INCREASING AND MAPPING GENETIC DIVERSITY IN UPLAND COTTON, *GOSSYPIUM HIRSUTUM*, THROUGH EXOTIC RELATIVES, MUTAGENESIS, AND HYBRIDIZATION

by

ELLEN REGINA ENGLERT SKELTON

(Under the Direction of Andrew H. Paterson)

ABSTRACT

Due to a repeated number of bottlenecks, including domestication, upland cotton, *Gossypium hirsutum*, has become extremely genetically impoverished. Cotton is a very important global crop, so it is important that fiber quality traits in cotton are able to improved to continue to be competitive against synthetic fibers. This research uses a variety of methods, incorporating exotic relatives, mutagenesis, and hybridization, to increase the genetic variation present in cotton. Exotic relatives were incorporated into elite lines using a nested association mapping population. Previously, elite cotton lines were treated with a mutagenesis chemical, and this research aims to map the mutations that conferred a fiber quality benefit. The lines created by mutagenesis were also used in hybridization with lines developed by QTL stacking. These methods successfully improved a variety of fiber quality characteristics which could be extremely useful as economic resources.

INDEX WORDS: Cotton, exotic relatives, mutagenesis, genetic diversity, *Gossypium hirsutum*, *G. hirsutum*, hybridization, QTL mapping, GWAS, QTL stacking.

INCREASING AND MAPPING GENETIC DIVERSITY IN UPLAND COTTON, *GOSSYPIUM HIRSUTUM*, THROUGH EXOTIC RELATIVES, MUTAGENESIS, AND HYBRIDIZATION

by

ELLEN REGINA ENGLERT SKELTON

B.S., The University of North Carolina at Wilmington, 2017

A Thesis Submitted to the Graduate Faculty of The University of Georgia in Partial Fulfillment of the Requirements for the Degree

MASTER OF SCIENCE

ATHENS, GEORGIA

2021

© 2021

Ellen Regina Englert Skelton

All Rights Reserved

INCREASING AND MAPPING GENETIC DIVERSITY IN UPLAND COTTON, *GOSSYPIUM HIRSUTUM*, THROUGH EXOTIC RELATIVES, MUTAGENESIS, AND HYBRIDIZATION

by

ELLEN REGINA ENGLERT SKELTON

Major Professor: Andrew H. Paterson
Committee: Peng W. Chee
Brian M. Schwartz

Electronic Version Approved:

Ron Walcott
Vice Provost for Graduate Education and Dean of the Graduate School
The University of Georgia
December 2021

# DEDICATION

I would like to dedicate this work to anyone and everyone that assisted in its creation, especially the members of the Plant Genome Mapping Laboratory at the University of Georgia.

# ACKNOWLEDGEMENTS

I would like to acknowledge my advisor, Dr. Andrew Paterson, for his continued support and advice. I would like to acknowledge my committee members, Dr. Peng Chee and Dr. Brian Schwartz, for all of their comments, advice and support. I would like to acknowledge all members of the Plant Genome Mapping Laboratory at the University of Georgia for all of their assistance, help, and friendship. I would especially like to acknowledge Dr. Jinesh Patel and Dr. Sameer Khanal for developing the variety of cotton lines that were used in this research. I would like to acknowledge Susan Aukland for her help in creating crosses in the Summer of 2018 and selfing in the Summer of 2019. I would also like to acknowledge Lisa Rainville for helping with selfing in 2019. I would like to acknowledge Jeevan Adhikari for his help in DNA extraction and creation of GBS libraries. I would also like to acknowledge Min Liu for her help in the creation and analysis of GBS libraries. Finally, I would like to acknowledge my family and friends, especially my boyfriend, Mark Zenoble, for all of their love and support.

# TABLE OF CONTENTS

Page

ACKNOWLEDGEMENTS .......... v

LIST OF TABLES .......... viii

LIST OF FIGURES .......... ix

CHAPTER

1 LITERATURE REVIEW .......... 1

Global Significance of Cotton .......... 1

Cotton Biology .......... 5

The Cotton Genome .......... 7

Fiber Quality .......... 12

Quantitative Trait Loci Mapping and Genotyping by Sequencing .......... 14

Mutagenesis .......... 18

References .......... 19

2 Using Exotic *Gossypium hirsutum* Lines to Increase Genetic Variation in Elite *G. hirsutum* Varieties .......... 36

Introduction .......... 36

Materials and Methods .......... 38

Results .......... 41

Discussion .......... 49

References .......... 88

3 Mapping Mutations in *Gossypium hirsutum* Caused by Ethyl Methanesulfonate.......95

Introduction.......95

Mutant S Population Background.......98

Materials and Methods.......98

Results.......102

Discussion.......114

References.......127

4 Hybridization of Cotton Lines Developed by Mutagenesis and Cotton Lines Developed by QTL Stacking.......135

Introduction.......135

Germplasm Background.......139

Materials and Methods.......141

Results.......144

Discussion.......150

Conclusion.......154

References.......163

# LIST OF TABLES

Page

Table 2.1: Botanical races of *G. hirsutum* ....................56

Table 2.2: Population List ....................57

Table 2.3: Parental heterozygosity ....................78

Supplemental Table 2.4: Phenotypic effects of all SNPs and alleles ....................79

Table 3.1: The LOD from the two-dimensional QTL scan for fiber elongation ....................123

Table 3.2: A list of pairs of chromosomes found for fiber elongation in the two-dimensional QTL analysis ....................124

Table 3.3: List of all QTLs found ....................125

Table 4.1: Average values for all traits in the first set of crosses between the mutant lines and the QTL stacked lines ....................156

Table 4.2: Average values for all traits in the second set of crosses between the mutant lines and the QTL stacked lines ....................157

Table 4.3: Average values for traits in the first set of crosses between different mutant lines ...158

Table 4.4: Average values for traits in the second set of crosses of different mutant lines.........159

Table 4.5: Average values for traits in the third set of crosses of different mutant lines ...........160

Supplemental Table 4.6: List of all pairs of progeny and parents that were significantly different from each other, and parents that were significantly different from each other ..............161

# LIST OF FIGURES

Page

Figure 2.1: Correlation matrix of fiber quality traits in 2015 ....................58

Figure 2.2: Boxplot of fiber length distribution by population....................59

Figure 2.3: Boxplot of fiber fineness distribution by population ....................60

Figure 2.4: Boxplot of fiber uniformity distribution by population ....................61

Figure 2.5: Boxplot of fiber strength distribution by population....................62

Figure 2.6: Boxplot of fiber elongation distribution by population....................63

Figure 2.7: Boxplot of short fiber content distribution by population....................64

Figure 2.8: Boxplot of lint percent distribution by population ....................65

Figure 2.9: Boxplot of fiber weight distribution by population....................66

Figure 2.10: Boxplot of fiber seed weight distribution by population....................67

Figure 2.11: Correlation matrix of fiber quality traits from 2015 and 2016....................68

Figure 2.12: Manhattan plot of p-values by chromosome for fiber length (UHM)....................69

Figure 2.13: Manhattan plot of p-values by chromosome for fiber fineness (MIC)....................70

Figure 2.14: Manhattan plot of p-values by chromosome for fiber uniformity (UI)....................71

Figure 2.15: Manhattan plot of p-values by chromosome for fiber strength (STR)....................72

Figure 2.16: Manhattan plot of p-values by chromosome for fiber elongation (ELO)....................73

Figure 2.17: Manhattan plot of p-values by chromosome for short fiber content (SFC) ....................74

Figure 2.18: Manhattan plot of p-values by chromosome for lint percent (LP)....................75

Figure 2.19: Manhattan plot of p-values by chromosome for fiber weight (FW) ....................76

Figure 2.20: Manhattan plot of p-values by chromosome for seed weight (SW)..........................77

Figure 3.1: Correlation matrix of all fiber quality traits .............................................................119

Figure 3.2: : Boxplots showing the distribution of six fiber traits..............................................120

Figure 3.3: Boxplots showing the distribution fiber quality traits across four environments......121

Figure 3.4: Heatmap of pairwise recombination fractions and LOD scores................................122

# CHAPTER 1

# LITERATURE REVIEW

## Global Significance of Cotton

Cotton has been revered as "white gold" for centuries due to its high economic value around the world. The main economic product is the long seed fibers, which are sturdy but soft and have a wide range of uses. While cotton is mainly grown to spin the fiber into yarn and fabric for the textile industry, of which it provides approximately 50% of the raw materials, other uses include hygiene products, medical and pharmaceutical products, recreational products such as tents and ropes, furniture and more, as well as using the seed and oil in animal feed.

There are 53 species recognized in the Gossypium genus, 45 diploids and seven tetraploids (Mansoor and Paterson, 2012). They are native to Africa, Central and South America, Asia, Australia, the Galapagos Islands and Hawaii, but are now grown in the tropical and subtropical regions of more than 80 different countries. Of the 54 species in the Gossypium genus, four are cultivated: two diploid species, Gossypium herbaceum and Gossypium arboreum, and two tetraploid species, Gossypium hirsutum and Gossypium barbadense. All four of these species were independently domesticated in different geographic regions by different prehistoric cultures (Wendel and Cronn, 2003). About 95% of globally cultivated cotton today is G. hirsutum, 3% is G. barbadense and the remaining is G. herbaceum and G. arboreum. This is because G. hirsutum has superior fiber yield and moderate fiber quality and G. barbadense has very high fiber quality, but lower yield.

In the 2018/2019 growing season, cotton was grown in approximately 30 countries around the world which resulted in 116 million bales produced, where one bale is about 218kg and in the 2019/2020 growing season approximately 125.3 million bales were produced (Zhang et al., 2019c). This results in around 30 million hectares of land dedicated to cotton production and 150 countries involved in import and export (Chen et al., 2007; Ulloa et al., 2020). This has caused cotton to be a major economic driver for many developing countries around the world, especially in Asia where about 60% of world cotton is produced (Shaban et al., 2018). For example, in Pakistan, cotton accounts for half of the total dollar amount of foreign export (Abbas et al., 2015). In India, there are 10 million farmers cultivating cotton and around 30 million people employed by the industry (Ramesh et al., 2019). They also have the largest cultivation area in the world, 12.3 million hectares, the largest production, 36.5 million bales, and is the 3rd largest exporter in the world (after the US and China) (Ramesh et al., 2019). Cotton is an extremely important and valuable cash crop around the world.

Cotton is a very important part of the economy of the United States as well. While the US is the third largest producer of cotton, it is the largest exporter of cotton in the world. In 2016, approximately 10 million bales were exported and in 2018, approximately 18.5 million bales were produced with 15 million bales exported (Zhang et al., 2019c). Cotton is a multi-billion dollar industry for the US alone. The industry especially benefits the southern and south-eastern states which are part of the "cotton belt," particularly Texas and Georgia, which are the top two producers in the country. In 2017, the entire cotton supply chain contributed $74 billion to the gross domestic product and directly generated over 1.3 million jobs (Lopes et al., 2020). Alone, the direct production, distribution and processing of cotton provides about $27 billion in direct business revenue and supports 200,000 jobs (Brown et al., 2019). This revenue provides

necessary income to the rural areas along the Cotton Belt of the United States. The US has continued to be competitive in the international market because of the higher fiber quality of the cotton produced (Zhang et al., 2019b). Therefore, it is extremely important that US breeders continue to improve cotton fiber quality.

Early in cotton breeding history, long, fine fiber, as well as yield was preferred with less emphasis on other fiber quality traits such as strength or elongation (Perkins, 1984). However, in recent history, more efficient technologies have been developed for spinning the cotton into yarn. These new machines and processes require a higher overall fiber quality which has increased the selection on previously less important fiber quality characteristics. In addition to new spinning technologies, the synthetic fiber market has also caused a push for higher fiber quality. This is because the synthetic market currently has acceptable quality with lower prices whereas the cost of cotton production has continued to increase (Zhang et al., 2019c). It is only recently that breeders have focused on improving quality with the advent of synthetic fibers and newer spinning technologies. Previously, breeders focused on improving yield while only maintaining an acceptable level of quality, but this is no longer working. Due to the limited genetic diversity, rapid genetic improvement of the crop does not seem feasible, so the focus has shifted to increasing fiber quality. However, the previous narrow focus on quality has likely contributed to a stagnation of quality improvement. This can be shown through the fact that yield has improved by 70% from 1975 to 2015 while the average upper half mean length (UHML) has only increased by 5.6% in the same time frame (Brown et al., 2019). This again suggests the narrow genetic diversity of G. hirsutum germplasm, especially in commercial germplasm (Brown et al., 2019).

While cotton fiber remains the more important part of the crop, it is important to note that cottonseed oil has been increasing in importance and popularity in recent years. Originally the seed was only used for replanting, fertilizer, and some animal feed, but most was thrown away. However, cottonseed in now considered to be the second most important potential source of plant protein and the fifth most important source of oil (Saravanan, 2010). For every 1kg of fiber that is produced, there is 1.6kg of cottonseed produced which can then be used to make oil and feed (Saha et al., 2020). Cottonseed is an important protein source for animal feed because is promotes growth, weight gain, and positive balance in nitrogen (Alford, 1996). The cottonseed, specifically the kernel, is compressed to extract the oil and to generate a cake, which is then exclusively used as feed for ruminants (Saha et al., 2020). Because cottonseed provides a high energy value to livestock, it is a more competitive option than other feeds, especially for lactating cows (Saha et al., 2020). In addition to being used as animal feed, cottonseed oil is regularly used as an ingredient in a variety of human foods such as potato chips and mayonnaise. Cotton is used as one of the most unsaturated edible oil in the US, partially due to its health benefits such as lowering cholesterol and losing weight (Saha et al., 2020). Cottonseed consists of 70% unsaturated fatty acids, 52% of which are polyunsaturated and is has a very well-balanced amino acid profile (Palle et al., 2013; Saha et al., 2020). In addition to providing nutrition, the byproducts of processing cottonseed have the potential to be raw materials for the production of biofuels (Bi et al., 2016). Because of these benefits, demands for cottonseed oil have increased from 35.76 million metric tons in 2015-2016 to 44.84 million metric tons in 2019-2020 around the world (Saha et al., 2020).

