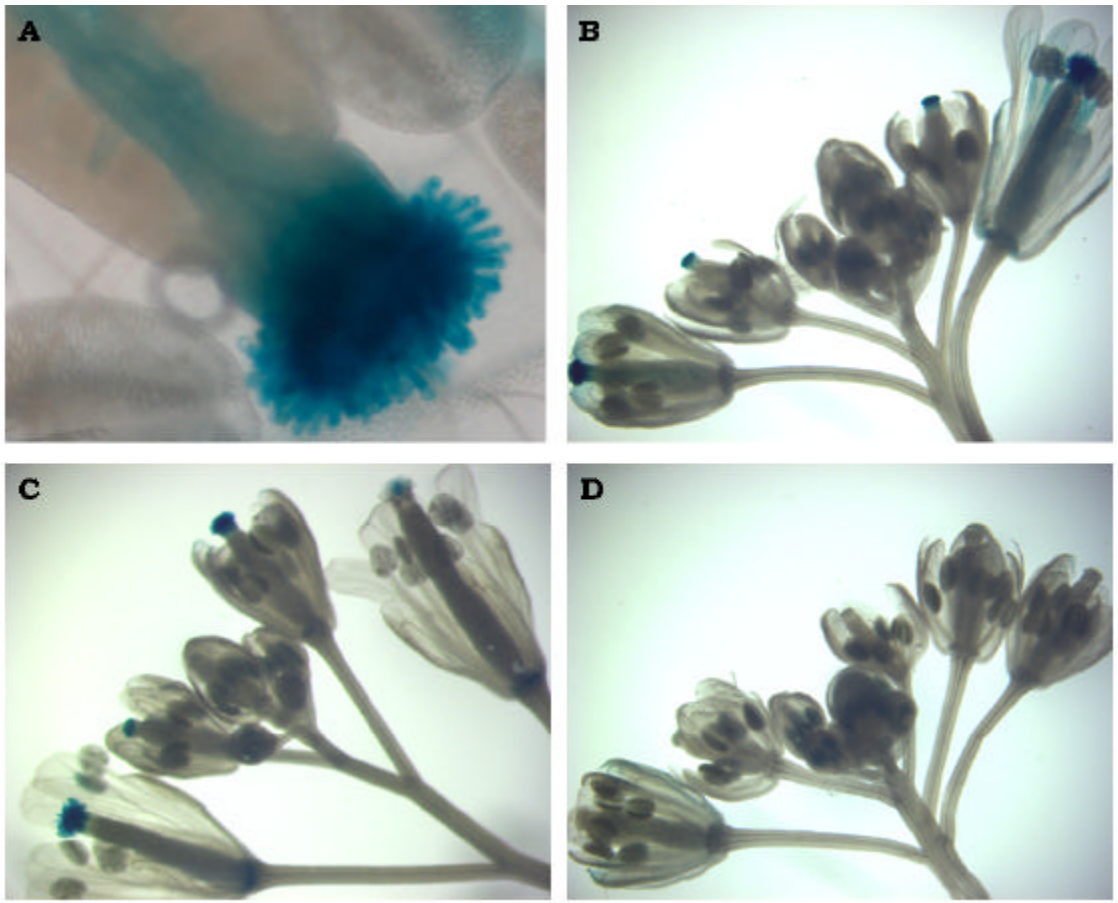

Figure 2.4. GUS staining (A) C24 *SRKb::uidA* pistil at stage +3. (B) C24 *SRKb::uidA* inflorescence. (C) Col *AtS1::uidA* inflorescence. (D) Col *SRKb::uidA* inflorescence.

The failure of GUS staining to detect *SRKb* promoter activity in the pistils of Col-0, Ler, and Rld *SRKb::uidA* transformants was surprising, given the fact that *SRK-SCR* transformants of these accessions express SI in -1 bud-stage stigmas. To investigate this discrepancy further, RT-PCR was used to compare the levels of GUS transcripts in the stigmas of Col-0 *SRKb::uidA* and Rld *SRKb::uidA* plants to that of C24 *SRKb::uidA* and Kas *SRKb:uidA* plants, which exhibit strong GUS staining. Figure 2.5 shows that GUS transcripts were present in Col-0 *SRKb::uidA* stigmas and Rld *SRKb::uidA* stigmas, albeit at a lower level than in C24. Expression in C24 is approximately 2.2 times higher than in Col-0 and about 1.4 times stronger than in Rld. The transcript was found to be expressed 1.7 times higher in Kas than in Col-0. The failure to observe GUS staining in Col-0 *SRKb::uidA* stigmas is therefore attributed to the relatively low sensitivity of histochemical GUS staining. Surprisingly however, there was no appreciable difference in GUS transcript levels in the transmitting tracts of Col-0 and C24, despite differences in GUS staining.

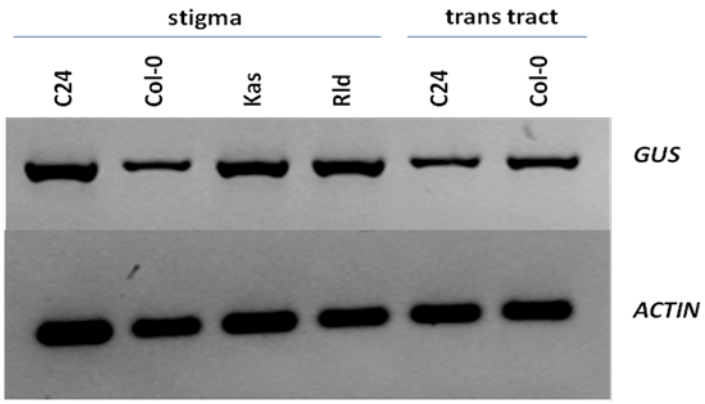

Figure 2.5. Expression of *SRKb::uidA* in the stigma and transmitting tract of various accessions of *A. thaliana*.

Table 2.2. GUS staining observed in *A. thaliana* accessions transformed with *SRKb::uidA*.

Accession	Total no. of T1 transformants	No. T1 transformants exhibiting GUS staining in stigma and transmitting tract	% GUS positive	Phenotype of *SRKb-SCRb* transformants
Col-0	21	0	0	transient SI
C24	54 [a]	18 [a]	33[a]	stable SI
Cvi	3	2	67	stable SI
Hodja	15	6	40	stable SI
Kas	20	12	60	stable SI
Ler	1	0	0	transient SI
Rld	9	2	22	transient SI
Sha	16	4	25	stable SI

[a] Data from Titima Tantikanjana

Search for putative transcription factor of NRPD1a

Since the point mutation in *nrpd1a-8* is in a predicted transcription factor-binding site, an effort was made to find the putative transcription factor. Candidates were chosen based on sequence similarity to *TFIIb*, similar expression to *NRPD1a*, and T-DNA availability (Table 2.3). T-DNA lines for *Arabidopsis TFIIb* homologs were obtained from the ABRC, screened for homozygosity using gene-specific primers (Table A.3), and homozygotes were crossed to Col-0 *SRKb-SCRb*. The known transcription factor for Pol II in *Arabidopsis* was excluded from this study. Gene At2g45100 was not analyzed further since it did not appear to contain a T-DNA insertion in the gene of interest. F2 individuals were screened for phenotype by pollination test. None of the individuals homozygous for the T-DNA insertions and harboring the *SRKb-SCRb* transgene were found to exhibit a breakdown of SI (15/35 for At3g10330 and 6/18 for At3g09360)

Table 2.3. Candidate transcription factors of NRPD1a based on sequence similarity to *TFIIB*.

Locus	SALK line tested	Other	Description
AT2G01280	-	expressed in stigma	MEE65 (maternal effect embryo arrest 65)
AT2G02955	-	expressed in stigma	MEE12 (maternal effect embryo arrest 12)
AT2G45100	SALK_146224	-	similar to RNA polymerase II transcription factor
AT3G09360	SALK_072770	similar expression to NRPD1a	similar to RNA polymerase II transcription factor
AT3G10330	SALK_086510	similar expression to NRPD1a	identical to Transcription initiation factor IIB-2
AT3G29380	-	expressed in stigma,	transcription factor IIB (TFIIB) family protein

Table 2.3 (continued)

AT3G57370	-	expressed in stigma, highly in pollen	transcription factor IIB (TFIIB) family protein
AT4G10680	-	-	transcription factor IIB (TFIIB) family protein
AT4G19550	-	expressed in stigma, highly in pollen	transcription regulator/ transcriptional activator/ zinc ion binding
AT4G35540	-	expressed in stigma, highly in pollen	zinc ion binding
AT4G36650	-	expressed in stigma, highly in pollen	transcription factor IIB (TFIIB) family protein;
AT5G39230	-	expressed in stigma, highly in pollen	transcription initiation factor-related

DISCUSSION

The results described in this chapter show that the *nrpd1a-8* mutation causes loss of SI in stigmas of the *A. thaliana* accession Col-0. Because *NRPD1a* functions in transgene silencing, disruption of this gene is expected to cause increased expression of the *SRKb* transgene. Unexpectedly, however, the stigmas of *nrpd1a SRKb* plants exhibited reduced *SRKb* transcript levels relative to WT *SRKb* plants. Interestingly, this reduction does not appear to be responsible for the loss of SI in *nrpd1a* mutants. Indeed, comparable reductions in *SRKb* transcripts were observed in *ago4*, and *nrpd1b*, mutants, all of which retain SI. Additionally, a similar reduction in *SRKb* transcript levels is observed in *rdr6* mutants that have enhanced SI[4]. Although reduced *SRKb* transcript levels may be attributed to disrupted methylation, analysis of two additional methylation mutants, *ago4* and *nrdp1b*, indicates that methylation status is not directly correlated with SI phenotype. All of these methylation mutants showed a loss of genome methylation based on analysis of the normally methylated *AtSN1* locus, but only *nrpd1a* and *rdr2* mutants exhibited a breakdown of SI. Surprisingly, *nrpd2a SRKb-SCRb* plants retained SI, even though NRPD2a, like NRPD1a, is a subunit of the PolIV complex[12], and a loss-of-function mutation in this gene should produce a similar phenotype to *nrpd1a* mutations. It would be worthwhile to determine conclusively if the T-DNA line test is indeed a null allele or perhaps test different T-DNA insertional mutants this gene.

Additional evidence for the lack of correlation between methylation status and SI is provided by the *nrpd1a-8* X C24 *SRKb-*

SCRb cross, in which all *nrpd1a-8* mutant progenies exhibited a loss of methylation, but some retained SI. One interpretation of these results is that *nrpd1a* and *rdr2* act on an as-yet unidentified locus involved in SI. This possibility is supported by the segregation ratios of approximately 3 SI:1 SC in *nrpd1a* mutants identified in crosses between *nrpd1a* and C24 *SRKb-SCR*. Furthermore, preliminary linkage analysis from F2 progeny of this cross implicates a locus on chromosome 5 in the breakdown of self-incompatibility in Col-0 *nrpd1a* mutants.

Analysis of *nrpd1a* and *rdr2* mutations revealed differences in the regulation of the *SRKb* gene in different floral tissues of Col-0 plants. In contrast to the stigmas of Col-0 *SRKb* plants homozygous for *nrpd1a-8* and other methylation mutants, which expressed reduced levels of full-length *SRKb* (but not *eSRKb*) transcripts, the petals, stamens, and transmitting tracts of these plants exhibited increased levels of both *SRKb* and *eSRKb* transcripts, in keeping with the abrogation of transgene silencing expected in *nrpd1a* loss-of-function mutants. The basis of this difference in regulation of the *SRKb* transgene and its two transcript species in different floral tissues is not understood. Undoubtedly, the unexpected reduction in full-length *SRKb* transcripts in stigmas is due to stigma-specific factors that might involve splicing problems, use of an alternative promoter, or silencing directed at the 3' end of *SRKb*.

NRPD1a seems to have no role in the regulation of *SRKb* in C24, a stably SI accession. By analyzing *nrpd1a* mutants in the C24 background crossed to C24 *SRKb-SCRb*, we were unable to recover any

SC individuals. This data supports the conclusion that the *nrpd1a* mutation results in SC plants due to factors found in the Col-0 background. Expression of *SRKb* should be analyzed in the C24 *nrpd1a* background to determine if there is a change comparable to that of the Col-0 *nrpd1a* mutant.

SRKb regulation varies by accession and the promoter appears to perform more effectively in some accessions of *A. thaliana* than others. In general, accessions exhibiting stable SI upon transformation with *SRKb-SCRb* show stronger GUS staining than accessions with transient SI. It would be worthwhile to quantify the expression of GUS across all the ecotypes more accurately using real-time RT-PCR. Variation in transmitting tract expression between the two accessions was not observable in our results, but it was evident based on visual observations that there was more GUS staining in the transmitting tracts of C24 as opposed to Col-0. A more reliable quantitative expression analysis would help to clear this issue. The variation amongst the accessions may be due to variation of certain transcription factors that are needed for proper *SRKb* expression.

Overall, it seems that there are many levels of *SRKb* regulation that vary substantially by ecotype, some of which likely involve epigenetic factors and others that may be involved in promoter interactions. Epigenetics appear to play a role in *SRKb* regulation in Col-0, since it is misregulated in methylation mutants. As suggested above, this misregulation could be the consequence of various factors. It is possible that NRPD1a is necessary for correct splicing regulation of *SRKb* in the stigma. It is known that splicing can be tissue

specific[16] and can be altered in the offspring of interspecific hybrids[17] . Mis-splicing also appears to occur in interspecific hybrids of the Brassicaceae[18] and perhaps this is due to variation and incompatibilities in methylation. Disruption of methylation patterns also typically occurs in the offspring of interspecific crosses[19] and even crosses within accessions of *A. thaliana* to a small degree[20]. It is also conceivable that altered methylation allows *SRKb* transcript to accumulate at an above average rate, which in turn alters splicing. A similar phenomena is observed in *FCA*, a gene controlling flowering time in plants[21]. Another possibility is that changes in methylation lead to the usage of an alternative promoter leading to altered expression. Alternative promoter usage has been found to occur in the mammal *Shc* locus due to demethylation[22]. Lastly, methylation changes could induce silencing directed at the 3' end of *SRKb*. A similar phenomenon occurs with *FLC*[23]. The silencing of the 3' end would need to occur through a pathway that is activated in the absence of functional *NRPD1a*. It is known that many *rdr2*-independent silencing RNAs are produced from the 3' end of □*SRK* in Col-0. It is possible that a change in methylation results in silencing or even splicing that down regulates the *SRKb* transcript. In addition to epigenetic regulation within an accession, variability in promoter efficiency also appears to occur between accessions, likely due to the activity of modifier loci. Modifiers of SI are known to vary between accessions of *A. thaliana*[3] and a characterized example of a modifier is *PUB8*. PUB8 is a modifier found to vary between the C24 and the Rld accessions and is known to be necessary for sufficient expression of

SRKb^{10} at later stages of stigma development. Since SI is believed to have been lost more than once along the *A. thaliana* lineage, elucidation of SI modifier loci of SI may be useful in studying the evolution of self-compatibility in this species.

Acknowledgements: Titima for *SRKb pr::uidA* construct, Col-0 and C24 transformants, EMS mutagenesis and identification of the *sc1* mutation; Tiffany Crispell for plant transformation, GUS assays, and plant care; and Craig Pikaard for C24 *nrpd1a* seed.

REFERENCES

1. Kusaba, M., Dwyer, K., Hendershot, J., Vrebalov, J., Nasrallah, J. B. and Nasrallah, M. E. Self-Incompatibility in the Genus *Arabidopsis*: Characterization of the S Locus in the Outcrossing *A. lyrata* and Its Autogamous Relative *A. thaliana*. Plant Cell 13, 627-643 (2001).
2. Nasrallah, M. E., Liu, P. and Nasrallah, J. B. Generation of Self-Incompatible *Arabidopsis thaliana* by Transfer of Two S Locus Genes from *A. lyrata*. Science 297, 247-249 (2002).
3. Nasrallah, M. E., Liu, P., Sherman-Broyles, S., Boggs., N. A., and Nasrallah, J. B. Natural variation in expression of self-incompatibility in *Arabidopsis thaliana*: Implications for the evolution of selfing. PNAS 101, 16070-16074 (2004).
4. Tantikanjana, T., Rizvi, N., Nasrallah, M. E. and Nasrallah, J. B. A Dual Role for the S-Locus Receptor Kinase in Self-Incompatibility and Pistil Development Revealed by an *Arabidopsis rdr6* Mutation. Plant Cell 21, 2642-2654 (2009).
5. Smyth, D. R., Bowman, J. L. and Meyerowitz, E. M. Early Flower Development in *Arabidopsis*. Plant Cell 2, 755-767 (1990).
6. Kho, Y. O. B., J. Observing pollen tubes by means of fluorescence. Euphytica 17, 298–302 (1968).
7. Doyle, J. J. a. J. L. D. Isolation of plant DNA from fresh tissue. Focus 12, 13-15 (1990).
8. The Arabidopsis Information Resource, www.arabidopsis.org

9. Hamilton, A., Voinnet, O., Chappell, L. and Baulcombe, D. Two classes of short interfering RNA in RNA silencing. EMBO J 21, 4671-4679 (2002).

10. Liu, P., Sherman-Broyles, S., Nasrallah, M. E. and Nasrallah, J. B. A Cryptic Modifier Causing Transient Self-Incompatibility in *Arabidopsis thaliana*. 17, 734 (2007).

11. Steven, J. C. and Andrew, F. B. Floral dip: a simplified method for Agrobacterium-mediated transformation of *Arabidopsis thaliana*. The Plant Journal 16, 735-743 (1998).

