Uncovering tasselsheath3. A Genomic and Phenotypic Analysis of a Maize Floral Mutant.

Uncovering *tasselsheath3*. A genomic and phenotypic analysis of a maize floral mutant.

A Thesis Presented

By

THOMPSON H. ZHANG

Submitted to the Graduate School of the

University of Massachusetts Amherst in partial fulfillment

of the requirements for the degree of

MASTER OF SCIENCE

September 2017

Plant Biology

UNCOVERING *TASSELSHEATH3*. A GENOMIC AND PHENOTYPIC ANALYSIS OF A MAIZE FLORAL MUTANT.

A Thesis Presented

By

THOMPSON H. ZHANG

Approved as to style and content by:

_____

Madelaine Bartlett, Chair

_____

Tobias Baskin, Member

_____

Elsbeth Walker, Member

_____

John Lopes, Department Head

Department of Plant Biology

ABSTRACT

UNCOVERING *TASSELSHEATH3*. A GENOMIC AND PHENOTYPIC ANALYSIS OF A MAIZE FLORAL MUTANT.

SEPTEMBER 2017

THOMPSON H. ZHANG, B.S., UNIVERSITY OF MASSACHUSETTS AMHERST

M.S., UNIVERSITY OF MASSACHUSETTS AMHERST

Directed by: Professor Madelaine Bartlett

In the modern era, maize has become the most successful crop grown in the United States. According to the USDA over 90 million acres of land are planted to corn and 96.2% of the U.S feed grain production is made up of the cereal. Part of the success of maize is due to its floral architecture, and its pollination technique in which the flower opens, exposing stamens containing pollen into the air. A unique organ called the lodicule functions as a release mechanism, forcing the flower to open. Lodicules from grasses and eudicot petals are homologous, yet there is little known of how lodicules are specified during development. Other examples of maize mutants with defects in the lodicule have been discovered including *silky1, bearded ear,* and *sterile tassel silky ear1*, but there has been no definitive pathway found that specifies the developmental characteristics of the lodicule. My work has focused

on a maize mutant, *tasselsheath3* (*tsh3*), which displays a floral phenotype in the lodicule whorl to better understand this organ.

Analysis of *tsh3* was separated into two sections: a quantitative phenotypic analysis of the *tsh3* floral mutant phenotype, compared to a previously unstudied floral phenotype of *tasselsheath1* (*tsh1*), as well as a *tsh1*; *tsh3* double mutant. I found that lodicule morphology and lodicule number was affected in *tsh1, tsh3,* and the *tsh1*; *tsh3* mutants. Section two was to identify the single gene that was disrupted in *tsh3* mutants. Through both fine mapping and next generation sequencing I was able to localize *tsh3* to a region between 148.1mbp and 152.8mbp on chromosome 6. This 4.7mbp region of interest contains 64 protein coding genes. As evidenced by the phenotyping data, *tsh3* plays a role specifying lodicule identity during development and has been localized to this region of chromosome 6 on the maize genome.

CONTENTS

Page

LIST OF TABLES

Table Page

LIST OF FIGURES

Figure Page

CHAPTER 1

INTRODUCTION

The goal of this research is to shed more light on the mechanisms behind floral development and organ organization in monocots, using maize as a model species. Floral diversity is extensive across the angiosperms, from showy roses (*Rosa rubiginosa*), to the tiny flowers of maize (*Zea mays*). Though roses and maize flowers differ in shape, color and even organ identity, they still follow certain rules of construction. Many angiosperm flowers, including roses and maize flowers, have organs arranged in whorls. Each whorl contains different organs making up the whole flower [Schrager-Lavelle et al., 2017; Specht and Bartlett, 2009].

Floral whorls are numbered from the outside in. In grasses, the first, outermost whorl contains the lemma and palea, bract-like structures that surround the rest of the flower. The second whorl contains the lodicules, which are likely homologous to petals [Whipple et al., 2007], but differ in form and function. The third whorl contains the stamens, and the innermost fourth whorl contains carpels. Maize has separate male and female flowers, this means that carpel growth is repressed in male flowers, and stamen growth is repressed in female flowers [Kellogg,2015] (fig. 1).

Although they are likely homologous organs, the function of grass lodicules is dramatically different from that of petals. At anthesis, the lodicules swell forcing the flower open and expose the stamens where wind carries pollen to female flowers for fertilization [Kellogg,2015]. In contrast, petals usually function to attract animal pollinators to flowers

[Schrager-Lavelle, 2017]. There are examples of wind pollinated eudicots including oak and birch, in which their catkins rupture, releasing pollen into the air, but do not rely on lodicules [Sharp and Chisman, 1961]. Although lodicule opening is critical for successful maize pollination and strong fruit set in maize [Whitford et al., 2013], very few genes that control lodicule development have been identified.

Genes that control both petal and lodicule development are the B-class MADS box genes, first described in the ABCE model of flower development [Krizek and Fletcher, 2005]. The ABCE model was developed using *Arabidopsis thaliana* (Arabidopsis) and *Antirrhinum majus* (snapdragon) and it refers to three classes of genes that act both independently and in conjunction with each other to specify organ identity among the different whorls of the flower [Krizek and Fletcher, 2005]. In the model, A class genes confer sepal identity, petal identity is conferred by A class and B class genes together, stamen identity is conferred by B and C class gene activity. C class genes independently control carpel development. E-class function is needed throughout the flower to confer floral organ identity [Krizek and Fletcher, 2005].

B-class gene function is deeply conserved in the angiosperms. In Arabidopsis, the two B-class genes, *APETALA3* (*AP3)* and *PISTILLATA* (*PI),* encode MADS-box transcription factors. In maize, *SILKY1 (SI1)* encodes the orthologue of *AP3*, and *STERILE TASSEL SILKY EAR1 (STS1)* encodes a maize homolog of *PI* [Ambrose, 2000; Bartlett et al, 2015]. Maize flowers in the tassel are male, thus have repressed carpels. In both *si1* and *sts1* mutants the stamens and lodicules are both replaced by bract like structures, something akin to the palea and lemma [Ambrose, 2000; Bartlett et al, 2015]. Thus, B-class

MADS-box genes confer both stamen and petal identity in Arabidopsis, and stamen and lodicule identity in maize.

Several other maize mutants have been shown to have developmental defects in the flower such as *bearded-ear* (*bde*) [Thompson et al., 2009]. The *bde* mutant, which encodes for a MADS-box transcription factor gets its name from the many silks that extend from the ear. In male *bde* flowers, there are is an increase in the number of lodicules per flower [Thompson et al., 2009]. Secondly, there were also instances of partially transformed lodicules into hybrid identity organs with traits of bracts [Thompson et al., 2009].

Apart from *bde* and the B-class genes, there are few described maize mutants where lodicule form is specifically disrupted. *tassel sheath3 (tsh3)*, however, is a novel maize mutant with disrupted lodicule development. In *tsh3* mutants, ectopic bracts occasionally form at the base of the tassel and spikelets and covers or sheaths them, hence the name *tassel sheath*. More pertinent to flower development, *tsh3* flowers display disorganization in the second, lodicule whorl (Whipple et al., 2010). Although *tsh3* was known to have a lodicule phenotype, this floral phenotype has not been fully described or quantified. In addition, the flowers of another *tassel sheath* mutant, *tsh1,* has not been quantified. Although *tsh1* is known to encodes a GATA-zinc finger transcription factor, the molecular identity of *tsh3* is unknown. My thesis work was focused on answering two questions: (1) what is the floral phenotype of *tsh3, tsh1,* and *tsh1; tsh3* double mutants and (2) what is the gene that has been disrupted in *tsh3* mutants.

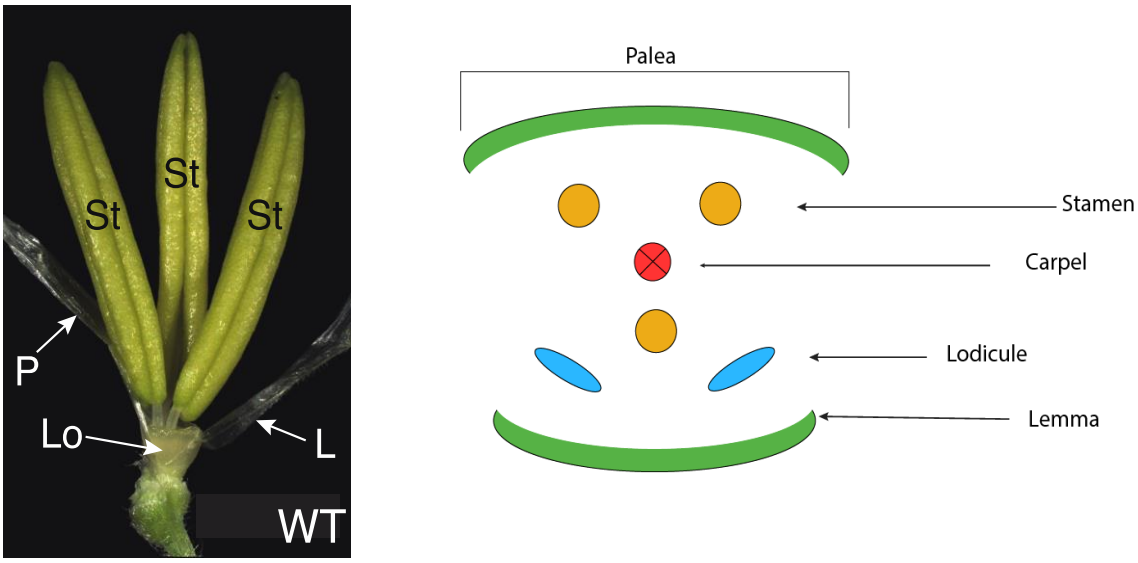

Figure 1: Left: wild type maize spikelet with labelled floral organs. St- stamens; Lo- lodicules; L- lemma; P-palea. Right: diagram of the floral organs, and the carpel is repressed in the male flower.

CHAPTER 2

MATERIALS AND METHODS

2.1 Determining the *tsh3, tsh1* and *tsh1; tsh3* double mutant phenotypes

I harvested male inflorescences (tassels) from maize plants pre-anthesis and labeled each with a number that was associated with the planting number. Tassels were put into Lawson pollinating bags and dried via a 55°C incubator for 24-48 hours. This process stops the tassel from rotting. After drying, I took a branch from each tassel at random and put into a beaker with 100mL water and soap. The beaker was then heated for 1 minute before being brought under the microscope.

The maize tassel has a central spike, surrounded by lateral branches [Kellogg, 2015]. Along these branches, and along the central spike, there are spikelets, which are paired. In each individual spikelet, there are two flowers. I examined individual spikelet pairs and the flowers within them using a 10x light microscope. Removal of the bracts surrounding each spikelet, the outer and inner glumes, revealed whether or not there were two flowers in each spikelet. Each individual flower was characterized visually to quantify for a number of traits and recorded on a table. The traits included stamen number, lodicule number, and hybrid lodicules. I defined hybrid lodicules as lodicules that appeared to have both lodicule characteristics, and palea/lemma-like characteristics. Often, these hybrid lodicules had a palea/lemma like extension at their tips. These traits of interest were used to quantify the differences between *tsh1* single mutant, *tsh3* single mutant and *tsh1; tsh3* double mutant

individuals. Data was then input Microsoft Excel and then exported to RStudio where the package ggplot2 was used to create stacked bar plots.