**Cotton Biology**

Cotton fibers are single-celled modified trichomes. Trichomes are uni- or multicellular epidermal structures that can be found on leaves and stems. However, cotton fibers are seed-borne epidermal trichomes that are unicellular and thickened (Ijaz et al., 2019). These seed trichomes can take two forms: there are the long lint fibers which easily detach from the seeds, and the fuzz fibers which strongly adhere to the seeds (Applequist et al., 2001). Cotton also has trichomes on their leaves and stems which are associated with environmental advantages such as drought tolerance (Espigares and Peco, 1995).

There are four stages of fiber development: initiation, elongation, secondary cell wall deposition and maturation, which occur in that order but usually overlap. Fiber cell development from flowering (anthesis) to harvest usually lasts around 60 days (Fang et al., 2020). The initiation stage begins a few days before flowers open. Elongation follows initiation and generally lasts about 25 days but depends on the environment (Haigler et al., 2012). Secondary cell wall deposition occurs 15-40 days post anthesis (DPA) and maturation occurs from 40-50 DPA (Niu et al., 2019). The transition from elongation to secondary cell wall deposition is accompanied with significant changes in physiological processes and cell wall contents (Zhou et al., 2019). Both phases play a role in determining fiber length and strength (Shi et al., 2006). At anthesis, approximately 15-25% of the seed epidermal layer cells start differentiating into lint fibers (Basra and Malik, 1984). Because of this, fiber yield depends on the initiation stage, as it determines the number of fibers present on each ovule (John and Keller, 1996). Zhang et al. (2011) found that increasing the amount of auxin present increases the number of cells that differentiate during the initiation phase, which in turn increases yield. During the elongation phase, each seed epidermal cell elongates until it becomes a fiber, thus the length of the

elongation phase is associated with fiber length (Deng et al., 2012; Shi et al., 2006). It has also been shown that the domestication of cotton has been associated with a longer elongation period which results in longer fibers (Applequist et al., 2001). Both pectin and ethylene are compounds that have been shown to be necessary for fiber elongation (Pang et al., 2010; Shi et al., 2006). In addition, calcium signaling may be important in regulating all fiber developmental processes (Ding et al.).

Since these developmental phases are so important to the economic qualities in cotton, they have been extensively studied. With recent technological developments, figuring out the genetic pathways of fiber development has become very important to breeding higher quality cotton. One of the major findings has been that the genes found to play a role in fiber development have orthologs in other plants (Haigler et al., 2005). This could be useful because since the cotton genome is tetraploid, it is extremely complex and hard to study, but we could potentially gain a greater understanding by studying these orthologs in plants that are easier to work with. One of the discoveries that has come out of this is that the fiber developmental regulatory shares a similar regulatory pathway to Arabidopsis trichomes (Ding et al., 2015). This has led to closer studies on the relationship between cotton fibers and cotton trichomes. Patel et al. (2016) found that the genes that control cotton stem trichome development are closely associated with fiber development and therefore could be used as a phenotypic marker. In addition, it was found that mutations resulting in trichome variation also often resulted in significant changes of at least one fiber quality trait (Patel et al., 2016). Another finding has shown that many transcription factor genes are differentially expressed during fiber initiation and the early developmental stages, suggesting that these are important controls in the process (Ijaz et al., 2019). While there are major genetic controls for the growth and development of cotton

and cotton fibers, it is important to note that environmental factors are also important. Since cotton is native to tropical habitats, low temperatures, in addition to high salinity, are major environmental factors that can affect early cotton growth, such as the germination and seedling stages (Sun et al., 2019).

**The Cotton Genome**

Of the 53 recognized species in the Gossypium genus, 45 are diploid (2n=26) and seven are tetraploid (4x=52). The Gossypium genus originated from paleo-hexaploidy of a eudicot progenitor and subsequent diversification into eight diploid genome groups, A-G and K (Paterson et. al 2012). The tetraploids, which are allotetraploids, are the result of hybridization and duplication of an A genome species which resembles the modern G. herbaceum, and a D genome species, which resembles the modern G. raimondii. This hybridization occurred approximately 1-2 million years ago. This event resulted in a genome size of approximately 2.5Gb with about 5,500 cM of recombination (Reinisch et al., 1994). While the two genomes combined after hybridizing, it is interesting to note that the cytoplasm of the AD genome species is only from the A parent (Wendel, 1989). It has also been found that, contrary to other tetraploid species, tetraploidy in cotton was not accompanied by extensive restructuring of the genome (Brubaker et al., 1999). In fact, the tetraploid A genome appears to differ from the diploid A genome primarily by only two reciprocal translocations (Endrizzi, 1985). However, polyploidization is often thought to “relax” selection on individual genes, due to having up to four copies of each allele (Wendel and Cronn, 2003). This allows for opportunities for compartmentalization of genes and for novel variation and genes, which has been seen associated with polyploidization in cotton (Chee et al., 2005a). In addition, polyploidization appears to have increased the recombination rate of cotton (Desai et al., 2006). There have also been

numerous changes in gene expression between the diploids and tetraploids recorded (Al-Ghazi et al., 2009; Claverie et al., 2012; Rapp et al., 2010; Xie et al., 2011). Working with and assembling alloypolyploid genomes like cotton can be difficult because of their large size and highly homologous subgenomes (Li et al., 2015). There has been a variety of studies to figure out differences between the A and D homologous genomes to figure out how each impacts fiber yield and quality.

There are two main differences between the A and D genomes, one being the size. The A subgenome is approximately twice the size of the D subgenome, however the D subgenome has a slightly higher recombination rate than A (Desai et al., 2006). The second prominent difference is that, while the A diploid species are able to produce spinnable fibers the diploid D species are not. However, it appears that the tetraploid D subgenome (Dt) still makes an important contribution to fiber production in the tetraploid species (Rong et al., 2007). It appears that the transcription factors that control fiber development are located on the Dt subgenome while the actual genes related to fiber development are on the At subgenome (Xu et al., 2010). This could be the result of polyploidization causing a compartmentalization of gene function or causing the fiber genes in the D genome to become derepressed (Mei et al., 2004; Xu et al., 2010). Even though the two subgenomes contain a lot of synteny between them (approximately 80% of gene order is conserved), both tetraploid subgenomes have experienced more rapid evolution than their diploid ancestors and may have experienced differential selection on homologous genes (Rong et al., 2007; Yang et al., 2019). It is hypothesized that more structural changes have occurred on the At (A-tetraploid) genome in the last 1-2 million years that over the previous 5-7 million years since the A and D diploid genomes diverged from each other (Desai et al., 2006). Some of this rapid evolution may be due to the fact that tetraploid can better handle the effects of

deleterious mutations or translocations since there are multiple copies of each gene to retain necessary functions (Desai et al., 2006). Some other differences between the subgenomes are that the Dt subgenome has higher frequency of genetic loss, a higher mutation rate and more active transposable elements (Li et al., 2015). In addition, the Dt genome seems to have harbored more nucleotide and allelic diversity than the At genome and carries more environmental resistance associated markers (Guo et al., 2007; Sun et al., 2019). While polyploidization played a large part in these differences, something else that contributed was human selection on plants during the domestication process.

As previously mentioned, each of the currently cultivated species of cotton were independently domesticated by separate cultures. The most commonly cultivated species, Gossypium hirsutum, was domesticated approximately 5000 years ago in the Yucatan peninsula (Brubaker et al., 1999). Domestication is associated with drastic changes to a crop species, often causing the crops we grow today to look almost nothing like their wild ancestors. Domestication in cotton is associated with longer, finer, and whiter fibers as well as a higher yield of individuals. Comparing cultivated varieties to wild relatives to understand the genetic changes associated with the domestication process is an important part of figuring out how fiber yield and quality are controlled. There have been numerous gene expression studies to find differences between wild and domestic cotton. Ji et al. (2003) and Rapp et al. (2010) both found that the most highly differentially expressed genes were profilin genes which help to regulate actin which is needed for growth and development. This gene family was studied in three independently domesticated species and it was found to be upregulated in all domestic cotton compared to its wild ancestors (Bao et al., 2011). Domestication has also appeared to increase the amount of coordination of expression among genes of related phenotype and caused dramatic alterations to

the fiber development transcriptome (Grover et al., 2020). Through all these changes from humans selecting for higher yield and quality, it is highly likely that cultivated cotton has lost many beneficial traits especially those related to resistance to abiotic or biotic factors (Feng et al., 2019). Not only have cultivated species lost specific beneficial traits, but the domestication process has also caused an extreme bottleneck which has resulted in an extremely decreased level of genetic diversity in cultivated cotton.

In addition to domestication, Upland cotton (G. hirsutum) has gone through a series of bottlenecks, such as polyploidization and migration, which has left it genetically impoverished. In addition to these bottlenecks, breeding and overexploitation of only a few genetic backgrounds have led to a plateau of yield and quality traits and extremely reduced the natural genetic diversity (Adhikari et al., 2017; Fang et al., 2013). These major contributions by a single or very few germplasms can lead to genetic uniformity and leave the crop vulnerable to several environmental stressors (Bowman et al., 1996). Numerous studies have confirmed that these factors have led to low genetic diversity. Reinisch et al. (1994) found that the gene pools of all cultivated species show only modest levels of DNA polymorphisms. (Lubbers, 2009) found that cotton has an extremely low heterozygosity value, 0.07, compared to other crop species. Abbas et al. (2015) found that levels of genetic similarity among genotypes ranges from 81.1% to 98.7% and used SSR (short sequence repeat) markers to confirm this limited diversity. This low genetic diversity is one of the major obstacles to cotton improvement (Paterson et al., 2004). However, many studies have shown that cultivated cotton does not make use of the total genetic variability that exists. Esbroeck (1998) estimated that potentially less than 1% of all genetic variability of cotton resides in the US commercial cultivars. This means that only a fraction of the total genetic base has been exploited (Bowman et al., 2003). The cotton genome remains

diverse, but the most useful parts of the gene are either unknown or not currently accessible in commercial cultivars (de Menezes et al., 2015). There are two potentially large sources to bring more diversity to the cotton genome. The first is synthetically via technologies such as the CRISPR/Cas9 complex and other genome editing, or from other organisms completely (Mansoor and Paterson, 2012). The second is to use the diversity that still exists in wild cotton relatives, both diploid and tetraploid.

Wild relatives of cotton, especially allotetraploids are rich sources of diversity for the improvement of commercial cotton. For example, G. tomentosum is a potential source to help enhance length and strength in domesticated tetraploid cotton (Mansoor and Paterson, 2012; Zhang et al., 2011). Wild G. hirsutum relatives have unexploited variation which can be used for introgressive breeding (Chandnani et al., 2017b). Truly wild forms of G. hirsutum (race yucatanense) are found as scattered populations in the coastal regions of tropical and subtropical zones of the Caribbean, northern South America, and Mesoamerica (d'Eeckenbrugge and Lacape, 2014). These wild races can be used to create race stocks. Race stocks are a primitive type of cultivated cotton that have undergone minimal selection by humans, but have formed under long-term natural selection from a variety of geographical conditions (Feng et al., 2019). These stocks may contain beneficial QTLs that were eliminated by heavy human selection on commercial cultivars (Feng et al., 2019). However, there are potentially some issues with this approach. For example, some regions of the cotton genome appear to be extremely resistance to introgression from wild relatives and the basis for this is not yet understood (Chandnani et al., 2017a). This would mean that if a specific beneficial allele is in this region of the genome, it would not introgress and therefore be inaccessible by standard breeding methods. Some other challenges of introgressing exotic alleles into elite cultivars are: the likely effect of linkage drag between

unwanted and beneficial alleles; the high level of epistasis that occurs in cotton; and the high likelihood of hybrid breakdown the would occur (Paterson et al., 2004). Despite these potential problems, introgressing beneficial alleles from wild relatives is still a method to be considered. There is so much variation in wild genotypes that would provide the much needed diversity to elite cultivars, and there is so much else we may be able to learn from these exotic types as well.

