Mapping the Exine Patterning Mutant *spot3* in *Arabidopsis thaliana*

Undergraduate Research Thesis

Presented in partial fulfillment of the requirements for graduation *with research distinction* in Molecular Genetics in the undergraduate colleges of The Ohio State University

By

Holly Welfley

The Ohio State University

April 2016

Project Advisor: Anna Dobritsa, Department of Molecular Genetics

Abstract

Exine, the outer cell wall of pollen, forms diverse patterns in different plant species and is important for plant reproduction. Exine is composed of sporopollenin, a material noted for its strength, elasticity, and chemical durability, making it attractive to biologists and material scientists. However, little is known about the developmental and biochemical aspects of exine formation. The aim of this project was to identify one of the genes involved in this process. In a previous genetic screen of *Arabidopsis thaliana*, a mutant with a spotty pollen phenotype that had largely disconnected exine elements was identified and named *spot3*. The *spot3* mutation was coarsely mapped to the bottom of chromosome 1. In order to locate the *spot3* mutation in the Arabidopsis genome, a F_2 mapping population of *spot3* mutant plants was created. *spot3* plants from that population were analyzed with molecular markers in order to reduce the interval containing the *spot3* mutation. Afterwards, candidate genes were sequenced and compared to wild-type sequences to identify any present mutations. Available mutant plants with T-DNA insertions in candidate genes were screened for the *spot3* phenotype. In addition to these approaches, bulked genomic DNA from *spot3* mutants was studied with next generation sequencing analysis. Use of these techniques has dramatically decreased the interval containing the *spot3* mutation from a 960 kb region with approximately 300 genes to 141 kb with 45 genes.

Introduction

Pollen, the male gametophyte of flowering plants, is surrounded by a remarkably tough and resistant pollen wall called exine. Exine is very different from other plant cell walls and is noted for its striking patterning designs, chemical durability, and its importance in reproduction. While the patterns created by exine are distinct between species, the general structure of the pollen wall is conserved between angiosperms (Dobritsa et al. 2011). Exine protects pollen against dry conditions, intense heat, bacterial infection, and water loss (Scott 1994; Ariizumi and Toriyama 2011). Exine also aids in pollen adhesion by providing species-specific factors that allow the female stigmas to recognize the pollen (Zinkl et al. 1999).

Amongst different pollen-producing plant species, there is a tremendous variety of exine patterns. The diversity of exine patterning can be attributed to its elaborate outer surface comprised of radially directed rods, called baculae, and tectum (Dobritsa et al. 2011). The tectum creates a roof-like connection between the baculae heads resulting in the patterning of exine on the surface of the pollen (Scott and Stead 1994). Studying the variation in exine patterning may provide an understanding of species-specific recognition mechanisms. Pollen grains from a countless number of plants have been imaged under high magnification and resolution to gain insight into the elaborate exine patterns (PalDat, https://www.paldat.org/). For example, *Arabidopsis thaliana* (thale cress) has a reticulate, network-like pattern while *Oryza sativa* L. (rice) pollen has a smooth exine surface (Figure 1). Other species may have even more elaborate designs, such as *Ilex aquifolium* (common holly), *Phyllanthus spec* (leafflower), *Agrostemma githago* (corn cockle), and *Dracunculus vulgaris* (dragon lily). The variation in exine morphology potentially contributes to speciation (Ariizumi and Toriyama 2011).

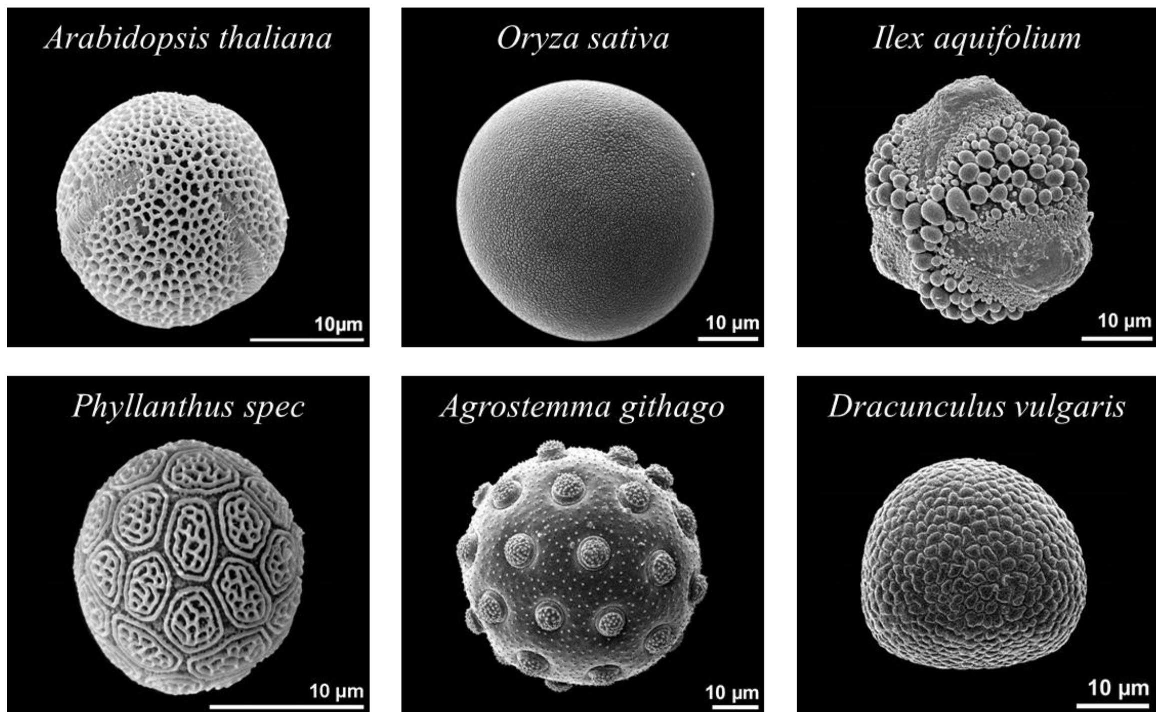

Figure 1: Exine patterning on pollen surface of different species. Images taken by scanning electron microscopy and provided by the PalDat website. Common species names listed from left to right-top: thale cress, rice, common holly; bottom: leafflower, corn cockle, dragon lily.