12. Herr, A. J., Jensen, M. B., Dalmay, T. and Baulcombe, D. C. RNA Polymerase IV Directs Silencing of Endogenous DNA. Science 308, 118-120 (2005).

13. Onodera, Y., Haag, J. R., Ream, T., Nunes, P. C., Pontes, O. and Pikaard, C. S. Plant Nuclear RNA Polymerase IV Mediates siRNA and DNA Methylation-Dependent Heterochromatin Formation. 120, 613 (2005).

14. Boggs, N. A., Nasrallah, J. B., and Nasrallah, M. E. Independent S-Locus Mutations Caused Self-Fertility in *Arabidopsis thaliana*. PLoS Genet 5 (2009).

15. Dwyer, K. G., Kandasamy, M. K., Mahosky, D. I., Acciai, J., Kudish, B. I., Miller, J. E., Nasrallah, M. E. and Nasrallah, J. B. A Superfamily of S Locus-Related Sequences in *Arabidopsis*: Diverse Structures and Expression Patterns. Plant Cell 6, 1829-1843 (1994).

16. Saiprasad Goud, P., Gul Shad, A. and Anireddy, S. N. R. Alternative splicing of pre-mRNAs of *Arabidopsis* serine/arginine-rich proteins: regulation by hormones and stresses. The Plant Journal 49, 1091-1107 (2007).

17. Scascitelli, M., Cognet, M. and Adams, K. L. An Interspecific Plant Hybrid Shows Novel Changes in Parental Splice Forms of Genes for Splicing Factors. Genetics, genetics.109.112557.

18. Nasrallah, J. B., Liu, P., Sherman-Broyles, S., Schmidt, R. and Nasrallah, M. E. Epigenetic Mechanisms for Breakdown of Self-Incompatibility in Interspecific Hybrids. Genetics 175, 1965-1973 (2007).

19. Jin, H., Hu, W., Wei, Z., Wan, L., Li, G., Tan, G., Zhu, L. and He, G. Alterations in cytosine methylation and species-specific transcription induced by interspecific hybridization between *Oryza sativa* and *O. officinalis*. TAG Theoretical and Applied Genetics 117, 1271 (2008).

20. Banaei Moghaddam, A., Fuchs, J., Czauderna, T., Houben, A. and Mette, M. Intraspecific hybrids of *Arabidopsis thaliana* revealed no gross alterations in endopolyploidy, DNA methylation, histone modifications and transcript levels. TAG Theoretical and Applied Genetics 120, 215.

21. Macknight, R., Bancroft, I., Page, T., Lister, C., Schmidt, R., Love, K., Westphal, L., Murphy, G., Sherson, S., Cobbett, C. and Dean, C. FCA, a Gene Controlling Flowering Time in *Arabidopsis*, Encodes a Protein Containing RNA-Binding Domains. 89, 737 (1997).

22. Ventura, A., Luzi, L., Pacini, S., Baldari, C. T. and Pelicci, P. G. The p66Shc Longevity Gene Is Silenced through Epigenetic Modifications of an Alternative Promoter. Journal of Biological Chemistry 277, 22370-22376 (2002).

23. Liu, F., Marquardt, S., Lister, C., Swiezewski, S. and Dean, C. Targeted 3' Processing of Antisense Transcripts Triggers Arabidopsis *FLC* Chromatin Silencing. Science 327, 94-97.

CHAPTER THREE:

RAPID EVOLUTION OF REPRODUCTIVE GENES IN *ARABIDOPSIS*

INTRODUCTION

The rapid evolution of genes involved in adaptive responses has recently been the focus of intense interest. The availability of sequence data from many different species has facilitated inter-genomic comparisons, which have led to the identification of rapidly evolving genes expressed in various tissue types in many different organisms. Rapid evolution was found to be associated with processes such as pathogen response[1, 2], environmental adaptation[3], and reproductive interactions. In particular, genes involved in sexual reproduction tend to be more divergent than genes expressed in other tissues[4, 5]. Rapidly-evolving reproductive genes have been found in many organisms including primates[6, 7], *Chlamydomonas reinhardtii*[8], marine invertebrates[9], *Drosophila*[10, 11], and *Brassica*[12] (see Swanson and Vacquier 2002a[13] and Chapter 1 for a review of rapidly-evolving reproductive genes). The identification and analysis of such genes may provide a better understanding of how interspecific reproductive incompatibility occurs.

In plants, the tissues directly involved in reproduction, and therefore most likely to function in mate selectivity, are the pollen grains, pollen tubes, and cells of the pistil that line the path of pollen tube growth. Only a few rapidly-evolving genes that are expressed in these tissues have been identified, namely the *S*-locus genes, which

function in self-incompatibility[14, 15,16, 17], and pollen coat proteins[18] (see Chapter 1 for a review). An evolutionary analysis of ESTs expressed in *Arabidopsis lyrata* inflorescences and their putative *Arabidopsis thaliana* orthologs identified 14 rapidly evolving genes[19]. However, most of these genes are likely not involved in reproduction, since the study used whole inflorescences, and the stigma, transmitting tract, and pollen would only represent a small portion of this tissue. The small number of rapidly evolving reproductive genes identified in plants may be partly attributed to the difficulty of isolating the subset of cells that are involved in reproduction. Furthermore, until recently, the sequence data necessary to perform the analyses required to identify such genes were not available. Fortunately, these problems have been largely overcome. An abundance of microarray data is now available that pinpoints gene expression in specific reproductive tissues. In addition, various sequencing projects have produced a wealth of data concerning variation among closely-related species. Recently, a study that utilized the newly sequenced *A. lyrata* genome, found pollen-specific genes to be the most rapidly evolving in comparison to whole flowers, seed, leaf, root, and stem-expressed genes[20], so it seems likely that genes involved in plant reproduction may evolve rapidly in comparison to genes expressed in nonreproductive tissues.

Here, we report on the results of a screen for candidate rapidly evolving reproductive genes using sequence information from two closely-related *Arabidopsis* species, *A. thaliana* and *A. lyrata,* which are estimated to have diverged from a common ancestor approximately 5

m.y.a.[21]. These two species are particularly suited for analysis of reproductive gene evolution because their complete genome sequences are available[22, 23]. Additionally, they exhibit the two different modes of mating that operate in the Brassicaceae family, with *A. lyrata* representing the ancestral self-incompatible outcrossing mating system and *A. thaliana* the derived self-fertilizing mating system. Therefore, identifying rapidly evolving genes by comparison of these two species will set the stage for future analysis of the levels and patterns of variation in these genes in self-fertilizing and outcrossing taxa.

For our study, the literature was scanned for *A. thaliana* data sets containing genes expressed in the stigma, transmitting tract, ovary, pollen, or pollen tube. These data sets were then pooled to perform a search for orthologous genes in *A. lyrata*. Candidate rapidly evolving genes were identified and a subset of these genes was selected for further analysis based on predicted function, expression data, presence of a signal peptide or transmembrane domain, and preliminary evidence of sequence differentiation between the species. Within-species sequence variation in these genes was investigated by isolating and sequencing alleles from various geographical accessions of *A. thaliana,* and further statistical analyses of positive selection were performed. Several rapidly evolving genes were identified which may play function in species-specific reproductive function.

MATERIALS AND METHODS

Identification of candidate rapidly-evolving genes

Exon sequence from *A. thaliana* accession Columbia for all genes in the reproductive tissue data sets was obtained courtesy of The *Arabidopsis* Information Resource (TAIR) curators and can be found at ftp://ftp.*Arabidopsis*.org/home/tair/Sequences/blast_data sets/TAIR9_blastsets. The TAIR8 release was used for the initial screen and the TAIR9 release was used for subsequent analysis. *A. lyrata* sequence was obtained from the DOE Joint Genome Institute (JGI) whole genome shotgun assembly with 8.3 x coverage. The sequence was pooled and locally queried against the *A. lyrata* sequence database using the BLASTn program[24]. Alignments were kept in frame with the reference *A. thaliana* Col-0 sequence. The codeml package of PAML[25] was used to estimate omega (ω) for each alignment. For genes with predicted ω>0.5, two additional models were used with ω fixed at 0.5 and 1.0 respectively. Likelihood ratio tests were performed to compare the free ratio model to the fixed-ratio model to test for a significant rate variation between species. Genes with a dS value greater than 0.25 were eliminated from the analysis since these were likely not orthologous alignments. Transposons and pseudogenes were also eliminated. All genes with a predicted ω >0.5 were selected for further analysis. The *A. lyrata* sequence for each gene was queried against the *A. thaliana* Col-0 genome using BLASTn to ensure orthology.

McDonald-Kreitman tests[26] as implemented in DNAsp version

5.10[27] were used to assess preliminary information about polymorphism within 20 *A. thaliana* accessions analyzed in another study[28]. All genes showing a low ratio of nonsynonymous to synonymous polymorphisms within *A. thaliana* were manually re-aligned using the entire coding sequence and all previously described statistics were recalculated based on these alignments. Gene Ontology (GO) classifications were determined using the TAIR server. GO term enrichment was calculated using the AmiGO server[28]. Signal P 3.0[29, 30] was used to determine which of the genes have a predicted signal peptide. Putative transmembrane domains were predicted using TMHMM[31]. Expression information was obtained from microarray data from other studies and from Genevestigator[32].

Amplification and sequencing of alleles

Seed from 10 geographical accessions, including Bay-0 (CS22633), Br-0 (CS22628), C24 (CS22620), Cvi-0 (CS22614), Est-1 (CS22629), Ler-1 (CS22618), Lov-0 (CS22575), Lz-0 (CS22615), Tamm-2 (CS22604), and Ts-1 (CS22647), was obtained from the Arabidopsis BioResource Center (ABRC) and plated on Murashige and Skoog (MS) media. Seed was vernalized for 3 days to synchronize germination. Plants were grown under long-day conditions (16 hour days/8 hour nights) at a temperature of 20°C in a controlled-environment chamber. Genomic DNA was prepared from leaf tissue using the CTAB method[33].

Sequencing was performed for 12 candidate positively-selected genes to obtain full gene coverage in alleles from all accessions analyzed. Gene-specific PCR primers were designed from the reference

Col-0 genome sequence to amplify overlapping regions of approximately 800 base pairs or less for each candidate gene and to anneal to exonic regions (Table A.4). The primers were used to amplify genomic DNA from each accession, and the resulting PCR products were purified using Qiagen Gel Extraction Kit (Qiagen, Valencia, CA) and sequenced at the Cornell Bioresource Center (Cornell University). Sequence for each gene was assembled into contiguous sequences using Sequencher 4.9 (Gene Codes Corporation, Ann Arbor, MI). At least two overlapping sequences were obtained for each nucleotide by sequencing both strands. Putative single nucleotide polymorphisms were manually checked by inspection of the chromatogram file. Whenever ambiguous base changes were observed, the segments were re-amplified and sequenced.

Molecular population genetics analysis

Sequences were aligned using Clustal W[34] and exons were predicted by using alignments to *A. thaliana* coding sequence. Bioedit[35] was used to manually edit the alignments.

DNAsp version 5.10[27] was used to calculate all summary statistics as well as to perform site frequency spectrum tests. These tests include Fu and Li's D[36], Fu and Li's F[36], Fay and Wu's H[37], and Tajima's D[38]. Distributions from a previous study were used to determine the significance of test results[39]. DNAsp was also used to perform McDonald-Kreitman tests[26]. The Neutrality Index (NI) was calculated as follows: (no. of polymorphic replacement sites/no. of fixed replacement sites)/(no of polymorphic silent sites/no. of fixed silent sites)[40].

DNA gel blot analyis

Genomic DNA from the Sha (CS22652), Ts-1, Ler-1, and Col-0 accessions was extracted using the CTAB method[41] and was digested overnight with *Eco*RI. Transfers were performed under alkaline conditions overnight onto Hybond N+ (GE Healthcare, Piscataway, NJ) membrane. Membranes were hybridized overnight with a radio-labeled probe specific either to gene At3g48231 or At3g57840. PCR primer pairs described below (Table A.4) were used to generate probe template from Col-0 genomic DNA. These PCR products were gel purified using the Qiagen Gel Extraction Kit (Hilden, Germany) then labeled using the Roche Random Priming Kit (Basel, Switzerland).

RESULTS

Sources of *A. thaliana* reproductive tissue-specific datasets

As a starting point in our search for rapidly-evolving reproductive genes, we used data from published studies of

reproductive tissues of the reference *A. thaliana* Col-0 accession. Pistil-expressed genes were represented by three data sets obtained by genome-wide transcriptional profiling. Two of these included genes expressed specifically in the stigma epidermis and the transmitting tract[42], an expression pattern that is suggestive of a role in pollen germination or tube growth. The third pistil data set comprised genes that were found to be differentially expressed between wild type ovules, *determinate infertile1* (*dif1*) ovules that lack an embryo sac, and/or *myb98* ovules that have aberrant synergid cells[43]. Because the embryo sac and synergids are important for pollen tube guidance and ultimately fertilization (see Chapter 1 for a review), some of these genes may function in intercellular communication between pollen tube and ovule.

Similarly, genes that function in pollen were represented by four data sets. Two data sets represent the transcriptome of the haploid pollen genome: of these, one set consisted of genes expressed in pollen grains[44], and another set consisted of genes expressed in pollen tubes that have interacted with the pistil or pollen tubes grown *in vitro* in the absence of pistil tissue[25]. Because many molecules that function in pollen-pistil interactions are located on the outer pollen coat and are encoded by the diploid genome of the tapetal cells, data obtained in two proteomic studies of pollen were used to supplement the pollen and pollen tube transcriptional data sets[18,45.]

Together, the data sets provide a thorough coverage of gene expression in all tissues involved in plant reproduction. This collective data set contained approximately 3,500 genes.

Initial screening for candidate rapidly-evolving genes

Positive selection does not often occur over all sites of a gene, which means when averaged over an entire coding region, ω typically falls well below 1 even though selection may be acting on some sites. This has been demonstrated in mammalian egg coat proteins[45] and the *S-locus Receptor Kinase* (*SRK*), a self-incompatibility gene[46]. We used an ω value of 0.5 or greater as a cut-off for rapidly evolving and positively selected genes, as this value was deemed a useful cut-off for identifying genes with sites under selection in previous studies[47, 48]. Large-scale preliminary alignments, estimation of ω, and removal of likely non-orthologous genes, including pseudogenes and transposons, identified 250 unique rapidly-evolving genes having an ω value of 0.5 or greater out of a total of 3,232 analyzable genes (Table A.5). Nine genes were found in more than one data set. Rapidly-evolving genes were found in all data sets analyzed, but the majority originated from the ovary, where 25.9% of genes in the data set had ω>1 (Table 3.1). The ovary-expressed genes also showed a much higher average ω value (average ω=0.5) than genes expressed in other reproductive structures. The average number of synonymous substitutions per synonymous site was calculated using all reproductive tissue-expressed genes in our study and found to be 0.14 (Table 1).

To determine if any of the putative 250 rapidly-evolving genes may be lineage-specific, we queried the results of a previous study to see which of our genes were present[49]. There were 52/250 (20.8%) genes determined to be specific to the Brassicaceae family and 16/250

(6.4%) genes found only in the *Arabidopsis* lineage. In comparison, genome-wide analysis of 26,862 *A. thaliana* genes suggest that 3.4% of genes in the *A. thaliana* genome are not found outside of the Brassicaceae and 4.9% are specific to the *Arabidopsis* genus[49]. For 21 *A. thaliana* genes in our analysis, a match in *A. lyrata* could not be identified (Table A.6) so these genes could not be used in later statistical analyses. It is of interest that 2 of these 21 genes were found to be Brassicaceae-specific and 7 are *Arabidopsis*-specific.

To determine if the candidate rapidly-evolving reproductive genes tend to belong to specific functional categories, the genes were categorized according to the molecular process in which they are predicted to be involved (Figure 3.1a). Based on Gene Ontology classification, the majority of these candidate rapidly evolving genes are uncharacterized and cannot yet be functionally classified. However, a number of enzymes and genes involved in binding nucleic acids or proteins were present in the data set. To determine if there was an enrichment of a particular class of genes, the rapidly-evolving genes were compared to all reproductive tissue-expressed genes in the analysis. The rapidly-evolving gene data set showed an enrichment of proteins predicted to be targeted to the endomembrane system ($p<0.01$): approximately 19.2% of the rapidly-evolving genes are believed to be associated with the endomembrane system, whereas only 5.2% of the collective reproductive tissue-expressed data set shows this same association. In concurrence with many of the gene products being membrane-associated, computational predictions suggest many candidate genes to be membrane-associated or secreted:

131 genes are predicted to have signal peptides and/or transmembrane domains out of the 250 genes exhibiting an estimated $\omega>0.5$ (Table A.5). The genes were also categorized according to the biological process in which they are involved Figure 1b). Again, most of the genes cannot yet be classified, but the largest categories of classifiable gene were gene involved in protein metabolism and transcription.

Population genetics and statistical analyses of selected candidates

In order to obtain more conclusive evidence for selection acting on the candidate genes, a subset of 12 genes showing a low-polymorphism nonsynonymous to synonymous substitution ratio and a high-divergence nonsynonymous to synonymous substitution ratio in comparison to the other candidate rapidly-evolving genes were selected for further analysis (Table 3.3). These genes were of interest because they show signatures of differentiation between species and may play an important role in defining reproductive differences. Besides possible species differentiation, these candidates also had supporting data that suggested they may play a role in cell-cell interactions in the reproductive phase of plant development. This assumption was based on presence of signal peptides, expression data, and putative function (Table A.5 and Table A.7). Any gene not likely to have a role in reproductive interactions in the two species analyzed was left out of this subset, including genes predicted to have premature stop codons in both *A. lyrata* and *A. thaliana,* and any gene product having substantial evidence for chloroplast or mitochondrial

Table 3.1. Omega estimates for individual data sets.