**Fiber Quality**

Fiber quality is one of the most important economic traits to measure in commercial cotton. It directly impacts how much money the cotton can be sold for and how competitive it is in the market, especially with the introduction of synthetic fibers. Fiber quality is defined as "the physical properties that relate to its spinnability into yarn, contribute to textile performance, and enhance end-product quality," (Chee et al., 2005b). It is measured by a few metrics: Length, strength, elongation, fineness, uniformity, and short fiber content (SFC). While fiber quality has not always been a top priority, some aspects have been previously selected for. These metrics include length, uniformity, and SFC. Longer fibers are easier to spin into yarn, while low uniformity and high SFC are associated with increased waste during manufacturing and decreased spinning efficiency (Chee et al., 2005b). Currently, with the advent of newer and more efficient manufacturing technologies, the primary properties for evaluating fiber quality are length, strength and fineness. Longer fibers can be processed at higher efficiency as shorter fibers require more twisting to spin into yarn (Chee et al., 2005b). Fiber usually ranges from 2.2-3.0 cm but can get up to 6.0 cm in G. barbadense (Chee et al., 2005b). Stronger fibers break less during rough spinning process and therefore produce less waste (Bradow, 2000). Finer mature fibers are able to be spun into yarns that have more fibers per cross-section, which results in stronger yarns and greater efficiency in the spinning process (Steadman, 1997). Fiber fineness is

a measurement of the diameter of an individual fiber. It is most commonly measured by micronaire (MIC), which estimates that amount of air resistance of a constant weight of fibers (Steadman, 1997). On average, the diameter of a single fiber measures around 11-22µm (Chee et al., 2005b). Fiber elongation is a trait that has only recently been under selective pressure from breeders (Zhang et al., 2011). According to Chee et al. (2005a), fiber elongation is determined by gradually stretching a bundle of combed fiber between two clamps and the distance traveled between them when the fibers break is measured and expressed as a percent. All of these metrics are important to the textile industry and producing good quality yarn and it is therefore important that breeders understand the mechanisms behind each of these traits.

Many researchers and breeders have done numerous studies to figure out the genetic controls that could lead to improved fiber quality. It has been shown that in general fiber quality traits tend to be highly heritable, which is useful for breeding (Fang et al., 2013). In addition, it appears that the genetic background of an individual appears to have a very strong influence on fiber quality, even more so than environmental effects (Shi et al., 2019). However, the environment can still play a large role in determining both fiber yield and quality (Si et al., 2017). While there have been minor improvements in fiber quality metrics, there appear to be signs of stagnation which may be due to a lack of genetic diversity (Esbroeck, 1998). What makes improvement harder is that fiber quality seems to be controlled largely by additive gene action, which means that numerous individual traits play a role (May, 1999). For example, length seems to be controlled by many minor additive genes, which makes it harder to find and select for ones that would make a lot of improvement to quality (Chee et al., 2005b). Fineness and elongation appear to have not quite reached the same plateau that length has, and this may be because they both have had a shorter history of selection (Chandnani et al., 2017a; Zhang et al.,

2011). Another challenge in improvement quality is that fiber yield and quality have been shown repeatedly to have a negative correlation (Culp and Harrell, 1973; Scholl and Miller, 1976; Zhou et al., 2020). It has been repeatedly shown that that there is a strong, consistent, negative association between fiber yield and strength, which could indicate the existence of linkage or pleiotropy (Scholl and Miller, 1976). In addition, there are negative associations between important fiber quality traits. Fineness has been shown to be negatively correlated with length and strength (Ulloa et al., 2020). These correlations make it harder to improve both yield and quality simultaneously with traditional breeding methods.

**Quantitative Trait Loci Mapping and Genotyping by Sequencing**

Fiber quality and yield characteristics are not simple mendelian traits. They are traits that are controlled by numerous genes throughout the genome and are mearsured on a continuous scale. These are referred to as quantitative traits and quantitative trait loci (QTLs) are regions of the genome that are associated with a particular phenotype. The regions that contribute to quantitative traits can be found via QTL mapping. QTL mapping is done by creating a population from at least two parents with distinct phenotypes for a trait of interest. Then, the genetic markers, such as single nucleotide polymorphisms (SNPs) or short sequence repeat (SSR) markers, that differ between the parents are found and statistical analysis is done on the whole population to see if any markers segregate with the trait of interest (Zeng, 1994). The markers which segregate are considered to be linked with a QTL for that phenotype (Zeng, 1994). These markers can then be used for marker assisted selection to introgress beneficial QTLs into existing lines for improvement (Wang et al., 2020). QTLs can have large or small effects on any given phenotype and the smaller the effect a QTL has, the harder it is to find. However, when large effect QTLs become fixed in a population, it becomes easier to detect

small effect QTL (Paterson et al., 1991). The success of QTL mapping directly depends on the number and usefulness of the markers used. The best marker would allow a breeder to maximize recombination between desirable and undesirable genes without losing the linkage between the desired trait and the marker (Paterson et al., 2004).

A more recent method of finding numerous markers for QTL mapping is through genotyping by sequencing (GBS) and next generation sequencing (NGS). Several types of marker systems have been developed, however, they tend to be constrained either by limited availability or high cost when needed for large scale analysis (Sonah et al., 2013). GBS uses pooled, barcoded samples to create libraries where multiple samples can be sequenced at the same time using NGS technologies such as Next-Seq or Mi-Seq. Through this method, thousands of SNPs can be detected and used as markers. Being able to sequence numerous individuals at once with drastically reduces the cost required to find a large number of SNP markers. This large marker density is very beneficial in QTL mapping. SNP markers have been shown to be universal since they are the most abundant form of polymorphisms among individuals within a species (Ramesh et al., 2019).

While fiber quality and yield are likely the most important qualitative traits in cotton, other important traits include disease resistance, pest resistance, environmental resistance, seed oil and protein content and leaf morphology (Said et al., 2013). So far, more than 1500 Fiber quality QTLs have been mapped in cotton (Said et al., 2015). However, their use has been limited. This is because genetic background appears to have profound effects on fiber quality QTL (Draye et al., 2005). This means that depending on the line or genotype that already exists in a population may cause a QTL to have less of an effect or no effect compared to the population in which it was found and deemed desirable. Utilizing the QTLs that have been

discovered can also be difficult due to the fact that many QTLs are effected by the environment, although there does appear to be a basal set of QTLs which seem to be relatively unaffected by the environment (Shen et al., 2007). Another issue is that, in many recent studies, the majority of QTLs identified have already been recorded and identified in previous studies (Zhang et al., 2019b). This suggested the close relatedness and the general narrow genetic base of elite US cottons (Zhang et al., 2019a).

QTL mapping is cotton has primarily focused on fiber quality traits as opposed to yield. This may be because of the recent pressures for increased fiber quality, as well as the fact that quality characteristics generally have higher heritabilities than yield (Shen et al., 2019). QTL mapping in cotton can also be difficult due to allopolyploidy and therefore the two highly homologous genomes. Therefore, in addition to improving the economic qualities, another goal of QTL mapping in cotton is to figure out difference in the At versus Dt genomes and how they each contribute to fiber quality. While there have been many QTLs found in each genome, the contributions of each are debated. Most studies show that there are overall more QTLs located in the Dt subgenome (Chandnani et al., 2018; Jiang et al., 1998; Rong et al., 2007; Ulloa et al., 2020) but other studies have shown no difference between the At and Dt subgenomes (Chee et al., 2005b; Grover et al., 2020) or more QTLs on the A subgenome (Ijaz et al., 2019; Mei et al., 2004). Wang et al. (2020) found that while the number of yield related QTLs are evenly distributed between the subgenomes, Dt had more quality related QTLs, which is interesting considering the D ancestor does not produce spinnable fiber. Other studies have categorized the QTLs further into the specific quality traits. Shi et al. (2019) found that more length and strength QTLs were on the At subgenome and there were more fineness and yield QTLs on the Dt subgenome. There is also disagreement among these findings, as Grover et al. (2020) found that

many of the QTLs in the D subgenome are associated with fiber length. The many contradictory results found in the literature illustrate the problems with finding stable, useful QTLs that work in many genetic backgrounds and environments. These discrepancies shown here is likely due to the variety of genetic backgrounds present across these experiments, which have been shown to greatly effects QTLs. In addition, since there are many quality and yield QTLs detected in both subgenomes, it suggests that fiber related traits may result from gene expression and interaction between the subgenomes (Wendel, 2000). In addition, QTLs in cotton tend to form clusters, sections of the genome where multiple QTLs (at least three), possibly for multiple traits are groups very closely together (Rong et al., 2007). These are especially common for QTLs associated with fiber development, which directly impact fiber quality and yield (Ulloa et al., 2020). Clusters have been helpful in figuring out stable QTLs, as there are many that are common between cultivars as well as different species of cotton (Said et al., 2015). However, one of the challenges with these clusters is that they also tend to harbor undesirable or negative QTLs within a number of beneficial QTLs, which makes it harder to improve all traits simultaneously (Shi et al., 2019).

The presence of negative QTLs in clusters is just one of the many challenges in QTL mapping in cotton. Additionally, interspecific crosses, which are often used in cotton QTL mapping, often generate negative correlations between quality and yield through linkage drag and hybrid breakdown (Zhang et al., 2014). Linkage drag also often occurs when breeding for increased stress resistance which is usually accompanied by lower yield and quality as well (Lopes et al., 2020). While linkage drag is a problem in cotton breeding, the major issue is the inconsistencies of QTLs reported. QTLs found with certain crosses or in certain populations often do not agree with results found in other populations (Grover et al., 2020). In addition,

interspecific crosses have been a common method for QTL discovery in cotton, however they are not the most useful since there appears to be a lack of stability among different populations as well as a localization of large genetic regions (Naoumkina et al., 2020). Also, it appears that the effects of certain introgressed segments may be controlled or dependent on the presence or absence of other chromosomal segments (Chee et al., 2005a). This means that even if the beneficial trait was successfully introgressed into an elite line, it might not have the same effect without the same control from a different part of the genome. All of these issues make the improvement of cotton more complicated.

**Mutagenesis**

As previously discussed, the low genetic variability of commercial cotton cultivars is a pressing concern among breeders. One way to potentially increase variation and allelic diversity is through mutagenesis. Mutagenesis is intentionally inducing mutations, usually point mutations, into a genome to see the effects that it might have. One way to accomplish this is through chemicals such as Ethyl methanesulfonate (EMS). Chemical mutagens are able to confer stable single nucleotide changes as opposed to deletions that have negative or lethal results (Parry et al., 2009). EMS is frequently favored as it is more efficient that irradiation and has a higher survival rate (Favret, 1960). EMS causes point mutations, which may generate a variety of novel alleles with altered function that will likely not be lethal to the plant (Talebi, 2012). These point mutations occur by EMS causing guanine to bond with thymine instead of cytosine. The cell will then try to repair these mismatches, but replace the guanine with an adenosine, resulting in an overall change from a G/C pair to an A/T pair (Greene et al., 2003). In addition, EMS is very easy to use, and its high mutation rate allows for smaller sample sizes to still be saturated with mutations (Talebi, 2012). Mutations caused by EMS can also be analyzed

using either forward or reverse genetics, as opposed to other methods, such as targeted genome editing, that require more prior knowledge about the gene of interest (Espina et al., 2018). EMS has been shown to cause beneficial mutations in peas, soybean, barley, tomatoes, wheat, Arabidopsis thaliana, and others (Blixt et al., 1966; Espina et al., 2018; Gichner and Ehrenberg, 1966; Hildering and Vanderve.Jh, 1966; Rao and Sears, 1962; Vanderve.Jh, 1967). There have also been a number of studies published where EMS mutagenesis led to the release of improved cotton lines (Auld et al., 2007; Bechere et al., 2007; Bechere et al., 2011; Brown et al., 2012; Herring et al., 2004). These results show that increasing diversity in the cotton genome through mutagenesis is successful and has led to an increase in quality of a number of lines. However, fewer studies have gone past releasing an improved line to figure out the genetic mechanism that was affected by the chemical mutagenesis.

**References**

Abbas A., Iqbal M.A., Rahman M.U., Paterson A.H. (2015) Estimating genetic diversity among selected cotton genotypes and the identification of DNA markers associated with resistance to cotton leaf curl disease. Turkish Journal of Botany 39:1033-1041. DOI: 10.3906/bot-1505-22.

Adhikari J., Das S., Wang Z.N., Khanal S., Chandnani R., Patel J.D., Goff V.H., Auckland S., Rainville L.K., Jones D., Paterson A.H. (2017) Targeted identification of association between cotton fiber quality traits and microsatellite markers. Euphytica 213. DOI: 10.1007/s10681-017-1853-0.

Al-Ghazi Y., Bourot S., Arioli T., Dennis E.S., Llewellyn D.J. (2009) Transcript Profiling During Fiber Development Identifies Pathways in Secondary Metabolism and Cell Wall

Structure That May Contribute to Cotton Fiber Quality. Plant and Cell Physiology 50:1364-1381. DOI: 10.1093/pcp/pcp084.

Alford B.B., Liepa, G. U., Vanbeber, A. D. (1996) Cottonseed protein: What does the future hold? Plant Food and Human Nutrition. DOI: https://doi.org/10.1007/BF01092517.

Applequist W.L., Cronn R., Wendel J.F. (2001) Comparative development of fiber in wild and cultivated cotton. Evolution & Development 3:3-17. DOI: 10.1046/j.1525-142x.2001.00079.x.

Auld D.L., Bechere E., Krifa M., Kebede H., Hequet E., Wright R., Misra S. (2007) Registration of 'Raider 276', a High-Yielding, Improved-Quality Upland Mutant Cotton Cultivar. Journal of Plant Registrations 1:115-116. DOI: 10.3198/jpr2007.01.0059crc.

Bao Y., Hu G.J., Flagel L.E., Salmon A., Bezanilla M., Paterson A.H., Wang Z.N., Wendel J.F. (2011) Parallel up-regulation of the profilin gene family following independent domestication of diploid and allopolyploid cotton (Gossypium). Proceedings of the National Academy of Sciences of the United States of America 108:21152-21157. DOI: 10.1073/pnas.1115926109.