In addition to its diverse and beautiful patterns, exine has intriguing biochemical aspects. Unlike other cell walls that are primarily made of cellulose and other carbohydrates, the primary component of exine is sporopollenin, a biopolymer that is comprised of fatty acids, phenolics, and tetraketide pyrones (Quilichini et al. 2015; Guilford et al. 1988; Ahlers et al. 1999). Sporopollenin is known for its tough, elastic, and resistant nature which makes it interesting not only to biologists, but also to material scientists. Another aspect that is unique to exine is its mode of formation. While the cell walls in other plant cells are produced by the cells themselves, in exine formation one of the major players is tapetum. Tapetum is the diploid secretory anther tissue surrounding microspores, the precursors to pollen grains (Piffanelli et al. 1998). The tapetal layer expresses multiple enzymes important for sporopollenin precursor synthesis

(Dobritsa et al. 2011). Once synthesized in the tapetum, these precursors are transported out of the tapetum and are deposited and assembled into exine on the microspore surface.

Analysis of the genetic mechanisms behind exine synthesis, assembly, and patterning could have biochemical implications for sporopollenin studies, as well as contribute to an understanding of developmental steps that contribute to exine formation (Dobritsa et al. 2011). A number of genes have been found to contribute to the formation of exine (Quilichini et al. 2015). Many of the genes identified encode proteins important for lipid metabolism, tetraketide pyrone synthesis, and surprisingly, carbohydrate metabolism. Other genes have been identified as playing a role in exine formation, including those that encode enzymes for fatty acids and phenylpropanoids (Morant et al. 2007; Dobritsa et al. 2009, 2010). However, most likely many of the genetic contributors to exine formation remain unknown (Dobritsa et al. 2011).

In order to better understand the biosynthesis, transport, and assembly of exine, a large-scale genetic screen was previously performed (Dobritsa et al. 2011). In this screen about 16,000 mutagenized *Arabidopsis thaliana* plants were analyzed under a dissecting microscope. The anthers and pollen were inspected from these plants for visual abnormalities in hopes of finding plants defective in exine formation and deposition. Pollen grains displaying unusual phenotypes under the low magnification were then analyzed at a higher magnification and higher resolution using a confocal microscope. Investigation of the pollen by confocal microscopy determined whether the abnormal pollen appearance was due to defects in exine. Forty-nine total mutants with varying exine defects were found using this approach, suggesting that a large number of genes are involved in exine formation. One of the mutants identified in this screen, 137-2-1, has a largely disconnected exine. Under confocal imaging the exine patterning appeared spotty, thus giving the mutant the name *spot3* (Figure 2). The disconnection of exine seen in *spot3* could be

attributed to defects in tectum formation where the roof connecting the baculae is not completely formed (Dobritsa et al. 2011). Previously, the location of *spot3* was coarsely mapped to the bottom of chromosome 1 using linkage analysis with markers NGA111 and NGA692 (Dobritsa et al. 2011).

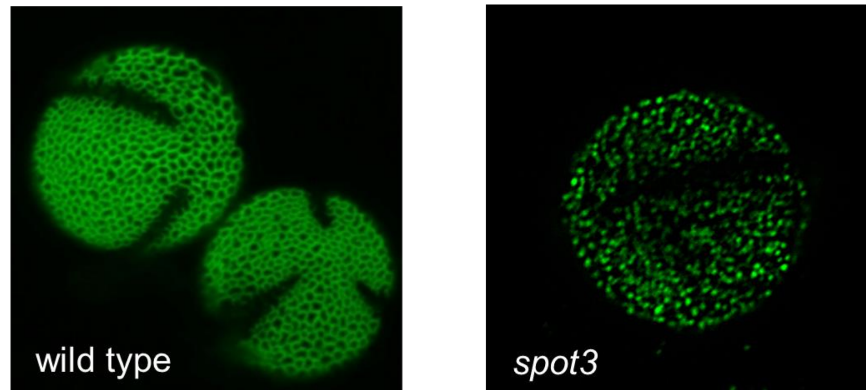

Figure 2: Confocal images of wild type pollen and *spot3* mutant pollen. Pollen grains stained with auramine O. Wild type grains have reticulate exine pattern, while exine of the *spot3* mutant appears spotty, possibly due to abnormalities in tectum formation.

The precise location of the *spot3* mutation, however, remained unknown. The purpose of this research study was to determine the location of *spot3* in the *Arabidopsis thaliana* genome. A large number of *spot3* mutant plants were used for positional cloning analysis, gene candidate sequencing, and next generation sequencing.

Results

Mapping Population

In order to locate the *spot3* mutation in the Arabidopsis genome, a mapping population was created (Figure 3). Generation of the mapping population began by crossing a homozygous mutant for *spot3* with a common ecotype, Landsberg. An ecotype is a geographic or ecological