2.2 Generating an F2 mapping population for identifying *tsh3*

tsh3 was identified in an ethyl methane sulfonate (EMS) mutagenesis screen of a maize inbred line called A619 [Whipple et al., 2010]. A *tsh3* mutant was then crossed to another inbred line, B73. A619 and B73 are significantly different from each other, both phenotypically and genetically (fig. 2). The resulting (F_1) progeny would have been entirely

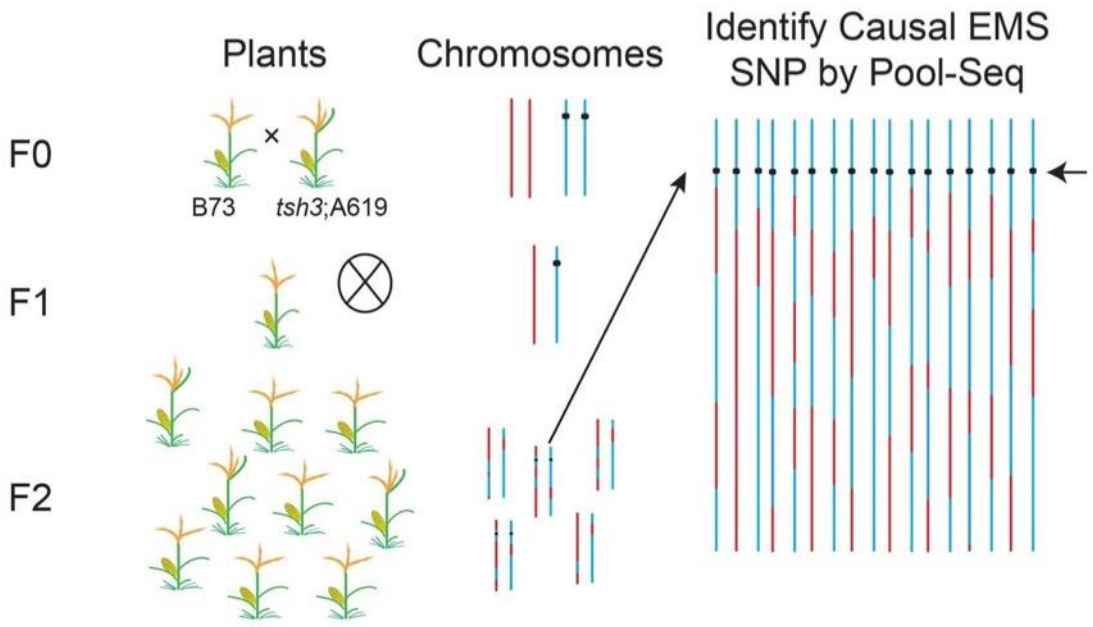

Figure 2: Bulk segregant analysis for mapping the tsh3 EMS mutation. Figure derived from Schlötterer et al. 2014 and Harry Klein

heterozygous at *tsh3,* as well as at every locus that differed between the A619 and B73 genomes. A few (8-10) F_1 individuals were then selfed to generate F_2 progeny. Mendelian genetics dictates that a heterozygous allele crossed to another heterozygous allele will have offspring will carry the genotypes in a 1:2:1 ratio. The 1s will be the two homozygous individuals and the 2 is proportion of heterozygous individuals. Phenotypically the heterozygous individuals will be wild type (if the allele is recessive). Thus F_2 mapping population approximately segregates the *tsh3* mutant phenotype at a 3:1 wild type: *tsh3* phenotypic ratio, according to Mendelian genetics, thus indicating that *tsh3* is a single locus. 80 *tsh3* mutants were identified from the F2 mapping population to be used for both bulked segregant analysis (BSA) and fine mapping.

2.3 Identifying *tsh3* using bulked-segregant analysis and Next-Generation Sequencing

The genomic location of *tsh3* was identified using bulked segregant analysis (BSA) [Michelmore et al., 1991]. The premise of BSA is to use genetic differences between different cultivars or lines, and DNA recombination to localize a gene. The process of BSA depends on recombination of DNA during the crossing over phase of meiosis. During self-cross of the F1 progeny, recombination occurs on homologous chromosomes where they will exchange DNA. As recombination occurs randomly along the chromosome, the one constant will be the presence of two copies of *tsh3* as well as a surrounding island of A619. In the *tsh3* mutants identified from the F2 mapping population, every locus in the genome that differs between A619 and B73 will be segregating at a 1:2:1 genotypic ratio (homozygous A619: heterozygous A619/B73: homozygous B73), except for *tsh3*, allowing us to detect recombination. At the *tsh3* locus, and in the region surrounding *tsh3, tsh3* mutants

will be preferentially homozygous for A619 alleles. By genotyping pools of wild type vs. mutant individuals at thousands of loci across the genome, BSA identifies these areas of increased homozygosity [Michelamore et al.,1991]. We used next-generation sequencing techniques to perform this genome-wide genotyping and localize *tsh3*.

In a separate pool, DNA from 18 *tsh3* individuals were used to generate a single sequencing library. The library was sequenced using an Illumina HiSeq2000 at University of Massachusetts Medical School Deep Sequencing Core (fig. 10). The resulting paired end reads were uploaded to Galaxy [Blankenberg et al., 2014] as fastq files using the FTP loader, CyberDuck. Once sequence was uploaded, the application FASTQGROOMER was used to change the format of the original fastq files to fastqsanger, so the sequencing data could be recognized and used in subsequent tools [Blackenberg et al., 2010] (fig. 10). Then the sequence was trimmed for quality using Trimmomatic [Blackenberg et al., 2010]. Trimmomatic uses a sliding window method to remove all sequences below a certain quality score threshold. The bases to average across was set to 4 and average quality required was 24. This meant that Trimmomatic started with 4 base pairs and continued along the read, moving 4bp at a time, until the quality score dropped below 24. Once the quality dropped below that threshold, the rest of the read was cut, leaving only reads that have passed this threshold.

Reads were concatenated into a single file, and aligned to the B73 genome using tools and software packages integrated into Galaxy [Blackenberg et al., 2014]. The individual datasets were concatenated using Concatenate datasets [Afgan et al., 2016] and aligned to the maize genome using Bowtie2 [Langmead & Salzberg, 2012]. Bowtie2, part of the Tuxedo suite, aligns short next-generation sequencing reads to a much longer reference genome. In

this instance, the reference genome is the version 4 maize genome, which I downloaded from the Gramene database (Tello-Ruiz et al., 2016) and the short reads are all of the reads post trimming [Tello-Ruiz et al., 2016]. The output is a binary alignment (BAM) file that has been sorted and indexed by Bowtie2. The BAM file is used as input for a program that is part of SAMTools, mPileup. SAMTools has two functions, the first is to change the BAM file to a pileup file. This mpileup file format is important to steps downstream of the pipeline. Secondly, it identifies positions in the alignment that vary from the reference genome [Li et al., 2009; Li, 2011]. The resultant mPileup is then filtered using the Galaxy tool filter Pileup, which removes read bases that are below a certain quality (score) threshold, and reports the number of reads at each position that differ from the B73 reference genome [Afgan et al., 2016]. This value was used to generate the frequency of reads that differed from B73 at each position (number of reads that differed from the reference/total number of reads). When the value of this frequency exceeded 0.99 (99% of the reads differ from B73), a position was interpreted to be homozygous A619.

Since *tsh3* was induced via an EMS mutation in A619, I am looking for a genomic region on a chromosome that does not match B73. The frequency values generated from the pileup file were therefore used to generate a histogram which visually showed where the sequencing data differed from the reference genome (B73) (i.e. regions of high homozygosity). This histogram, generated using the R package ggplot2, visually illustrates those differences across the length of each chromosome. The histograms were made for each chromosome with the y-axis depicting the count, or number of instances deviation fulfilled the criteria and the x-axis is the base pair number in million base pair bins.

The histogram, though helpful in localizing the location of *tsh3* to region on a single chromosome, there were approximately 20mbp in that region. Thus I used another tool, Varscan to identify the genes in the newly defined region of interest. Varscan reports variants at a specified read depth and will designate it heterozygous and homozygous respectively [Koboldt et al., 2012]. Using the Mpileup file from Galaxy, Varscan was run with a low minimum read depth (3) in order to include as many single nucleotide polymorphisms (SNPs) as possible since there were so few individuals that made up the original library. After Galaxy finishes running Varscan, the file was downloaded locally and the unix command "awk" was used to extract the region of interest where there was high homozygosity (148mbp-152mbp on chromosome 6). The data was then uploaded to the MGHPCC server where SNPEff was run. SNPEff will annotate and predict effects of SNPs while Varscan acted as a filter to remove any SNPs that low read depth (1 or 2 reads). The result is a list of SNPs within a region that have predicted effects as well as whether or not the SNP is homozygous or heterozygous. The output file was downloaded and separated based on the predicted effect the SNP would have on genes in the region.

2.4 Fine mapping of *tsh3*

In parallel with BSA, and using my next-generation sequencing results to design markers, I also undertook fine mapping of *tsh3*. Leaf samples were collected from 80 *tsh3* individuals from the same F_2 population described above. DNA was extracted from the leaf samples using a TrisHCl buffer made up of 200mM TrisHCl, 250mM NaCl, 25mM EDTA and 0.5% SDS by weight. The concentrations of the individual samples were measured using

a Thermo Scientific *NANODROP 2000* spectrophotometer and diluted to 50ng/uL. The spectrophotometer measures the concentration of nucleic acid based on the absorption of light. However, this method is limited due to the nano-drop indiscriminately measuring both DNA and RNA concentration.

Two previously described markers were identified that likely flanked *tsh3*, umc1352 and umc1805. These are short sequence repeat (SSR) markers catalogued at the Maize Genome Database [Sen et al., 2010]. SSR markers are primer pairs that amplify a genetic difference between A619 and B73 inbreds. umc1352 and umc1805 are on chromosome 6, located at 136,195,104bp and 152,846,900bp, respectively. PCR amplification was achieved using New England Biolabs (NEB) Phusion polymerase due to the high concentration of guanine and cytosine in maize DNA. 10μL reactions contained 50ng of DNA, 1x GC buffer, 10% DMSO, 250μM dNTPs, 0.5μM forward and reverse primers and 1 unit of NEB Phusion polymerase. All 80 individuals were amplified with both markers, loaded on 3.5% agarose gel made with a 1:1 mixture of 3:1 agarose and low electroendosmosis (LE) agarose by weight. 3:1 agarose was used as it allows for higher resolution with smaller band size. The amplified band size difference was no larger than 50bps for both umc1805 and umc1352. The agarose was mixed with appropriate the volume of 0.5X Tris/Borate/EDTA (TBE) buffer. Each gel also had 4 controls: a negative control (H_2O), A619, B73 and an artificial heterozygous mixture of A619 and B73 made with a 1:1 mixture of A619 and B73 DNA. The gel was run for 105-120 minutes at 100V and imaged using a UV transilluminator. 100bp ladder was added to the gel to make sure that the PCR product is the right size, as well as to approximate the size difference between the B73 and A619 bands.

There were no more useful SSR markers in the *tsh3* region already catalogued in the Maize Genome Database. Therefore, I designed cleaved amplified polymorphic sequence (CAPS) genotyping assays using SNPs identified in my next-generation sequencing data analysis. Genomic differences between B73 and A619 induce or remove restriction enzyme cut sites. CAPS assays amplify these specific regions using PCR and then cleave resultant product using restriction endonucleases (enzymes) to determine genotype. Both the SSR and CAPS markers were used to genotype *tsh3* individuals. All PCR was performed on genomic DNA with Phusion polymerase and GC buffer to account for the high GC content in maize. PCR product of CAPS markers were run at 110V for 35 minutes on a 1% agarose gel made with LE agarose by weight and 0.5X TBE buffer by volume. If banding showed expected size, the remaining PCR product was digested using a restriction endonuclease. Restriction digest solution contained 1x CutSmart Buffer, 2μL of PCR product and 5 units of restriction enzyme. The new solution was incubated at 37°C for 1 hour and run on a 1% agarose gel. The resulting bands would then indicate whether the DNA is homozygous for A619, B73 or heterozygous for A619 and B73 thus indicating recombination somewhere between *tsh3* and that locus.