Data set	No. of orthologous genes with ω>0.5[1]	Total data set size	% of genes with ω>0.5[1]	Average dN	Average dS	Average ω
Ovary	119 (9)	459	25.9 (1.9)	0.059	0.136	0.496
Pollen	21 (2)	906	2.3 (0.2)	0.016	0.136	0.134
Pollen Coat Proteins	6 (2)	187	3.2 (1.1)	0.011	0.134	0.107
Pollen Tube	103 (19)	1746	5.9 (1.1)	0.025	0.140	0.213
Stigma	4 (0)	148	2.7 (0)	0.040	0.140	0.186
Transmitting Tract	6 (1)	58	10.3 (1.7)	0.035	0.150	0.296
Total (Unique)	250 (33)	3233	7.8 (1.0)	0.028	0.139	0.231

[1]Numbers in parenthesis use number of genes with a p-value<0.05 for having ?>0.5.

Table 3.2. Omega estimates for genes enriched in individual data sets.

Data set	No. of orthologous genes with w>0.5	Total	% of genes with w>0.5	Average w
Ovary-enriched	21	134	15.7	0.362
Pollen-specific	10	372	2.7	0.158
Pollen Coat Proteins	-	-	-	-
Pollen Tube[1]	36	353	10.1	0.297
Stigma	1	38	2.63	-
Transmitting Tract	0	18	0	-

[1]All pollen tube-expressed genes that were differentially expressed between pollen tubes grown *in vitro* and pollen tubes grown through the style.

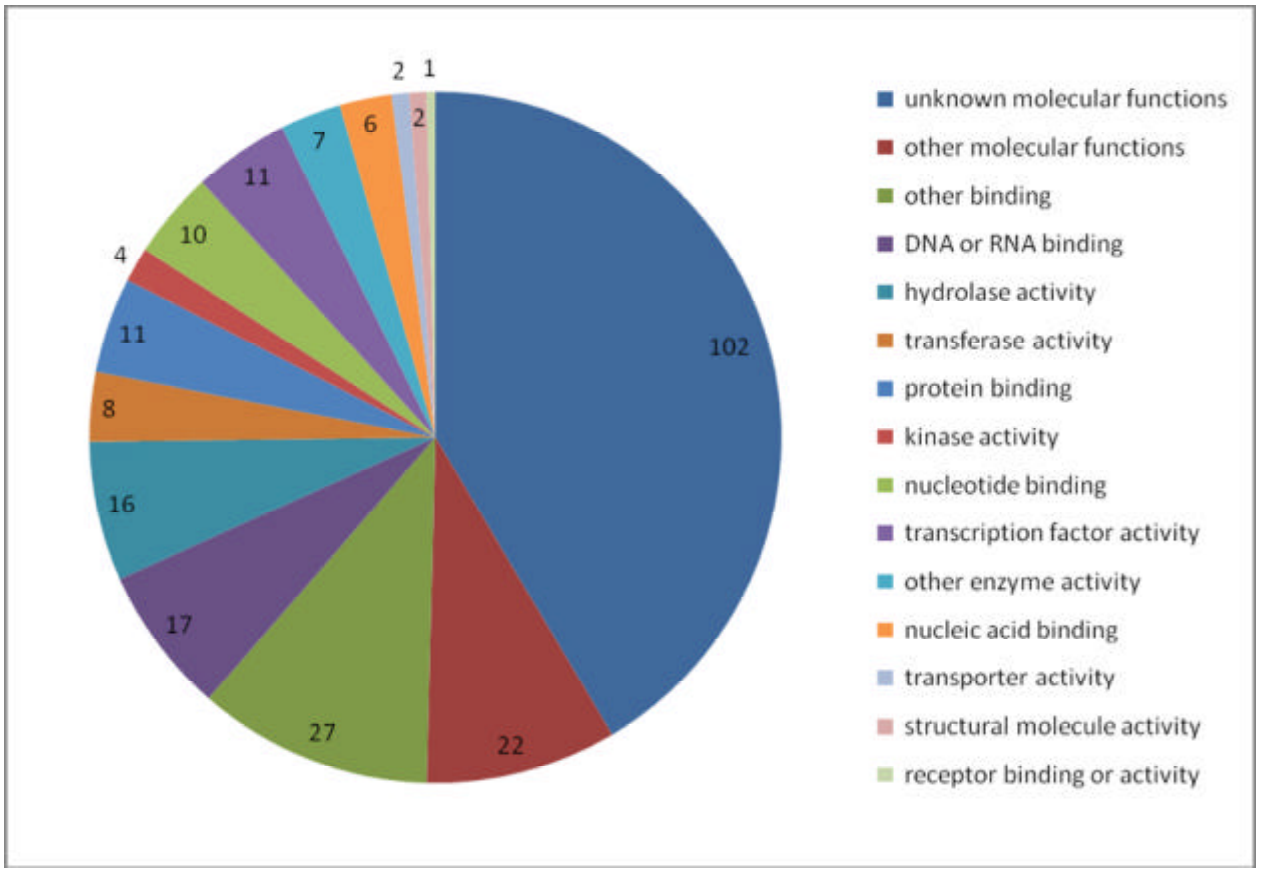

Figure 3.1a. Functional classification of candidate rapidly-evolving genes. Numbers represent the number of candidate genes that belong to each functional category. Some genes belong to more than one category.

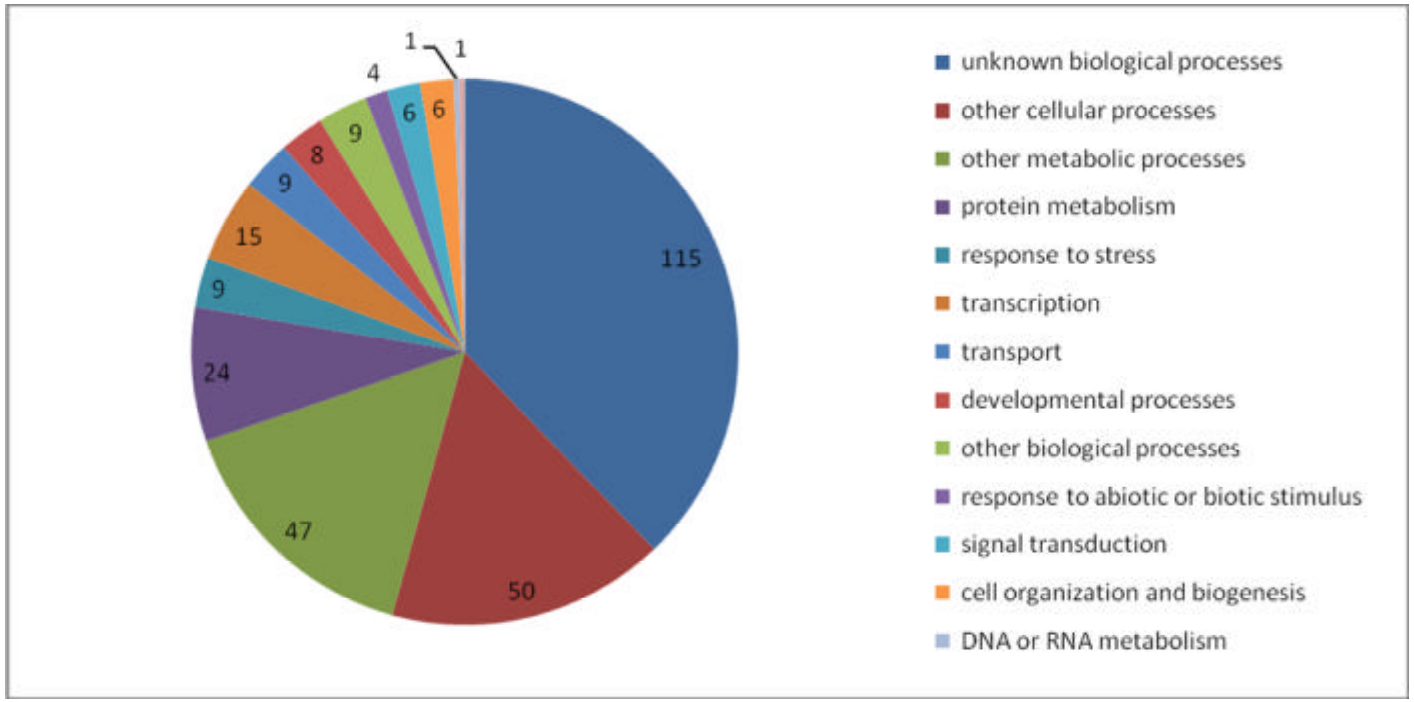

Figure 3.1b. Classification of candidate rapidly-evolving genes according to their biological process. Numbers represent the number of candidate genes that belong to each category. Some genes belong to more than one category.

localization. Of the 12 candidate genes, eight were found only in the ovary data set, while two were found only in the pollen tube data set. Two genes had overlapping expression in two data sets: At4g21326 was found in both the pollen and the pollen tube data sets, while At5g09370 was found in the ovary and pollen tube data sets.

For the selected candidate genes, efforts were made to obtain polymorphism data by sequencing alleles of each gene from ten *A. thaliana* accessions. Summary statistics based on the site frequency spectrum were generated for each gene and tests for deviation from neutrality were performed (Table 3.5). In particular, we calculated Tajima's D[38], since it can be useful for detecting positive selection. We also used Fay and Wu's H[37], because the demographic issues present in *Arabidopsis* may be less of a problem in this test since it is influenced by high and intermediate frequency variants[39]. We used distributions generated from a previous study of 334 randomly distributed genomic regions in 12 accession of *A. thaliana* to determine significance of the tests[39].

Two *A. thaliana* genes, At3g59460 (F-box) and At3g50020 (pathogenesis-related), were difficult to align to *A. lyrata*. Therefore, some sections of the alignments were not used in the analysis for tests involving an outgroup. For At3g59460, the first exon and intron of *A. lyrata* were left out. Similarly, the first (and only) intron of At3g50020 did not align conclusively with *A. lyrata*, so this region of *A. lyrata* was excluded in some tests.

Five genes showed signs of selection using site-frequency spectrum-based tests (Table 3.4). At5g46960, an enzyme inhibitor,

had significant results ($p<0.05$) in more than one test. Four genes, At1g20730 (F-box), At2g20597 (plant thionin), At4g21326 (subtilase), and At5g09370 (protease inhibitor), had a significantly negative ($p<0.05$) Fay and Wu's H, suggesting a recent selective event may have acted on these genes.

McDonald-Kreitman tests (Table 3.5 and Table A.8) were performed since they may be more robust to some of the violations of assumptions that occur within *Arabidopsis*[50]. Six genes (At1g20730, At2g06090, At3g50020, At4g21326, At5g09370, and At5g46960) showed convincing evidence of variation in the ratios of nonsynonymous-to-synonymous polymorphism and divergence. None of the genes had a neutrality index above 1, suggesting that all genes with a significant McDonald-Kreitman test have an excess of amino acid evolution between species. A low level of polymorphism in At3g48231, At3g57840, At3g59460, At5g09370, At5g18403, and At5g23035 resulted in tests that had little power to detect deviations from neutrality. At3g57840 (self-incompatibility-related) and At3g48231 (cysteine-rich) were particularly unusual in that they showed no polymorphism in any of the accessions. However, the inability to generate PCR products with our Col-0-specific primers and DNA gel blot analysis suggest these two genes may be absent from some accessions (Figure A.1). In two accessions (Lz-0 and Ts-1) the genes appear to have been deleted. At3g48231 also may have been deleted in C24 and Est. Additional results suggested that At3g57840 is also missing in the Sha accession.

Eight of the candidate genes have a predicted ω that exceeds 1.0

over their entire coding sequence alignments (Table 3.5). Five of the genes, At1g20730, At3g48231, At3g57840, At5g09370, and At5g23035, are likely to have an omega value greater than 0.5 ($p<0.05$), based on likelihood ratio tests, and are therefore good candidates for having some positively-selected sites. Additionally, At1g20730 and At5g23035 are both likely to have an $\omega>1$ based on likelihood ratio tests ($p<0.1$). All 12 candidates had a dN value greater than the data set average of 0.04. Our estimates of dS varied greatly among the 12 genes analyzed and several of the genes fell below the genome-wide average of dS (0.12).

Overall, only At3g59460 failed to show a significant departure from neutrality in any of the tests. The remaining genes all showed signatures of positive selection in at least one of the measures implemented in our study. Of particular interest, one gene, At5g09370, shows significant deviations from neutrality in both site frequency spectrum tests and divergence-based methods.

Table 3.3. Selected candidate rapidly evolving genes.

AGI Number	Gene Model Description	Data Set	Signal Peptide	Predicted Helices
AT1G20730	unknown, F-box-associated Type 1 domain	ovary	N	1
AT2G06090	self-incompatibility protein-related	ovary	Y	0
AT2G20597	Encodes a Plant thionin family protein	ovary	Y	1
AT3G48231	Encodes a member of a family of small, secreted, cysteine rich protein with sequence similarity to the PCP (pollen coat protein) gene family	ovary	Y	0
AT3G50020	PR (pathogenesis-related) peptide that belongs to the PR-6 proteinase inhibitor family.	ovary	N	0
AT3G57840	self-incompatibility protein-related	ovary	Y	2
AT3G59460	similar to F-box family protein	pollen tube	Y	0
AT4G21326	ATSBT3.12; subtilase	pollen - pollen tube	Y	0
AT5G09370	protease inhibitor/seed storage/lipid transfer protein family protein	ovary-pollen tube	Y	0
AT5G18403	Encodes a defensin-like (DEFL) family protein.	ovary	Y	0
AT5G23035	Encodes a defensin-like (DEFL) family protein.	ovary	Y	0
AT5G46960	invertase/pectin methylesterase inhibitor family protein	ovary	Y	0

Table 3.4. Summary statistics for top candidate rapidly evolving genes and tests for deviation from neutrality based on the site frequency spectrum.

AGI Number	Length (bp)	Protein Length (a.a)	S	q	p	Fu and Li's D[a]	Fu and Li's F[a]	Fay and Wu's H[a]	Tajima's D
AT1G20730	4406	527	115	0.010	0.009	-0.24	-0.38	-8.3 [b]	-0.3
AT2G06090	408	135	8	0.007	0.006	-0.2	-0.3	0.8	-0.1
AT2G20597	425	84	12	0.012	0.008	-0.5	-1.0	-6.2[b]	-1.4
AT3G48231	616	80	2	0.002	0.002	-0.2	-0.3	0.5	-0.3
AT3G50020	523	96	7	0.003	0.005	-1.0	-1.2	1.0	-1.0
AT3G57840	465	154	0	0.000	0.000	NA	NA	NA	NA
AT3G59460	811	106	4	0.002	0.002	0.7	1.0	0.2	1.1
AT4G21326	3680	754	51	0.006	0.005	0.0	-0.2	-5.0 [b]	-0.6
AT5G09370	1183	158	8	0.003	0.002	0.7	0.3	-6.49 [b]	-1.4
AT5G18403	373	87	6	0.006	0.007	1.1	1.3	-0.2	0.7
AT5G23035	436	88	2	0.002	0.002	-0.5	-0.3	0.1	0.2
AT5G46960	607	174	9	0.006	0.003	-2.4 [b]	-2.7 [b]	-0.5	-1.8

[a]Calculated using *A. lyrata* as an outgroup.
[b]Significant at 0.05 level.

Table 3.5. Results of divergence-based tests of positive selection.

AGI Number	Fisher's Exact p-value [a]	N. I.	dN	dS	w	p w >0.5	p w>1.0
AT1G20730	0.05[b] (0.02[b])	0.31 (0.47)	0.087	0.047	1.851	0.001[b]	0.002[b]
AT2G06090	0.03[b] (.03[b])	0.18 (0.17)	0.145	0.190	0.761	0.230	1.000
AT2G20597	0.28 (0.32)	0.23 (0.44)	0.098	0.085	1.146	0.190	0.533
AT3G48231	NA (0.55)	NA (0)	0.100	0.064	1.892	0.044[b]	0.405
AT3G50020	0.08[c] (NA)	0.14 (NA[d])	0.096	0.056	1.726	0.223	0.359
AT3G57840	NA (NA)	NA (NA)	0.197	0.161	1.224	0.010[b]	0.557
AT3G59460	0.4 (1.0)	0 (0)	0.115	0.170	0.677	0.182	1.000
AT4G21326	0.03[b] (<0.01[b])	0.38 (0.33)	0.054	0.086	0.630	0.232	1.000
AT5G09370	NA (0.02)[b]	NA (0)	0.098	0.061	1.612	0.006[b]	0.242
AT5G18403	0.43 (0.12)	0.29 (0.20	0.274	0.158	1.104	0.109	0.242
AT5G23035	1 (0.19)	NA (NA)	0.183	0.060	3.050	0.001[b]	0.051[c]
AT5G46960	0.05[c] (0.06[c])	0.20 (0.21)	0.118	0.167	0.707	0.289	1.000

[a] Numbers in parenthesis include both coding and noncoding regions
[b] significant at 0.05 level
[c] significant at 0.1 level
[d] noncoding region for this gene could not be aligned

DISCUSSION

Initial Screen Results

While the approach used for initial screening of candidate rapidly evolving genes no doubt overlooked some interesting candidates, it seems to have been effective based on its ability to select previously-identified positively-selected candidate genes. One group of genes that is known to evolve rapidly that our screen successfully identified is the group of GRP proteins found in the pollen coat[18,54]. These genes have been implicated in species recognition and four of them were identified as having an $\omega>0.5$ in our screen (At5g07510, At5g07520, At5g07540, and At5g07550). Additionally, many other genes known to exhibit signatures of adaptive evolution such as self-incompatibility-related genes and pathogenesis-related genes showed preliminary evidence of positive selection acting on some sites.