Basra A.S., Malik C.P. (1984) DEVELOPMENT OF THE COTTON FIBER. International Review of Cytology-a Survey of Cell Biology 89:65-113. DOI: 10.1016/s0074-7696(08)61300-5.

Bechere E., Auld D.L., Cantrell R.G., Hequet E., Krifa M., Misra S., Smith C.W. (2007) Registration of TTU 0774-3-3 and TTU 0808-1-6-1 Upland Cotton Germplasm Lines with Improved Fiber Length and Strength. Journal of Plant Registrations 1:58-59. DOI: 10.3198/jpr2006.09.0587crg.

Bechere E., Auld D.L., Krifa M., Smith C.W., Cantrell R.G. (2011) Registration of TTU 0782, an Upland Cotton Germplasm Line with Superior Fiber Quality. Journal of Plant Registrations 5:207-210. DOI: 10.3198/jpr2010.07.0404crg.

Bi C.W., Paterson A.H., Wang X.L., Xu Y.Q., Wu D.Y., Qu Y.S., Jiang A.N., Ye Q.L., Ye N. (2016) Analysis of the Complete Mitochondrial Genome Sequence of the Diploid Cotton Gossypium raimondii by Comparative Genomics Approaches. Biomed Research International. DOI: 10.1155/2016/5040598.

Blixt S., Ehrenberg L., Gelin O. (1966) STUDIES OF INDUCED MUTATIONS IN PEAS .18. EMS-TREATMENT OF DIFFERENT DURATION OF VARIETIES WEITOR AND WITHAM WONDER. Agri Hortique Genetica 24:138-+.

Bowman D.T., May O.L., Calhoun D.S. (1996) Genetic base of upland cotton cultivars released between 1970 and 1990. Crop Science 36:577-581. DOI: 10.2135/cropsci1996.0011183X003600030008x.

Bowman D.T., May O.L., Creech J.B. (2003) Genetic uniformity of the US upland cotton crop since the introduction of transgenic cottons. Crop Science 43:515-518. DOI: 10.2135/cropsci2003.0515.

Bradow J.M., Davidonis, G. H. . (2000) Quantitation of fiber quality and the cotton production-processing interface: a physiologist’s perspective. Journal of Cotton Science 4:34-64.

Brown A., Paterson A.H., Li L. (2012) Genomics and Breeding in Food Crops.

Brown N., Kumar P., Singh R., Lubbers E., Campbell B.T., Myers G.O., Wright R.J., Subramani J., Jones D., Paterson A.H., Chee P.W. (2019) Evaluation of a Chromosome Segment from Gossypium barbadense Harboring the Fiber Length QTL qFL-Chr.25 in Four

Diverse Upland Cotton Genetic Backgrounds. Crop Science 59:2621-2633. DOI: 10.2135/cropsci2019.05.0321.

Brubaker C.L., Paterson A.H., Wendel J.F. (1999) Comparative genetic mapping of allotetraploid cotton and its diploid progenitors. Genome 42:184-203. DOI: 10.1139/gen-42-2-184.

Chandnani R., Kim C., Guo H., Shehzad T., Wallace J.G., He D.H., Zhang Z.S., Patel J.D., Adhikari J., Khanal S., Paterson A.H. (2018) Genetic Analysis of Gossypium Fiber Quality Traits in Reciprocal Advanced Backcross Populations. Plant Genome 11:15. DOI: 10.3835/plantgenome2017.06.0057.

Chandnani R., Wang B.H., Draye X., Rainville L.K., Auckland S., Zhuang Z.M., Lubbers E.L., May O.L., Chee P.W., Paterson A.H. (2017a) Segregation distortion and genome-wide digenic interactions affect transmission of introgressed chromatin from wild cotton species. Theoretical and Applied Genetics 130:2219-2230. DOI: 10.1007/s00122-017-2952-y.

Chandnani R., Zhang Z.S., Patel J.D., Adhikari J., Khanal S., He D.H., Brown N., Chee P.W., Paterson A.H. (2017b) Comparative genetic variation of fiber quality traits in reciprocal advanced backcross populations. Euphytica 213. DOI: 10.1007/s10681-017-2029-7.

Chee P., Draye X., Jiang C.X., Decanini L., Delmonte T.A., Bredhauer R., Smith C.W., Paterson A.H. (2005a) Molecular dissection of interspecific variation between Gossypium hirsutum and Gossypium barbadense (cotton) by a backcross-self approach: I. Fiber elongation. Theoretical and Applied Genetics 111:757-763. DOI: 10.1007/s00122-005-2063-z.

Chee P.W., Draye X., Jiang C.X., Decanini L., Delmonte T.A., Bredhauer R., Smith C.W., Paterson A.H. (2005b) Molecular dissection of phenotypic variation between Gossypium hirsutum and Gossypium barbadense (cotton) by a backcross-self approach: III. Fiber length. Theoretical and Applied Genetics 111:772-781. DOI: 10.1007/s00122-005-2062-0.

Chen Z.J., Scheffler B.E., Dennis E., Triplett B.A., Zhang T.Z., Guo W.Z., Chen X.Y., Stelly D.M., Rabinowicz P.D., Town C.D., Arioli T., Brubaker C., Cantrell R.G., Lacape J.M., Ulloa M., Chee P., Gingle A.R., Haigler C.H., Percy R., Saha S., Wilkins T., Wright R.J., Van Deynze A., Zhu Y.X., Yu S.X., Abdurakhmonov I., Katageri I., Kumar P.A., Mehboob ur R., Zafar Y., Yu J.Z., Kohel R.J., Wendel J.F., Paterson A.H. (2007) Toward Sequencing cotton (Gossypium) Genomes. Plant Physiology 145:1303-1310. DOI: 10.1104/pp.107.107672.

Claverie M., Souquet M., Jean J., Forestier-Chiron N., Lepitre V., Pre M., Jacobs J., Llewellyn D., Lacape J.M. (2012) cDNA-AFLP-based genetical genomics in cotton fibers. Theoretical and Applied Genetics 124:665-683. DOI: 10.1007/s00122-011-1738-x.

Culp T.W., Harrell D.C. (1973) BREEDING METHODS FOR IMPROVING YIELD AND FIBER QUALITY OF UPLAND COTTON (GOSSYPIUM-HIRSUTUM L). Crop Science 13:686-689. DOI: 10.2135/cropsci1973.0011183X001300060030x.

d'Eeckenbrugge G.C., Lacape J.M. (2014) Distribution and Differentiation of Wild, Feral, and Cultivated Populations of Perennial Upland Cotton (Gossypium hirsutum L.) in Mesoamerica and the Caribbean. Plos One 9. DOI: 10.1371/journal.pone.0107458.

de Menezes I.P.P., da Silva J.O., Malafaia G., Silveira R.D.D., Barroso P.A.V. (2015) Natural hybridization between Gossypium mustelinum and exotic allotetraploid cotton species. Genetics and Molecular Research 14:14177-14180. DOI: 10.4238/2015.October.29.40.

Deng F.L., Tu L.L., Tan J.F., Li Y., Nie Y.C., Zhang X.L. (2012) GbPDF1 Is Involved in Cotton Fiber Initiation via the Core cis-Element HDZIP2ATATHB2. Plant Physiology 158:890-904. DOI: 10.1104/pp.111.186742.

Desai A., Chee P.W., Rong J., May O.L., Paterson A.H. (2006) Chromosome structural changes in diploid and tetraploid A genomes of Gossypium. Genome 49:336-345. DOI: 10.1139/g05-116.

Ding M.Q., Cao Y.F., He S., Sun J., Dai H.Q., Zhang H., Sun C.D., Jiang Y.R., Paterson A.H., Rong J.K. GaHD1, a candidate gene for the Gossypium arboreum SMA-4 mutant, promotes trichome and fiber initiation by cellular H2O2 and Ca2+ signals. Plant Molecular Biology. DOI: 10.1007/s11103-020-01000-3.

Ding M.Q., Ye W.W., Lin L.F., He S.E., Du X.M., Chen A.Q., Cao Y.F., Qin Y., Yang F., Jiang Y.R., Zhang H., Wang X.Y., Paterson A.H., Rong J.K. (2015) The Hairless Stem Phenotype of Cotton (Gossypium barbadense) Is Linked to a Copia-Like Retrotransposon Insertion in a Homeodomain-Leucine Zipper Gene (HD1). Genetics 201:143-+. DOI: 10.1534/genetics.115.178236.

Draye X., Chee P., Jiang C.X., Decanini L., Delmonte T.A., Bredhauer R., Smith C.W., Paterson A.H. (2005) Molecular dissection of interspecific variation between Gossypium hirsutum and G-barbadense (cotton) by a backcross-self approach: II. Fiber fineness. Theoretical and Applied Genetics 111:764-771. DOI: 10.1007/s00122-005-2061-1.

Endrizzi J.E., Turcotte, E. L., Kohel, R. J. . (1985) Genetics, cytology, and evolution of Gossypium. Advanced Genetics 23:271-375.

Esbroeck G.B., DT (1998) Cotton germplasm diversity and its importance to cultivar development. Journal of Cotton Science 2:121-129.

Espigares T., Peco B. (1995) MEDITERRANEAN ANNUAL PASTURE DYNAMICS - IMPACT OF AUTUMN DROUGHT. Journal of Ecology 83:135-142. DOI: 10.2307/2261157.

Espina M.J., Ahmed C.M.S., Bernardini A., Adeleke E., Yadegari Z., Arelli P., Pantalone V., Taheri A. (2018) Development and Phenotypic Screening of an Ethyl Methane Sulfonate Mutant Population in Soybean. Frontiers in Plant Science 9. DOI: 10.3389/fpls.2018.00394.

Fang D.D., Hinze L.L., Percy R.G., Li P., Deng D., Thyssen G. (2013) A microsatellite-based genome-wide analysis of genetic diversity and linkage disequilibrium in Upland cotton (Gossypium hirsutum L.) cultivars from major cotton-growing countries. Euphytica 191:391-401. DOI: 10.1007/s10681-013-0886-2.

Fang D.D., Naoumkina M., Thyssen G.N., Bechere E., Li P., Florane C.B. (2020) An EMS-induced mutation in a tetratricopeptide repeat-like superfamily protein gene (Ghir_A12G008870) on chromosome A12 is responsible for the li(y) short fiber phenotype in cotton. Theoretical and Applied Genetics 133:271-282. DOI: 10.1007/s00122-019-03456-4.

Favret E.A. (1960) Somatic mutations of four genes for albinism in barley induced by X-rays and ethyl methane-sulphonate. Hereditas 46:622-634.

Feng L.C., Zhang S.W., Xing L.T., Yang B., Gao X., Xie X.M., Zhou B.L. (2019) QTL analysis for yield and fibre quality traits using three sets of introgression lines developed from three Gossypium hirsutum race stocks. Molecular Genetics and Genomics 294:789-810. DOI: 10.1007/s00438-019-01548-w.

Gichner T., Ehrenberg L. (1966) INFLUENCE OF POST-TREATMENT STORAGE ON FREQUENCY OF EMS-INDUCED CHROMOSOMAL ABERRATIONS IN BARLEY. Biologia Plantarum 8:256-+. DOI: 10.1007/bf02930787.

Greene E.A., Codomo C.A., Taylor N.E., Henikoff J.G., Till B.J., Reynolds S.H., Enns L.C., Burtner C., Johnson J.E., Odden A.R., Comai L., Henikoff S. (2003) Spectrum of chemically induced mutations from a large-scale reverse-genetic screen in arabidopsis. Genetics 164:731-740.

Grover C.E., Yoo M.J., Lin M., Murphy M.D., Harker D.B., Byers R.L., Lipka A.E., Hu G.J., Yuan D.J., Conover J.L., Udall J.A., Paterson A.H., Gore M.A., Wendel J.F. (2020) Genetic Analysis of the Transition from Wild to Domesticated Cotton (Gossypium hirsutum L.). G3-Genes Genomes Genetics 10:731-754. DOI: 10.1534/g3.119.400909.

Guo W.Z., Cai C.P., Wang C.B., Han Z.G., Song X.L., Wang K., Niu X.W., Wang C., Lu K.Y., Shi B., Zhang T.Z. (2007) A microsatellite-based, gene-rich linkage map reveals genome structure, function and evolution in gossypium. Genetics 176:527-541. DOI: 10.1534/genetics.107.070375.

Haigler C.H., Betancur L., Stiff M.R., Tuttle J.R. (2012) Cotton fiber: a powerful single-cell model for cell wall and cellulose research. Frontiers in Plant Science 3. DOI: 10.3389/fpls.2012.00104.

Haigler C.H., Zhang D.H., Wilkerson C.G. (2005) Biotechnological improvement of cotton fibre maturity. Physiologia Plantarum 124:285-294. DOI: 10.1111/j.1399-3054.2005.00480.x.

Herring A.D., Auld D.L., Ethridge M.D., Hequet E.F., Bechere E., Green C.J., Cantrell R.G. (2004) Inheritance of fiber quality and lint yield in a chemically mutated population of cotton. Euphytica 136:333-339. DOI: 10.1023/b:euph.0000032747.97343.54.

Hildering G.J., Vanderve.Jh. (1966) MUTUAL INDEPENDENCE OF M1-FERTILITY AND MUTANT YIELD IN EMS TREATED TOMATOES. Euphytica 15:412-+. DOI: 10.1007/bf00022188.