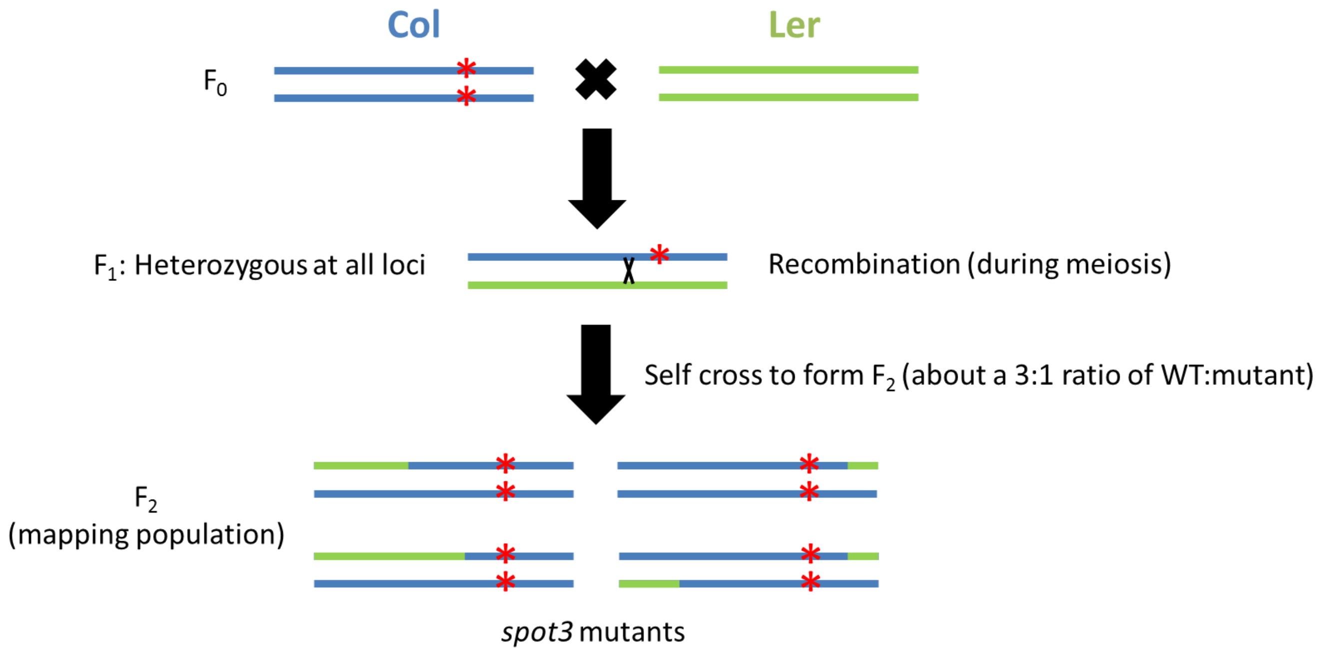

Figure 2: Illustration representation of mapping population formation. Blue indicates Columbia ecotype DNA, green indicates Landsberg DNA, red asterisk represents the *spot3* mutation.

variety of the species *Arabidopsis thaliana* that is genetically distinct from other varieties. The *spot3* mutant was originally identified in the ecotype Columbia. The genetic differences, called polymorphisms, make Columbia DNA distinguishable from Landsberg DNA. The resulting first generation of progeny, the F_1 generation, yielded heterozygous plants at all loci for Columbia and Landsberg DNA. The F_1 plants were allowed to self-pollinate to yield the F_2 generation. The F_2 generation was then screened for the *spot3* phenotype. Screening of these plants was done under a dissecting microscope, which allowed for successful differentiation of wild type plants from *spot3* plants. Flowers were removed from F_2 plants so that anthers could be isolated. Pollen was scraped off of the anthers in order to visualize its phenotype. Wild type pollen grains are easily separated from each other, while *spot3* pollen is notably stickier and clumps together (Figure 4). From this heterozygous cross, a Mendelian genetic ratio of three wild type plants to

one *spot3* mutant plant was expected and observed. About 1,600 plants were screened from the F_2 population and 536 of these plants were found to display the *spot3* phenotype. All wild type plants were discarded, while the plants displaying the *spot3* phenotype were kept for use as the mapping population and their tissues were collected.

Figure 4: Appearance of WT and *spot3* pollen under a dissecting microscope. Red ovals indicate the location of pollen. Note that WT pollen is easily separated into individual grains, while *spot3* pollen appears clumpy and sticks together.

Molecular Markers

Afterwards, genomic DNA (gDNA) was extracted from the mapping population *spot3* tissue samples. The isolated gDNA was used to analyze molecular markers by polymerase chain reaction (PCR) reactions. Molecular markers are fragments of DNA associated with a specific region in the genome. The molecular markers chosen take advantage of the distinguishing polymorphisms between Columbia and Landsberg ecotypes. In order to identify or create markers for use in this project, information was obtained from the following websites: Arabidopsis Mapping Platform (http://amp.genomics.org.cn/), SALK Arabidopsis 1,001

Genomes (http://signal.salk.edu/atg1001), and The Arabidopsis Information Resource (https://www.arabidopsis.org/).

Two different types of molecular markers were used in this study (Figure 5). The first type allowed for visualization of length differences between Columbia and Landsberg sequences. The differences in sequence length were caused by insertions or deletions (INDELs) or by Simple Sequence Length Polymorphisms (SSLPs). In the case of INDELs, one ecotype would have either an insertion or deletion in a region of the genome that is not present in the other ecotype. SSLPs are repetitive sequences of varying base length. The number of bases in a unit, as well as the number of unit repeats can vary. The length of repeated units is important in distinguishing the differences in Columbia and Landsberg SSLP sequences. Once the molecular marker region is amplified through PCR, the DNA can be separated based on length on an agarose gel via gel electrophoresis. Particular INDELs or SSLPs were chosen because their base pair length difference can be easily resolved by ultraviolet light imaging of an agarose gel. Visualization of the gel can identify which ecotype DNA is present at the given position in the genome.

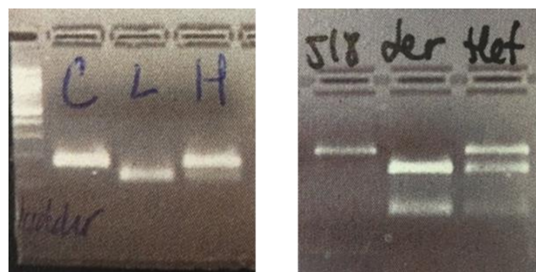

Figure 5: Agarose gel images of the two types of markers used. An example of marker separated based on sequence length shown with a ladder, Columbia, Landsberg, and Heterozygous gDNA respectively (left). An example of a marker requiring restriction digest before analysis shown with sample 518 which is an example of Columbia, Landsberg, and Heterozygous gDNA respectively (right).