Our estimate of dS (0.14), despite being biased by the inclusion of only reproductive-expressed genes, is similar to the genome-wide synonymous substitution distance of 0.12 estimated from a much larger comparison of 22,564 orthologous gene pairs that include most of the *Arabidopsis* coding regions[51]. This is not surprising considering that synonymous substitutions should be nearly neutral and their rates are expected to be similar between different tissue types. Our estimate of ω for pollen-specific genes is however quite different from a previous estimate[20]. This study estimates pollen-specific genes to have an ω of approximately 0.39, much larger than our estimate of 0.134, possibly due to a difference in data sets and sample sizes. The ovule appears to have the highest prevalence of expressed genes that are

candidates for rapid evolution, possibly due to the large number of genes having similarity to defense-related or self-incompatibility-related genes. The pollen tube and transmitting tract also had comparatively higher values of ω. Since the *A. thaliana* stigma allows the hydration and germination of pollen grains from other species within the Brassicaceae, the stigma may not be a major site for inter-specific pollen recognition in this species. Rather, it is possible that factors in the ovary as well as in the pollen tube and transmitting tract could be responsible for between-species recognition, making genes expressed in these tissues more likely to be positively selected.

Genes that may have sites subjected to positive selection appear to be more likely than others to be lineage-specific. Approximately 27% of genes with an $\omega>0.5$ were lineage-specific genes. This observation means these genes are not found outside of the Brassicaceae and in some instances are not present outside of the *Arabidopsis* clade. This estimate is much higher than the whole-genome average of ~8%. These particular rapidly evolving genes are so greatly diverged between lineages, that they can no longer be detected outside of the Brassicaceae based on sequence homology. Alternatively, they might have been gained in some lineages or lost in others. In addition, 21 genes in our analysis did not have any significant matches in the *A. lyrata* genome sequence, meaning they are either absent from *A. lyrata* entirely or are present in the species, but were not part of the sequence database. It may be worthwhile to investigate these lineage-specific genes further as they may also shed light on interspecific variation.

We found that genes expressed specifically in reproductive tissues generally do not seem to be evolving more rapidly than ubiquitously expressed genes. In fact, ovary-specific genes appear to be somewhat less positively selected than those expressed in ovary and other tissues, possibly due to expression in ovary tissue of several functionally constrained genes involved in seed development.

It is important to note that an absence of synonymous substitutions can artificially inflate ω, possibly leading to incorrect evolutionary interpretations[52]. This may be especially apparent in comparisons of closely-related taxa such as *A. lyrata* and *A. thaliana* in which there may be few synonymous substitutions[52]. Additionally, as in any screen, some of the criteria used may have excluded important candidates. In particular, averaging ω across the gene has the effect of overlooking positive selection that may be occurring in subdomains of the protein. In an attempt to compensate for this effect, we chose an ω value below that expected for positive selection over the entire gene. However, *FERONIA*, a gene believed to be involved in inter-specific recognition in *Arabidopsis*, has a value of ω below our cut-off when averaged over the gene[56]. Although this particular gene was not included in our analysis, it is important to note that in some cases, only a few amino-acid changes may be contributing to a particular adaptive function, making the signatures of selection hard to detect without additional information. Furthermore, other genes that had been implicated in species discrimination, such as *ANUXR1*[53], and *ANUXR2*[53], were not included in the initial data sets because they were not represented in the microarrays.

Analysis of Selected Candidates

By utilizing publicly available polymorphism, expression and function data, as well as using prediction software to characterize interesting gene features, we selected a subset of 12 candidates for further analysis. Genes with possible reproduction-related functions that showed preliminary signatures of rapid evolution and differentiation between *A. lyrata* and *A. thaliana* were prime candidates for further analysis. Out of these 12 candidates, the most interesting genes based on our statistical analyses are genes predicted to encode a subtilase (At4g21316) and two enzyme inhibitors (At5g09370 and At5g46960). These three genes produced significant McDonald-Kreitman values that suggest differentiation between the lineages and site-frequency spectrum tests are in concordance with this conclusion. They also had an elevated ω that seems driven by an above-average rate of nonsynonymous substitutions. These genes are likely candidates for proteins that have undergone functional divergence between *A. thaliana* and *A. lyrata*. Two additional genes, a self-incompatibility-related protein (At2g06090) and a pathogenesis-related peptide (At3g50020) showed signs of differentiation between the two Arabidopsis species tested, but tests of the site-frequency spectrum were not significant (however, see below). A lack of substantial polymorphism in our sample for the self-incompatibility-related At3g57840, the F-box domain-containing At3g59460, and the defensin-like At5g18403 and At5g23035 genes makes it difficult to draw any firm conclusions about these genes. They appear to be

rapidly evolving based on ω, and their lack of polymorphism suggests that they may have undergone a recent selective sweep within the *A. thaliana* lineage. The addition of more samples may help clarify this issue.

When interpreting these cumulative results, it is important to note that a gene may show different signatures of selection when different types of analysis are used. Tests involving the site-frequency spectrum, typically only detect recent adaptation events, whereas divergence methods, i.e. McDonald-Kreitman and ω estimates, evaluate past recurrent selection[50]. Therefore, it is quite possible to detect selection in only some of the tests.

Many of the candidates have indels either in some of the accessions or in *A. lyrata* (Table A.9). Most of these indels are in frame and do not disrupt the amino-acid sequence, suggesting that there may be some constraint on the proteins that they encode. However, in three genes, At1g20730, At2g06090, At5g18403, and At5g23035, there is a frameshift mutation that disrupts the downstream sequence. It is important to realize that this may not mean the gene is non-functional either in these accessions or in the accessions that do not have the frameshift. In all cases but one, the frameshift occurs near the 3' end of the gene, such that the protein may still retain its function. Additionally, the At3g57840 and At3g48231 genes are deleted in some of the accessions analyzed. Indels can be an important means of adaptive evolution. A deletion of the gene, *RPS5*, has been found to segregate in *A. thaliana* and is believed to have undergone balancing selection due to differing pathogen selective pressures[54, 55]. Also, a

variable repeat found in the intron of the *ISOPROPYL MALATE ISOMERASE LARGE SUB UNIT1* gene has been found to be involved in environmental adaptation[56]. Since many of the candidate genes are members of gene families, it is possible that duplication and differential adaptation has occurred between paralogs, such that the candidate gene no longer has the same function in different species or in different accessions.

When interpreting our results, several issues were kept in mind. As mentioned previously, *Arabidopsis spp.* are known to deviate from assumptions of the standard neutral model in several ways. Besides the fact that *A. thaliana* is a selfing species, there is also population structure[57] as witnessed by continuous genetic isolation by distance[58]. Additionally, both *Arabidopsis* species in this study have a small population size, which can obscure evidence of past adaptive divergence by inflating levels of nonsynonymous to synonymous polymorphism[59]. This can be observed in *A. thaliana* as an excess of amino acid polymorphism relative to divergence[60]. There is also a predominance of low-frequency variants in *A. thaliana* that results in a negative Tajima's D average for this species[57]. This is believed to be due to background selection working on mildly deleterious sites[57]. By comparing our results to results from distance methods, as well as by using multiple tests of the site frequency spectrum, we may have more accurately detected true signatures of positive selection. We also based our conclusions on previously generated distributions from 12 accessions of *A. thaliana* to help ameliorate the effects of demographics[39]. It would be useful to compare our test results to

genome-wide empirical distributions based on the accession used in this study (to be discussed in Chapter 4).

Our top candidates can be grouped into functional categories. Two ovary-expressed genes, At2g06090 and At3g57840, share sequence similarity with the S1 self-incompatibility protein from *Papaver rhoeas*. The inhibition of incompatible pollen in *Papaver* occurs in the stigma and is caused by an interaction between a glycosylated protein, called the S-protein, which is secreted by stigmatic cells, and a membrane-associated protein expressed in pollen and pollen tubes [61, 62]. As mentioned earlier, self-incompatibility genes are known to undergo adaptive evolution, and our results indicate that genes related to these genes, although not involved in self-incompatibility, may also be subject to similar selective pressures. In particular, the presence/absence polymorphism that we detected for these genes upon DNA gel blot analysis may be indicative of balancing selection causing extreme sequence divergence. At3g57840 is of particular interest, since it was found to be down-regulated in *myb98* ovules and is likely to be expressed in synergid cells[43] (Table A.7).

Four of the candidate genes are small, cysteine-rich peptides (At2g20597, At3g48231, At5g18403, and At5g23035). Two genes falling in this category, At5g18403, a synergid-expressed gene (Table A.7)[43], and At5g23035, are classified as defensin-like genes. Defensin-like genes are prone to duplication[63] and are known to evolve rapidly while their tertiary structure, which is normally maintained by disulfide bridges between conserved cysteines, remains intact[64].

Defensin-like genes have a wide variety of functions in plants such as involvement in self-incompatibility, inhibition of pathogens and plant root growth, and inhibition of enzyme and protein translation activities (see Carvalho et al., 2009 for review)[65]. Of particular relevance to our study, the defensin-like S-locus cysteine-rich (SCR) protein, which is a component of the pollen coat, functions as the male determinant of specificity in the self-incompatibility response of the Brassicaceae[66]. Additionally, small defensin-like molecules expressed in *Torenia* ovules have been reported to function in pollen tube guidance towards the ovule and as possible species-specific pollen-tube attraction cues[70].

Two genes, At1g20730 and At3g59460, having regions similar to F-box domains were selected as positive candidates. F-box domains are thought to be involved in protein-protein interactions and protein degradation, but most have unknown biological functions[67]. One family of *A. thaliana* F-box genes is involved in auxin regulation[68]. It has been proposed that some F-box proteins with potential functions in proteolysis and pathogen defense are positively selected and undergo rapid birth-death evolution[69, 70]. Many of these genes evolved by indels, making alignment and analyses based on coding sequence evolution difficult[70].

It is hoped that the candidates produced as a result of this screen will lead to a better understanding of the mechanisms underlying intra-specific and inter-specific selectivity in mate choice and the mode of diversification of reproductive genes in plants. Information may be gained on how reproductive isolation may occur through the effects of mutation in rapidly-evolving reproductive gene.

Currently, few candidate speciation genes have been identified in plants, so this work may provide another view on this process. The functional analysis of divergent genes in the female reproductive tract and pollen or pollen tube may also lead to the identification of new genes involved in pollination and fertilization processes. T-DNA knockouts of the candidate can be analyzed for defects in pollination and fertilization processes (Table A.10). Furthermore, since few rapidly-evolving genes have been characterized, and even fewer have been proven to be speciation genes, none of which are in plants, this work may be useful for understanding mechanisms of reproductive isolation and speciation across organisms.

Special thanks to Nathan Clark for high-throughput DNA sequence alignments, analysis, and helpful suggestions, to Suganthi Kandasamy for help with the generation of DNA sequence from *A. thaliana* accessions, and to Tiffany Crispell for plant care.

REFERENCES

1. Maxwell, A. I., Morrison, G. M. and Dorin, J. R. Rapid sequence divergence in mammalian beta-defensins by adaptive evolution. Mol Immunol 40, 413-21 (2003).
2. Watanabe, Y., Maruoka, T., Walter, L. and Kasahara, M. Comparative genomics of the Mill family: a rapidly evolving MHC class I gene family. Eur J Immunol 34, 1597-607 (2004).
3. Yang, L. and Gui, J. F. Positive selection on multiple antique allelic lineages of transferrin in the polyploid *Carassius auratus*. Mol Biol Evol 21, 1264-77 (2004).
4. Vacquier, V. D. Evolution of gamete recognition proteins. Science 281, 1995-8 (1998).
5. Singh, R. S., and Kulathinal, R. J. Sex gene pool evolution and speciation: A new paradigm. Genes and Genetic Systems Vol. 75, pp.119-130 (2000).
6. Makalowski, W. and Boguski, M. S. Evolutionary parameters of the transcribed mammalian genome: an analysis of 2,820 orthologous rodent and human sequences. Proc Natl Acad Sci U S A 95, 9407-12 (1998).
7. Wyckoff, G. J., Wang, W. and Wu, C. I. Rapid evolution of male reproductive genes in the descent of man. Nature 403, 304-9 (2000).
8. Ferris, P. J., Pavlovic, C., Fabry, S. and Goodenough, U. W. Rapid evolution of sex-related genes in *Chlamydomonas*. Proc Natl Acad Sci U S A 94, 8634-9 (1997).
9. Biermann, C. H. The molecular evolution of sperm bindin in six species of sea urchins (Echinoida: Strongylocentrotidae). Mol Biol Evol 15, 1761-71 (1998).

10. Tsaur, S. C. and Wu, C. I. Positive selection and the molecular evolution of a gene of male reproduction, *Acp26Aa* of *Drosophila*. Mol Biol Evol 14, 544-549 (1997).
11. Tsaur, S. C., Ting, C. T. and Wu, C. I. Positive selection driving the evolution of a gene of male reproduction, *Acp26Aa*, of *Drosophila*: II. Divergence versus polymorphism. Mol Biol Evol 15, 1040-1046 (1998).
12. Sainudiin, R., Shuk, W., Wong, W., Yogeeswaran, K., Nasrallah, J. B., Yang, Z., and Nielsen, R. Detecting site-specific physiochemical slective pressures. Journal of Molecular Evolution 60, 315-226 (2005).
13. Swanson, W. J. and Vacquier, V. D. The rapid evolution of reproductive proteins. Nature Reviews Genetics 3, 137 -144 (2002).
14. Nishio, T. and. Kusaba, M. Sequence Diversity of SLG and SRK in *Brassica oleracea* L. Annals of Botany 85, 141-146 (2000).
15. Stein, J. C., Howlett, B., Boyes, D. C., Nasrallah, M. E. and Nasrallah, J. B. 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, 8816-20 (1991).
16. Schopfer, C. R., Nasrallah, M. E. and Nasrallah, J. B. The male determinant of self-incompatibility in *Brassica*. Science 286, 1697-700 (1999).
17. Watanabe, M., Ito, A., Takada, Y., Ninomiya, C., Kakizaki, T., Takahata, Y., Hatakeyama, K., Hinata, K., Suzuki, G., Takasaki, T., Satta, Y., Shiba, H., Takayama, S. and Isogai, A. Highly divergent sequences of the pollen self-incompatibility (S) gene in class-I S haplotypes of *Brassica campestris* (syn. rapa) L. FEBS Lett 473, 139-44 (2000).
18. Mayfield, J. A., Fiebig, A., Johnstone, S. E. and Preuss, D. Gene families from the *Arabidopsis thaliana* pollen coat proteome. Science 292, 2482-5 (2001).

19. Barrier, M., Bustamante, C. D., Yu, J. and Purugganan, M. D. Selection on Rapidly Evolving Proteins in the *Arabidopsis* Genome. Genetics 163, 723-733 (2003).
20. Slotte, T., St. Onge, K., Wright, S., Morton, B. (unpublished).
21. Koch, M. A., Haubold, B. and Mitchell-Olds, T. Comparative evolutionary analysis of chalcone synthase and alcohol dehydrogenase loci in *Arabidopsis*, *Arabis*, and related genera (Brassicaceae). Mol Biol Evol 17, 1483-98 (2000).
22. The *Arabidopsis* Genome Initiative. Analysis of the genome sequence of the flowering plant *Arabidopsis thaliana*. Nature 408, 796 (2000).
23. *Arabidopsis lyrata* Sequencing Consortium. (Unpublished data).
24. Altschul, S. F., Gish, W., Miller, W., Myers, E. W. and Lipman, D. J. Basic local alignment search tool. J Mol Biol 215, 403-10 (1990).
25. Yang, Z. Phylogenetic Analysis by Maximum Liklihood (PAML). University College, London, version 2.0 (1999).
26. McDonald, J. H. and Kreitman, M. Adaptive protein evolution at the *Adh* locus in *Drosophila*. Nature 351, 652-654 (1991).
27. Rozas, J., Sanchez-DelBarrio, J. C., Messeguer, X. and Rozas, R. DnaSP, DNA polymorphism analyses by the coalescent and other methods. Bioinformatics 19, 2496-2497 (2003).
28. Carbon, S., Ireland, A., Mungall, C. J., Shu, S., Marshall, B., Lewis, S., the Ami, G. O. H. and the Web Presence Working, G. AmiGO: online access to ontology and annotation data. Bioinformatics 25, 288-289 (2009).
29. Bendtsen, J. D., Nielsen, H., von Heijne, G., and Brunak, S. Improved prediction of signal peptides: SignalP 3.0. J Mol Biol 340, 783-95 (2004).

30. Nielsen, H., and Krogh, A. Prediction of signal peptides and signal anchors by a hidden Markov model. Proceeding of the Sixth International Conference on Intelligent Systems for Molecular Biology, 122-130 (1998).
31. Krogh, A., Larsson, B., von Heijne, G., and Sonnhammer, E. L. Predicting transmembrane protein topology with a hidden Markov model: application to complete genomes. J Mol Biol 19, 567-580 (2001).
32. Hruz T, L. O., Szabo G, Wessendorp F, Bleuler S, Oertle L, Widmayer P, Gruissem W and Zimmermann, P. Genevestigator V3: a reference expression database for the meta-analysis of transcriptomes. Advances in Bioinformatics (2008).
33. Doyle, J. J. and Doyle, J. D. Isolation of plant DNA from fresh tissue. Focus 12, 13-15 (1990).
34. Thompson, J., Higgins, D. and Gibson, T. CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. Nucleic Acids Research 22, 4673 - 4680 (1994).
35. Hall, T. A. BioEdit: a user-friendly biological sequence alignment editor and analysis program for Windows 95/98/NT. Nucleic Acids Symposium Series 41, 95-98 (1999).
36. Fu, Y. X. and Li, W. H. Statistical Tests of Neutrality of Mutations. Genetics 133, 693-709 (1993).
37. Fay, J. C. and Wu, C.-I. Hitchhiking Under Positive Darwinian Selection. Genetics 155, 1405-1413 (2000).
38. Tajima, F. Statistical method for testing the neutral mutation hypothesis by DNA polymorphism. Genetics 123, 585-595 (1989).