CHAPTER 3

RESULTS AND DISCUSSION

3.1 *tsh1* and *tsh3* affect lodicule identity and lodicule number

Both *tsh1* and *tsh3* were grown from previous seed stocks, while *tsh1;tsh3* double mutants were generated through a cross of *tsh1* mutants x *tsh3* mutants, leading to fully heterozygous offspring at both loci. These F1 plants were self-crossed and this led to the segregating F2 generation which was analyzed. There were 4 potential phenotypes, WT, *tsh1*, *tsh3* and *tsh1*; *tsh3* that segregated in a 9:3:3:1 ratio. These four phenotypes were the focus of my assays of floral morphology. To analyze the effect of the *tasselsheath* mutations on the male inflorescence, and to identify genetic interactions between *tsh1* and *tsh3*, I looked at 187 individual florets, 77 wild type florets, 43 *tsh3* florets, 36 *tsh1* florets, and 30 *tsh1; tsh3* florets. These individuals were picked via tassel phenotype, which is distinct in *tsh1* and *tsh3*, and extremely enhanced in double mutants (fig. 3). The wild type tassel was used as a baseline for the double mutant tassels. The tassels for *tsh1* display strongly reduced branching down to a single central spike. There are also multiple bracts that form at the base of the tassel above the flag leaf. The tassel phenotype of *tsh3* is much more subtle; there is not as much reduced branching as *tsh1* (fig. 3). The most visible phenotype is the presence of bracts that form at the base of spikelet pairs. The *tsh1*; *tsh3* double mutant looked like a mix of the single mutants, but exacerbated (fig. 3). There are bracts that cover the spikelets, single spike and bracts that cover the entire tassel. These were used for comparative analysis of the number of lodicules, number of stamens and number of hybrid identity organs per floret. Hybrid

identity organs were defined as any organ that had characteristics of two or more wild type organs.

I found that, as expected, wild type florets contain 3 stamens and 2 lodicules with very few exceptions. There were 2 instances where the florets had either 1 or 3 lodicules (fig. 4). However, the overall majority was two lodicules per floret. There were very few instances of hybrid identity organs in wild type individuals. Out of the 77 florets, I observed hybrid identity organs 5 times (fig. 4).

.

Figure 3: Tassels of maize showing differences between wild type, tsh1, tsh3, tsh1; tsh3.

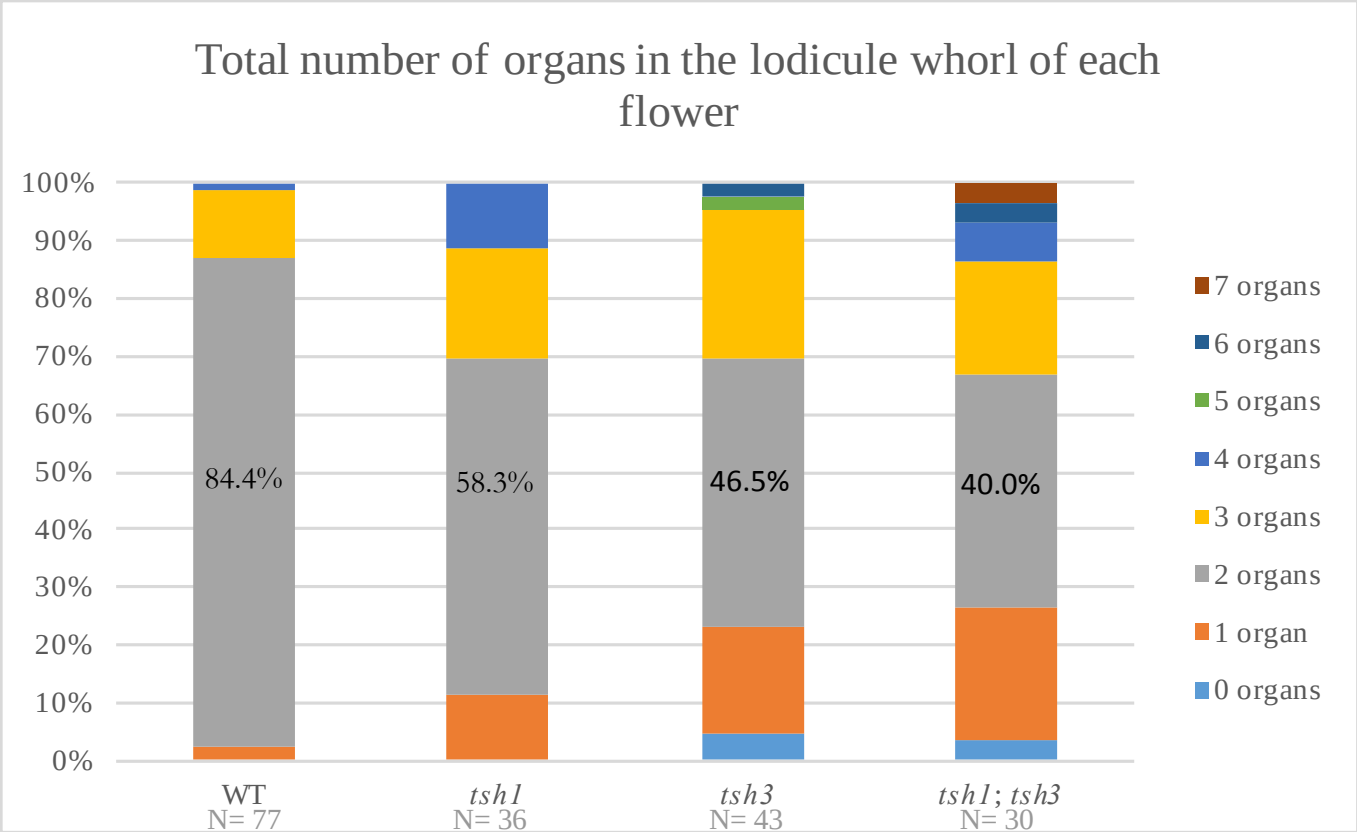

Figure 4: Stacked bar plots showing the percentage of the number of instances of the total number of organs in the lodicule whorl per total number of each phenotype. Percentages found by number of instances over total number of flowers of the phenotype.

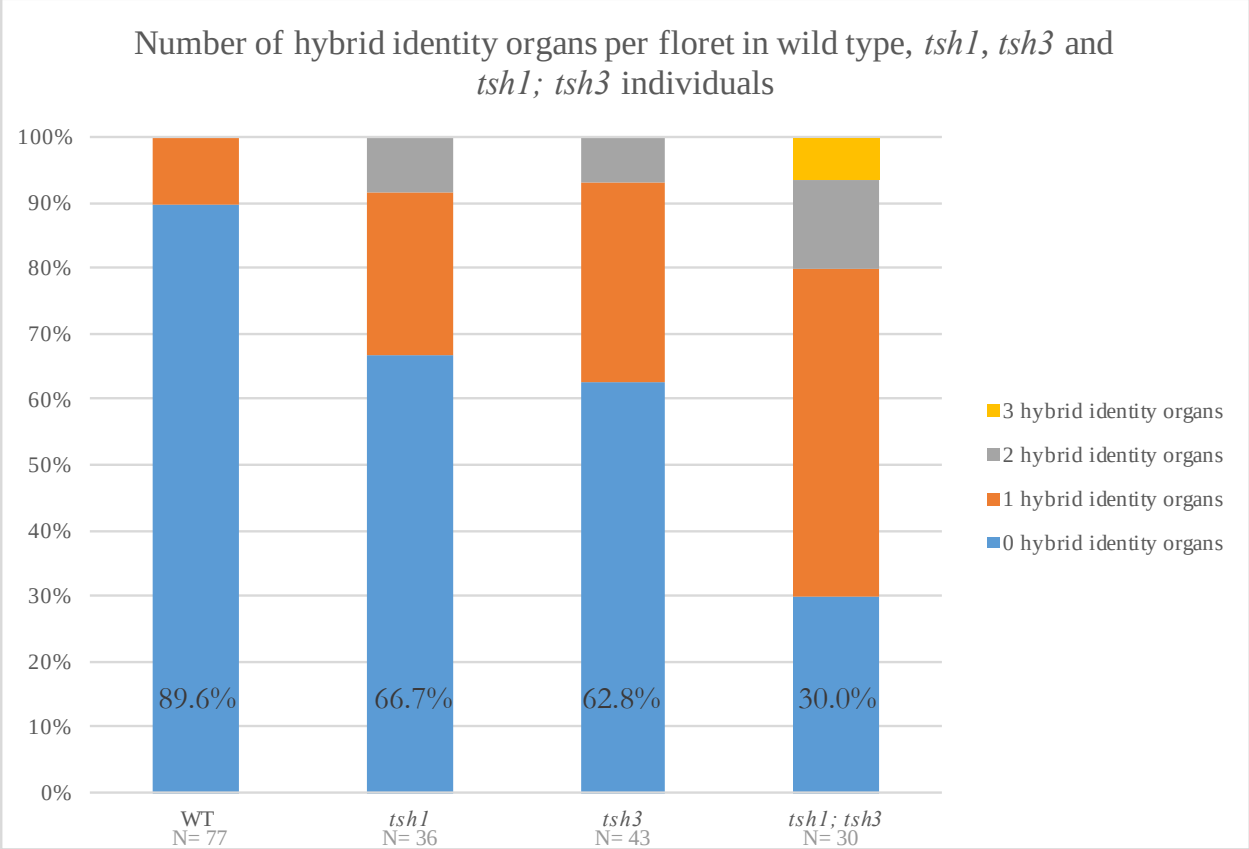

Figure 5: Stacked bar plots showing the percentage of the number of instances of hybrid identity organs in the lodicule whorl per total number of each phenotype. Percentages found by number of instances over total number of flowers of the phenotype.

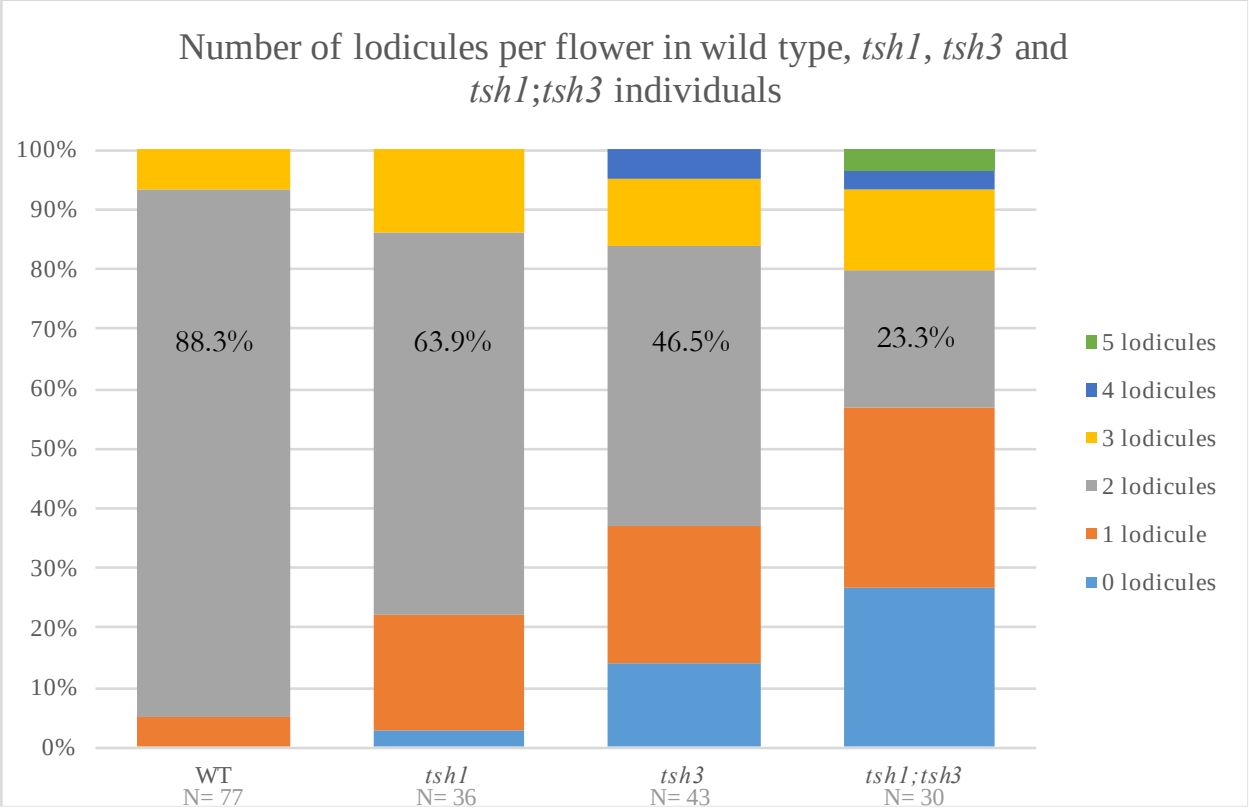

Figure 6: Stacked bar plots showing the percentage of the number of instances of the number of lodicules per total number of each phenotype. Percentages found by number of instances over total number of flowers of the phenotype.