Another type of marker used in this study required restriction digestion before analysis. Either Cleaved Amplified Polymorphic Sequences (CAPS) or Derived Cleaved Amplified Polymorphic Sequences (dCAPS) were used (Konieczny and Ausubel 1993; Neff et al. 1998). Both CAPS and dCAPS markers can be used for identifying polymorphisms that cannot be visualized on a gel by simple length difference. A single nucleotide polymorphism (SNP), in which only a single nucleotide differs between the ecotypes, is an example of such a polymorphism. CAPS markers recognize a restriction site directly caused by a SNP that only exists in one of the ecotypes studied. Alternatively, dCAPS markers are used if the SNP does not innately produce a restriction site in the sequence. In order to create a restriction site, one of the primers for a dCAPS marker contains one or more base pair mismatches to the template DNA strand. Once replicated, the mismatches are incorporated into the newly synthesized strand. The combination of the mismatch insertion with the existing SNP, present in only one of the studied ecotypes, creates a restriction site. Small INDELs can also potentially introduce or remove a restriction site, and can be used for creating a CAPS marker. Thus, the molecular markers for these SNPs or INDELs not only have to be PCR amplified, but also digested by the appropriate enzyme. Only the polymorphic site of one ecotype will be digested and two fragments, as opposed to a single, undigested band in another ecotype, will be produced.

Positional Cloning

Analysis of the ecotype DNA present at a given molecular marker assists in positional cloning of *spot3*. Due to recombination events occurring during male and female meiosis in F_1 heterozygous plants, each F_2 plant found in the mapping population has unique recombination signatures that give rise to many different combinations of Columbia and Landsberg genomic DNA. Since *spot3* is known to be in the Columbia background, markers closely linked to *spot3*

are expected to be highly enriched in Columbia sequences. If analysis of a given marker in a *spot3* mutant from the mapping population generates sequences from both ecotypes in one reaction, the mutant was heterozygous at that marker (Figure 6). Therefore, this indicates that the mutation causing the *spot3* phenotype must not be completely linked to that particular marker.

Figure 6: Agarose gel dyed with ethidium bromide and imaged under ultraviolet light to visualize DNA using marker F10A5-1. White asterisks indicate samples that were at this position, heterozygous. All other samples had the Columbia sequence at this position. Controls are in the last row.

As previously determined, *spot3* was mapped to the bottom of chromosome 1 between the markers F3N23-1 (27.41 Mb) and F10A5-1 (28.37 Mb). The region between these two markers spanned approximately 960 kb and contained approximately 300 predicted genes. 879 *spot3* mutant plants (536 plants from the first screen, 215 previously collected samples, and 128 plants from a second screen described later) were analyzed to narrow this region. Molecular

markers were first used starting at the known borders, but new markers inside the borders were generated and used to narrow the interval further. Mutant samples that displayed heterozygosity at the previous molecular marker were analyzed for the next marker to determine whether heterozygous DNA was still present. Heterozygous DNA at the next marker is an indication that a recombination event removing Landsberg DNA had not occurred at that position, and the *spot3* mutation must still be further inside the interval from the given marker. The samples in which only Columbia DNA was present at the borders of the interval were non-informative and were no longer used to narrow the interval. These samples were not of interest because the likelihood of a new recombination event occurring between the marker and the mutation that would introduce Landsberg DNA would be nearly impossible (Figure 7). This process was repeated multiple times with new sets of DNA samples until the region was narrowed to 180,393 bp between the markers F3N23-5 (27.46 Mb) and spot3-13 (27.64 Mb) (Figure 8).

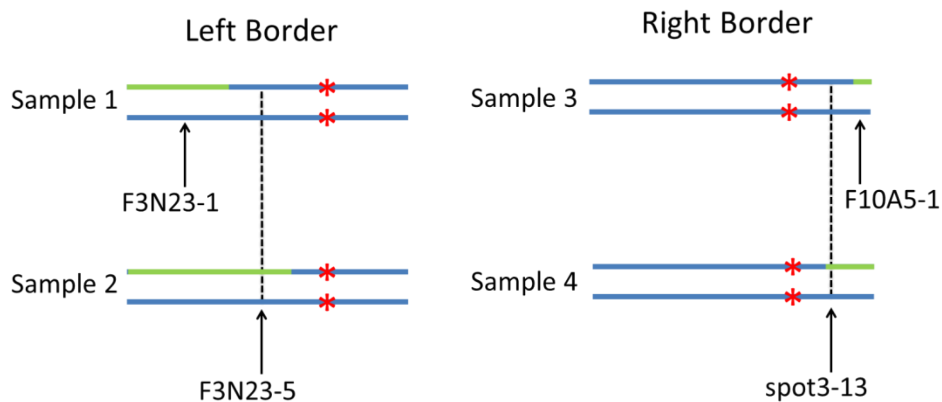

Figure 7: Figure representation of marker analysis. At the left border, Sample 1 is Het at F3N23-1 and Col at F3N23-5. Sample 2 is Het at both markers. At the inside marker (F3N23-5), Sample 1 is no longer informative, while Sample 2 can be used further with a new marker to the right of F3N23-5. Similar logic is shown at the right border. Sample 3 is Het at F10A5-1 and Col at spot3-13. Sample 4 is Het at both markers. Sample 3 is not informative at spot3-13.