39. Schmid, K. J., Ramos-Onsins, S., Ringys-Beckstein, H., Weisshaar, B. and Mitchell-Olds, T. A Multilocus Sequence Survey in *Arabidopsis thaliana* Reveals a Genome-Wide Departure From a Neutral Model of DNA Sequence Polymorphism. Genetics 169, 1601-1615 (2005).

40. Rand, D. and Kann, L. Excess amino acid polymorphism in mitochondrial DNA: contrasts among genes from *Drosophila*, mice, and humans. Mol Biol Evol 13, 735-748 (1996).

41. Doyle, J. J. and. Doyle, J. L. Isolation of plant DNA from fresh tissue. Focus 12, 13-15 (1990).

42. Tung, C.-W., Dwyer, K. G., Nasrallah, M. E. and Nasrallah, J. B. Genome-Wide Identification of Genes Expressed in *Arabidopsis* Pistils Specifically along the Path of Pollen Tube Growth. Plant Physiol. 138, 977-989 (2005).

43. Jones-Rhoades, M. W., Borevitz, J. O. and Preuss, D. Genome-Wide Expression Profiling of the *Arabidopsis* Female Gametophyte Identifies Families of Small, Secreted Proteins. PLoS Genet 3, e171 (2007).

44. Honys, D. and Twell, D. Comparative Analysis of the *Arabidopsis* Pollen Transcriptome. Plant Physiol. 132, 640-652 (2003).

45. Swanson, W. J., Yang, Z., Wolfner, M. F. and Aquadro, C. F. Positive Darwinian selection drives the evolution of several female reproductive proteins in mammals. Proceedings of the National Academy of Sciences of the United States of America 98, 2509-2514 (2001).

46. Sato, K., Nishio, T., Kimura, R., Kusaba, M., Suzuki, T., Hatakeyama, K., Ockendon, D. J. and Satta, Y. Coevolution of the S-Locus Genes SRK, SLG and SP11/SCR in *Brassica oleracea* and *B. rapa*. Genetics 162, 931-940 (2002).

47. Swanson, W. J., Wong, A., Wolfner, M. F. and Aquadro, C. F. Evolutionary Expressed Sequence Tag Analysis of *Drosophila* Female Reproductive Tracts Identifies Genes Subjected to Positive Selection. Genetics 168, 1457-1465 (2004).

48. Turner, L. M., Chuong, E. B. and Hoekstra, H. E. Comparative Analysis of Testis Protein Evolution in Rodents. Genetics 179, 2075-2089 (2008).

49. Slotte, T., Foxe, J. P., Hassouri, K. M. and Stephen I. Wright. Genome-wide evidence for efficient positive and purifying selection in *Capsella grandiflora*, a plant species with a large effective population size. Molecular Biology and Evolution (2010).

50. Nielsen, R. Molecular signatures of natural selection. Annual Review of Genetics 39, 197-218 (2005).

51. Pedro, J., and Van de Peer, Y. (unpublished).

52. Wolf, J. B. W., Kunstner, A., Nam, K., Jakobsson, M. and Ellegren, H. Nonlinear Dynamics of Nonsynonymous (dN) and Synonymous (dS) Substitution Rates Affects Inference of Selection. Genome Biol Evol 2009, 308-319 (2009).

53. Boisson-Dernier, A., Roy, S., Kritsas, K., Grobei, M. A., Jaciubek, M., Schroeder, J. I. and Grossniklaus, U. Disruption of the pollen-expressed *FERONIA* homologs *ANXUR1* and *ANXUR2* triggers pollen tube discharge. Development 136, 3279-3288 (2009).

54. Stahl, E. A., Dwyer, G., Mauricio, R., Kreitman, M. and Bergelson, J. Dynamics of disease resistance polymorphism at the *Rpm1* locus of *Arabidopsis*. Nature 400, 667 (1999).

55. Tian, D., Araki, H., Stahl, E., Bergelson, J. and Kreitman, M. Signature of balancing selection in *Arabidopsis*. Proceedings of the National Academy of Sciences of the United States of America 99, 11525-11530 (2002).

56. Sureshkumar, S., Todesco, M., Schneeberger, K., Harilal, R., Balasubramanian, S. and Weigel, D. A Genetic Defect Caused by a Triplet Repeat Expansion in *Arabidopsis thaliana*. Science 323, 1060-1063 (2009).

57. Nordborg, M., Hu, T., Ishino, Y., Jhaveri, J., Toomajian, C., Zheng, H., Bakker, E., Calabrese, P., Gladstone, J., Goyal, R., Jakobsson, M., Kim, S., Morozov, Y., Padhukasahasram, B., Plagnol, V., Rosenberg, N., Shah, C., Wall, J., Wang, J., Zhao, K., Kalbfleisch, T., Schulz, V., Kreitman, M. and Bergelson, J. The Pattern of Polymorphism in *Arabidopsis thaliana*. PLoS Biology 3, e196 (2005).

58. Platt A, H. M., Huang Y. S., Li Y, Anastasio A. E., Mulyati N. W., Agren J, Bossdorf, O, B. D., Donohue K, Dunning M, Holub E. B, Hudson A, Le Corre V, Loudet O, Roux, F, W. N., Weigel D, Rivero L, Scholl R, Nordborg M, Bergelson J, and Borevitz, J. O. The scale of population structure in *Arabidopsis thaliana*. PLoS Genet. (2010).

59. Fay, J. C. and Wu, C.-I. The neutral theory in the genomic era. Current Opinion in Genetics and Development 11, 642 (2001).

60. Foxe, J. P., Dar, V. U., Zheng, H., Nordborg, M., Gaut, B. S. and Wright, S. I. Selection on amino acid substitutions in *Arabidopsis*. Mol Biol Evol 25, 1375-83 (2008).

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

62. Wheeler, M. J., Franklin-Tong, V. E., and Franklin F. C. H. The Molecular and Genetic Basis of Pollen-Pistil Interactions. New Phytol 151, 565-584 (2001).

63. Patil, A., Hughes, A. L. and Zhang, G. Rapid evolution and diversification of mammalian {alpha}-defensins as revealed by comparative analysis of rodent and primate genes. Physiol. Genomics 20, 1-11 (2004).

64. Zhu, S., Gao, B. and Tytgat, J. Phylogenetic distribution, functional epitopes and evolution of the CSaß superfamily. Cellular and Molecular Life Sciences 62, 2257 (2005).

65. Carvalho, A. d. O. and Gomes, V. M. Plant defensins--Prospects for the biological functions and biotechnological properties. Peptides 30, 1007 (2009).

66. Schopfer, C. R., Nasrallah, M. E. and Nasrallah, J. B. The Male Determinant of Self-Incompatibility in *Brassica*. Science 286, 1697-1700 (1999).

67. Kipreos, E. and Pagano, M. The F-box protein family. Genome Biology 1, reviews3002.1 - reviews3002.7 (2000).

68. Dharmasiri, N., Dharmasiri, S., Weijers, D., Lechner, E., Yamada, M., Hobbie, L., Ehrismann, J. S., Jürgens, G. and Estelle, M. Plant Development Is Regulated by a Family of Auxin Receptor F Box Proteins. Developmental Cell 9, 109 (2005).

69. Thomas, J. H. Adaptive evolution in two large families of ubiquitin-ligase adapters in nematodes and plants. Genome Research 16, 1017-1030 (2006).

70. Xu, G., Ma, H., Nei, M. and Kong, H. Evolution of F-box genes in plants: Different modes of sequence divergence and their relationships with functional diversification. Proceedings of the National Academy of Sciences 106, 835-840 (2009).

71. Qin, Y., Leydon, A. R., Manziello, A., Pandey, R., Mount, D., Denic, S., Vasic, B., Johnson, M. A. and Palanivelu, R. 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, (2009).

CHAPTER 4:

FUTURE DIRECTIONS

New questions have been generated as a result of this research and in this chapter these will be identified and discussed. Methods for further elucidating factors involved in both intraspecific and interspecific recognition will be addressed.

FURTHER ELUCIDATION OF REGULATORY MECHANISMS EFFECTING *SRK*

In chapter 2, it was observed that methylation mutants have reduced *SRKb* expression in the stigma. To determine if a change in methylation status of the *SRKb* transgene is occurring in *nrpd1a* mutants several strategies could be used. A simple method of determining methylation at specific sites would be to use methylation-specific enzymes that have recognition sites within the transgene. Specifically, it would be of interest to assess the promoter and 3' end of the gene. Primers would then be designed for these specific regions to determine if a change in methylation occurs between wild type and the mutant. A more detailed way of determining methylation status of the SRKb transgene would be to perform bisulfite sequencing. This would allow the specific bases that are associated with methylation to be identified regardless of whether an enzyme recognition site is present. To determine if altered splicing of *SRKb* is occurring in the *nrpd1a* mutants it may be worthwhile to transform the mutants with *SRKb* constructs that are incapable of making *eSRK*. This would be

possible by removing the first intron of *SRKb*. By transforming with such a construct, it may be possible to recover normal expression of *SRKb* in *nrpd1a* mutants if deviations in alternative splicing are responsible for the reduction of the full length transcript in the stigma. To assess if silencing directed at the 3' end of *SRKb* is occurring in *nrpd1a* mutants, sRNA blots could be screened for the presence of small RNAs with sequence similarity to this region.

The C24 *nrpd1a* mutants did not exhibit a breakdown of SI unlike Col-0. To determine if misregulation of the *SRKb* transcript is still occurring, real-time RT-PCR could be used to quantify expression of *SRKb* and *eSRK*. It is possible that this accession has a reduction in *SRKb* transcript, but still maintains self-incompatibility as seen in some of the methylation mutants screened in this study.

Since the misregulation of *SRKb* does not appear to attribute to the breakdown of SI in *nrpd1a* mutants, a search for factors involved this breakdown should be initiated. By fine mapping using the *nrpd1a* x C24 *SRKb-SCRb* F2 individuals, it may be possible to determine other factors necessary for SI in Col-0. The *nrpd1* x Sha *SRKb-SCRb* cross may also prove useful in this search.

The search for a transcription factor necessary for NRPD1a function was not successful. More T-DNA lines having insertions in genes homologous to *TFIIb* could be crossed and screened to determine if any of them exhibit a breakdown of SI similar to *nrpd1a*.

The promoter of *SRKb* was studied in several accessions of *A. thaliana* by transformation with *SRKb::GUS*. Most stable SI accessions had GUS staining in the stigmas, whereas the transient accessions did

not. In one case, the Sha accession that is normally fully SI did not have any GUS staining in any of the transformants observed. More transformants are necessary to determine if this is the general trend in this accession. Additionally, the Rld accession exhibited staining in the stigmas, albeit at a lower level that the full SI accessions. Real-time RT-PCR to analyze GUS expression levels across all of the accessions in the study would help to quantitate GUS to see if it is found at a lower level in transient SI accessions.

SUPPLEMENTARY STATISTICAL ANALYSES AND FUNCTIONAL CHARACTERIZATION OF RAPIDLY EVOLVING REPRODUCTIVE GENES

In Chapter 3 a number of putatively positively selected genes were identified. To gain further evidence of positive selection in the 12 candidate genes, it may be worthwhile to analyze linked regions that flank each gene. A selective sweep should leave reduced neutral variation in these regions if it occurred recently. Additionally, this analysis will help narrow down the site of selection. For the site-frequency spectrum tests, it is possible that a nearby region is under selection, causing the gene of interest to appear positively selected. Furthermore, dS was found to be quite variable when considering our top candidates. It may be that some of the candidates are found in areas of elevated or depressed synonymous substitutions. Looking at surrounding regions in the *A. thaliana* to *A. lyrata* comparisons may help to determine why dS is variable in these candidates.

A lack of substantial polymorphism in our sample for the self-

incompatibility-related At3g57840, the F-box domain-containing At3g59460, and the defensin-like At5g18403 and At5g23035 genes makes it difficult to draw any firm conclusions about these genes. They appear to be rapidly evolving based on ω, and their lack of polymorphism suggests that they may have undergone a recent selective sweep within the *A. thaliana* lineage. The addition of more samples and analysis of surrounding regions may help clarify this issue.

Demographic factors can make accurately detecting selection in *A. thaliana* difficult[40,51,52], so care must be taken when interpreting tests based upon a null assumption of neutrality. In an attempt to lessen these effects, distributions for all summary statistics based on a subset of accessions from a genome-wide sequencing effort[39] can be generated. Only sequence data for the accessions used in our analysis would be used to obtain distributions that were directly comparable to our selected subset of accessions.

To better address factors that may be involved in specifying mating system differences within the Brassicaceae, it would be worthwhile to sequence alleles of the candidate genes from *A. lyrata*. These sequences could then be analyzed in comparison to *Capsella* by McDonald-Krietman tests. It would be expected that genes involved in mating system differences between *A. thaliana* and *A. lyrata* would exhibit more similar patterns of evolution between two self incompatible species.

To determine specific sites under selection in the 12 candidate rapidly evolving and positively selected genes, sequence from

additional species must be obtained. *Arabidopsis halleri* should be included due to its close relation to *A. thaliana*. *Capsella* sequence may be available soon as a result of a genome-wide sequencing effort. Additionally, by including species from *Cardamine*, *Arabis*, and possibly *Lesquerella*, it should be possible to gain enough statistical power to detect positive selection at specific sites within the candidate genes. Polyploidy may complicate these sequencing efforts, so it will be important to confirm orthology by BLAST searches of the *A. thaliana* and *A. lyrata* genomes, DNA gel blot analysis, and the used of available information concerning synteny of the various genomes.

Analysis of mRNA would be useful for each of the candidate genes. Coding regions of the 12 candidate genes were predicted using mainly computational methods for the non-reference sequence generated. To assure these predictions are accurate, it would be useful to sequence cDNA prepared for each gene from each of the *A. thaliana* accessions and *A. lyrata*. Additionally, gene expression could be determined by RT-PCR in each of the accessions and species used to ensure that it is similar to that found in the reference genome.

T-DNA lines are available for most of the 12 candidate genes (Table A.10). For candidates with no available T-DNA lines, RNAi may be used to knockdown gene expression. These mutants could be analyzed on the basis of phenotype to see if they have inefficient pollinations/fertilization, as determined by low seed set, reduced pollen tube number, or disrupted pollen tubes guidance. Mutants may also allow interspecific pollinations more readily although this seems unlikely since pollen tubes seem to get to their target mainly

through guidance cues, not active blocks to fertilization (except for the block to polyspermy witnessed on the egg. Some of the candidate genes may have functionally redundant paralogs (Table A.10), in which case it will be necessary to perform crosses to generate double mutants.

APPENDIX

Table A.1. Primers used for transgene screening and mapping of *NRPD1a*.