Ijaz B., Zhao N., Kong J., Hua J.P. (2019) Fiber Quality Improvement in Upland Cotton (Gossypium hirsutum L.): Quantitative Trait Loci Mapping and Marker Assisted Selection Application. Frontiers in Plant Science 10. DOI: 10.3389/fpls.2019.01585.

Ji S.J., Lu Y.C., Feng J.X., Wei G., Li J., Shi Y.H., Fu Q., Liu D., Luo J.C., Zhu Y.X. (2003) Isolation and analyses of genes preferentially expressed during early cotton fiber development by subtractive PCR and cDNA array. Nucleic Acids Research 31:2534-2543. DOI: 10.1093/nar/gkg358.

Jiang C.X., Wright R.J., El-Zik K.M., Paterson A.H. (1998) Polyploid formation created unique avenues for response to selection in Gossypium (cotton). Proceedings of the National Academy of Sciences of the United States of America 95:4419-4424. DOI: 10.1073/pnas.95.8.4419.

John M.E., Keller G. (1996) Metabolic pathway engineering in cotton: Biosynthesis of polyhydroxybutyrate in fiber cells. Proceedings of the National Academy of Sciences of the United States of America 93:12768-12773. DOI: 10.1073/pnas.93.23.12768.

Li F.G., Fan G.Y., Lu C.R., Xiao G.H., Zou C.S., Kohel R.J., Ma Z.Y., Shang H.H., Ma X.F., Wu J.Y., Liang X.M., Huang G., Percy R.G., Liu K., Yang W.H., Chen W.B., Du X.M., Shi C.C., Yuan Y.L., Ye W.W., Liu X., Zhang X.Y., Liu W.Q., Wei H.L., Wei S.J., Huang G.D., Zhang X.L., Zhu S.J., Zhang H., Sun F.M., Wang X.F., Liang J., Wang J.H., He Q., Huang L.H., Wang J., Cui J.J., Song G.L., Wang K.B., Xu X., Yu J.Z., Zhu Y.X., Yu S.X. (2015) Genome sequence of cultivated Upland cotton (Gossypium hirsutum TM-1) provides insights into genome evolution. Nature Biotechnology 33:524-U242. DOI: 10.1038/nbt.3208.

Lopes C.M.L., Suassuna N.D., Cares J.E., Gomes A., Perina F.J., Nascimento G.F., Mendonca J.S.F., Moita A.W., Carneiro R. (2020) Marker-assisted selection in Gossypium spp. for Meloidogyne incognita resistance and histopathological characterization of a near immune line. Euphytica 216. DOI: 10.1007/s10681-020-2554-7.

Lubbers E.L., Chee, P. W. . (2009), in: A. H. Paterson (Ed.), Genetics and Genomics of Cotton Plant Genetics and Genomics: Crops and Models Springer. pp. 23-52.

Mansoor S., Paterson A.H. (2012) Genomes for jeans: cotton genomics for engineering superior fiber. Trends in Biotechnology 30:521-527. DOI: 10.1016/j.tibtech.2012.06.003.

May O.L. (1999) Genetic variation for fiber quality, in: A. S. Basra (Ed.), Cotton fibers-developmental biology, quality improvement, and textile processing, Food Products Press, New York. pp. 183-229.

Mei M., Syed N.H., Gao W., Thaxton P.M., Smith C.W., Stelly D.M., Chen Z.J. (2004) Genetic mapping and QTL analysis of fiber-related traits in cotton (Gossypium). Theoretical and Applied Genetics 108:280-291. DOI: 10.1007/s00122-003-1433-7.

Naoumkina M., Zeng L.H., Fang D.D., Wang M.J., Thyssen G.N., Florane C.B., Li P., Delhom C.D. (2020) Mapping and validation of a fiber length QTL on chromosome D11 using two independent F-2 populations of upland cotton. Molecular Breeding 40. DOI: 10.1007/s11032-020-01111-1.

Niu Q., Tan K.L., Zang Z.L., Xiao Z.Y., Chen K.J., Hu M.Y., Luo M. (2019) Modification of phytosterol composition influences cotton fiber cell elongation and secondary cell wall deposition. Bmc Plant Biology 19. DOI: 10.1186/s12870-019-1830-y.

Palle S.R., Campbell L.M., Pandeya D., Puckhaber L., Tollack L.K., Marcel S., Sundaram S., Stipanovic R.D., Wedegaertner T.C., Hinze L., Rathore K.S. (2013) RNAi-mediated Ultra-low gossypol cottonseed trait: performance of transgenic lines under field conditions. Plant Biotechnology Journal 11:296-304. DOI: 10.1111/pbi.12013.

Pang C.Y., Wang H., Pang Y., Xu C., Jiao Y., Qin Y.M., Western T.L., Yu S.X., Zhu Y.X. (2010) Comparative Proteomics Indicates That Biosynthesis of Pectic Precursors Is Important for Cotton Fiber and Arabidopsis Root Hair Elongation. Molecular & Cellular Proteomics 9:2019-2033. DOI: 10.1074/mcp.M110.000349.

Parry M.A.J., Madgwick P.J., Bayon C., Tearall K., Hernandez-Lopez A., Baudo M., Rakszegi M., Hamada W., Al-Yassin A., Ouabbou H., Labhilili M., Phillips A.L. (2009) Mutation discovery for crop improvement. Journal of Experimental Botany 60:2817-2825. DOI: 10.1093/jxb/erp189.

Patel J.D., Wright R.J., Chandnani R., Goff V.H., Ingles J., Paterson A.H. (2016) EMS-mutated cotton populations suggest overlapping genetic control of trichome and lint fiber variation. Euphytica 208:597-608. DOI: 10.1007/s10681-015-1614-x.

Paterson A.H., Boman R.K., Brown S.M., Chee P.W., Gannaway J.R., Gingle A.R., May L., Smith C.W. (2004) Reducing the genetic vulnerability of cotton. Crop Science 44:1900-1901. DOI: 10.2135/cropsci2004.1900.

Paterson A.H., Tanksley S.D., Sorrells M.E. (1991) DNA MARKERS IN PLANT IMPROVEMENT. Advances in Agronomy 46:39-90.

Perkins H.E., DE; Bragg, CK. (1984) Fiber American Society of Agronomy:438–508.

Ramesh U.M., Methre R., Kumar N.V.M., Katageri I.S., Gowda S.A., Adiger S., Yadava S.K., Lachagari V.B.R. (2019) Genome mapping and molecular markers identification for yield, yield component and fibre quality traits in tetraploid cotton. Plant Breeding 138:880-896. DOI: 10.1111/pbr.12728.

Rao S.H.K., Sears E.R. (1962) EMS-INDUCED MUTATIONS IN HEXAPLOID WHEAT. Genetics 47:983-&.

Rapp R.A., Haigler C.H., Flagel L., Hovav R.H., Udall J.A., Wendel J.F. (2010) Gene expression in developing fibres of Upland cotton (Gossypium hirsutum L.) was massively altered by domestication. Bmc Biology 8. DOI: 10.1186/1741-7007-8-139.

Reinisch A.J., Dong J., Brubaker C.L., Stelly D.M., Wendel J.F., Paterson A.H. (1994) A DETAILED RFLP MAP OF COTTON, GOSSYPIUM-HIRSUTUM X GOSSYPIUM BARBADENSE - CHROMOSOME ORGANIZATION AND EVOLUTION IN A DISOMIC POLYPLOID GENOME. Genetics 138:829-847.

Rong J., Feltus E.A., Waghmare V.N., Pierce G.J., Chee P.W., Draye X., Saranga Y., Wright R.J., Wilkins T.A., May O.L., Smith C.W., Gannaway J.R., Wendel J.R., Paterson A.H. (2007) Meta-analysis of polyploid cotton QTL shows unequal contributions of

subgenomes to a complex network of genes and gene clusters implicated in lint fiber development. Genetics 176:2577-2588. DOI: 10.1534/genetics.107.074518.

Saha S., Bellaloui N., Jenkins J.N., McCarty J.C., Stelly D.M. (2020) Effect of chromosome substitutions fromGossypium barbadenseL.,G. tomentosumNutt. Ex Seem andG. mustelinumWatt intoG. hirsutumL. on cottonseed protein and oil content. Euphytica 216. DOI: 10.1007/s10681-020-02644-4.

Said J.I., Lin Z.X., Zhang X.L., Song M.Z., Zhang J.F. (2013) A comprehensive meta QTL analysis for fiber quality, yield, yield related and morphological traits, drought tolerance, and disease resistance in tetraploid cotton. Bmc Genomics 14. DOI: 10.1186/1471-2164-14-776.

Said J.I., Song M.Z., Wang H.T., Lin Z.X., Zhang X.L., Fang D.D., Zhang J.F. (2015) A comparative meta-analysis of QTL between intraspecific Gossypium hirsutum and interspecific G. hirsutum x G. barbadense populations. Molecular Genetics and Genomics 290:1003-1025. DOI: 10.1007/s00438-014-0963-9.

Saravanan N.A., Ravikesavan, R., Raveendran, T. S. . (2010) Combining ability analysis for yield and fibre quality parameters in intraspecific hybrids of G. hirsutum L. . Elect J Pl Br 4:856-863.

Scholl R.L., Miller P.A. (1976) GENETIC ASSOCIATION BETWEEN YIELD AND FIBER STRENGTH IN UPLAND COTTON. Crop Science 16:780-783. DOI: 10.2135/cropsci1976.0011183X001600060010x.

Shaban M., Miao Y.H., Ullah A., Khan A.Q., Menghwar H., Khan A.H., Ahmed M.M., Tabassum M.A., Zhu L.F. (2018) Physiological and molecular mechanism of defense in

cotton against Verticillium dahliae. Plant Physiology and Biochemistry 125:193-204. DOI: 10.1016/j.plaphy.2018.02.011.

Shen C., Wang N., Huang C., Wang M.J., Zhang X.L., Lin Z.X. (2019) Population genomics reveals a fine-scale recombination landscape for genetic improvement of cotton. Plant Journal 99:494-505. DOI: 10.1111/tpj.14339.

Shen X.L., Guo W.Z., Lu Q.X., Zhu X.F., Yuan Y.L., Zhang T.Z. (2007) Genetic mapping of quantitative trait loci for fiber quality and yield trait by RIL approach in Upland cotton. Euphytica 155:371-380. DOI: 10.1007/s10681-006-9338-6.

Shi Y.H., Zhu S.W., Mao X.Z., Feng J.X., Qin Y.M., Zhang L., Cheng J., Wei L.P., Wang Z.Y., Zhu Y.X. (2006) Transcriptome profiling, molecular biological, and physiological studies reveal a major role for ethylene in cotton fiber cell elongation. Plant Cell 18:651-664. DOI: 10.1105/tpc.105.040303.

Shi Y.Z., Liu A.Y., Li J.W., Zhang J.F., Zhang B.C., Ge Q., Jamshed M., Lu Q.W., Li S.Q., Xiang X.H., Gong J.W., Gong W.K., Shang H.H., Deng X.Y., Pan J.T., Yuan Y.L. (2019) Dissecting the genetic basis of fiber quality and yield traits in interspecific backcross populations of Gossypium hirsutum x Gossypium barbadense. Molecular Genetics and Genomics 294:1385-1402. DOI: 10.1007/s00438-019-01582-8.

Si Z.F., Chen H., Zhu X.F., Cao Z.B., Zhang T.Z. (2017) Genetic dissection of lint yield and fiber quality traits of G. hirsutum in G. barbadense background. Molecular Breeding 37. DOI: 10.1007/s11032-016-0607-3.

Sonah H., Bastien M., Iquira E., Tardivel A., Legare G., Boyle B., Normandeau E., Laroche J., Larose S., Jean M., Belzile F. (2013) An Improved Genotyping by Sequencing (GBS)

Approach Offering Increased Versatility and Efficiency of SNP Discovery and Genotyping. Plos One 8. DOI: 10.1371/journal.pone.0054603.

Steadman R.G. (1997) Cotton testing, in: J. A. Smirfitt (Ed.), Textile progress, The Textile Institute, Manchester.

Sun H., Meng M.H., Yan Z.H., Lin Z.X., Nie X.H., Yang X.Y. (2019) Genome-wide association mapping of stress-tolerance traits in cotton. Crop Journal 7:77-88. DOI: 10.1016/j.cj.2018.11.002.

Talebi A.B., Shahrokhifar, B. (2012) Ethyl methane sulphonate (EMS) induced mutagenesis in Malaysian rice (cv. MR219) for lethal dose determination. American Journal of Plant Sciences.

Ulloa M., De Santiago L.M., Hulse-Kemp A.M., Stelly D.M., Burke J.J. (2020) Enhancing Upland cotton for drought resilience, productivity, and fiber quality: comparative evaluation and genetic dissection. Molecular Genetics and Genomics 295:155-176. DOI: 10.1007/s00438-019-01611-6.

Vanderve.Jh. (1967) EMS INDUCED NON-RECESSIVE EMBRYONIC LETHALS IN ARABIDOPSIS THALIANA. Genetica 38:61-&.

Wang F.R., Zhang J.X., Chen Y., Zhang C.Y., Gong J.W., Song Z.Q., Zhou J., Wang J.J., Zhao C.J., Jiao M.J., Liu A.Y., Du Z.H., Yuan Y.L., Fan S.J., Zhang J. (2020) Identification of candidate genes for key fibre-related QTLs and derivation of favourable alleles in Gossypium hirsutum recombinant inbred lines with G. barbadense introgressions. Plant Biotechnology Journal 18:707-720. DOI: 10.1111/pbi.13237.