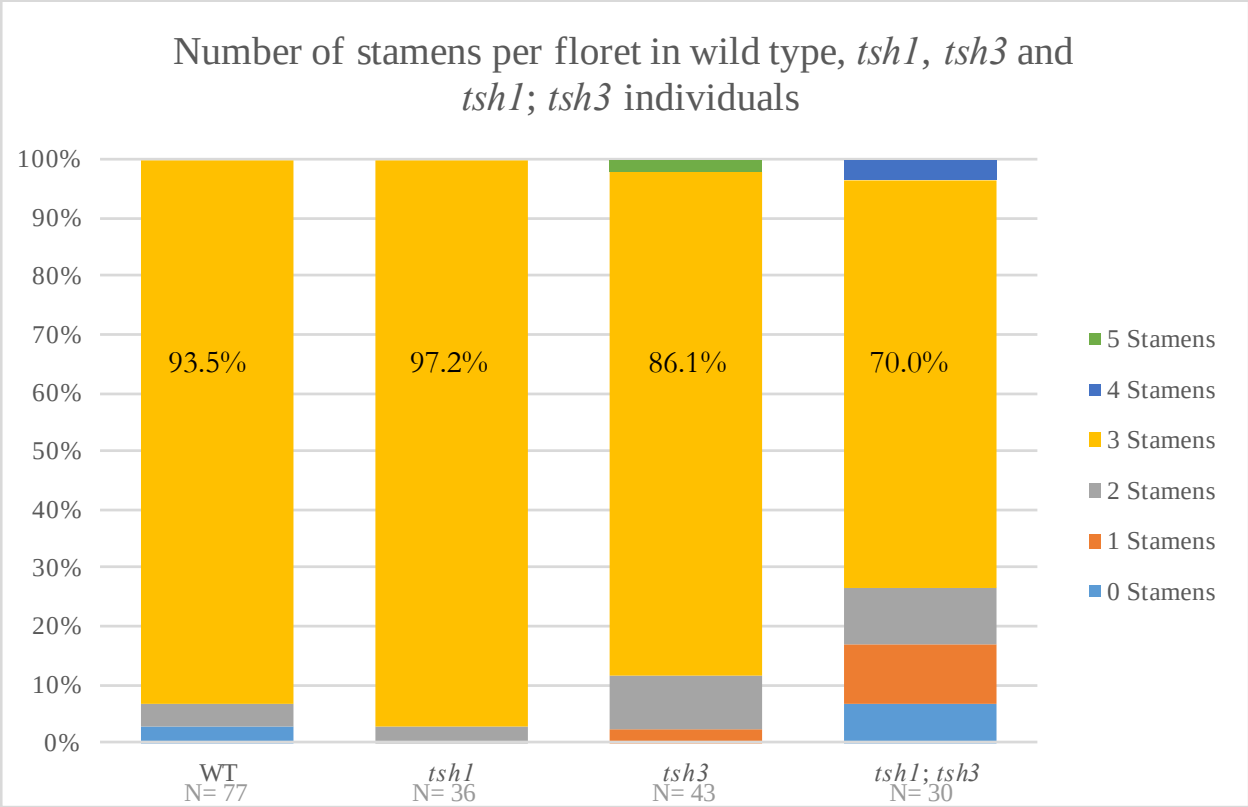

Figure 7: Stacked bar plots showing the percentage of the number of instances of the total number of organs in the lodicule whorl per total number of each phenotype Percentages found by number of instances over total number of flowers of the phenotype.

The *tsh1* florets were analyzed as well and though the tassel structure and size deviated dramatically from wild type [Whipple et al., 2010], the number of floral organs remained constant. There were generally still two lodicules present, as well as three stamens (fig. 6, fig. 7). However, there was a significant increase in the number of hybrid identity organs found. Out of the 36 florets analyzed 12 contained at least 1 hybrid identity organ and 3 out of the 12 contained two hybrid identity organs. Overall total organ number is more variable in *tsh1* mutants, although the average second whorl organ number is still 2 (fig. 4).

The causative gene of *tsh1* was previously identified as a GATA zinc-finger protein, a transcription factor with a homologue in Arabidopsis known as *HANABA TARANU* (*HAN*) [Whipple et al., 2010]. The Arabidopsis *han* mutant phenotype has been analyzed and compared to that of the wild type variants. Since petals have been shown to be homologous to lodicules, we can compare petal number and lodicule numbers. Both of the two wild type Arabidopsis variants (Columbia and Landsberg erecta) have 4 petals per flower with no variation [Ding et al., 2015]. There was also little variation in the number of lodicules in the wild type maize flowers. Comparing the *han* mutant to its wild type counterpart, it was possible to see a significant reduction in the number of petals [Ding et al., 2015]. Not only was there a significant decrease in the number of petals, there were changes in petal identity as well, which similarly occurred in the *tsh1* mutant.

Out of the 43 *tsh3* florets, there was a greater degree of variation than was seen in *tsh1* in all aspects; the number of lodicules, the number of hybrid identity lodicules and the number of total organs (fig 4, fig 5, fig. 6). The frequency of changes in the number of stamens was not changed greatly in the single mutants compared to wild type (fig. 7). Out of the 77 wild type flowers observed, 93% of them had 3 stamens (fig. 7). In the *tsh1* and *tsh3* single mutants, they had 97.2% and 86.1% with 3 stamens respectively. There is slightly higher variation in the *tsh3* mutant. In the *tsh1*/*tsh3* double mutant 70% of individuals had 3 stamens which is less than that of the single mutants and the wild type (fig. 7). Though there were a high percentage of flowers with 2 lodicules per floret, there was significant variation in lodicule number (fig 5). While the number of normal lodicules decreased, hybrid identity organs increased in number. This result indicated that *tsh3* has a role in regulating organ identity, more specifically in the lodicules.

The presence of hybrid identity organs in both the *tsh1* and *tsh3* individuals showed that the underlying genes affected the same whorl. As a result, it was of interest to see how both genes combined would impact organ differentiation. To answer this question, I examined florets from the *tsh1*; *tsh3* double mutant individuals. The hybrid identity organ and lodicule counts were definitely exacerbated in the *tsh1; tsh3* individuals. The range of values for hybrid identity organs and for number of lodicules were much greater. There is a significant variability in the number of lodicules (fig. 6). The total organ number was the most variable in the *tsh1*; *tsh3* double mutant (fig. 4).

The *tsh3* mutant phenotype differs considerably from that of known B-Class mutants such as *si1*. *si1* mutants are completely sterile in the male floret. They have bracts in the place of stamens and bracts in the place of lodicules. However, there are no changes in organ number in the floret, while there are distinct changes in the organ number in *tsh3* mutants. Compared to the frequency of changes in lodicule identity in the *si1* mutants, the *tsh3* mutant had changes in lodicule morphology less often. However, instead of wholesale changes in identity, as is seen in *si1,* I would say it is more of a fusion of whorls between the palea and lodicules.

3.2 Bulked Segregant Analysis using Next-generation Sequencing

The primary goal of the next generation sequencing was to pilot the use of whole-genome next-generation sequencing for bulked segregant analysis in maize, and to identify the causative *tsh3* locus. DNA from 18 *tsh3* individuals was pooled, and a single sequencing library was constructed and sequenced on an Illumina MiSeq instrument at UMass Medical

School. This sequencing run produced 341,632,923 paired end 100bp reads. After trimming and quality control [Blackenberg et al., 2010], 97.5% of these reads mapped to the v4 B73 reference maize genome using Bowtie2 [Langmead & Salzberg, 2012]. 46,746,135 of these mapped reads mapped to chromosome 6 of the reference genome. Average coverage of the maize genome was 8X.

The next series of steps were sequence processing and quality control allowing the sequencing data to be read and manipulated with the Galaxy tools. The Trimmomatic procedure was the only step in which there was active input, dictating the sliding window with a quality score of 24. With mPileup, I was able to list all the compiled SNPs along all of the chromosomes. They are combined with the script resulting in a file that can be further edited.

I constructed a histogram that showed the length of the chromosome vs the number of reads that show differences between the reference genome and said reads. The length of the chromosome is binned in to 1,000,000bp sections as each maize chromosome is approximately 200,000,000bp long. Chromosome 1 is the longest at approximately 300,000,000bp and chromosome 10 is the shortest at approximately 140,000,000. If the next generation sequencing reads in a particular bin match with the reference genome, then there will be a low count. If the reads in a bin do not match, then there will be a high count and a high bar at that bin position. Due to *tsh3* existing in an A619 background and all 18 of the individuals were phenotypically *tsh3,* there should be an area of high homozygosity where there is significant deviation from the reference genome (B73).

I compiled histograms for each chromosome of the maize genome (fig. 7). Chromosome 2, 3, 7, 9 and 10 all had negligible amounts of homozygosity, thus ruling out *tsh3* being on any of these chromosomes. On chromosome 1, there was a pileup of lower count numbers along the first 55,000,000-60,000,000bp. Then there is a negligible amount of homozygosity until a sharp spike at the 298,000,000bp bin. The pileup data of chromosome 4 had low homozygosity until approximately 195,000,000bp (fig. 8). Then there is a higher level until the end of the chromosome. Chromosome 5 and 8 showed something similar to chromosome 1. These peaks at the ends of chromosomes 1,4,5, and 8 are common in BSA results from our lab (data not shown), and may represent regions of the B73 seed stock that the lab uses that differs from the B73 stock that was sequenced as the 'reference genome' (Schnable, et al., 2009). As a result, it is unlikely that these peaks contains *tsh3*. This leaves the broad peak on chromosome 6 as the likely site of *tsh3*.

There is an interval on chromosome 6 where there is a significant difference from the reference genome (fig. 8). The main peak is about 20,000,000bp long and at its maximum height, it is approximately three to six times higher than the other peaks on the other chromosomes (fig. 8; fig. 9). That implies that this is the "island" of A619 DNA that contains the *tsh3* SNP. The large peak begins at approximately 140mbp and extends to 160mbp.