Primer Name	Location (bp)	Primer 1 Sequence	Primer 2 Sequence
SRKbHVRIII	NA	TGGGTTGGGATGTCAAGAAAG	CAACTTCATCTTTCTCAGGCACAA
Chromosome 1			
NGA63	3,200,000	ACCCAAGTGATCGCCACC	AACCAAGGCACAGAAGCG
CIW12	9,621,344	AGGTTTTATTGCTTTTCACA	CTTTCAAAAGCACATCACA
SO392	10,860,075	GTTGATCGCAGCTTGATAAGC	TTTGGAGTTAGACACGGATCTG
T27K12-SP6	15,926,702	GGAGGCTATACGAATCTTGACA	GGACAACGTCTCAAACGGTT
GAPB.2	16,129,978	CACTATGTTCAGTGCTGCG	GATCACTTGCAGCTATGGC
CIW1	18,367,549	ACATTTTCTCAATCCTTACTC	GAGAGCTTCTTTATTTGTGAT
NGA280	20,939,226	GGCTCCATAAAAAGTGCACC	CTGATCTCACGGACAATAGTGC
F8A5a	22,294,831	GCAGAGCATAAAGCCATAAACA	CGTGCATGTTGTTGGAATCT
T13M11b	22,751,771	AAGTCCACAAAGGAGGAGAAA	CCCTTCCTAAGCCTAGATTTTGT
F24O1a	23,105,819	TGCTACCTGTTGCAACCTCA	TTTGCTTTCGATGTGAAATGA
F23N19a	23,235,281	GGAAGGAGCCTGAGGTAGAGA	CAGCATTCCCCAACTCTTTC
F16P17a	23,283,086	TGAGACATTTCCTACATTATTTATTTG	AATTACATAAGTTTGATGTTTGTGCT
F16P17b	23,307,069	CTTTGGCTTGGTCACTAATATGTTCA	TTCAATGCTGGTTCACAGAAA
F16P17c	23,352,777	CCTGATGTTGTTTCATTTCGATTAG	GTGCTCTTGGAATGGCTTGT
F16M19a	23,414,690	TCAGTAATCAGAGAGAGATTTAGGG	AACTCGCCATTGACACAACA
F9N12c_col	23,448,992	GGACATTATGTTGAACCTCGTTT	TTCAACAACCTTTTGGTAAATAAGA
F9N12c_ler	23,448,992	GACATTATGTTGAACCTCGTGA	TTCAACAACCTTTTGGTAAATAAGA
F9N12b	23,490,088	ACTGAGGACAAATGTTTTGTAGC	TCGAAACAAAGCAGGAGGTT
F2K11b	23,508,251	TCCAAATGATGATCCTGCAA	CCGGCTTATGCAGTAAGAAA
F2K11a	23,554,102	TGGTCGTCACTCTCGTTCAG	CAGATGAGACAATTCAGGGAGA

Table A.1 (continued)

F24D7a	23,630,160	TCACCCAATCTCCCCATAAG	CGAGCGCCTTACTCTGTGAT
F24D7b	23,633,080	AACCAGTTTTCAAATCAACTGAAG	TGGTGGTGTTGAGGTACCAA
T12P18b	23,677,880	CGTGCGCAGTTATCTCCTTT	TTATCACAGTTTTCATTTACCAAAAA
T12P18a	23,712,789	AGATAAATGCATCAACAAATTGAC	ACCCACCTCACACTCTCTCC
NGA111	27,418,736	TGTTTTTTAGGACAAATGGCG	CTCCAGTTGGAAGCTAAAGGG
Chromosome 2			
NGA1145	683,625	GCACATACCCACAACCAGAA	CCTTCACATCCAAAACCCAC
CIW3	6,409,928	GAAACTCAATGAAATCCACTT	TGAACTTGTTGTGAGCTTTGA
NGA1126	11,703,391	GCACAGTCCAAGTCACAACC	CGCTACGCTTTTCGGTAAAG
NGA168	16,298,919	GAGGACATGTATAGGAGCCTCG	TCGTCTACTGCACTGCCG
Chromosome 3			
NGA172	786,303	CATCCGAATGCCATTGTTC	AGCTGCTTCCTTATAGCGTCC
NGA162	460,8284	CTCTGTCACTCTTTTCCTCTGG	CATGCAATTTGCATCTGAGG
CIW11	9,775,545	CCCCGAGTTGAGGTATT	GAAGAAATTCCTAAAGCATTC
T27C7-SP6	13,047,019	ATGCCTAACTATTCGCTGAC	TTCTGTAGTTCTTTGTGAGTGC
F1P2-TGF	17,552,222	TTTGTCTGAAGATGTGGAGAGAGAG	CAAAACCCCACTCTTCATTATTGTT
NGA6	23,042,025	ATGGAGAAGCTTACACTGATC	TGGATTTCTTCCTCTCTTCAC
Chromosome 4			
CIW5	737,954	GGTTAAAAATTAGGGTTACGA	AGATTTACGTGGAAGCAAT
NGA8	5,628,810	TGGCTTTCGTTTATAAACATCC	GAGGGCAAATCTTTATTTCGG
CIW7	11,524,362	AATTTGGAGATTAGCTGGAAT	CCATGTTGATGATAAGCACAA
NGA1107	18,096,131	CGACGAATCGACAGAATTAGG	GCGAAAAAACAAAAAAATCCA
Chromosome 5			
CTR1.2	979,763	CCACTTGTTTCTCTCTCTAG	TATCAACAGAAACGCACCGAG
CA72	4,254,762	CCCAGTCTAACCACGACCAC	AATCCCAGTAACCAAACACACA
NGA139	8,428,136	GGTTTCGTTTCACTATCCAGG	AGAGCTACCAGATCCGATGG
ATHPHYC	14,025,127	CTCAGAGAATTCCCAGAAAAATCT	AAACTCGAGAGTTTTGTCTAGATC

Table A.1 (continued)

CIW9	17,061,229	CAGACGTATCAAATGACAAATG	GACTACTGCTCAAACTATTCGG
MTH12	24,100,521	GTAAAATTTTCTATTGCA	ATGTCCTCCTGTTCTGTCCA

Table A.2. Primers used in T-DNA screening of methylation mutants.

T-DNA line	Associated gene	Primer 1	Primer 2
SALK_143437	*nrpd1a*	TTTGGAACTGGATGCTATTCG	AACCTGATGGCATGCATAAAG
SALK_128428	*nrpd1a*	AAAAGGGATCAAAACGAGACG	TTAATGTTCTTCATGCGGGAC
nrpd1a (rdm5)	*nrpd1a*	ATTCAGAAGCGGCGAAGTT	TTCCGCTCAAGAAATGCTATC
SALK_005512	*dcl3*	TTGGTTCAAAACTAAGAAATGGAAG	TTGGTTCAAAACTAAGAAATGGAAG
SALK_059661	*rdr2*	CTGATCGCGAGATTTCAGTTC	AGAAGATTGGAGCAAGCTTCC
SALK_071772	*ago4*	TTCTCCAGCTGGCTAGCTATG	CCCACAAAATCAAAGTGAAGAAG
SALK_021316	*drm1*	GAGCCGTCTCATCAAACTGAC	GGCAAGCACTGAAGCTAGTTC
SALK_039030	*drm2*	TGACAAGCAACCCAAGAGATC	ATAAACTTCTCGCTCGTCACG
SALK_150863	*drm2*	CTATTTCCAGGCACTTGTTCG	AAGATCCTCTCATCCTCGCAC
SALK_132061	*drd1*	AACGCTTAACCAAATCCATCC	TGGTAATATGACATTAATGTGTTTAACC
SALK_076137	*nrpd1b*	ATTTCACATCAGGCATAAGGC	AACTGACTCAGAACCTGCTGC
SALK_076129	*nrdp1b*	ATTTCACATCAGGCATAAGGC	AACTGACTCAGAACCTGCTGC
SALK_095689	*nrdp2a*	TGTTTAAACTTGGAGACTTTGGC	TCGGATAACCCTTACCGAATC
SALK_057148	*cmt3*	TCGCTTCAGATTCTGGAATTG	TTGTGGGAATTTATTCAAATCC

Table A.3. Primers used for transcription factor candidate screening.

T-DNA Line	Associated Gene	Primer 1	Primer 2
SALK_146224	AT2G45100	CGATGAAATGTAACTTACGAGAGC	TTACACCACACCATAGCTCCC
SALK_072770	AT3G09360	ACGCAATGGTTACACCATACC	AAAGCGATGAGAAACTGAGGC
SALK_086510	AT3G10330	ACCCATAATCCCAATCGAAAC	TGCTGATGGTTTCTTGGAATC

Table A.4. Primers used for PCR and sequencing of selected candidates.

Primer Name	Forward Primer	Primer Name	Reverse Primer
1g20730aF	GAGTCCTCTTTACTCCTTACTCAATCA	1g20730aR	CAATCCCGTGAGAATCCAGA
1g20730bF	GCCGAGTGTCGTACTCTTGA	1g20730bR	TGCACGTTCCAGTCATCATT
1g20730cF	ACGACAGCATTGGTTGTGAG	1g20730bcR	TTCATACGCAAAACTTGATCACT
1g20730dF	GAATTTTTATCAAACAAATCACCA	1g20730dR	TTCGTAATTGGACTGGGACA
1g20730eF	TGAATTTCCTAGGAGGTTTAAAATAA	1g20730eR	GATTTGCTTACCAGCTTCTTGTT
1g20730fF	TGGGAAGAGTGTCCTTGAAT	1g20730fR	CAGGTTTCATCCATGTCGTG
1g20730gF	ACGGGTTCGAGATCAGTGAC	1g20730gR	TCATCAGCTTTTGGTACATTTGTT
2g06090F	AAGATTTTACCAACCATATATTTTCAC	2g06090R	ATTATGAATAATCTTTTTGTTTTGCTC
2g20597F	GGATTCTAAGACGATCAAAGATGG	2g20597aR	TGGGCACATCGCTCTGCACATT
2g20597bF	GCAGGAAACACGACGACGGC	2g20597R	CAACAAATCTACGACGAAAATG
3g48231F[a]	GTTTTAATAGACCATAGTACGGATCA	3g48231R[a]	TCTTCAATAGACTTCAAAACCAAAAA
3g50020aF	CGGCTTAATTTGTTCAATAAGCAT	3g50020aR	TGAATCGTAATTGTCCGCTCT
3g50020bF	CTCATCCGATGTGAGGAAAA	3g50020bR	TCTTTTTGGATGAAACCCATT
3g57840F[b]	GAATAAAAGATGGCTTCCCAAA	3g57840aR	AGCGAGTTCGTCCCCAGTAAGT
3g57840bF	TCGGAGGTGGTTTGGTGCTACA	3g57840R[b]	AATTGGTTTATTTATCCCAAGCA
3g59460F	ATAATTAACTGTCACACAAACATGG	3g59460R	GCAATACATGGTATTATCCACCA
4g21326aF	AGAATATTGGGAATTGTGAAGG	4g21316aR	TTTGCTGGGTCAAAATCCTC
4g21326bF	GGTTTTCTTGACTCAGGAGTATGG	4g21316bR	TGCCATTTCCCAATCTTCTT
4g21326cF	TGAACACTATAAAACCGACACCTC	4g21316cR	GTTCTCCCTCCGCAAAGATT
4g21326dF	TTGTTATTATAGCATGGAAGACTGA	4g21316dR	CGAATTGAGGAATCTTCTTATTCAA
5g09370F	CTGTCGGGTTTTAGATACGG	5g09370R	TAAAACATGGCTTACTTCTCCACA
5g18403F	TTCTAAGGAAAACAAAATGAAGACAA	5g18403R	CAATGACAAAGTTTTTGTAACTACGG
5g23035F	AACACGTAAATTTGGTTGTGAA	5g23035R	TGGAGAAAATGATGTTGAGCA
5g39570aF	TCCACCTTCCGATGAGACTT	5g39570aR	TTCCTACGCTGCTCATCATC

Table A.4 (continued)

5g39570bF	GAGTTACAGGAAGCAGCCTAGC	5g39570bR	GATTCAAACAAACACAACACCAA
5g46960F	ACAACATATTAGTAAAAGCCAAAGGA	5g46960R	AACTTTCTCTGCTTCTTCTCACTCTC
5g46960aF	CAATCGATGTTATATGAAACACTTTG	5g46960aR	GAACATTAAATCGCGTGATTATCAA
5g46960bF	TCGCAATTGCCTGGTACATC	5g46960bR	GCTTCTTCTCAAAGTGAAAACA

[a]This primer pair was used to generate a probe template for At3g48231.
[b]This primer pair was used to generate a probe template for At3g57840.

Table A.5. Candidate rapidly evolving genes (ω>0.5).

AGI number	Gene Model Description	Data Set	Sig Pep	TM Dom	dN	dS	w	p w>0.5	p w>1.0
Cysteine-rich/Defensin-like									
AT1G30972	Encodes a Plant thionin family protein	ovary	Y	N	0.044	0.039	1.102	0.14	0.86
AT1G35537	Encodes a defensin-like (DEFL) family protein.	ovary	Y	N	0.079	0.198	0.590	0.77	1.00
AT1G53282	Encodes a Plant thionin family protein	ovary	Y	Y	0.072	0.221	0.837	0.34	1.00
AT1G58055	Encodes a defensin-like (DEFL) family protein.	ovary	Y	Y	0.102	0.048	4.083	0.07	0.23
AT1G60985	SCRL6 (SCR-Like 6)	ovary	Y	N	0.198	0.122	1.334	0.06	0.60
AT1G73603	LCR64 (Low-molecular-weight cysteine-rich 64)	ovary	Y	Y	0.174	0.219	0.540	0.88	0.88
AT1G73607	LCR65 (Low-molecular-weight cysteine-rich 65)	ovary	Y	Y	0.064	0.061	1.078	0.19	0.90
AT2G12475	Encodes a defensin-like (DEFL) family protein.	ovary	Y	N	0.045	0.098	0.686	0.66	1.00
AT2G15535	LCR10 (Low-molecular-weight cysteine-rich 10)	pollen tube	Y	Y	0.182	0.109	1.673	0.04	0.39
AT2G20070	similar to LCR81 (Low-molecular-weight cysteine-rich 81)	ovary	Y	Y	0.078	0.072	1.123	0.28	0.88
AT2G20597	Encodes a Plant thionin family protein	ovary	Y	Y	0.113	0.101	1.255	0.22	0.78
AT2G21465	Encodes a DEFL protein	ovary	Y	Y	0.099	0.195	0.720	0.46	1.00

Table A.5 (continued)

AT2G24693	Encodes a defensin-like (DEFL) family protein.	ovary	Y	Y	0.115	0.112	1.036	0.22	0.95
AT2G25185	Encodes a defensin-like (DEFL) family protein.	ovary	Y	N	0.076	0.119	0.660	0.61	1.00
AT3G04540	Encodes a defensin-like (DEFL) family protein.	ovary	N	N	0.113	0.118	0.669	0.67	1.00
AT3G07005	LCR43 (Low-molecular-weight cysteine-rich 43)	ovary	Y	Y	0.042	0.120	1.033	0.52	0.98
AT3G48205	Encodes a Plant thionin family protein	ovary	Y	N	0.083	0.182	0.800	0.32	1.00
AT3G48231	LCR48 (Low-molecular-weight cysteine-rich 48)	ovary	Y	N	0.100	0.064	1.892	0.03	0.32
AT3G61182	LCR54 (Low-molecular-weight cysteine-rich 54)	ovary	Y	N	0.151	0.097	1.538	0.07	0.51
AT4G08875	Encodes a defensin-like (DEFL) family protein.	ovary	Y	N	0.065	0.192	0.537	0.91	1.00
AT4G13955	Encodes a defensin-like (DEFL) family protein.	ovary	Y	Y	0.071	0.193	0.707	0.53	1.00
AT4G17713	Encodes a defensin-like (DEFL) family protein.	ovary	Y	N	0.060	0.111	0.639	0.67	1.00
AT4G17718	Encodes a defensin-like (DEFL) family protein.	ovary	Y	Y	0.080	0.145	0.827	0.29	1.00
AT4G30067	LCR63 (Low-molecular-weight cysteine-rich 63)	ovary	Y	Y	0.100	0.167	0.746	0.50	1.00
AT4G30070	LCR59 (Low-molecular-weight cysteine-rich 59)	ovary	Y	Y	0.081	0.088	0.740	0.32	1.00
AT4G39917	LCR45	ovary	Y	N	0.099	0.117	0.546	0.90	1.00

Table A.5 (continued)

AT5G01870	Predicted to encode a PR (pathogenesis-related) protein.	stigma	Y	N	0.102	0.112	1.013	0.09	0.09
AT5G08315	Encodes a defensin-like (DEFL) family protein.	ovary	Y	N	0.021	0.000	2.136	0.02	0.02
AT5G18403	Encodes a defensin-like (DEFL) family protein.	ovary	Y	N	0.156	0.143	1.254	0.18	0.25
AT5G23035	Encodes a defensin-like (DEFL) family protein.	ovary	Y	N	0.087	0.042	2.556	0.00	0.12
AT5G36805	Encodes a Plant thionin family protein	pollen tube	Y	Y	0.108	0.115	0.937	0.20	1.00
AT5G37474	Encodes a defensin-like (DEFL) family protein.	ovary	Y	N	0.063	0.106	0.515	0.96	0.96
AT5G43285	Encodes a defensin-like (DEFL) family protein.	ovary	Y	N	0.045	0.198	0.578	0.78	1.00
AT5G48515	Encodes a defensin-like (DEFL) family protein.	ovary	Y	N	0.132	0.106	1.250	0.10	0.70
AT5G48595	Encodes a defensin-like (DEFL) family protein.	ovary	Y	Y	0.052	0.079	0.582	0.78	1.00
AT5G60553	Encodes a defensin-like (DEFL) family protein.	ovary	Y	N	0.061	0.126	0.572	0.78	1.00
AT5G63063	Encodes a Plant thionin family protein	ovary	Y	Y	0.150	0.094	1.601	0.13	0.27
AT5G63065	Encodes a Plant thionin family protein	ovary	N	Y	0.105	0.108	0.960	0.27	1.00
AT5G63087	Encodes a Plant thionin family protein	ovary	Y	Y	0.098	0.078	1.498	0.27	0.70

Table A.5 (continued)

DNA/RNA/Protein-binding									
AT1G08680	ZIGA4 (ARF GAP-LIKE ZINC FINGER-CONTAINING PROTEIN ZIGA4)	pollen tube	N	N	0.044	0.067	0.655	0.24	1.00
AT1G15940	binding	pollen tube	N	N	0.053	0.065	0.818	0.02	1.00
AT1G17790	DNA-binding bromodomain-containing protein	pollen tube	N	N	0.034	0.060	0.563	0.75	1.00
AT1G26540	agenet domain-containing protein	ovary	N	N	0.042	0.084	0.528	0.78	1.00
AT1G30814	contains domain Glucocorticoid receptor-like	ovary	N	N	0.083	0.126	0.696	0.37	1.00
AT1G65300	AGAMOUS-LIKE 38 (AGL38);	pollen tube	N	N	0.080	0.095	0.835	0.10	1.00
AT1G70895	CLE17 (CLAVATA3/ESR-RELATED 17); receptor binding	pollen tube	N	Y	0.105	0.080	1.312	0.13	0.69
AT1G77250	PHD finger family protein	pollen	N	N	0.040	0.070	0.578	0.60	1.00
AT2G20400	myb family transcription factor	pollen tube	N	N	0.068	0.095	0.716	0.26	1.00
AT2G32370	DNA binding / transcription factor	pollen tube	N	N	0.083	0.114	0.730	0.04	1.00
AT2G34440	AGL29 (AGAMOUS-LIKE 29); transcription factor	pollen	N	N	0.047	0.056	0.829	0.30	1.00
AT2G35670	FIS2 (FERTILIZATION INDEPENDENT SEED 2); transcription factor	ovary	N	N	0.043	0.163	0.734	0.08	1.00