	F3N23-1	F3N23-5	spot3-12	T9L24-1	spot3-5	spot3-13	F6D5-1	spot3-6	F25P22-1	F2P9-4	F2P9-3	F2P9-1	spot3-10	F1M20-1	F1B16-1	F10A5-1
Plant	27.41 Mb	27.46 Mb	27.55 Mb	27.57 Mb	27.59 Mb	27.64 Mb	27.65 Mb	27.65 Mb	27.71 Mb	27.78 MB	27.8 MB	27.82 Mb	27.935 Mb	27.98 Mb	28.26 Mb	28.37 Mbs
Cbus3-235	C															C
Cbus3-236	C															C
Cbus3-237	H	H	C	C?	C			C?				C				C
Cbus5-380	C				C		C	C	?	C?	C?	H				H
Cbus6-457	C				C			C				C				H
Cbus7-510	C	C			C	H	H	??		H?	H?	H				H

Figure 8: Example of a data sheet used to maintain the records of marker analysis results. Molecular marker analysis narrowed the *spot3*-containing interval to a region between the markers F3N23-5 (27.46 Mb) and spot3-13 (27.64 Mb).

Further use of molecular markers inside this new interval proved difficult, because its smaller size meant that fewer molecular markers were present between Columbia and Landsberg. The few remaining molecular markers failed to yield clearly resolved bands on an agarose gel.

In order to continue analyzing samples for heterozygosity, a new approach was taken. In this approach, samples that still exhibited heterozygosity at either F3N23-5 or spot3-13 were sequenced at SNP sites located inside this interval. A sample of mutant DNA that was heterozygous at the marker border was PCR amplified using primers that flank a region of DNA with a SNP between Columbia and Landsberg DNA. Sanger sequencing provided chromatogram data of the sequences which were then interpreted. A chromatogram utilizes four different colors, each color representing a different base, to provide a visual representation of the DNA sequence. A peak of a certain color represents the presence of the respective base at that position. When looking at the position in which the SNP is located, heterozygosity can be determined if there are two peaks present at the same position corresponding to each polymorphic base (Figure 9). By sequencing the SNPs, the right border was moved approximately 40 kb inward to a position in the gene At1g73430 at 27.60 Mb. Only Columbia DNA was present at the first SNP to the right of the left border when the single remaining heterozygous sample was sequenced. As a result, movement of the left border could not be completed by marker analysis. As a result of the

mapping approach, the region containing *spot3* was significantly narrowed to an interval of 141,577 bp containing 45 candidate genes (Figure 10).

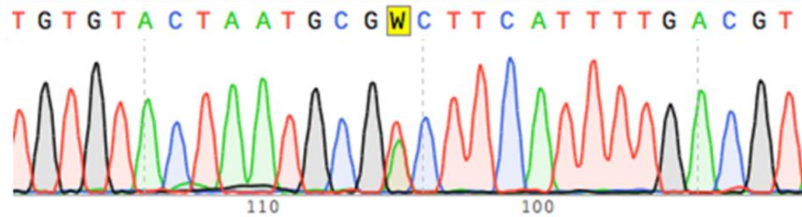

Figure 9: Chromatogram data indicated a heterozygous site at the position "W." Note the green (A) and red (T) peaks present at this position, showing both nucleotides.

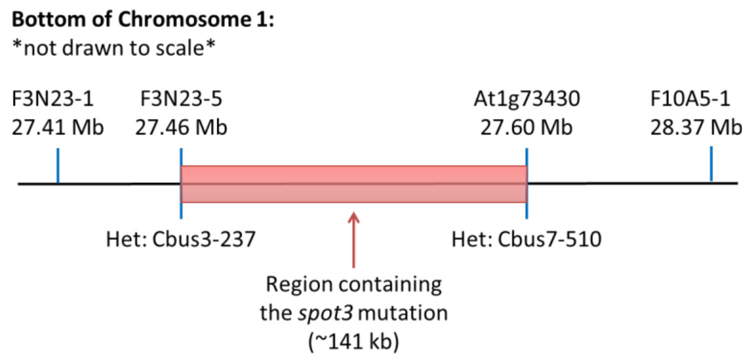

Figure 10: Visual representation of interval containing the *spot3* mutation. Polymorphism and marker analysis have narrowed the interval to 141,577 bp. Within this region, there are 45 candidate genes.

Candidate Gene Analysis

After the region was sufficiently narrowed, investigation of these 45 candidate genes was necessary. The Arabidopsis Information Resource (TAIR) database provided information about these predicted genes. After reviewing the descriptions and putative functions of the gene candidates, open reading frames of 11 promising genes were sequenced. The 11 candidate genes analyzed were chosen because they either stood out as potentially being involved in exine production and assembly or had no known function (Figure 11). For example, one of the

candidate genes At1g73370, also known as Sucrose Synthase 6, was annotated to be involved in sucrose metabolic processes. SUS6 stood out because of its potential importance for sucrose metabolism in cell wall formation. Another candidate gene, At1g73160, is a UDP-Glycosyltransferase family protein. Two UDP-Glycosyltransferases have been previously identified as important for exine formation (Dobritsa et al. 2011). These types of enzymes create glycosidic linkages in sugars which are important components of the cell wall. In addition, cytochromes P450 (CYP450s) have been previously identified as being essential for exine development and sporopollenin synthesis (Dobritsa et al. 2009; Morant et al. 2007). CYP450s are a diverse family of enzymes that catalyze a vast array of reactions and are conserved throughout the plant kingdom (Dobritsa et al. 2009; Morant et al. 2007). To uncover any present mutations, DNA from a mutant sample was PCR amplified and the PCR fragments were treated with Exonuclease I (ExoI) and Shrimp Alkaline Phosphatase (SAP) and then directly sequenced with the primers used for amplification. The sequencing results were compared to known wild-type Columbia sequences (Figure 12). However, none of the sequenced genes displayed mutations in their coding regions and introns when compared to the wild-type sequence.