Wendel J.F. (1989) NEW WORLD TETRAPLOID COTTONS CONTAIN OLD-WORLD CYTOPLASM. Proceedings of the National Academy of Sciences of the United States of America 86:4132-4136. DOI: 10.1073/pnas.86.11.4132.

Wendel J.F. (2000) Genome evolution in polyploids. Plant Molecular Biology 42:225-249. DOI: 10.1023/a:1006392424384.

Wendel J.F., Cronn R.C. (2003) Polyploidy and the evolutionary history of cotton. Advances in Agronomy, Vol 78 78:139-186. DOI: 10.1016/s0065-2113(02)78004-8.

Xie F.L., Sun G.L., Stiller J.W., Zhang B.H. (2011) Genome-Wide Functional Analysis of the Cotton Transcriptome by Creating an Integrated EST Database. Plos One 6. DOI: 10.1371/journal.pone.0026980.

Xu Z.Y., Yu J.Z., Cho J.M., Yu J., Kohel R.J., Percy R.G. (2010) Polyploidization Altered Gene Functions in Cotton (Gossypium spp.). Plos One 5. DOI: 10.1371/journal.pone.0014351.

Yang Z.E., Ge X.Y., Yang Z.R., Qin W.Q., Sun G.F., Wang Z., Li Z., Liu J., Wu J., Wang Y., Lu L.L., Wang P., Mo H.J., Zhang X.Y., Li F.G. (2019) Extensive intraspecific gene order and gene structural variations in upland cotton cultivars. Nature Communications 10. DOI: 10.1038/s41467-019-10820-x.

Zeng Z.B. (1994) PRECISION MAPPING OF QUANTITATIVE TRAIT LOCI. Genetics 136:1457-1468.

Zhang J.F., Percy R.G., McCarty J.C. (2014) Introgression genetics and breeding between Upland and Pima cotton: a review. Euphytica 198:1-12. DOI: 10.1007/s10681-014-1094-4.

Zhang K., Kuraparthy V., Fang H., Zhu L.L., Sood S., Jones D.C. (2019a) High-density linkage map construction and QTL analyses for fiber quality, yield and morphological traits using

CottonSNP63K array in upland cotton (Gossypium hirsutum L.). Bmc Genomics 20. DOI: 10.1186/s12864-019-6214-z.

Zhang Z., Liu A.Y., Huang Z., Fan S.M., Zou X.Y., Deng X.Y., Ge Q., Gong J.W., Li J.W., Gong W.K., Shi Y.Z., Fan L.Q., Zhang Z.B., Jiang X., Lei K., Yuan Y.L., Xu A.X., Shang H.H. (2019b) Quantitative trait locus mapping for plant height and branch number in an upland cotton recombinant inbred line with an SNP-based high-density genetic map. Euphytica 215. DOI: 10.1007/s10681-019-2428-z.

Zhang Z., Zou X.Y., Huang Z., Fan S.M., Qun G., Liu A.Y., Gong J.W., Li J.W., Gong W.K., Shi Y.Z., Fan L.Q., Zhang Z.B., Liu R.X., Jiang X., Lei K., Shang H.H., Xu A.X., Yuan Y.L. (2019c) Genome-wide identification and analysis of the evolution and expression patterns of the GATA transcription factors in three species of Gossypium genus. Gene 680:72-83. DOI: 10.1016/j.gene.2018.09.039.

Zhang Z.S., Rong J.K., Waghmare V.N., Chee P.W., May O.L., Wright R.J., Gannaway J.R., Paterson A.H. (2011) QTL alleles for improved fiber quality from a wild Hawaiian cotton, Gossypium tomentosum. Theoretical and Applied Genetics 123:1075-1088. DOI: 10.1007/s00122-011-1649-x.

Zhou H., Deng H.Q., Su J.Q., Sun Y., Jiang L.N., Liu Z.L., Tang C.M. (2020) Breeding high-yield upland cotton (Gossypium hirsutum) lines with high-quality fibre by mutation. Plant Breeding 139:1201-1212. DOI: 10.1111/pbr.12858.

Zhou X.Y., Hu W.R., Li B., Yang Y., Zhang Y., Thow K., Fan L., Qu Y.Y. (2019) Proteomic profiling of cotton fiber developmental transition from cell elongation to secondary wall deposition. Acta Biochimica Et Biophysica Sinica 51:1168-1177. DOI: 10.1093/abbs/gmz111.

# CHAPTER 2

# USING EXOTIC *GOSSYPIUM HIRSUTUM* TO INCREASE GENETIC VARIATION IN ELITE *G. HIRSUTUM* VARIETIES

## Introduction

Cotton (*Gossypium hirsutum*), often referred to as white gold, is one of the most important crops in the United States and around the world. While there are more than fifty species in the *Gossypium* genus, only four are cultivated and of those, *G. hirsutum* accounts for approximately 95% of global production (Mansoor and Paterson, 2012). One hundred and fifty countries are involved in cotton import and export and it is a major economic driver for developing countries (Chen et al., 2007). In the 2019-2020 growing season, an estimated 125.3 million bales were produced (Saha et al., 2020). In the US, cotton production, distribution, and processing provides approximately $27 billion in revenue and supports around 200,000 jobs (Zhang et al., 2019). Recently, the onset of synthetic fibers and more efficient spinning technologies have negatively affected the cotton market because of a lower cost (Brown et al., 2019). In order for the US to remain competitive in the global market, the focus has had to shift from increasing fiber yield to improving fiber quality (Brown et al., 2019).

Previously, fiber quality characteristics focused on length and fineness with little to no emphasis on strength, uniformity, or elongation, but with improved processing, these qualities are becoming more important to consider (Chee et al., 2005). However, upland cotton, *G. hirsutum*, is genetically impoverished due to a series of bottlenecks from polyploidization, domestication, and migration (Paterson et al., 2004). In addition, overexploitation of few genetic

backgrounds have led to plateaus in yield and fiber quality traits and reduced the standing genetic diversity of the elite cotton gene pool (Abbas et al., 2015; Adhikari et al., 2017). This narrow genetic base is one of the major obstacles to cotton improvement (Paterson et al., 2004). Therefore, finding ways to add or increase genetic diversity back into the cotton gene pool is extremely important to increasing fiber yield and quality simultaneously.

One potential way to increase genetic diversity is bring in new alleles to elite varieties from wild relatives. Wild relatives of cotton are a rich source of diversity for improvement (Wang et al., 2017). Wild *G. hirsutum* has unexploited variation which can be used for introgressive breeding and *G. tomentosum* is a potential source to enhance strength and length in domesticated cotton (Mansoor and Paterson, 2012; Zhang et al., 2011). Previously, wild or exotic relatives have been used to find QTLs associated with increased yield in soybean, increase disease resistance in oat, improve a number of economically important traits in sunflower, and map complex traits in sorghum (Jan et al., 2020; Mace et al., 2021; Palomeque et al., 2009; Warburton et al., 2017). In addition, exotic cotton germplasm has been evaluated in a number of studies which show that exotic lines have traits which can be helpful for disease resistance, drought tolerance, improved fiber quality characteristics, as well as an overall increase in genetic diversity (Abdurakhmonov et al., 2008; Quisenberry et al., 1981; Seyoum et al., 2018; Zaidi et al., 2020; Zeng and Meredith, 2009). These benefits from exotic and wild cotton could have profound impacts if they are able to be introgressed into current elite cultivars.

One way to find and incorporate beneficial alleles from numerous exotic lines is through the creation of nested association mapping populations (NAM). NAM populations were first developed in maize and are a result of a multi-cross scheme where diverse parents are mated with a common parent that is often an elite or adapted variety (Bajgain et al., 2021; Perumal et

al.). NAM populations mitigate the weaknesses of developing RIL (recombinant inbred lines) populations by creating numerous RILs that share a common recurrent parent which allows for a more complete picture of the genetic architecture of traits by increasing the sampling of segregating genetic variation (Brock et al., 2020). Understanding the genetic architecture of complex traits is very important in the development of lines and to design effective breeding strategies (Brock et al., 2020; Olatoye et al., 2020). These populations have a higher mapping resolution and power than bi-parental segregation populations due to the ability to access multiple gene alleles as opposed to just two (Ali et al., 2020; Bu et al., 2021; Perumal et al.). The NAM approach also increases overall allelic diversity and the number of recombination events while decreasing confounding population structure (Gage et al., 2020). NAM populations have been used to identify pathogen resistance genes from exotic sources in maize, barley, and wheat (Buttner et al., 2020). While NAM populations have increased in popularity in recent decades, there are extremely few instances of their occurrence in cotton. This article aims to use one of the first NAM populations developed in cotton to find SNP markers from exotic lines with a genome wide association study (GWAS).

**Materials and Methods**

Fourteen exotic *G. hirsutum* lines were chosen for hybridization representing different botanical races of cotton. The exotic parents that represent each race and the populations that were created from them can be seen in Table 2.1. Race stocks, or semi-wild lines, are lines formed by long term natural selection under different geographical conditions and may have retained numerous advantageous traits (Feng et al., 2019). The selected lines have been previously selected for day neutral flowering. The exotic lines were hybridized with elite *G. hirsutum* varieties, either DES 56 or Acala Maxxa, in 2010 to create a variety of unique

populations, twenty-four of which (P001-P023, and P025) were propagated by selfing through the F6 generation. The list of populations and the crosses that were used to create them can be seen in Table 2.2. The F5 generation was grown at the former Plant Sciences Farm in Bogart, GA in 2015 and the F6 generation was grown at the University of Georgia Iron Horse Farm in Watkinsville, GA in 2016. The F6 generation was grown in a randomized complete block design with two replications. Fiber weight and seed weight were collected from the whole plot after harvesting and after ginning. A 10g sample was sent to Cotton Inc. in Cary, NC for fiber quality analysis using a high-volume instrument (HVI). All parents were also grown in the same location in 2016 and fiber yield and quality data were also collected for comparison. All phenotypic data was analyzed in the R statistical program. Differences among the populations and between the parental and progeny populations were analyzed using linear regression, ANOVA and Tukey HSD (honest significant difference) tests. Correlations between traits were also tested using R. Phenotypic data from 2015 was analyzed separately to compare to the genotype analysis, and then phenotypic data from both years was analyzed together to look for any interactions between population and years.

DNA was extracted from individual plants in 2015 using the methods described in Paterson (1993). This method uses fresh, young leaves and a combination of buffers to extract DNA of a size and quality that is suitable for analysis. DNA concentration was checked on a nanodrop to make sure it was of high enough quality for sequencing library creation. Libraries were created according to a modified version of the methods in (Poland and Rife, 2012). DNA samples were digested with CutSmart® Buffer and the enzymes PstI-HF and MspI and run in a thermocycler at 37°C for three hours and then at 80°C for twenty minutes. Barcodes were then added to the DNA samples with T4 DNA Ligase buffer and T4 DNA Ligase (Enzymatics). To

ligate the barcodes to the samples the plate was run in the thermocycler at 25°C for two hours and 65°C for twenty minutes. The DNA samples were then pooled together and cleaned using a Qiagen PCR clean-up kit. The DNA was then run on a 2% TAE size selection gel at 80V for one and a half hours. The fragment between 200-500bp was cut out and used for gel elution. Gel elution was done using a Qiagen gel elution kit and then the DNA concentration was checked on Qubit. Afterward, the DNA samples were amplified using 2X GoTaq Master mix and forward and reverse primers. The plate was run for 18 cycles or PCR, then pooled and the concentration taken again. Another size selection gel and gel elution step were completed to remove any DNA strands that were shorter than 200bp. Samples were then sent to the Georgia Genome Facility for fragment analysis and sequencing using Illumina NextSeq.

The sequences from Illumina NextSeq were processed in the TASSEL 5.0 standalone pipeline (Bradbury et al., 2007). This pipeline calls the distinct barcodes associated with each sample, and aligns each tag with a reference genome. The reference genome used was the *Gossypium hirsutum* (AD1) 'TM-1' genome UTX_v2.1 genome alignment. The individual sequences were aligned to the reference genome using the bwa alignment tool. SNPs were then called from the sequences using the ProductionSNPCallerPlugin in TASSEL. The resulting SNPs were output to a HapMap file for use in analysis.

A total of 54,159 SNPs were called by TASSEL. After filtering so that each site had a minor allele frequency (MAF) of at least 0.01 and all sites had at least 20% of the taxa with non-missing data, there were 8,728 SNPs left. This was done to remove low coverage and monomorphic sites. A genome wide association analysis (GWAS) was also done in TASSEL. First, a kinship matrix was created using the Normalized_IBS model which is best for analysis of complex traits. Principle coordinate analysis was completed using R and the first five PCs were

included as covariates in the model. To include both the kinship and PCs, a mixed linear model was used for analysis. SNPs were considered significant if they had a LOD (logarithm of odds) score of greater than three which equates to a p-value of less than 0.001. Four markers, *SA02_104284095, SA02_107270426, SA11_14167635,* and *SA12_104595483*, had extreme LOD values and were therefore included as covariates. The allelic, or genotypic, effects were estimated in TASSEL. The allelic effects show the effect of each different allele that was seen within each SNP in reference to the genotype that was set to 0. For example, if at a marker there were three genotypes seen, AA, CC, and AC, and AA was set to 0. Then CC has a negative effect and AC has a positive effect, that would mean that the genotype CC contributed a lower phenotypic value than AA and AC contributed a higher phenotypic value than AA. Specifically, these effects show the phenotypic difference between genotypes. A positive effect may or may not be preferred depending on the trait, as lower values are preferred for fineness and short fiber content. It is important to note that for the fiber weight and seed weight analysis the parental phenotypes were not included in the analysis because they were collected using a different method than the progeny populations. Heterozygosity was also measured for all parents by counting the number of heterozygous genotypes at each SNP.