Table A.5 (continued)

AT2G46240	BAG6	pollen tube	N	N	0.073	0.139	0.523	0.76	1.00
AT2G46610	arginine/serine-rich splicing factor, putative	pollen tube	N	N	0.041	0.069	0.588	0.68	1.00
AT3G13590	DC1 domain-containing protein	pollen tube	N	N	0.089	0.137	0.652	0.30	1.00
AT3G19510	homeobox protein (HAT 3.1)	pollen tube	N	N	0.039	0.057	0.695	0.20	1.00
AT3G43480	nucleic acid binding / zinc ion binding	ovary	N	Y	0.059	0.084	0.795	0.28	1.00
AT3G57370	transcription factor IIB (TFIIB) family protein	pollen tube	N	N	0.104	0.118	0.885	0.28	1.00
AT4G00870	basic helix-loop-helix (bHLH) family protein	ovary	Y	N	0.051	0.113	0.545	0.72	1.00
AT4G14530	similar to AGL97	pollen tube	N	N	0.137	0.207	0.663	0.51	1.00
AT4G25530	FWA (FWA); DNA binding / transcription factor	ovary	N	N	0.085	0.055	1.157	0.00	0.47
AT4G30480	tetratricopeptide repeat (TPR)-containing protein	pollen tube	N	N	0.046	0.082	0.561	0.77	1.00
AT5G46550	DNA-binding bromodomain-containing protein	pollen tube	N	N	0.092	0.146	0.632	0.32	1.00
AT5G60140	transcriptional factor B3 family protein	ovary	N	N	0.024	0.128	0.971	0.05	1.00
AT5G66980	B3 family protein	pollen tube	N	N	0.076	0.146	0.522	0.88	1.00

Table A.5 (continued)

ECA1									
AT3G01327	Encodes a ECA1 gametogenesis related family protein	ovary	Y	N	0.090	0.136	0.524	0.90	1.00
AT3G01331	Encodes a ECA1 gametogenesis related family protein	ovary	Y	N	0.145	0.094	1.638	0.02	0.37
AT3G30247	Encodes a ECA1 gametogenesis related family protein	ovary	Y	N	0.074	0.229	0.513	0.94	1.00
AT5G36310	Encodes a ECA1 gametogenesis related family protein	ovary	Y	Y	0.074	0.107	0.693	0.44	1.00
AT5G36340	Encodes a ECA1 gametogenesis related family protein	ovary	Y	Y	0.051	0.194	0.693	0.44	1.00
AT5G36370	Encodes a ECA1 gametogenesis related family protein	ovary	Y	Y	0.074	0.107	0.647	0.54	1.00
AT5G42567	Encodes a ECA1 gametogenesis related family protein	ovary	Y	Y	0.042	0.127	0.664	0.56	1.00
AT5G60964	Encodes a ECA1 gametogenesis related family protein	ovary	Y	N	0.068	0.195	0.592	0.75	1.00

Table A.5 (continued)

Enzymes/Inhibitors									
AT1G09370	enzyme inhibitor/ pectinesterase	ovary	Y	N	0.094	0.173	0.695	0.45	1.00
AT1G09720	kinase interacting family protein	pollen tube	N	N	0.053	0.096	0.555	0.57	1.00
AT1G34460	cyclin-dependent protein kinase regulator	pollen tube	N	N	0.054	0.091	0.588	0.54	1.00
AT1G47780	acyl-protein thioesterase-related	ovary	N	N	0.174	0.162	1.173	0.17	0.81
AT1G50325	enzyme inhibitor/ pectinesterase	ovary	Y	Y	0.136	0.101	1.193	0.00	0.58
AT1G53690	DNA-directed RNA polymerases I, II, and III 7 kDa subunit, putative	pollen tube	N	N	0.075	0.001	-	0.03	0.12
AT1G62130	AAA-type ATPase family protein	pollen tube	N	N	0.075	0.109	0.687	0.05	1.00
AT2G05850	Serine carboxypeptidase S10 family protein	pollen-pcp	Y	N	0.078	0.135	0.576	0.58	1.00
AT2G07698	ATP synthase alpha chain, mitochondrial, putative	pcp	Y	Y	0.010	0.007	1.555	0.14	0.14
AT2G36325	hydrolase, acting on ester bonds	ovary	Y	N	0.029	0.086	0.580	0.58	1.00
AT2G40500	protein kinase family protein	pollen tube	N	N	0.100	0.168	0.593	0.58	1.00
AT2G40560	protein kinase family protein	pollen tube	N	N	0.106	0.163	0.648	0.34	1.00

Table A.5 (continued)

AT2G47690	NADH-ubiquinone oxidoreductase-related	pollen	N	N	0.026	0.049	0.537	0.94	1.00
AT3G17150	pectinesterase inhibitor	ovary	Y	N	0.020	0.153	0.674	0.36	1.00
AT3G17225	invertase/pectin methylesterase inhibitor family protein	ovary	Y	N	0.063	0.061	1.043	0.09	0.93
AT3G17230	invertase/pectin methylesterase inhibitor family protein	ovary	Y	N	0.128	0.116	1.056	0.00	0.84
AT3G30540	(1-4)-beta-mannan endohydrolase family	ovary	Y	N	0.024	0.134	0.517	0.90	1.00
AT3G56540	serine carboxypeptidase, putative	pollen tube	Y	N	0.041	0.074	0.557	0.80	1.00
AT4G12960	gamma interferon responsive lysosomal thiol reductase	pollen	Y	N	0.076	0.118	0.644	0.40	1.00
AT4G16970	kinase	pollen tube	N	N	0.052	0.094	0.558	0.56	1.00
AT4G21326	ATSBT3.12 (SUBTILASE 3.12); identical	pollen-pollen tube			0.053	0.085	0.626	0.23	1.00
AT4G23710	VAG2; hydrolase, acting on acid	pollen	N	N	0.026	0.050	0.522	0.95	1.00
AT4G32090	galactosyltransferase	ovary	Y	Y	0.055	0.204	0.516	0.95	1.00
AT5G01780	oxidoreductase, 2OG-Fe(II) oxygenase family protein	pollen tube	N	N	0.079	0.152	0.520	0.88	1.00

Table A.5 (continued)

AT5G17100	CONTAINS InterPro DOMAIN/s: Cystatin-related, plant	ovary	N	N	0.098	0.203	0.570	0.72	1.00
AT5G18990	pectinesterase family protein	ovary	Y	N	0.018	0.199	0.521	0.88	1.00
AT5G43580	serine-type endopeptidase inhibitor	ovary	N	N	0.076	0.061	1.411	0.33	0.76
AT5G46960	invertase/pectin methylesterase inhibitor family protein	ovary	Y	N	0.019	0.104	0.924	0.09	1.00
AT5G58400	peroxidase, putative	ovary	Y	Y	0.127	0.142	0.534	0.93	1.00
AT5G58820	subtilase family protein	ovary	Y	N	0.029	0.101	0.580	0.43	1.00
AT5G58830	subtilase family protein	ovary	Y	Y	0.015	0.120	0.508	0.94	1.00
F-box									
AT1G20730	Protein of unknown function DUF833 (InterPro:IPR008551), F-box associated type 1 (InterPro:IPR017451	ovary	N	Y	0.107	0.038	1.971	0.00	0.02
AT2G17690	F-box family protein	ovary	N	N	0.240	0.108	1.601	0.01	0.29
AT2G43270	F-box family protein	pollen tube	N	Y	0.126	0.248	0.508	0.94	1.00
AT3G26930	similar to F-box family protein	pollen tube	N	N	0.093	0.175	0.532	0.79	1.00
AT3G49450	F-box family protein	pollen tube	N	N	0.099	0.171	0.576	0.50	1.00

Table A.5 (continued)

AT3G59460	similar to F-box family protein	pollen tube	Y	N	0.109	0.176	0.619	0.62	1.00
AT3G61340	F-box family protein	pollen tube	N	N	0.055	0.108	0.508	0.95	1.00
AT5G38390	F-box family protein	pollen tube	N	N	0.100	0.184	0.545	0.68	1.00
Glycine-rich									
AT2G05380	glycine-rich protein 3 short isoform (GRP3S) mRNA, complete	trans tract	Y	Y	0.086	0.145	0.894	0.14	1.00
AT3G04640	glycine-rich protein	pollen tube	Y	N	0.089	0.169	0.527	0.87	1.00
AT5G07510	encodes a glycine-rich protein that is expressed in low abundance in stems and leaves, and very low abundance in flowers.	pcp	Y	Y	0.080	0.056	1.436	0.02	0.42
AT5G07520	encodes a glycine-rich protein that is expressed only in flowers during a specific developmental stage (flower stage 12).	pcp	Y	Y	0.041	0.068	0.600	0.69	1.00
AT5G07540	encodes a glycine-rich protein that is expressed only in flowers during a specific developmental stage (flower stages 11 and 12).	pcp	Y	Y	0.100	0.105	0.954	0.05	1.00

Table A.5 (continued)

Glycoprotein									
AT1G63530	similar to hydroxyproline-rich glycoprotein family protein	pollen tube	N	N	0.085	0.091	0.929	0.01	1.00
AT1G63540	hydroxyproline-rich glycoprotein family protein	pollen tube	N	N	0.078	0.089	0.878	0.06	1.00
AT3G13520	AGP12 (ARABINOGALACTAN PROTEIN 12)	pollen tube	Y	Y	0.032	0.042	0.777	0.61	1.00
AT3G57690	AGP23 (ARABINOGALACTAN-	pollen	Y	Y	0.017	0.032	0.535	0.95	1.00
AT3G60380	similar to hydroxyproline-rich glycoprotein family protein	pollen tube	Y	Y	0.068	0.104	0.653	0.23	1.00
AT5G53250	AGP22/ATAGP22 (ARABINOGALACTAN PROTEINS 22)	pollen tube	Y	Y	0.023	0.020	1.122	0.46	0.92
Metal Ion Binding									
AT1G07600	metallothionein, binds to and detoxifies excess copper and other metals, limiting oxidative damage.	pollen-trans tract	N	N	0.043	0.000	-	0.02	0.07
AT1G63950	heavy-metal-associated domain-containing protein	ovary	N	N	0.050	0.048	1.025	0.15	0.96
AT1G79800	plastocyanin-like domain-containing protein	ovary	Y	Y	0.160	0.182	0.676	0.65	1.00

Table A.5 (continued)

AT4G31360	selenium binding	pollen tube	N	N	0.056	0.066	0.853	0.36	1.00
AT5G52760	heavy-metal-associated domain-containing protein	pollen tube	N	N	0.068	0.130	0.523	0.93	1.00
Protease									
AT1G72290	trypsin and protease inhibitor family protein / Kunitz family protein	ovary-trans tract	Y	N	0.067	0.162	0.660	0.43	1.00
AT2G28010	aspartyl protease family protein	ovary	Y	Y	0.065	0.133	0.600	0.43	1.00
AT2G38870	protease inhibitor, putative	pollen tube	N	N	0.063	0.117	0.535	0.90	1.00
AT3G50020	serine protease inhibitor, potato inhibitor I-type family protein	ovary	N	N	0.201	0.079	2.814	0.00	0.07
AT3G57310	protease inhibitor/seed storage/lipid transfer protein (LTP) family protein	ovary	Y	N	0.244	0.159	2.021	0.05	0.34
AT4G00690	Ulp1 protease family protein	pollen tube	N	N	0.065	0.098	0.665	0.44	1.00
AT4G22050	aspartyl protease family protein	ovary	Y	N	0.121	0.244	0.915	0.17	1.00
AT5G09370	protease inhibitor/seed storage/lipid transfer protein (LTP) family protein	ovary-pollen tube	Y	N	0.142	0.064	2.084	0.00	0.16

Table A.5 (continued)

Ribosomal									
AT1G15810	ribosomal protein S15 family protein	pollen tube	N	N	0.056	0.091	0.610	0.53	1.00
AT3G11120	60S ribosomal protein	pollen tube	N	N	0.000	0.000	0.518	1.00	100
AT1G54740	BEST Arabidopsis thaliana protein match is: structural constituent of ribosome (TAIR:AT1G22110.1)	pollen	N	N	0.088	0.176	0.501	1.00	1.00
Proline-rich									
AT3G49307	similar to proline-rich family protein	ovary	Y	N	0.082	0.077	1.191	0.25	0.83
AT4G19200	proline-rich family protein	pollen	N	N	0.039	0.078	0.508	0.97	1.00
AT5G01280	similar to proline-rich family protein	pollen tube	N	N	0.069	0.108	0.643	0.35	1.00
AT5G26080	proline-rich family protein	pollen tube	Y	Y	0.088	0.017	5.178	0.01	0.10
Self-incompatibility									
AT1G04645	self-incompatibility protein-related	ovary	Y	Y	0.000	0.125	0.923	0.20	1.00

Table A.5 (continued)

AT1G56385	BEST Arabidopsis thaliana protein match is: self-incompatibility protein-related	ovary	Y	N	0.110	0.088	2.141	0.17	0.49
AT2G06090	self-incompatibility protein-related	ovary	Y	N	0.035	0.182	0.678	0.39	1.00
AT3G10460	self-incompatibility protein-related	pollen tube	Y	Y	0.107	0.142	0.751	0.29	1.00
AT3G24065	contains InterPro domain Plant self-incompatibility S1; (InterPro:IPR010264)	ovary	Y	Y	0.048	0.024	1.554	0.01	0.33
AT3G24068	contains InterPro domain Plant self-incompatibility S1; (InterPro:IPR010264)	ovary	Y	Y	0.013	0.079	0.573	0.72	1.00
AT3G26880	self-incompatibility protein-related	pollen tube	Y	N	0.095	0.152	0.626	0.59	1.00
AT3G57840	self-incompatibility protein-related	ovary	Y	Y	0.088	0.041	2.644	0.05	0.30
AT5G12060	self-incompatibility protein-related	ovary	Y	N	0.070	0.239	0.884	0.19	1.00
AT5G12070	self-incompatibility protein-related	pollen tube	Y	N	0.069	0.062	1.112	0.08	0.82
Ubiquitination									
AT1G36340	UBC31 (ubiquitin-conjugating enzyme 3); ubiquitin-protein ligase	ovary	N	N	0.043	0.228	0.781	0.23	1.00

Table A.5 (continued)

AT2G03160	ASK19 (ARABIDOPSIS SKP1-LIKE 19); ubiquitin-protein ligase	pollen tube	N	N	0.142	0.206	0.690	0.45	1.00
AT5G18340	U-box domain-containing protein	pollen tube	N	N	0.124	0.108	1.149	0.00	0.52
Zinc Finger									
AT2G24460	similar to zinc finger (C3HC4-type RING finger) family protein [Arabidopsis thaliana] (TAIR:AT2G24480.1)	pollen tube	N	N	0.077	0.041	1.861	0.19	0.56
AT3G23060	zinc finger (C3HC4-type RING finger) family protein	pollen tube	N	N	0.087	0.119	0.734	0.09	1.00
AT3G23140	zinc finger (C2H2 type) family protein	ovary	N	N	0.052	0.104	0.537	0.89	1.00
AT3G43470	similar to zinc knuckle (CCHC-type) family protein	ovary	N	N	0.106	0.092	1.545	0.05	0.47
AT4G25380	zinc finger (AN1-like) family protein	pollen tube	N	N	0.049	0.073	0.672	0.57	1.00
Other									
AT1G02430	ADP-ribosylation factor D1B); GTP binding	pollen tube	N	N	0.050	0.079	0.629	0.58	1.00

Table A.5 (continued)

AT1G10710	PHS1 regulates recombination and pairing of homologous chromosomes during meiotic prophase by controlling transport of RAD50 from cytoplasm to the nucleus.	pollen tube	N	N	0.065	0.077	0.841	0.06	1.00
AT1G10745	Encodes a Maternally expressed gene (MEG) family protein	ovary	Y	N	0.034	0.049	0.569	0.88	1.00
AT1G12180	similar to heat shock protein-related	pollen tube	N	N	0.174	0.153	1.138	0.07	0.78
AT1G35467	RALFL5 (RALF-LIKE 5)	ovary	Y	Y	0.096	0.128	0.938	0.24	1.00
AT1G51915	cryptdin protein-related	pollen tube	Y	N	0.047	0.063	0.750	0.54	1.00
AT1G54300	catalytic/ cobalamin binding];[AT1G54300, similar to catalytic/ cobalamin binding	pollen tube	N	N	0.115	0.099	1.163	0.01	0.65
AT1G59660	nucleoporin family protein	pollen tube	N	N	0.051	0.086	0.593	0.35	1.00
AT1G70030	paired amphipathic helix repeat-containing protein	ovary	N	N	0.066	0.098	0.663	0.50	1.00

Table A.5 (continued)

AT1G70782	Upstream open reading frames (uORFs) are small open reading frames found in the 5' UTR of a mature mRNA, and can potentially mediate translational regulation of the largest, or major, ORF (mORF). CPuORF28 represents a conserved upstream opening reading frame relative to major ORF AT1G70780.1	pollen tube	N	N	0.010	0.021	0.504	1.00	1.00
AT1G72800	nuM1-related	pollen tube	N	N	0.042	0.065	0.652	0.63	1.00
AT1G77300	Encodes a protein with histone lysine N-methyltransferase activity required specifically for the trimethylation of H3-K4 in FLC chromatin (and not in H3-K36 dimethylation).	pollen	N	N	0.045	0.082	0.548	0.49	1.00
AT2G03020	[AT4G16540, heat shock protein-related];[AT2G03020, heat shock protein-related]	pollen tube	N	N	0.081	0.109	0.745	0.25	1.00