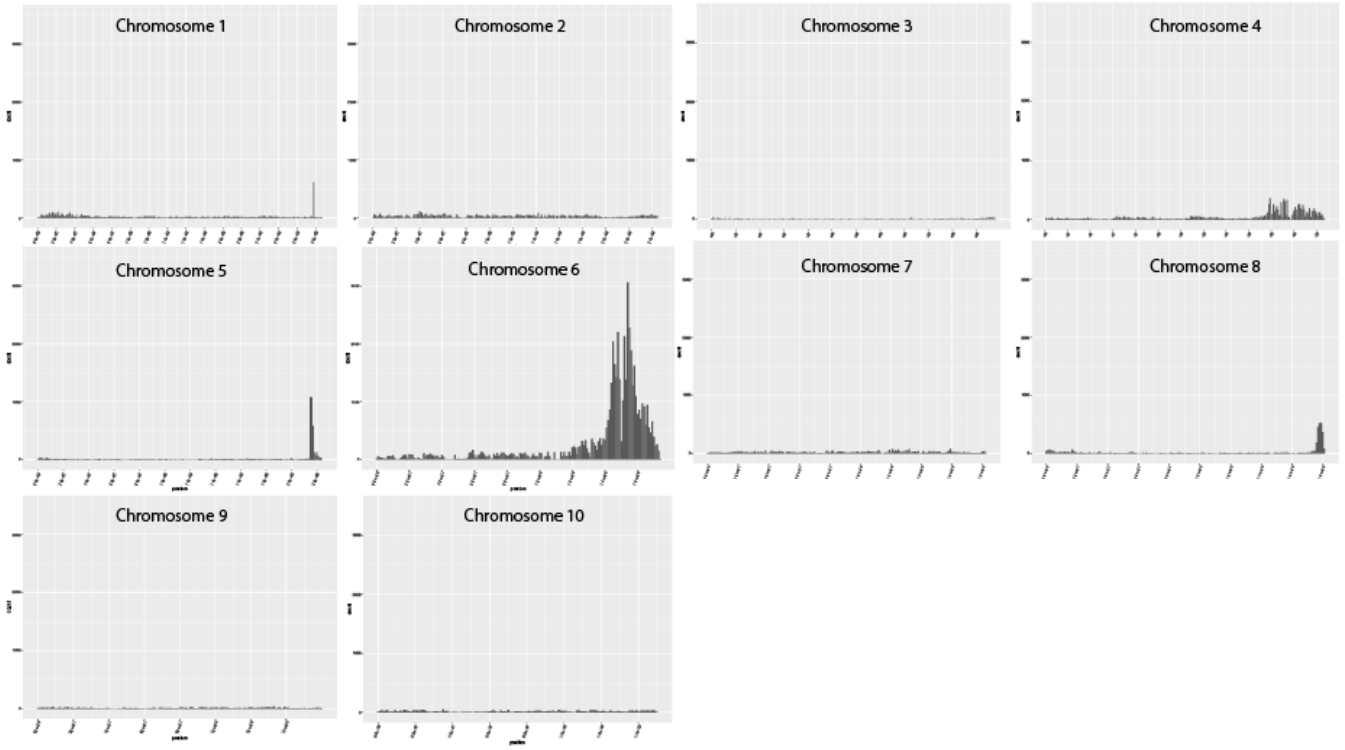

Figure 8: Histograms showing the number of reads that differ from the maize genome in 1Mbp bins. The x-axis dictates the location on the chromosome. The y-axis dictates the number of reads in the 1Mbp bin that differ from the reference genome (B73) greater than or equal to 99%. Chromosome 6 shows tentative location of tsh3 by the large number of counts that differ from the reference genome (B73). The peak in chromosome 6shows and extended length that has a high number of reads that deviate from the B73 reference genome.

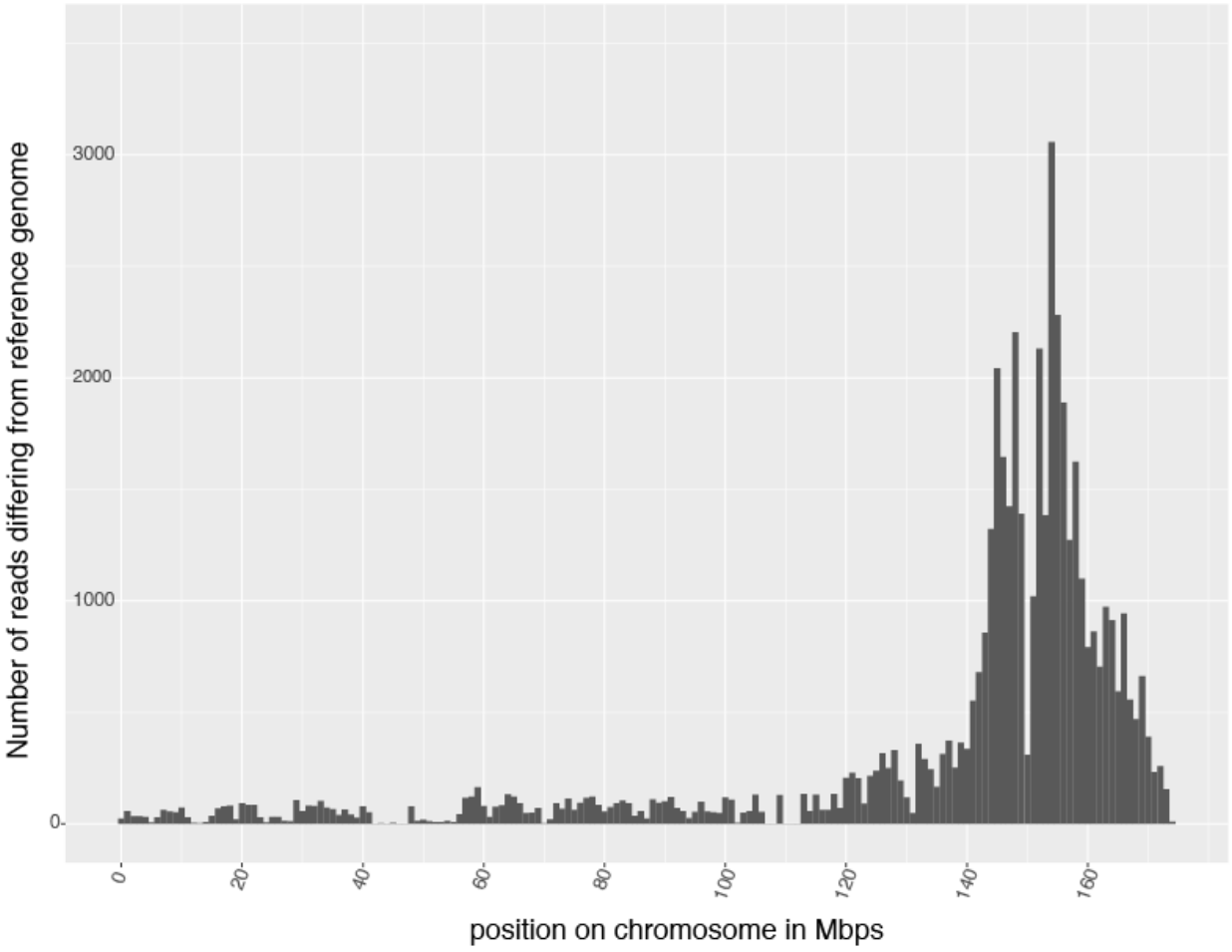

Figure 9: Histogram of chromosome 6 showing number of reads that differ from the reference genome. Starting at approximately 140mbp the count of reads dramatically increases until approximately 160mbp. The x-axis dictates the location on the chromosome. The y-axis dictates the number of reads in the 1Mbp bin that differ from the reference genome (B73) greater than or equal to 99%. Chromosome 6 shows tentative location of tsh3 by the large number of counts that differ from the reference genome (B73).

Next, I looked for individual genes in the region that had canonical EMS SNPS that might disrupt gene function. These SNPs were identified using the programs VarScan and SNPeff [Koboldt et al., 2009; Koboldt et al., 2012; Cingolani et al., 2012]. VarScan converts the mpileup file produced by samtools into a variant call format (vcf) file that SNPEff can read. This vcf file is essentially a list of positions in the B73 reference genome, and the genotype of the samples sequenced at each position. SNPEff consolidates this list of SNPs that VarScan outputs and annotates the predicted effect of each (i.e. identifies missense, nonsense, and loss of function mutations). Prior to running VarScan, there are filters that I can change to make the criteria more or less stringent. I decided to put the required read depth at 3 to be recognized as a SNP. This allowed me to take as many SNPs as I could and analyze them without them being masked or removed. This was done in order to prevent read depth from affecting the output of SNPs and their causative effects (fig. 10).

By using SnpEff and the low-depth VarScan in conjunction, I was able to narrow down the list of candidate genes that exist in my region of interest. The output was 4108 SNPs in the region with varying effects. Focusing on the three types of effects, being loss of function (LOF), high (HIGH), and moderate (MOD). There were 134 unique genes in the region of interest. This new list also shows the change in nucleotide sequence. I focused on canonical EMS SNPs in my analysis, meaning they were either G-A or C-T mutations. Once the list was trimmed down using that filter, I was left with 433 EMS SNPs that were annotated to have an effect on genes in the area. I then separated the list of genes via the severity of impact. As a result, there were 30 unique SNPs that are categorized as LOF, 43

unique HIGH SNPs and 360 MOD SNPs. The 30 unique LOF SNPs were matched with their causative gene and likely candidate genes were Sanger sequenced (table 2).

Picking through genes to sequence were very difficult as I attempted to create a series of filters in order slim down the large list of SNPs post analysis. Firstly, the genes were separated into the three different characterizations of predicted severity, LOF [loss of function], HIGH, MOD [moderate]. The resultant individuals were then separated to whether or not they were canonical EMS SNPs. The majority of EMPs SNPs are A/T and oppositely C/G. This dramatically shortened the list of possible genes. Then the genes were searched against known maize genetic variants [Chia et al., 2012, to remove any SNPs that occur naturally in maize inbreds. Finally, the search was looked through manually in order to select for genes that are likely. Ideally, the next step would to take a look at gene ontology terms and try to incorporate that into the data collected. What this allows me to do is look at homology with other organisms.

There were some genes that I had individually sanger sequenced as they were potentially viable candidates in my region of interest [table 2]. Zm00001d038207, located at 151,568,742bp has no known name in the maize genome database. However, the gene has an ortholog in Arabidopsis that codes for the CUC transcription factor, which has a number of roles in plant development [Tello-Ruiz et al.,2016]. However, no lesions that could change or disrupt gene function were identified in this gene.