## Results

### Phenotypic Analysis: 2015

Correlations between all traits can be seen in Figure 2.1. There were two strongly negative correlations, which were between SFC and UI and seed weight and lint percent. These relationships make sense as uniformity is used to calculate SFC, and seed weight is used to calculate lint percent. There were four strong positively correlated pairs, which were UHM and STR, UHM and UI, STR and UI, and fiber weight and lint percent. There were four moderately

negatively correlated pairs which were UHM and MIC, UHM and ELO, UHM and SFC, and STR and SFC. There were two positive moderately correlated pairs which are between lint percent and MIC and lint percent and STR.

For fiber length (UHM), no progeny populations were significantly different from their exotic parent and only one population was significantly different from its elite parent. P017 was significantly longer than its elite parent Acala Maxxa, but not its exotic parent T18. This population was also significantly better than 13 of the other progeny populations (P001, P002, P004, P005, P010, P013, P014, P016, P018, P019, P020, P021, P023). P011 was not significantly different from either of its parents, DES56 or T63, but was significantly longer than ten of the other progeny populations (P001, P004, P005, P013, P016, P018, P019, P020, P021, P023). A boxplot showing the distributions of fiber length by population is shown in Figure 2.2.

For fiber fineness (MIC) there were several differences between the parents and the progeny populations. Populations P006, P007, P009, P014, P015, P017, and P021 were significantly better (lower value is preferred) than their elite parent Acala Maxxa and populations P008, P011, and P019 were significantly better than their elite parent DES56. Populations P021 and P019 were also significantly better than their exotic parent T246 and P015 was also significantly better than its exotic parent, T171. P021 was also significantly better than eight other progeny populations (P001, P002, P003, P005, P010, P013, P018, and P023). A boxplot showing the distributions of fiber fineness by population is shown in Figure 2.3.

For uniformity index (UI) no progeny populations were significantly different than their elite parents. One population, P020, was significantly worse than its exotic parent T368. P021 was also significantly worse than P011, which was the only difference among the progeny

populations. A boxplot showing the distributions of fiber uniformity by population is shown in Figure 2.4.

For fiber strength, there were no progeny populations that were significantly different than their elite parent. There were also no statistically significant differences between any of the progeny populations. There were eight different progeny populations that were significantly worse than their exotic parents. P004 and P023 were significantly worse than their exotic parent, MDN257. P005 and P013 were significantly worse than their exotic parent, T45. P016 and P020 were significantly worse than their exotic parent, T368. P021 was significantly worse than its exotic parent, T246, and P008 was significantly worse than its exotic parent T18. A boxplot showing the distributions of fiber strength by population is shown in Figure 2.5.

For fiber elongation (ELO) none of the progeny populations were significantly different from their exotic parent. There was one progeny population, P004, that was significantly better than its elite parent, DES56, but again was not different from the exotic parent, MDN257. Populations P004 and P022 were also significantly better than four other progeny populations (P003, P007, P008, and P016). A boxplot showing the distributions of fiber elongation by population is shown in Figure 2.6.

There were no differences between the progeny and either of the parents for short fiber content (SFC). There were also no differences between any of the progeny populations for SFC. A boxplot showing the distributions of short fiber content by population is shown in Figure 2.7.

For lint percent (LP), all progeny populations except for P007 were significantly lower than both the elite parents and the exotic parents. In addition, populations P001 and P007 were significantly better than seven other progeny populations (P011, P012, P013, P015, P019, P021,

and P025). P001 was also significantly better than P022. A boxplot showing the distributions of lint percent by population is shown in Figure 2.8.

For fiber weight and seed weight, only differences between progeny populations were analyzed. For fiber weight, populations P001, P007, P010, P016, and P017 were significantly higher than at least one other progeny population. P016 and P017 were only significantly better than P006. P007 was significantly better than three other populations (P006, P011, and P012). P001 was significantly better than five other populations (P006, P011, P012, P015, and P025). Finally, P010 was significantly better than nine other populations (P006, P009, P011, P012, P015, P018, P019, P022, and P025). A boxplot showing the distributions of fiber weight by population is shown in Figure 2.9.

For seed weight, populations P004, P010, P012, and P013 were significantly higher than at least one other progeny population. P004 was significantly higher than P018. P010 was significantly higher than three other populations (P001, P007, and P018). P012 was significantly higher than two populations (P007 and P018). P013 was significantly higher than nine other populations (P001, P006, P007, P016, P018, P020, P022, P023, and P025). A boxplot showing the distributions of seed weight by population is shown in Figure 2.10.

**Phenotypic analysis: 2015 and 2016**

For the 2015 and 2016 data together, correlations between all traits can be seen in Figure 2.11. There were three strong negative correlations, which were between SFC and UI, seed weight and fiber weight, and seed weight and lint percent. There were four strong positively correlated pairs, which were UHM and STR, UHM and UI, MIC and LP, and fiber weight and lint percent. There were eight moderately negatively correlated pairs, which were UHM and MIC, MIC and seed weight, STR and seed weight, fiber weight and ELO, UHM and SFC, ELO

and STR, ELO and lint percent, and STR and SCF. There were six positive moderately correlated pairs, which are between fiber weight and MIC, fiber weight and STR, STR and UI, ELO and seed weight, STR and lint percent, and SFC and lint percent.

For the F5 and F6 generations together, we were looking for differences between parents and progeny and between the progeny populations, as well as any genotype by environment (GxE) interactions, which in this case is population by year. GxE effects were seen in MIC, UHM, UI, FW, SW, and lint percent. For MIC, the populations that had significant effects between years were P015, P019, and P021. For UHM, the populations that had significant effects between years were P016, P019, P020, and P021. For UI, the only population that had significant effects between years was P021. For FW, the populations that had significant effects between years were P004, P005, P006, P009, P011, P012, P014, P015, P018, P019, P020, P022, and P025. For seed weight, the populations that had significant effects between years were P002, P003, P004, P005, P008, P009, P010, P011, P012, P013, P014, P017, P019, and P021. For lint percent, the populations that had significant effects between years were P001, P002, P003, P007, P016, P017, P018, P019, and P021.

In MIC, the only significant difference that occurred was P015 was significantly better than its exotic parent, T171. For UHM, one population, P017, was significantly longer than its elite parent Acala Maxxa. P017 was also significantly better than all other progeny populations except for P011. P005, P013, and P020 were significantly shorter than their exotic parents, T45 and T368. P011 was also significantly longer than its exotic parent, T63, and longer than progeny populations P004, P005, P013, P018, P019, and P020. For UI, there were no statistically significant differences among the progeny populations. However, four populations, P005, P013, P016, and P020, were significantly lower than their exotic parents, T45 for P005

and P013, and T368 for P016 and P020. For STR, again there were no differences between any of the progeny populations. However, six populations, P004, P005, P013, P016, P020, and P023, were significantly lower than their exotic parents, MDN257, T45, and T368. For ELO and SFC, there were no significant differences between any parents and their progeny or between any of the progeny populations.

For lint percent, P002, P003, P004, p008, P010, P011, P012, P013, P019, P022, and P025 were all significantly lower than their elite parent DES56. Populations P004, P019, and P021 were also significantly lower than their exotic parents, MDN257 and T246. There were no differences among any of the progeny for lint percent. For fiber weight, there were no differences in any of the progeny populations. For seed weight, P018 was significantly lower than P004, P008, P010, P012, P013, and P019. P007 was significantly lower than P004, P010, and P013. P013 was also significantly higher than P001, P007, P018, P020, P022, and P023.

**GWAS Results**

Genome wide association analysis was carried out with a mixed linear model using a kinship matrix and five PCs and four markers as covariates. In total, 101 unique SNPs were found to be associated with at least one of the nine traits studied. Twelve of these SNPs were associated with more than one fiber quality trait, for a total of 113 SNP marker-trait associations discovered. There were trait-associated SNPs found on 23 out of the 26 chromosomes, with D02, D03, and D06 not having any SNPs associated with traits. The chromosomes that had the most trait-associated SNPs were D11 with ten SNPs and D12 with thirteen SNPs.

For fiber length (UHM), three associated SNPs were found across three chromosomes, A07, D04, and D11. One of the SNPs was also associated with fiber strength. These SNPs explained 4-7% of the phenotypic variation. In the allelic, or genotypic, effects, the rare alleles

at all three SNPs contributed positive effects. The effects for all genotypes and the overall additive model for all traits can be seen in supplementary table 2.4. A Manhattan plot showing the p-values for all genome wide SNPs are shown in Figure 2.12.

For fiber fineness (MIC), there were three associated SNPs on three different chromosomes, A04, A12, and D04. These SNPs were not associated with any other traits and explained approximately 5-7% of the phenotypic variation seen. Lower values in MIC are preferred, so a negative effect here is beneficial. One rare allele contributed negative effects in the allelic, or genotypic, effects and two contributed positive effects. A Manhattan plot showing the p-values for all genome wide SNPs are shown in Figure 2.13.

For fiber uniformity index (UI), there were eight associated SNPs across six chromosomes. Four SNPs were found on each subgenome. All these SNPs were associated with SFC, and one was also associated with STR. These SNPs explained 6-12% of the phenotypic variation observed. In the allelic, or genotypic, effects all but one of the rare alleles contributed much larger negative effects than any of the more common alleles. For example, in SNP *SD01_23337801*, the rare allele contributed a negative effect of -7.84, while the common allele contributed a positive effect of 0.397. The rare allele of one SNP, *SD11_2869210*, contributed positive effects. A Manhattan plot showing the p-values for all genome wide SNPs are shown in Figure 2.14.

For fiber strength (STR), there were 28 associated SNPs across 18 chromosomes. Fourteen SNPs were in the D genome and 14 were in the A genome. The chromosome that had the most SNPs was D11 with four. These SNPs explained 4-11% of the phenotypic variation. One of these SNPs was associated with UHM, and another SNP was associated with UI and SFC. In the allelic, or genotypic, effects, nine of the rare alleles contributed negative effects and

18 of the rare alleles contributed positive effects. A Manhattan plot showing the p-values for all genome wide SNPs are shown in Figure 2.15.

For fiber elongation (ELO), there were no associated SNPs. A Manhattan plot showing the p-values for all genome wide SNPs are shown in Figure 2.16.

For short fiber content (SFC), there were 21 associated SNPs, across ten chromosomes. Ten SNPs were in the D genome and eleven were in the A genome. The chromosomes that had the most were D12, A10 and A11 with four SNPs each. These SNPs explained 4-21% of the observed variation. Eight of these SNPs were also associated with uniformity index and one with STR. In the allelic, or genotypic, effects, twelve of the rare alleles contributed positive effects, and nine contributed negative effects. It is important to note that lower values are also preferred for SFC, so negative effects are beneficial. A Manhattan plot showing the p-values for all genome wide SNPs is shown in Figure 2.17.

For lint percent (LP), there were 37 associated SNPs across eighteen chromosomes. Seventeen SNPs were in the A genome and 20 were in the D genome. The chromosome with the most SNPs was D12 with six SNPs, and chromosomes A05, A12, D05, D07, and D11 all had three SNPs each. Two of these SNPs were also associated with fiber weight. These SNPs explained 2-7% of the phenotypic variation observed. In the allelic, or genotypic, effects, 32 of the rare alleles contributed positive effects, and only five contributed negative effects. A Manhattan plot showing the p-values for all genome wide SNPs is shown in Figure 2.18.

For fiber weight, in which the parental phenotypes were not included, there were nine associated SNPs across seven chromosomes. Five SNPs were in the A genome and four were in the D genome. Two of these SNPs were also associated with lint percent. These SNPs explained approximately 3-9% of the phenotypic variation observed. In the allelic, or genotypic,

effects, eight rare alleles contributed positive effects, and one contributed negative effects. A Manhattan plot showing the p-values for all genome wide SNPs is shown in Figure 2.19.

For seed weight, which also did not include the parental phenotypes, there were four SNPs on four chromosomes, two in the A genome and two in the D genome. None of these SNPs were associated with any other traits. These SNPs explained approximately 5-8% of the phenotypic variation observed. In the allelic, or genotypic, effects, three rare alleles contributed positive effects and one rare allele contributed negative effects. A Manhattan plot showing the p-values for all genome wide SNPs is shown in Figure 2.20.

The results of the heterozygosity percentages for each parent are shown in Table 2.3. Heterozygosity was measured for all called SNPs and for all significant SNPs. Each parent's heterozygosity percentage increased in the significant SNPs showing that heterozygous genotypes may play a significant role in improving fiber quality.

**Discussion**

Even though 101 SNPs were found to be associated with at least one trait, there appears to be only three instances of quantitative trait locus (QTL) discovery, based on compelling evidence from multiple SNPs. In short fiber content, on chromosome D12, there is a cluster of four SNPs near the end of the chromosome spanning approximately 10cM. In lint percent, also on chromosome D12, there is a cluster of six SNPs within approximately 16cM near the end of the chromosome. These two peaks also overlap. A peak of significant p-values, such as these two, is often indicative of the presence of a QTL. These SNPs for both traits ranged from approximately 45Mbp to 61Mbp on the chromosome. Based on information from the Cotton QTL database, it is possible that these SNPs are within the QTL qLP-c25-2 for lint percent and could be used as markers (Said et al., 2013; Said et al., 2015). There is currently not a QTL

reported for short fiber content in this region, so this could potentially be a novel QTL. For short fiber content, all four of the SNPs that made up this potential QTL contributed large positive effects. However, lower values for short fiber content are preferred, so these are not beneficial.