At#	Gene name	Predicted function	Process/Phenotype	Sequencing Information
At1g73030	CHMP1A	ESCRT-related	mediates multivesicular body sorting of auxin carriers and is required for plant development (embryonic axis)	all well
At1g73040		Mannose-binding lectin superfamily	carbohydrate binding	All seq well
At1g73090		unknown	unknown	All seq well (184-2654)
At1g73120		unknown	response to oxidative stress	All seq well (87-877)
At1g73140	TBL31	TRICHOME BIREFRINGENCE-LIKE family	Two members of TBL family are involved in the synthesis and deposition of secondary wall cellulose, presumably by influencing the esterification state of pectic polymers.	all seq well
At1g73160		UDP-Glycosyltransferase	transferring glycosyl groups	all seq well
At1g73220	OCT	organic cation/carnitine transporter1	carnitine transporter activity, carbohydrate transmembrane transporter activity	All seq well (184-2213 bp)
At1g73250		bifunctional 3, 5-epimerase-4-reductase	converts GDP-D-mannose to GDP-L-fucose in vitro along with MUR1 (GDP-D-mannose 4,6-dehydratase). It is expressed in all tissues examined, but most abundantly in roots and flowers.	All seq well
At1g73320		methyl transferase	unknown	all seq well (91-1862)
At1g73340		CYP450	unknown	All seq well (173-2570)
At1g73370	SUS6	sucrose synthase 6, UDP-glycosyltransferase	production of sucrose and/or UDP-glucose	All seq well (179-4192)

Figure 11: The 11 candidate genes sequenced. Listed are their gene names, predicted functions (if known) and the sequencing results.

```
73040       CAAGGTGATAAGAATCTGACAGTGTTCGTAGGACCCTGGGGAGGAAATGGAGGAACCACT 145
73030-AF    CAAGGTGATAAGAATCTGACAGTGTTCGTAGGACCCTGGGGAGGAAATGGAGGAACCACT 102
73040-AR    CAAGGTGATAAGAATCTGACAGTGTTCGTAGGACCCTGGGGAGGAAATGGAGGAACCACT 137
            ************************************************************

73040       TGGGATGATGGGATTTATGATGGTGTCCGTGAGATCAGACTTGTTTATGACCATTGCATT 205
73030-AF    TGGGATGATGGGATTTATGATGGTGTCCGTGAGATCAGACTTGTTTATGACCATTGCATT 162
73040-AR    TGGGATGATGGGATTTATGATGGTGTCCGTGAGATCAGACTTGTTTATGACCATTGCATT 197
            ************************************************************
```

Figure 12: Example of a sequencing alignment. Top row (73040) depicts wild-type sequence. Second row (73030-AF) and third row (73040-AR) show the sequences from a *spot3* mutant sample sequenced on both strands. Asterisks represent sequence alignment.

An alternative approach was then used to investigate the candidate genes. Available mutants with T-DNA insertions affecting some of the genes in the *spot3* interval were obtained from the Arabidopsis seed stock center (ABRC, Arabidopsis Biological Research Center) (Figure 13). Mutants with T-DNA insertions in the exons, introns, or the promoter region of a given gene candidate were chosen in hope of disrupting the gene function. Plants with predicted mutations in 26 genes were grown and their pollen was inspected under a dissecting microscope. Pollen was studied in the same manner as during the previous screen of the mapping population. If the *spot3* phenotype was observed, one could infer that the *spot3* mutation would likely affect the same gene as the T-DNA mutation. After careful inspection of all mutant seed stocks ordered, a *spot3* phenotype was not seen (Figure 14).

Figure 13: Example of a T-DNA insertion chosen for analysis (signal.salk.edu). Top green bars represent exons of a gene. The red vertical line shows the annotated site of insertion of the chosen T-DNA mutant line into the exon of the candidate gene.

Gene	Stock Number	Number of Plants	Description
At1g73200	CS822663	16	All wt
At1g73030	CS824699	13	All wt
At1g73240	CS829119	12	All wt
At1g73250	CS855305	21	All wt
At1g73360	CS856540	21	All wt
At1g73380	SALK_008555	19	All wt
At1g73020	SALK_012541C	9	All wt
At1g73350	SALK_023207C	12	All wt
At1g73480	SALK_025681C	7	All wt
At1g73220	SALK_032805C	4	All wt
At1g73230	SALK_043673C	6	All wt

Figure 14: Chart showing results of some of the 26 genes analyzed with T-DNA insertions.

Next Generation Sequencing

After investigation of candidate genes, next generation sequencing was performed using bulked DNA of a subset of mutants from a mapping population. A second batch of F_2 progeny from the same Columbia and Landsberg cross used in the generation of the mapping population was screened. Approximately 600 F_2 plants were screened for the *spot3* phenotype and 128 *spot3* plants were found. Two tissues samples from each of the *spot3* mutant plants were collected. One of the tissue samples was used to prep gDNA for molecular marker analysis, utilizing the same procedure that followed the first screen. The second sample was then used for next generation sequencing. All 128 *spot3* mutant tissues were pooled together for gDNA extraction. The isolated gDNA was then sent for next generation sequencing to the Genomics Shared Resource facility at OSU. The sequencing data was analyzed by Brigitte Hofmeister and Robert Schmitz at the University of Georgia. They particularly concentrated on the sequencing reads mapped to the *spot3* region. The results of this analysis did not yield an obvious candidate, such

as a small insertion, small deletion, or a SNP for the *spot3* mutation. The program used by our collaborators at the University of Georgia was designed to recognize such types of mutations. Manual alignment of sequencing reads may uncover a chromosomal translocation, large insertion, large deletion, or another type of mutation causing *spot3* that could not be recognized by the program used to interpret the next generation sequencing data.

Discussion

The aim of this study was to map the location of the exine patterning mutation *spot3* in *Arabidopsis thaliana*. The *spot3* mutation causes spotty exine patterning on the surface of the pollen, possibly as a result of abnormalities in tectum formation. As a result of this study, the interval in which the *spot3* mutation is located has been narrowed substantially from 960 kb to 141 kb (from about 300 genes to 45 genes), but the precise location of *spot3* remains unknown. Identification of the *spot3* mutation in the Arabidopsis genome will allow characterization of the affected gene, and, consequently, better understanding of the genetic mechanisms of exine patterning, production, and assembly.