Another gene of interest was Zm00001d038175, located at 150,542,742bp. This gene is labeled as an auxin response protein or auxin IAA [Tiwari et al., 2001]. Auxin is well known for having a large role in development and growth in the meristem [Tiwari et al.,

2001], making this gene a strong candidate for *tsh3*. Out of all the candidates SNPs identified in my next-gen sequencing analysis, none of the sequencing came back with positive results showing the EMS SNP. In light of that, I used my next-gen data to identify *tsh3* using traditional fine-mapping techniques (Gallavoti and Whipple, 2015)

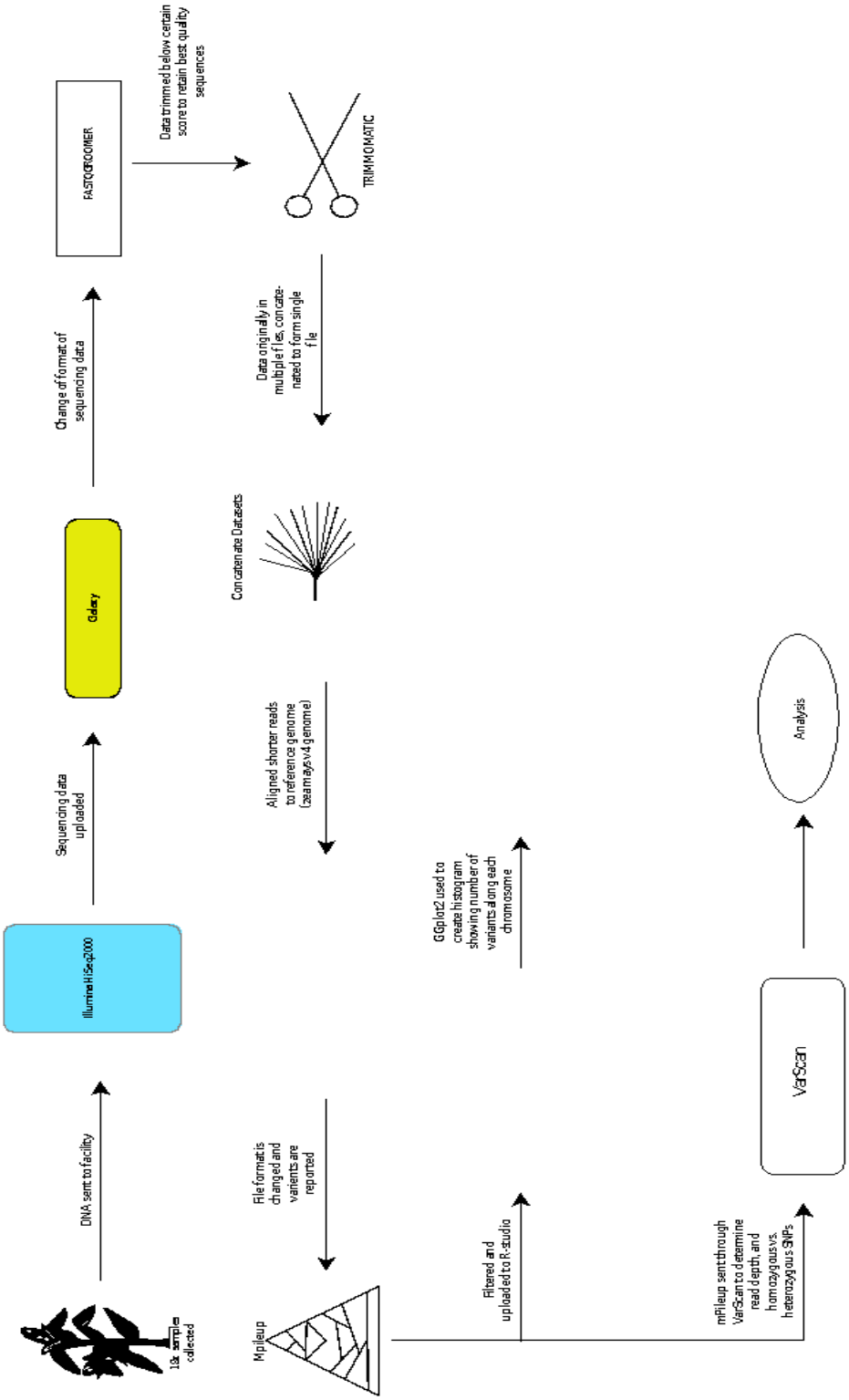

Figure 10: Flow-chart delineating the created pipeline for next generation sequencing analysis.

3.3 Fine Mapping Narrows *tsh3* Region

My next-generation sequencing results localized *tsh3* to the long arm of chromosome 6 . My first step was validating this result using known short sequence repeat (SSR) genetic markers from the Maize Genetics Stock Center database, maizegdb (Sen, 2010). Each SSR marker begins with a primer pair that amplifies a region of the genome using PCR. SSR works by utilizing the differences in genomic sequence between inbreds of the maize genome. The resulting PCR product amplifies products of different sizes which is shown using gel electrophoresis. Marker umc1352, found at maizegdb, is located at ~136,195,104bp on chromosome 6. The umc1352 primer pair amplifies a region of 147bp in the B73 reference genome, while it amplifies a region approximately 20bp larger in the A619 genome (fig. 11). This size difference, detected on a high-percentage agarose gel, can be utilized to tell the genotype of a particular *tsh3* mutant plant at that locus. Out of the 80 *tsh3* individuals that were tested, meaning 160 chromosomes, there were 35 instances of recombination at umc1352. This implies that *tsh3* is 21.88cM away from umc1352 (35/160 =21.88%). The other SSR marker I had to validate was further towards the end of chromosome 6, umc1805, located at 152,846,298. The umc1805 primers amplify a band of ~65bp in A619 and 50bp in B73 (fig. 12). I detected 21 instances of recombination at this marker, tested against the same 80 *tsh3* individuals (160 chromosomes) used for umc1352. This recombination frequency (22/160 =13.75%) means that *tsh3* is 13.75cM away from umc1805. There were 12 plants recombinant at both umc1352 and umc1805. Taken together, my results clearly demonstrate that *tsh3* is linked to both umc1805 and umc1352, and likely lies in-between these two markers.

After confirming that *tsh3* likely lies between 136,195,104bp and 152,846,877bp on chromosome 6, the next step was to reduce the *tsh3* interval from 16Mbp and 99 known protein coding genes. Since there were no more well-located SSR markers polymorphic between A619 and B73 available at maizegdb [Sen et al.,2010], I designed a series of Cleaved Amplified Polymorphic Sequence (CAPS) markers using my next-generation sequencing data [Konieczny, 1997].

CAPS assays work by once again utilizing the difference between inbred genomes. An area of interest is amplified with a key single nucleotide polymorphism (SNP). This SNP dictates either the presence or absence of a restriction enzyme cut site; after digestion with the aforementioned enzyme, the solution is run out on a gel to see visual difference in size allowing me to see parental genome at that locus. The first marker I developed overlapped the gene GRMZM2G889326, located at ~141.5mbp. The GRMZM2G889326 primer pair amplifies a 500bp sequence that contains a known sequence difference between the two parental genomes, in this case A619 and B73 at a BsaWI restriction enzyme cut site. BsaWI, cuts the A619 PCR product into two 200bp and a 300bp fragments, but does not cut the B73 PCR product (fig. 13). I used this marker to genotype 80 individuals (160 chromosomes) and found 22 recombinants (22/160=13.75%=13.75cM).

The next marker I designed was in a gene candidate for *tsh3*- GRMZM2G044150 – located at 145,148,378bp. GRMZM2G044150 codes for a lateral organ boundary gene that we thought fit with the *tsh3* phenotype because of the hybrid lodicule phenotype [Shuai, Reynaga-Peña, and Springer, 2002]. The lateral organ boundary gene family has been found to have alterations in structure of plants. More specifically the gene *asymmetric leaves2* has been shown to change the leaf structures in Arabidopsis; in monocots, we thought this might

explain the presence of bracts on tassels in *tsh3* mutants [Lin, 2003]. The gene was sequenced, but no SNPs that might inactivate the gene were found in the coding sequence. However, there was a difference in a SacI cut site that allowed me to design another CAPS assay. SacI cuts GRMZM2G044150 in A619, but not in B73. I designed a primer pair amplifying a PCR product that contained this SacI cut site (fig. 14). I used this marker to genotype all the individuals I had identified as recombinant at umc1352, umc1805 at GRMZM2G044150. I found evidence for recombination between *tsh3* and GRMZM2G044150, implying that this gene is not *tsh3*, but is instead ~19cM away (14/72 chromosomes = 19.44%).

The final marker I designed was in another sequenced gene, Zm00001d038104, located at 148,044,310bp, encoding a glycoside hydrolase protein. In Arabidopsis, GLYCOSIDE HYDROLASE 1 functions throughout the plant as a form of pathogen defense and as a stress response gene [Xu et al., 2004]. Though there are no necessarily recorded functions in terms of organ development and floral organization, there was a potential loss of function SNP in this gene, identified in my next-gen sequencing analysis. However, this SNP was also seen in our A619 material, eliminating the possibility of this being the causative SNP. Variation in the gene did allow me to make a CAPS marker. Plants were genotyped at this marker, utilizing the restriction enzyme HphI. HphI cuts a 287bp product in B73 but not in A619 (fig. 15). Out of the 14 prior recombinants at other markers, there were 5 recombinants. This indicates that Zm00001d038104 is 17.8cM away from *tsh3* (5/28 = 17.8%).

The results of my fine mapping revealed a region on chromosome 6 that contains 64 protein genes, none of which is a clear candidate *tsh3* (fig. 16). Further fine mapping efforts will be required to reduce this *tsh3* interval further.

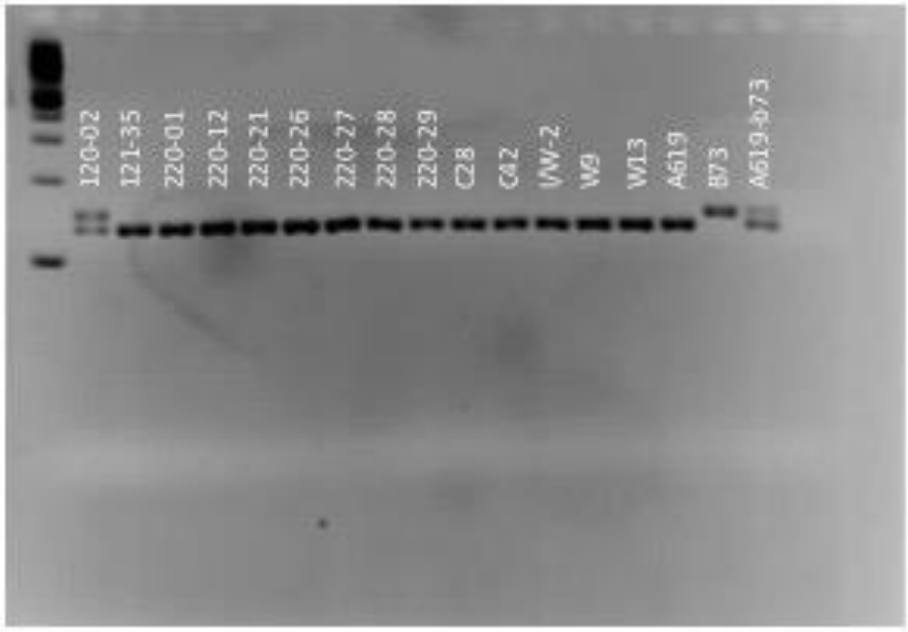

Figure 11: Gel electrophoresis umc1352: SSR PCR, 3.5% agarose, 100V, 110 min. 100bp ladder left. Two bands of amplified DNA indicate recombination.

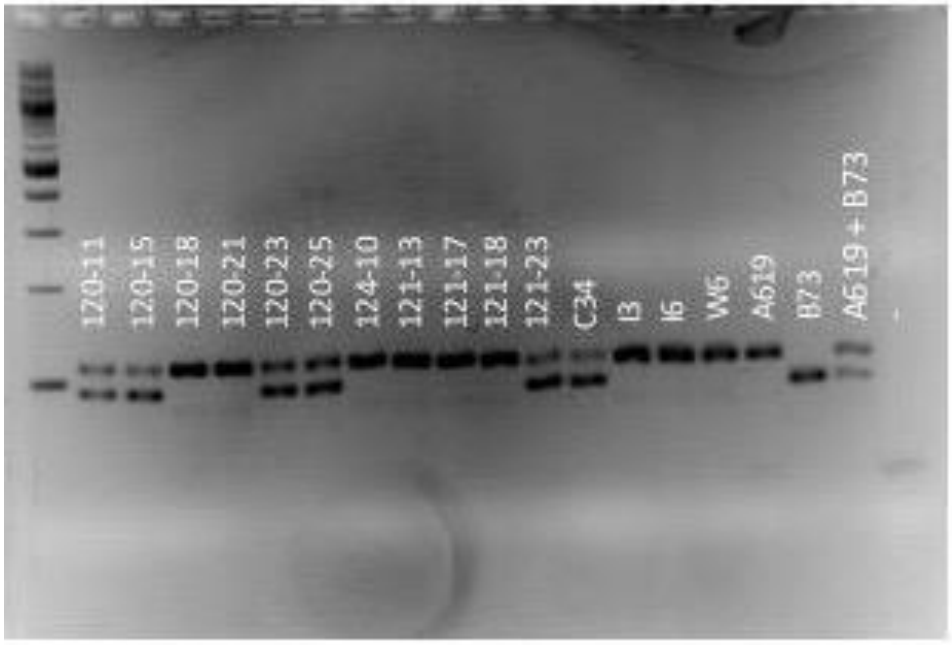

Figure 12: Gel electrophoresis, with marker umc1805: SSR PCR, 3.5% agarose, 100V, 120 min. 100bp ladder left. Two bands of amplified DNA indicate recombination.