Table A.5 (continued)

AT3G04160	similar to SWAP (Suppressor-of-White-APricot)/surp domain-containing protein	ovary	N	N	0.045	0.102	0.667	0.17	1.00
AT3G05770	catalytic/ cobalamin binding	pollen tube	N	N	0.113	0.097	1.162	0.00	0.56
AT3G49860	ATARLA1B (ADP-ribosylation factor-like A1B); GTP binding	pollen tube	N	N	0.065	0.090	0.722	0.38	1.00
AT3G57250	emsy N terminus domain-containing protein / ENT domain-containing protein	ovary	N	N	0.035	0.230	0.882	0.12	1.00
AT3G63180	ATTKL/TKL (TICKLE)	pollen tube	N	N	0.045	0.090	0.501	1.00	1.00
AT4G00760	APRR8 (PSEUDO-RESPONSE REGULATOR 8); transcription regulator	pollen tube	N	N	0.052	0.037	1.406	0.01	0.42
AT4G02380	SAG21 (SENESCENCE-ASSOCIATED	pollen	N	N	0.010	0.015	0.664	0.82	1.00
AT4G11510	RALFL28 (RALF-LIKE 28)	ovary	Y	Y	0.048	0.145	0.550	0.85	1.00
AT4G18300	eIF4-gamma/eIF5/eIF2-epsilon domain-containing protein	pollen tube	N	N	0.068	0.125	0.542	0.66	1.00
AT4G35370	transducin family protein / WD-40 repeat family protein	pollen tube	N	N	0.061	0.074	0.823	0.08	1.00
AT4G35890	La domain-containing protein	pollen	N	N	0.049	0.093	0.527	0.82	1.00

Table A.5 (continued)

AT5G07550	member of Oleosin-like protein family	pcp	Y	Y	0.057	0.059	0.952	0.21	1.00
AT5G23470	[AT2G02290, NLI interacting factor (NIF) family protein];[AT5G23470, NLI interacting factor (NIF) family protein]	pollen tube	N	N	0.098	0.156	0.628	0.38	1.00
AT5G27495	ion channel inhibitor	ovary	Y	Y	0.183	0.160	1.255	0.13	0.71
Unknown									
AT1G06420	unknown protein	ovary	N	N	0.071	0.048	1.762	0.25	0.62
AT1G14455	unknown protein	ovary	Y	N	0.099	0.092	1.088	0.16	0.89
AT1G17780	unknown protein	ovary	N	N	0.043	0.093	0.602	0.57	1.00
AT1G20680	unknown protein	ovary	N	N	0.111	0.083	1.180	0.00	0.59
AT1G21950	unknown protein	ovary	N	Y	0.044	0.038	1.116	0.05	0.80
AT1G22890	unknown protein	stigma	Y	Y	0.082	0.135	0.597	0.72	1.00
AT1G27610	unknown protein	pollen tube	N	N	0.104	0.186	0.557	0.83	1.00
AT1G51670	unknown protein	pollen tube	N	N	0.071	0.101	0.708	0.34	1.00
AT1G52910	unknown protein	ovary	Y	Y	0.054	0.178	0.570	0.76	1.00
AT1G56415	expressed protein	ovary	Y	N	0.082	0.151	0.528	0.92	1.00
AT1G56418	unknown protein	ovary	Y	N	0.031	0.054	0.614	0.77	1.00
AT1G59835	unknown protein	ovary	Y	N	0.026	0.019	1.139	0.02	0.73
AT1G67350	unknown protein	pollen	N	Y	0.064	0.053	1.212	0.08	0.72
AT1G77525	unknown protein	ovary	Y	N	0.205	0.224	0.817	0.33	1.00
AT2G14800	unknown protein	pollen tube	N	N	0.157	0.069	2.268	0.00	0.14

Table A.5 (continued)

AT2G19700	unknown protein	pollen tube	N	N	0.173	0.122	1.422	0.01	0.37
AT2G20820	unknown protein	pollen	Y	N	0.029	0.049	0.585	0.82	1.00
AT2G25250	unknown protein	pollen tube	N	N	0.118	0.225	0.524	0.90	1.00
AT2G27315	unknown protein	ovary	Y	Y	0.060	0.247	0.504	0.98	1.00
AT2G32190	unknown protein	pollen tube	N	N	0.051	0.081	0.635	0.70	1.00
AT2G36695	unknown protein	ovary	Y	N	0.038	0.088	0.974	0.27	1.00
AT2G39435	unknown protein	pollen tube	N	N	0.056	0.102	0.555	0.71	1.00
AT2G41440	unknown protein	pollen	N	N	0.072	0.065	1.105	0.02	0.78
AT3G03828	unknown protein	ovary	Y	Y	0.025	0.052	0.506	0.98	1.00
AT3G11880	unknown protein	pollen tube	N	N	0.124	0.032	3.851	0.00	0.00
AT3G22240	unknown protein	stigma -trans tract	Y	N	0.067	0.054	1.268	0.16	0.77
AT3G28780	unknown protein	pollen tube	Y	Y	0.055	0.099	0.553	0.77	1.00
AT3G28790	unknown protein	pollen tube	Y	Y	0.066	0.113	0.583	0.59	1.00
AT3G28830	unknown protein	pollen tube	Y	Y	0.050	0.099	0.507	0.96	1.00
AT3G28840	unknown protein	pollen tube	Y	N	0.071	0.122	0.585	0.59	1.00

Table A.5 (continued)

AT3G45800	unknown protein	pollen tube	N	N	0.106	0.113	0.938	0.02	1.00
AT3G49490	unknown protein	pollen tube	N	N	0.057	0.107	0.527	0.82	1.00
AT3G56610	unknown protein	ovary	Y	Y	0.088	0.054	1.995	0.06	0.37
AT3G56730	unknown protein	pollen tube	N	N	0.149	0.034	4.376	0.28	0.62
AT3G58540	unknown protein	pollen tube	N	N	0.079	0.123	0.638	0.76	1.00
AT4G02880	unknown protein	pollen	N	N	0.052	0.103	0.501	1.00	1.00
AT4G05631	unknown protein	ovary	N	N	0.119	0.127	0.927	0.24	1.00
AT4G12990	unknown protein	pollen	N	Y	0.056	0.055	1.016	0.61	0.99
AT4G14810	unknown protein	pollen tube	N	N	0.157	0.195	0.805	0.50	1.00
AT4G15053	unknown protein	ovary	Y	Y	0.092	0.138	0.685	0.23	1.00
AT4G22960	unknown protein	pollen tube	N	N	0.091	0.068	1.344	0.00	0.36
AT4G23350	unknown protein	ovary	Y	Y	0.016	0.073	0.664	0.22	1.00
AT4G23360	unknown protein	ovary	Y	Y	0.016	0.167	0.896	0.00	1.00
AT4G24030	unknown protein	pollen tube	Y	N	0.078	0.013	6.210	0.00	0.04
AT4G26410	unknown protein	pollen	N	N	0.039	0.073	0.532	0.86	1.00
AT4G28775	unknown protein	ovary	N	N	0.097	0.093	1.038	0.16	0.94
AT4G35725	unknown protein	ovary	Y	Y	0.074	0.172	0.595	0.72	1.00
AT4G35940	unknown protein	pollen	N	N	0.049	0.077	0.644	0.46	1.00
AT4G38980	unknown protein	pollen tube	N	N	0.081	0.086	0.947	0.05	1.00

Table A.5 (continued)

AT5G08270	unknown protein	pollen tube	Y	N	0.054	0.095	0.565	0.64	1.00
AT5G22970	unknown protein	ovary	Y	N	0.056	0.145	0.906	0.14	1.00
AT5G38400	unknown protein	pollen tube	N	N	0.092	0.159	0.576	0.72	1.00
AT5G39570	unknown protein	pollen tube	Y	Y	0.065	0.013	5.015	0.00	0.01
AT5G46115	unknown protein	trans tract	N	N	0.078	0.023	3.787	0.02	0.14
AT5G51105	unknown protein	ovary	Y	Y	0.094	0.080	1.342	0.05	0.58
AT5G51850	unknown protein	ovary	N	N	0.023	0.233	0.591	0.47	1.00
AT5G52130	unknown protein	ovary	N	N	0.060	0.183	0.575	0.79	1.00
AT5G52960	unknown protein	pollen tube	N	N	0.043	0.080	0.544	0.84	1.00
AT5G53820	unknown protein	pollen tube	N	N	0.010	0.000	-	0.15	0.25
AT5G57760	unknown protein	pollen tube	N	N	0.063	0.087	0.722	0.55	1.00
AT5G60260	unknown protein	pollen tube	N	N	0.097	0.067	1.455	0.11	0.59
AT5G61720	unknown protein	stigma -trans tract	N	N	0.039	0.079	0.520	0.90	1.00
AT5G63905	unknown protein	pollen tube	N	N	0.024	0.012	2.111	0.20	0.51

Table A.6 *A. thaliana* genes with no *A. lyrata* match.

AGI Number	Description	Data set	Lineage Specificity
AT1G47320	Transposable element; similar to nucleic acid binding/ribonuclease H	ovary	-
AT1G50290	unknown protein	pollen tube	*Arabidopsis*
AT1G51300	acyl-protein thioesterase-related	ovary	-
AT1G77093	Encodes a defensin-like (DEFL) family protein.	ovary	Brassicaceae
AT2G11405	unknown protein	ovary	*Arabidopsis*
AT2G13550	unknown protein	ovary	-
AT2G13940	transposable element	pollen	-
AT2G18440	Encodes a noncoding RNA	pollen	-
AT2G18938	unknown protein	ovary	-
AT2G20660	RALFL14 (RALF-LIKE 14)	ovary	-
AT2G28980	transposable element gene	pollen	-
AT3G30720	QUA-QUINE Starch (QQS)	pollen tube	*Arabidopsis*
AT3G42473	LCR47 (Low-molecular-weight cysteine-rich 47)	ovary	Brassicaceae
AT3G44430	unknown protein	pollen tube	*Arabidopsis*
AT3G44700	Unknown protein	pollen tube	-
AT4G16040	unknown protein	pollen tube	*Arabidopsis*
AT4G17505	unknown protein	ovary	-
AT4G38330	unknown protein	pollen tube	*Arabidopsis*
AT5G20460	unknown protein	pollen tube	*Arabidopsis*
AT5G22220	E2F1; transcription factor	pollen tube	-
AT5G48780	disease resistance protein (TIR-NBS class), putative	pollen tube	-

Table A.7 Summary expression data of selected candidates from relevant microarrays.

AGI Number	WT log_2 Exp[43]	*dif1* Exp[43]	*myb98* Exp[43]	Ovary-enriched[43]	Avg signal SIV PT[71]	SIV PT vs dry[71]	SIV PT-enriched[71]	Up-reg in 4hr PT vs 0.5 PT[71]	SIV PT vs 4 H PT[71]
AT1G20730	3.48	up	-	-	-	-	-	-	-
AT2G06090	6.55	down	-	yes	-	-	-	-	-
AT2G20597	6.42	down	-	-	-	-	-	-	-
AT3G48231	6.90	down	down reg	-	-	-	-	-	-
AT3G50020	6.13	down	-	-	-	-	-	-	-
AT3G57840	5.78	down	down	yes	-	-	-	-	-
AT3G59460	-	-	-	-	9.15	up	no	yes	up
AT4G21326	-	-	-	-	12.13	up	no	no	-
AT5G09370	6.85	down	-	-	9.67	up	no	no	up
AT5G18403	7.04	down	down	-	-	-	-	-	-
AT5G23035	7.44	down	-	-	-	-	-	-	-
AT5G46960	6.52	down	-	-	-	-	-	-	-

Table A.8 McDonald-Kreitman test results for comparisons between *A. thaliana* accessions and *A. lyrata*.

AGI Number	Fixed Nonsyn Diff	Fixed Syn Diff [a]	Nonsyn Polymorph	Syn Polymorph[a]	N. I.	Fisher's Exact p-value [a]
AT1G20730	115	21 (434)	12	7 (96)	0.31 (0.47)	0.05[b] (0.02[b])
AT2G06090	41	12 (12)	3	5 (5)	0.18 (0.17)	0.03[b] (.03[b])
AT2G20597	13	3 (19)	3	3 (10)	0.23 (0.44)	0.28 (0.32)
AT3G48231	14	4 (27)	0	0 (2)	NA (0)	NA (0.55)
AT3G50020	19	4 (54)	2	3 (5)	0.14 (NA)	0.08[c] (NA[d])
AT3G57840	55	14 (14)	0	0 (0)	NA (NA)	NA (NA)
AT3G59460	22	13 (35)	0	1 (1)	0 (0)	0.4 (1.0)
AT4G21326	79	44 (93)	11	16 (39)	0.38 (0.33)	0.03[b] (<0.01[b])
AT5G09370	30	8 (36)	0	0 (8)	NA (0)	NA (0.02[b])
AT5G18403	24	7 (19)	1	1 (4)	0.29 (0.20	0.43 (0.12)
AT5G23035	24	2 (33)	2	0 (0)	NA (NA)	1 (0.19)
AT5G46960	40	16 (17)	3	6 (6)	0.20 (0.21)	0.05[c] (0.06[c])

[a] Numbers in parenthesis include both coding and noncoding regions.
[b] significant at 0.05 level
[c] significant at 0.1 level
[d] Intron cannot be aligned

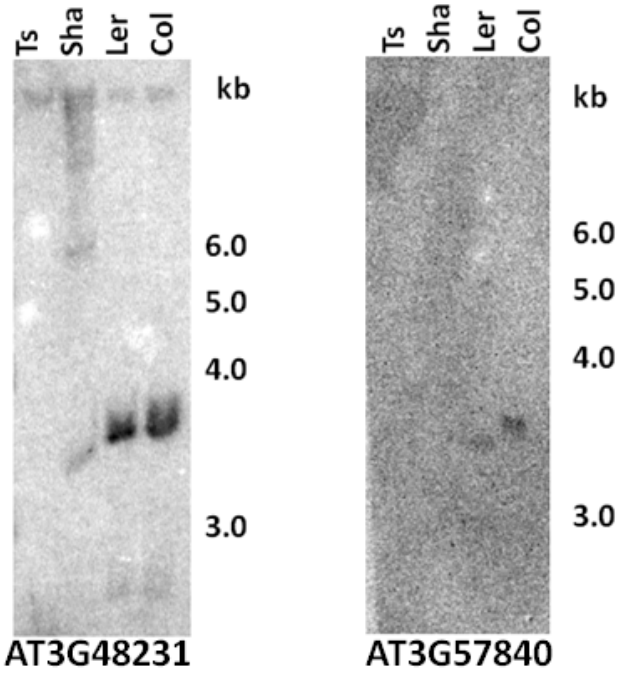

Figure A.1. DNA gel blot analysis of candidate genes At3g48231 and At3g57840 to determine presence/absence of gene in different *A. thaliana* accessions.

Table A.9 Significant coding sequence changes between *A. lyrata* and *A. thaliana* accessions.

AGI Number	Stop Codons in ORF?	Other
AT1G20730	Yes, *A.lyrata* amino acid 5	variable introns, Lz has a deletion of nucleotides 1248-1457, frameshift at amino acid 419; insertion of 9 nucleotides at position 1480 in Bay, Br, C24, Est, Ler, Lov, Lz, and Ts; frameshift at amino acid 471 in Bay, Br, Est, Ler, goes back in frame at amino acid 534; several deletions in *A. lyrata* around nucleotide 160 causing frameshift
AT2G06090	Yes, C24 at amino acid 111	frameshift in C24 at nucleotide 285
AT2G20597	no	several indels in *A. lyrata*, but no frameshift
At3G48231	Yes, *A. lyrata* amino acid 1	none
AT3G50020	no	cannot align intron with *A. lyrata*
AT3G57840	no	no polymorphism

Table A.9 (continued)

AT3G59460	no	cannot align first exon and intron with *A. lyrata*, in frame deletion of nucleotides 561-576 in Est-0, Ler-1, Lov-0, and Tamm
AT4G21326	no	start codon is attg in gDNA but is ATG in mRNA, several small *A. lyrata* insertions but no frameshift
AT5G09370	no	several small *A. lyrata* insertions but no frameshift
AT5G18403	no	frameshift in *A. lyrata* at amino acid 60, several insertions upstream but no frameshift
AT5G23035	no	in frame deletion in Cvi-0 at 340 bp, frameshift insertion in Lov-0 at nucleotide 42
AT5G46960	no	*A. lyrata* indel frameshift at nucleotide 290 then goes back in frame at nucleotide 299

Table A.10. T-DNA lines useful in functional analysis and putative paralogs of top candidates.

AGI Number	Stock Name	Insert Location	cds length	Paralog	cds match	T-DNA Line for Paralog?	Location
AT1G20730	SALK_081415C CS806527	promoter intron	1704	AT1G20740	309/599	Yes	Exon
AT2G06090	CS828274	promoter	408	AT5G27238	324/380	No	-
AT2G20597	SALK_031094	promoter	255	AT2G20619	99/100	Yes	promoter
AT3G48231	-	-	-	AT2G33233	224/243	No	-
AT3G50020	SALK_105937	promoter	291	-	-	-	-
AT3G57840	SALK_108664	exon	465	AT3G57850	221/238	Yes	promoter
AT3G59460	CS828443	intron	321	-	-	-	-
AT4G21326	SALK_130445	exon	2265	-	-	-	-
AT5G09370	SALK_128996C	exon	447	-	-	-	-
AT5G18403	SALK_076371	exon	264	-	-	-	-
AT5G23035	SALK_127795C	exon	267	-	-	-	-
AT5G46960	SALK_108076C	promoter	525	AT5G46950	491/525	-	-