Several reasons could account for the failure to find the *spot3* mutation. First, when analyzing candidate sequences, only 11 of the total 45 genes were sequenced. While these genes represented the candidates we judged to be the most likely, it is possible that the *spot3* mutation is located in one of the genes that was not sequenced. Additionally, only the coding region and introns of these 11 genes were compared to the wild-type Columbia sequence. If a mutation were present in the promoter region or in the untranslated regions it would have been missed in my analysis. Third, the T-DNA insertion mutations analyzed for 26 genes in the interval were not guaranteed to disrupt the gene function. When ordering Arabidopsis lines with T-DNA

mutations, I preferred to choose the insertions of the T-DNA sequence in the exons of candidate genes since exon insertions would have a higher chance to disrupt the gene function. However, T-DNA insertions in exons were not available for all candidate genes. In those cases, insertions within the promoter regions and introns were chosen on the possibility that they might potentially also disrupt gene function. However, insertions into the promoter regions or introns of genes are less reliable in affecting gene expression. In addition, while most of the lines that were ordered were annotated as having the homozygous T-DNA insertions, the investigation to ensure the presence of the homozygous T-DNA insertions for each studied candidate was not completed. Thus, it is formally possible, that some of the lines may not have had a homozygous mutant.

Fourth, the type of the next generation sequencing analysis performed so far might have missed the *spot3* mutation. The program used to interpret the next generation sequencing data was designed to preferentially recognize small insertions, small deletions, or SNPs. This design set-up would be biased against picking up other types of mutations, such as chromosomal translocations, large insertions, or large deletions.

What can be done in the future to identify the *spot3* mutation? One approach would be to compare expression levels of the genes within the interval in wild-type and *spot3* plants. This would identify any transcriptional issues occurring as a result of the *spot3* mutation. *spot3* mutant seeds have been planted and upon flowering, when proteins required for exine formation and assembly are expressed, bud samples will be collected for RNA extraction. Reverse transcription polymerase chain reaction (RT-PCR) techniques could quantify the RNA transcript levels present in *spot3* mutant plants. If gene transcription or transcript stability is affected as a result of mutations in non-coding regions, this approach can potentially pick up these defects. In

addition to this approach, the candidate genes can be disrupted using the CRISPR/Cas technique (Cong et al. 2013; Hwang et al. 2013). Guide RNAs (gRNAs) will be engineered to target candidate gene sequences and, along with Cas, would function to cleave the targeted DNA. As a result, the gene function would be lost and the plants with these targeted mutations could be screened for the *spot3* phenotype. It is possible to design constructs containing two guide RNAs (that would create a deletion that would be easy to identify) and that would affect more than one gene, thus speeding up the analysis of genes in the interval.

Another approach to find *spot3* will be to further analyze the next generation sequencing results. This analysis will include investigation of the NGS data outside of the theoretical interval, in case errors were made in determining the borders. In addition, manual interpretation of the raw NGS data could help in identifying other types of mutations in addition to small insertions, small deletions, or SNPs. In addition, there is still one remaining sample that is heterozygous at the current right border. Therefore in order to narrow the interval further, this sample can be sequenced at polymorphic sites left of the current border. If the sample presents heterozygous DNA at the next marker, the border can then be moved. On the other hand, if only Columbia DNA is present then the current interval will remain the same.

Once the *spot3* mutation is located, complementation analysis with a transgene will be performed. Little is known about the developmental and biochemical aspects of exine formation and the complete set of genes involved in this process remains to be discovered. Identification of the gene affected by the *spot3* mutation would contribute to the understanding of how exine is synthesized and assembled.

Materials and Methods

Mapping Population:

spot3 homozygous mutant in the Columbia ecotype background was crossed with Landsberg ecotype. The resulting F_1 progeny were heterozygous for Columbia and Landsberg DNA. The F_1 progeny were self-crossed yielding a F_2 population in which *spot3* mutants were segregated at a 3:1 ratio (WT:mutant). Plants were grown in either a greenhouse supplemented with light or in growth chambers under long-day conditions (18 hours light:6 hours dark). The F_2 plants were screened under a dissecting microscope for their pollen appearance. Wild type pollen grains separated easily from each other, while the *spot3* pollen formed clumps and stuck together. Tissue samples from the *spot3* plants were collected to be used for gDNA isolation. Out of the initial screen of approximately 1,600 plants, and 536 plants displaying the *spot3* phenotype were isolated. A second screen at a later date of about 600 plants yielded 128 *spot3* plants. 215 previously collected *spot3* gDNA samples were also used for mapping, thus resulting in a mapping population consisting of a total of 879 mutant plants.

Genomic DNA isolation:

Tissue samples of *spot3* plants from the mapping population were flash frozen with liquid nitrogen and lyophilized for 24 hours. Zirconia-silica beads (2.3 mm diameter, Biospec) were added and samples were vortexed for several minutes to grind up the tissues into a fine powder. Dr. Raha's buffer (per 100 mL stock: 2 mL 5 M NaCl, 500 µL 2 M Tris-HCl (pH 8.0), 200 µL 0.5 M EDTA, 10 mL 10% SDS, and 2 mL Triton-X 100) and phenol:chloroform:isoamyl alcohol solution (25:24:1) were added to the samples and vortexed for 30 seconds. Samples were

centrifuged at 8,000 rpm for 10 minutes. 80 µL of the supernatant was transferred to new tubes and then 1:5 dilutions of the gDNA were prepared.