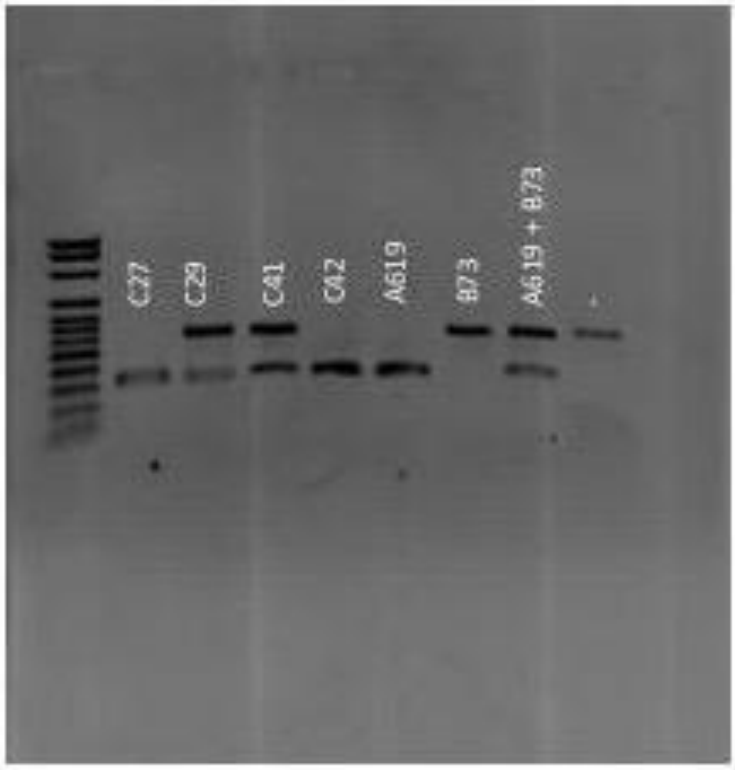

Figure 13: Gel electrophoresis, CAPS marker, 1% agarose, 30 min, 120V Sac1 endonuclease, 100bp ladder left. Two bands of amplified DNA indicate recombination

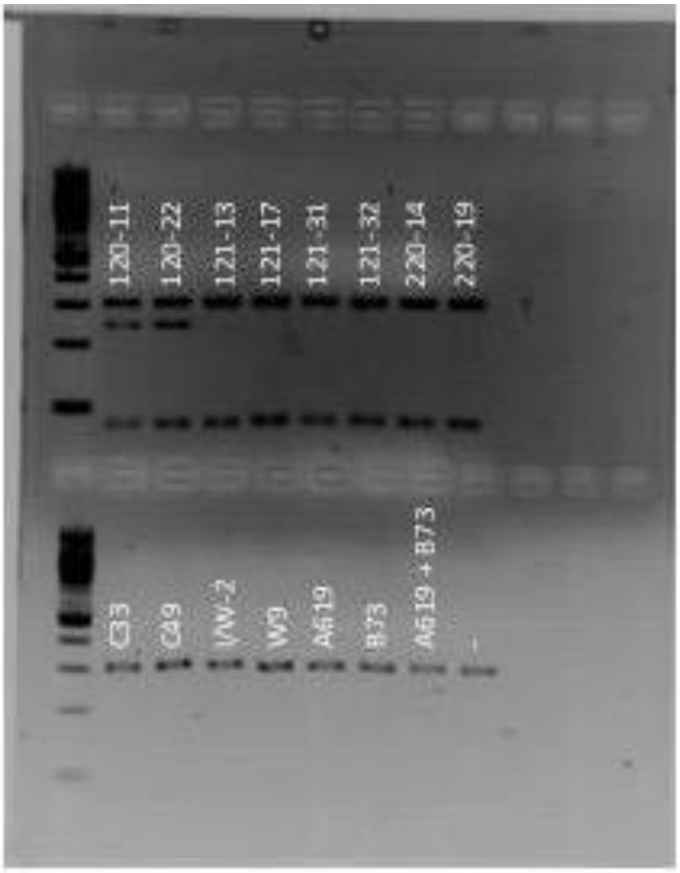

Figure 14: Gel electrophoresis, CAPS marker, 1% agarose, 30 min, 120V Sac1 endonuclease, 100bp ladder left. Two bands of amplified DNA indicate recombination.

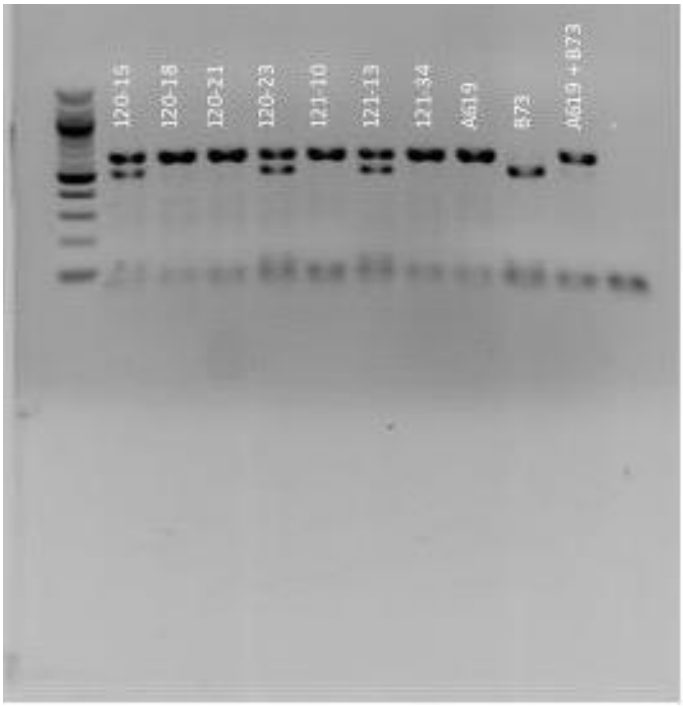

Figure 15: Gel electrophoresis, CAPS marker, 1% agarose, 35 min, 110V HphI endonuclease, 100bp ladder left. Two bands of amplified DNA indicate recombination.

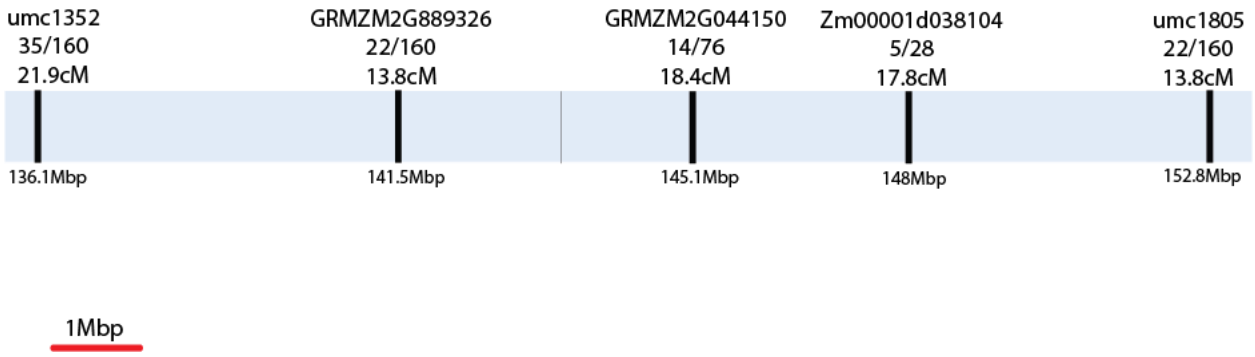

Figure 16: Fine mapping localized tsh3 to a region between 148Mbp and 152.8Mbp. Each marker is labelled with the gene ID, location on the chromosome and the recombination frequency.

CHAPTER 4

CONCLUSIONS AND FUTURE DIRECTIONS

There are two major conclusions that can be drawn from these results. Firstly, both *tsh1* and *tsh3* mutants have distinct floral phenotypes that change lodicule morphology and second whorl organ number. The *tsh1* floral phenotype, although quite mild, has never been described before. The *tsh3* phenotype is more variable in the lodicule whorl, leading to greater number of lodicules or hybrid identity organs. This can potentially be explained by a repression of a boundary gene, similar to *HANABA TARANU* (*HAN*) in Arabidopsis [Ding et al., 2015]. Since lemma and lodicules are on different whorls, the repression of a boundary gene could lead to hybrid identity organs.

The second conclusion is the confirmation that the mutation of *tsh3* is located on the long arm of chromosome 6, between 148.1Mbp and 152.8Mbp. Unfortunately, I was not able to identify the underlying gene, but I was able to shorten the region by approximately 75% of the original distance to a region with 64 genes in it.

To find the locus of *tsh3*, the most important thing to do would be to continue the fine mapping, which requires 2 steps. First, increase the number of *tsh3* individuals. With only 80 individuals and a total of 160 chromosomes, the closer I got to the causative locus, the less instances of recombination I had. At the marker created at the gene Zm00001d038104 located at approximately 148Mbp, individuals tested previously showed recombination at marker umc1805 and/or marker GRMZM2G044150. Out of the 28

chromosomes tested, 5 showed recombination at that locus. This implies that *tsh3* is between Zm00001d038104 and umc1805, which is approximately 5Mbp.

Next generation sequencing was similar in that the library was made with only 18 individuals. I think this small number of individuals meant that the peak identified using next-gen sequencing was broad, with many genes in it. However, even with this small number of individuals, the next-generation sequencing results provided ample information on genetic variation, allowing for rapid marker design. The BSA technique that I used, coupled to fine mapping, has the potential to accelerate the gene discovery process in maize.

CHAPTER 5

TABLES

Primer Name	Primer Sequence (5'-3')	Enzyme	Location (maize v4)	Overlapping Gene	Recombination Frequency
UMC 1805 forward	AGTGCACCAGCTTTTAATCACCTC	N/A (SSR)	152846877	N/A	21/160
UMC 1805 reverse	TGTGACCTGTGTGGTCTGTGG	N/A (SSR)	152846950	N/A	21/160
UMC 1352 Forward	GTGACGAGATGGTGCAGAAAGAT	N/A (SSR)	136195104	Zm00001d037745	35/160
UMC1352 Reverse	CCTGGAGGTGGAAGGAGAGG	N/A (SSR)	136195232	Zm00001d037745	35/160
BsaWI_F	CTCGCCAGTGGTTCCACACGCTGAG	BsaWI	141534473	Zm00001d037901	22/150
BsaWI_R	CATTGAACAGGACCGCGTAGTAGTC	BsaWI	141535063	Zm00001d037901	22/160
GRMZM2G044150_F	CACACACATGCAGAGTTGACTTAAT	SacI	145148308	GRMZM2G044150	14/76
GRMZM2G04	TGCTCGATGGATTG	SacI	145149568	GRMZM2	14/76

4150_NR	GACAGT			G044150	
ZM00001D038104_P_2_F	CCGCACATGTACTTGTTATGGT	HphI	148045942	ZM00001D038104	5/28
ZM00001D038104_P_2_R	TGGGACACTCCATTATCCACG	HphI	148046303	ZM00001D038104	5/28

Table 1: List of primers used in fine mapping. Names of primers are listed as named in the lab database. All primer sequences are displayed from the 5'-3' direction for both forward and reverse primers. If enzymes were associated with the primers as restriction digests, they are listed as well. The locations along chromosome 6 are based on where the 5' end of the primer begins according to the version 4 genome. Finally, recombination frequency is listed based on the number of chromosomes.