Molecular Marker Analysis:

Per 25 µL PCR reaction 5 µL of E2-5x buffer, 2 µL of dNTPs (2.5 mM each), 0.5 µL of 10 µM forward primer, 0.5 µL of 10 µM reverse primer, 0.25 µL of E2 Tak (Takara), and 2 µL of 1:5 gDNA were combined. Samples were amplified through 35 cycles that included steps of denaturation (98°C, 15 seconds), primer annealing (temperature varied between 51.5°C to 58°C depending on the primers, 15 seconds), and elongation (72°C, time varied between 30 seconds and 1 minute depending on fragment length). For CAPS and dCAPS markers, restriction digestion of PCR products with an appropriate restriction enzyme was performed. 20 µL restriction reactions containing 10 µL of PCR product, 2 µL of appropriate 10x New England Biolabs buffer, and 0.4 µL of appropriate New England Biolabs enzyme were incubated for 1 hour at 37°C. 25 µL of PCR product for length-based markers or 20 µL digest product for restriction site-based markers were combined with 3 µL of DNA loading dye and then ran on an agarose gel stained with ethidium bromide. Agarose gel percentage varied from 1% to 3.5% based on the size of fragments and the expected length difference between Col and Ler. DNA was visualized under ultraviolet light.

Sequencing:

For sequencing, fragments of candidate genes were amplified in 50 µL PCR reactions. For this, 10 µL of E2-5x buffer, 4 µL of dNTPs (2.5 mM each), 1 µL of 10 µM forward primer, 1 µL of 10 µM reverse primer, 0.5 µL of E2 Tak (Takara), and 2 µL of 1:5 gDNA were combined. PCR conditions followed the protocol above, again varying the annealing temperature and elongation

time appropriately. 5 μL of PCR product were ran on an agarose gel to ensure that amplification has worked. To remove primers that were used for PCR, 5 μL of PCR product was incubated with 0.2 μL ExonucleaseI, and 0.4 μL of Shrimp Alkaline Phosphatase in 10 μL reactions for 20 minutes at 37°C followed by incubation at 80°C for 15 minutes. Sequencing samples were then prepared using 2.5 μL of ExoSAP purified PCR product, 1 μL of 2 μM primer, and 2.5 μL dH_2O and sent for Sanger sequencing analysis to Plant-Microbe Genomics Facility at OSU.

T-DNA Insertion Analysis:

Available seed stocks with T-DNA insertions were ordered from Arabidopsis Biological Resources Center at OSU. Seeds were grown either in a greenhouse or growth chambers under the long-day conditions. Once they were flowering, pollen was analyzed under a dissecting microscope for the *spot3* phenotype.

Next generation sequencing:

F_2 plants were screened and 128 *spot3* mutants were found. Samples were split pooled together and gDNA was prepared with the Qiagen MaxiPrep plant kit following the manufacturer's instructions. DNA libraries were prepared and next generation sequencing was performed at the James Comprehensive Cancer Center at OSU. Interpretation of the data results was completed by affiliates Brigitte Hofmeister and Robert Schmitz at the University of Georgia.

Works Cited

Ahlers F, Thom L, Lambert J, Kuckuk R, Wiermann R (1999) 1H NMR analysis of sporopollenin from Typha angustifolia. Phytochemistry 5: 1095–1098

Ariizumi T., Toriyama K. Genetic regulation of sporopollenin synthesis and pollen exine development. Annu. Rev. Plant Biol. 2011;62:437-460.

Cong, Le, et al. "Multiplex genome engineering using CRISPR/Cas systems." Science 339.6121 (2013): 819-823.

Dobritsa, A., et al. "A large-scale genetic screen in Arabidopsis to identify genes involved in pollen exine production." Plant physiology 157.2 (2011): 947-970.

Dobritsa, A., et al. "CYP704B1 is a long-chain fatty acid ω-hydroxylase essential for sporopollenin synthesis in pollen of Arabidopsis." Plant physiology 151.2 (2009): 574-589.

Dobritsa, A., et al. "LAP5 and LAP6 encode anther-specific proteins with similarity to chalcone synthase essential for pollen exine development in Arabidopsis thaliana." Plant physiology (2010): pp-110.

Guilford WJ, Schneider DM, Labovitz J, Opella SJ (1988) High resolution solid state 13C NMR spectroscopy of sporopollenins from different plant taxa. Plant Physiol 86: 134–136

Hwang, Woong Y., et al. "Efficient genome editing in zebrafish using a CRISPR-Cas system." Nature biotechnology 31.3 (2013): 227-229.

Konieczny A, Ausubel FM. A procedure for mapping Arabidopsis mutations using co-dominant ecotype-specific PCR-based markers. Plant J. 1993 Aug;4(2):403-10. PMID: 8106085

Morant, Marc, Jørgensena K, Schallerb H, Pinotb F, Lindberg Møllera B, Werck-Reichhartb B and Baka S. "CYP703 is an ancient cytochrome P450 in land plants catalyzing in-chain hydroxylation of lauric acid to provide building blocks for sporopollenin synthesis in pollen." The Plant Cell 19.5 (2007): 1473-1487.

Neff MM, Neff JD, Chory J, Pepper AE. dCAPS, a simple technique for the genetic analysis of single nucleotide polymorphisms: experimental applications in Arabidopsis thaliana genetics. Plant J. 1998 May;14(3):387-92. PMID: 9628033

Qin P1, Tu B, Wang Y, Deng L, Quilichini TD, Li T, Wang H, Ma B, Li S. ABCG15 encodes an ABC transporter protein, and is essential for post-meiotic anther and pollen exine development in rice. Plant Cell Physiol 54, 138–54 (2013).

Piffanelli P, Ross JHE, Murphy DJ (1998) Biogenesis and function of the lipidic structures of pollen grains. Sex Plant Reprod 11: 65–80

Quilichini, TD., Grienenberger E, and Douglas CJ. "The biosynthesis, composition and assembly of the outer pollen wall: A tough case to crack."Phytochemistry 113 (2015): 170-182.

Scott, R. J., and A. D. Stead. Molecular and Cellular Aspects of Plant Reproduction. Cambridge: New York, 1994. Print.

Scott RJ, Spielman M, Dickinson HG (2004) Stamen structure and function. Plant Cell (Suppl) 16: S46–S60

Zinkl GM, Zwiebel BI, Grier DG, Preuss D (1999) Pollen-stigma adhesion in Arabidopsis: a species-specific interaction mediated by lipophilic molecules in the pollen exine. Development 126: 5431-5440