Gene ID (V4)	Location start	Location end	Gene Name (based on orthologues)
Zm00001d038148	149403737	149406493	Val1
Zm00001d038175	150542749	150543931	Auxin Responsive Protein (AuxIAA)
Zm00001d038104	148044310	148054129	Glycoside hydrolase family 2
Zm00001d038207	151568742	151571658	NAC Transcription Factor

Table 2: List of genes Sanger sequenced after the next generation sequencing.

REFERENCES

Afgan, E., Baker, D., van den Beek, M., Blankenberg, D., Bouvier, D., Čech, M., Chilton, J., Clements, D., Coraor, N., Eberhard, C., et al. (2016). The Galaxy platform for accessible, reproducible and collaborative biomedical analyses: 2016 update. Nucleic Acids Res *44*, W3–W10.

Ambrose, B.A., Lerner, D.R., Ciceri, P., Padilla, C.M., Yanofsky, M.F., and Schmidt, R.J. (2000). Molecular and genetic analyses of the *silky1* gene reveal conservation in floral organ specification between eudicots and monocots. Molecular Cell *5*, 569–579.

Bartlett, M.E., Williams, S.K., Taylor, Z., DeBlasio, S., Goldshmidt, A., Hall, D.H., Schmidt, R.J., Jackson, D.P., and Whipple, C.J. (2015). The Maize *PI/GLO* Ortholog *Zmm16/sterile tassel silky ear1* interacts with the zygomorphy and sex determination pathways in flower development. The Plant Cell Online *27*, 3081–3098.

Blankenberg, D., Gordon, A., Von Kuster, G., Coraor, N., Taylor, J., Nekrutenko, A., and Galaxy Team (2010). Manipulation of FASTQ data with Galaxy. Bioinformatics *26*, 1783–1785.

Blankenberg, D., Von Kuster, G., Bouvier, E., Baker, D., Afgan, E., Stoler, N., Taylor, J., and Nekrutenko, A. (2014). Dissemination of scientific software with Galaxy ToolShed. Genome Biology *15*, 403.

Chia, J.-M., Song, C., Bradbury, P.J., Costich, D., de Leon, N., Doebley, J., Elshire, R.J., Gaut, B., Geller, L., Glaubitz, J.C., et al. (2012). Maize HapMap2 identifies extant variation from a genome in flux. Nat Genet *44*, 803–807.

Chory, J., Nagpal, P., and Peto, C.A. (1991). Phenotypic and genetic analysis of det2, a new mutant that affects light-regulated seedling development in *Arabidopsis*. The Plant Cell online *3*, 445–459.

Cingolani, P., Platts, A., Wang, L.L., Coon, M., Nguyen, T., Wang, L., Land, S.J., Lu, X., and Ruden, D.M. (2012a). A program for annotating and predicting the effects of single nucleotide polymorphisms, SnpEff. Fly *6*, 80–92.

Cingolani, P., Platts, A., Wang, L.L., Coon, M., Nguyen, T., Wang, L., Land, S.J., Lu, X., and Ruden, D.M. (2012b). A program for annotating and predicting the effects of single nucleotide polymorphisms, SnpEff: SNPs in the genome of *Drosophila melanogaster* strain w1118; iso-2; iso-3. Fly (Austin) *6*, 80–92.

Ding, L., Yan, S., Jiang, L., Zhao, W., Ning, K., Zhao, J., Liu, X., Zhang, J., Wang, Q., and Zhang, X. (2015). HANABA TARANU (HAN) bridges meristem and organ primordia boundaries through PINHEAD, JAGGED, BLADE-ON-PETIOLE2 and CYTOKININ OXIDASE 3 during flower development in *Arabidopsis*. PLOS Genetics *11*, e1005479.

Drews, G.N., Bowman, J.L., and Meyerowitz, E.M. (1991). Negative regulation of the *Arabidopsis* homeotic gene AGAMOUS by the APETALA2 product. Cell *65*, 991–1002.

Field, S., and Thompson, B. (2016). Analysis of the maize dicer-like1 mutant, fuzzy tassel, implicates microRNAs in anther maturation and dehiscence. PLOS ONE *11*, e0146534.

Gallavotti, A., and Whipple, C.J. (2015). Positional cloning in maize (*Zea mays* subsp. mays, Poaceae). Applications in Plant Sciences *3*, 1400092.

Hu, H., Dai, M., Yao, J., Xiao, B., Li, X., Zhang, Q., and Xiong, L. (2006). Overexpressing a NAM, ATAF, and CUC (NAC) transcription factor enhances drought resistance and salt tolerance in rice. PNAS *103*, 12987–12992.

Jack, T., Brockman, L.L., and Meyerowitz, E.M. (1992). The homeotic gene APETALA3 of Arabidopsis thaliana encodes a MADS box and is expressed in petals and stamens. Cell *68*, 683–697.

Kellogg EA. Flowering Plants. Monocots: Poaceae (Vol. 13). New York, USA: Springer; 2015.

Koboldt, D.C., Chen, K., Wylie, T., Larson, D.E., McLellan, M.D., Mardis, E.R., Weinstock, G.M., Wilson, R.K., and Ding, L. (2009). VarScan: variant detection in massively parallel sequencing of individual and pooled samples. Bioinformatics *25*, 2283–2285.

Koboldt, D.C., Zhang, Q., Larson, D.E., Shen, D., McLellan, M.D., Lin, L., Miller, C.A., Mardis, E.R., Ding, L., and Wilson, R.K. (2012). VarScan 2: Somatic mutation and copy number alteration discovery in cancer by exome sequencing. Genome Res. *22*, 568–576.

Konieczny, A., and Ausubel, F.M. (1993). A procedure for mapping *Arabidopsis* mutations using co-dominant ecotype-specific PCR-based markers. Plant J. *4*, 403–410.

Krizek, B.A., and Fletcher, J.C. (2005). Molecular mechanisms of flower development: an armchair guide. Nat. Rev. Genet. *6*, 688–698.

Krizek, B.A., and Meyerowitz, E.M. (1996). The *Arabidopsis* homeotic genes APETALA3 and PISTILLATA are sufficient to provide the B class organ identity function. Development *122*, 11–22.

Langmead, B., and Salzberg, S.L. (2012). Fast gapped-read alignment with Bowtie 2. Nat Meth *9*, 357–359.

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

Li, H. (2011). A statistical framework for SNP calling, mutation discovery, association mapping and population genetical parameter estimation from sequencing data. Bioinformatics *27*, 2987–2993.

Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., Marth, G., Abecasis, G., and Durbin, R. (2009). The sequence alignment/map format and SAMtools. Bioinformatics *25*, 2078–2079.

Lin, W., Shuai, B., and Springer, P.S. (2003). The *Arabidopsis* LATERAL ORGAN BOUNDARIES–domain gene ASYMMETRIC LEAVES2 functions in the repression of KNOX gene expression and in adaxial-abaxial patterning. The Plant Cell online *15*, 2241–2252.

Michelmore, R.W., Paran, I., and Kesseli, R.V. (1991). Identification of markers linked to disease-resistance genes by bulked segregant analysis: a rapid method to detect markers in specific genomic regions by using segregating populations. PNAS *88*, 9828–9832.

Miller, C.A., Qiao, Y., DiSera, T., D'Astous, B., and Marth, G.T. (2014). bam.iobio: a web-based, real-time, sequence alignment file inspector. Nat Meth *11*, 1189–1189.

Nakashima, K., Takasaki, H., Mizoi, J., Shinozaki, K., and Yamaguchi-Shinozaki, K. (2012). NAC transcription factors in plant abiotic stress responses. Biochimica et Biophysica Acta (BBA) - Gene Regulatory Mechanisms *1819*, 97–103.

Pautler, M., Eveland, A.L., LaRue, T., Yang, F., Weeks, R., Lunde, C., Je, B.I., Meeley, R., Komatsu, M., Vollbrecht, E., et al. (2015). *FASCIATED EAR4* encodes a bZIP transcription factor that regulates shoot meristem size in maize. The Plant Cell online *27*, 104–120.

Schlötterer, C., Tobler, R., Kofler, R., and Nolte, V. (2014). Sequencing pools of individuals — mining genome-wide polymorphism data without big funding. Nat Rev Genet *15*, 749–763.

Schnable, P.S., Ware, D., Fulton, R.S., Stein, J.C., Wei, F., Pasternak, S., Liang, C., Zhang, J., Fulton, L., Graves, T.A., et al. (2009). The B73 maize genome: complexity, diversity, and dynamics. Science *326*, 1112–1115.

Schrager-Lavelle, A., Klein, H., Fisher, A., and Bartlett, M. (2017). Grass flowers: An untapped resource for floral evo-devo. Journal of Sytematics Evolution n/a-n/a.

Sen, T.Z., Harper, L.C., Schaeffer, M.L., Andorf, C.M., Seigfried, T.E., Campbell, D.A., and Lawrence, C.J. (2010). Choosing a genome browser for a model organism database: surveying the maize community. Database (Oxford) *2010*, baq007.

Sharp, W.M., and Chisman, H.H. (1961). Flowering and fruiting in the white oaks. I. Staminate flowering through pollen dispersal. Ecology *42*, 365–372.

Shuai, B., Reynaga-Peña, C.G., and Springer, P.S. (2002). The lateral organ boundaries gene defines a novel, plant-specific gene family. Plant Physiology *129*, 747–761.

Specht, C.D., and Bartlett, M.E. (2009). Flower evolution: The origin and subsequent diversification of the angiosperm flower. Annual Review of Ecology, Evolution, and Systematics *40*, 217–243.

Tello-Ruiz, M.K., Stein, J., Wei, S., Preece, J., Olson, A., Naithani, S., Amarasinghe, V., Dharmawardhana, P., Jiao, Y., Mulvaney, J., et al. (2016). Gramene 2016: comparative plant genomics and pathway resources. Nucleic Acids Res. *44*, D1133-1140.

Thompson, B.E., Bartling, L., Whipple, C., Hall, D.H., Sakai, H., Schmidt, R., and Hake, S. (2009). bearded-ear encodes a MADS box transcription factor critical for maize floral development. Plant Cell *21*, 2578–2590.

Tiwari, S.B., Wang, X.-J., Hagen, G., and Guilfoyle, T.J. (2001). AUX/IAA proteins are active repressors, and their stability and activity are modulated by auxin. The Plant Cell online *13*, 2809–2822.

Weigel, D., and Meyerowitz, E.M. (1994). The ABCs of floral homeotic genes. Cell *78*, 203–209.

Whipple, C.J., Zanis, M.J., Kellogg, E.A., and Schmidt, R.J. (2007). Conservation of B class gene expression in the second whorl of a basal grass and outgroups links the origin of lodicules and petals. PNAS *104*, 1081–1086.

Whipple, C.J., Hall, D.H., DeBlasio, S., Taguchi-Shiobara, F., Schmidt, R.J., and Jackson, D.P. (2010). A conserved mechanism of bract suppression in the grass family. Plant Cell *22*, 565–578.

Whitford, R., Fernandez, A., Groodt, R.D., Ortega, E., and Hilson, P. (2008). Plant CLE peptides from two distinct functional classes synergistically induce division of vascular cells. PNAS *105*, 18625–18630.

Whitford, R., Fleury, D., Reif, J.C., Garcia, M., Okada, T., Korzun, V., and Langridge, P. (2013). Hybrid breeding in wheat: technologies to improve hybrid wheat seed production. J Exp Bot *64*, 5411–5428.

Xu, Z., Escamilla-Treviño, L., Zeng, L., Lalgondar, M., Bevan, D., Winkel, B., Mohamed, A., Cheng, C.-L., Shih, M.-C., Poulton, J., et al. (2004). Functional genomic analysis of *Arabidopsis thaliana* glycoside hydrolase family 1. Plant Mol Biol *55*, 343–367.