There also may be evidence for a QTL for fiber strength on chromosome D11. Here there are three SNPs within a cluster of approximately 13cM from 6Mbp to 19Mbp on the chromosome. This is potentially within the known QTLs qFS24.1, QTLFS12.f for fiber strength. A large percentage of the rare alleles in the SNPs that significantly contributed to fiber strength were in one exotic parent, T368. This exotic line did have the highest average strength of all parents and progeny populations with 35 g/tex, however, it does not appear that this trait/QTL was passed down to the progeny populations, P016 and P020, as the statistical analysis found that these populations were significantly worse in fiber strength that T368. This is also reflected in the genotypic data, as the progeny populations do not have the positive rare alleles of the parent. Often, this parent had the only instance of the rare beneficial allele in all individuals sampled. This could be due to the fact that previous studies have found that only approximately 0.5% of the exotic DNA segments from wild/exotic cotton is transmitted into the genome of each introgression line (He et al., 2011). Additionally, the exotic parent T368 is from the race of *G. hirsutum* known as marie-galante. This race is thought to include introgressions from *G. barbadense*, which is known to have higher fiber strength, and which could also impact the ability of these alleles to introgress into elite *G. hirsutum* lines (d'Eeckenbrugge and Lacape, 2014). However, these SNPs could potentially be used as markers for future use in populations especially those created from this parent. This information is also useful because it helps to gain insight into what alleles may be positively associated with fiber strength that could potentially be used in genome editing technology such as CRISPR/Cas. This is a preliminary study with a

small sample size, however, and would need more repetition and verification in order to confirm these alleles as beneficial. The present GWAS lacks power to discern how these SNPs may interact with each other within an individual, which may also contribute significantly to the outcome of a trait. The other two exotic parents that did significantly better than their progeny with regard to fiber strength, T45 and MDN257, did not have the same alleles as T368 that contributed significant benefits. However, occasionally they had alleles that were not the common alleles that generally contributed negative effects, but just did not appear to contribute any effects at all.

Overall, 52 of the SNPs (46%) found were in QTLs that were associated with the same trait that the SNPs were associated with. Also, 89 out of 101 SNPs (88%) were within known gene sequences based on the NCBI assembly Gossypium_hirsutum_v2.1 and the UTX assembly *Gossypium hirsutum* (AD1) 'TM-1' genome UTX_v2.1 (Chen et al., 2020; Yu et al., 2014). Of those 89 SNPs within genes, 82 (92%) were within exons of the genes, and seven were within introns. In addition, only sixteen of these genes were ones where the function is unknown. All other SNPs were within genes with known functions. Also, 28 SNPs were within previously known markers according to the Cotton Marker Database (Yu et al., 2014).

Forty-eight out of 101 of the SNP markers (47.5%) were found in the A genome. This means that the number of SNPs in both the A and D genome was approximately equal. This reinforces finding from previous studies that show that even though the D genome ancestors are not able to produce spinnable fiber, the Dt genome makes an important contribution to fiber production of *G. hirsutum* (Adhikari et al., 2017). The D subgenome has also been hypothesized to harbor higher levels of nucleotide and allelic diversity, so finding more potential SNPs to use

as markers in genes that potentially control a significant amount of fiber development could be very useful to exploit that diversity (Guo et al., 2007).

From the statistical analysis, we can see that some of the exotic parents did provide beneficial traits to progeny populations, such as an increase in length in P017 over its elite parent. However, this does not appear to be reflected in the genotypic data. In the three SNPs associated with fiber length, the exotic parent T18 had one allele that was different from the elite parent, Acala Maxxa, but none of the individuals in the progeny population had this allele from the exotic parent. Another example is in fiber elongation. Progeny population P004 was significantly better than DES56, but not different than its exotic parent MDN257. In the marker data, generally the individuals of P004 had the same genotype as the exotic parent, but there were no SNPs found at all. In addition, SFC had a large number (21) of associated SNPs, but yet there were no statistical differences in the phenotypic data at all. These discrepancies could be due to the fact that each progeny population had only ten individuals each, which may not be enough power to detect these individual associated alleles. One potential reason for this is because the minor allele frequency for the association study is set at 0.01, so any beneficial allele would have to be present in at least three individuals to be included in the analysis. Since this is a small population, it is possible that an allele could only be in one individual and therefore would not have been included, or have enough power, in this analysis. Since we can see significant differences in phenotypes in the statistical data, it may be beneficial to take the best progeny populations for each trait and evaluate them in larger number for a bi-parental QTL mapping study, instead of as part of a larger study. This will increase the power to potentially be able to find specific QTLs or alleles that have been introgressed into the progeny from the exotic parents. While this specific study was unable to find some specific beneficial alleles, it was able

to identify some hybrid combinations that had a greater heterotic effect, which could be beneficial in future population or selection studies (de Carvalho et al., 2017).

A result that might be of economic interest, is if there are any patterns that emerge based on the different races of exotic parents used in this study. There were a few patterns that emerged from the phenotypic data. Generally, populations with the same exotic parent (not necessarily in the same race) did not differ from each other even though they were crossed to different elite parents. In fiber fineness, at least one progeny population representative of each of the seven races used was better than the elite parent when the elite parent was Acala Maxxa. Three races, latifolium, richmondii, and marie-galante, also had progeny populations that were better than their elite parent DES56. In addition, when both Acala and Des56 were crossed to T246, which was from race marie-galante, the progeny populations were better than both parents, suggesting this may be an additive trait. For fiber strength, a number of exotics were better than the progeny they produced, regardless of whether they were crossed to Acala or DES56. The races that were represented by stronger exotics were morrilli, punctatum, richmondii, and marie-galante. For the race marie-galante, both exotic parents, T368 and T246, had significantly higher fiber strength but this trait was not passed down to the progeny populations as mentioned above. In addition, T368 also had a high uniformity which was not passed down to its progeny populations either. For fiber elongation, the two best populations were from two different races, morrilli and richmondii, but were both crossed to DES56. Overall, there were only a few patterns that emerged based on the botanical races represented, but this informations could be useful for targeting certain races to use for certain traits, such as marie-galante for fiber strength.

One thing that was interesting to look at was the diversity of the exotic parents versus the elite parents. One measure of diversity could be heterozygosity, which measures the number of

heterozygous genotypes within an individual or a population, depending on the population structure. It has been found that in general, elite genotypes have low levels of heterozygosity, and therefore diversity, due to intense inbreeding. So, one would expect exotic genotypes to have increased levels of heterozygosity. However, looking at this data the elite genotypes had around 1-2% heterozygous genotypes and the exotic genotypes had anywhere from 0.8 to 3% heterozygous genotypes, which is not drastically higher. In the SNPs that were found to be significant, the elite genotypes had around 4-5% heterozygous genotypes and the exotic genotypes had anywhere from 2 to 11.5% heterozygous genotypes, which is higher but still not as much as one might expect. Overall, there was a higher percentage of heterozygous genotypes in the significant SNPs which could suggest this diversity plays an important role in fiber quality improvement.

Additionally, since DES56 and Acala Maxxa are two lines that are very widely utilized, it may be of interest to breeders to know how these two elite lines are different. In the total of approximately 54,000 SNPs Acala and DES56 differed in approximately 1000 sites (not including any sites where one or both SNPs were missing), which is about 1.8%. In the SNPs that were found to be significantly associated with traits, 7.9% or eight SNPs in total were different between Acala and DES56. The traits that were associated with these SNPs were fiber weight, lint percent, short fiber content, fiber strength, fiber length, and seed weight. The percentage of differences did increase in the SNPs that were significant, showing that these differences in the elite lines may significantly contribute to the fiber quality of each.

Not only are exotics useful from improving economically valuable traits but learning and evaluating traits from wild relatives of a variety of crops, not just cotton, could have a profound benefit in the future as the climate continues to change. These wild relatives exist in a variety of

environments that could potentially tell us how our most important crops and food sources may change as local environments experience more droughts, or floods, or heat. "Natural selection has already tested more options than humans ever will," (Cortes and Lopez-Hernandez, 2021). This knowledge could prove to be invaluable as the world continues to change.

In conclusion, this study shows that NAM populations can be successfully utilized to incorporate genetic variation and diversity from exotic relatives and that these exotic relatives have beneficial alleles that can be introgressed into elite cultivars. This is one of few studies in cotton that has used the diversity present in exotic relatives to incorporate, improve and map economically important traits in elite cotton lines. More in-depth mapping needs to be done on a larger number of the best individual bi-parental populations in order to fully understand individual beneficial alleles and their effects within each genetic background.

**Table 2.1: Botanical races of *G. hirsutum*.** A list of the botanical races of *G. hirsutum* included in this experiment, the exotic parents that represent that race, and the progeny populations resulting from each exotic parent.

| Race | Exotic Parent | Progeny Populations |
|---|---|---|
| richmondii | T281 | P001, P018 |
| | T18 | P008, P017 |
| | T463 | P022 |
| yucatanense | T1046 | P002, P006 |
| palmeri | T326 | P003 |
| | T347 | P009, P012 |
| morrilli | MDN257 | P004, P023 |
| | T171 | P015 |
| punctatum | T45 | P005, P013 |
| | T165 | P010, P014 |
| latifolium | T63 | P007, P011 |
| | MDN117 | P025 |
| marie-galante | T368 | P016, P020 |
| | T246 | P019, P021 |

**Table 2.2: Population List.** A list of all progeny populations and the elite and exotic parents that created them.

| Population | Elite Parent | Exotic parent |
|---|---|---|
| P001 | DES56 | T281 |
| P002 | DES56 | T1046 |
| P003 | DES56 | T326 |
| P004 | DES56 | MDN257 |
| P005 | Acala Maxxa | T45 |
| P006 | Acala Maxxa | T1046 |
| P007 | Acala Maxxa | T63 |
| P008 | DES56 | T18 |
| P009 | Acala Maxxa | T347 |
| P010 | DES56 | T165 |
| P011 | DES56 | T63 |
| P012 | DES56 | T347 |
| P013 | DES56 | T45 |
| P014 | Acala Maxxa | T165 |
| P015 | Acala Maxxa | T171 |
| P016 | Acala Maxxa | T368 |
| P017 | Acala Maxxa | T18 |
| P018 | Acala Maxxa | T281 |
| P019 | DES56 | T246 |
| P020 | DES56 | T368 |
| P021 | Acala Maxxa | T246 |
| P022 | DES56 | T463 |
| P023 | Acala Maxxa | MDN257 |
| P025 | DES56 | MDN117 |

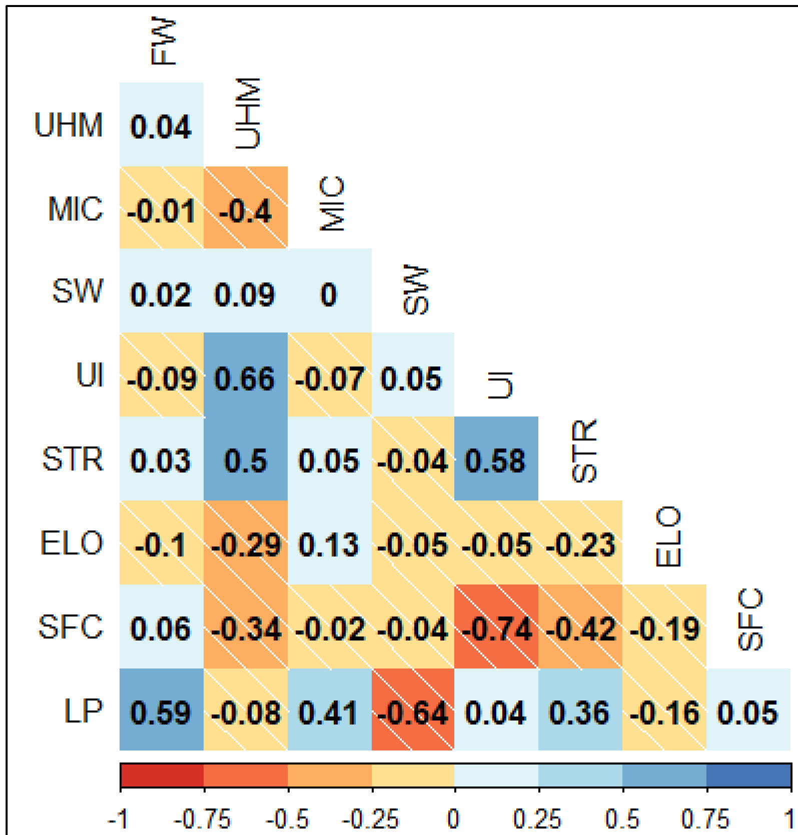


**Figure 2.1**: **Correlation matrix of fiber quality traits in 2015**. This is for all fiber quality traits in the F5 generation. Strong positive correlations are shown in darker blue and strong negative correlations are shown in darker red. Abbreviations shown are as follows fiber weight (FW), fiber length (UHM), fiber fineness (MIC), seed weight (SW), uniformity index (UI), fiber strength (STR), fiber elongation (ELO), short fiber content (SFC), and lint percent (LP).

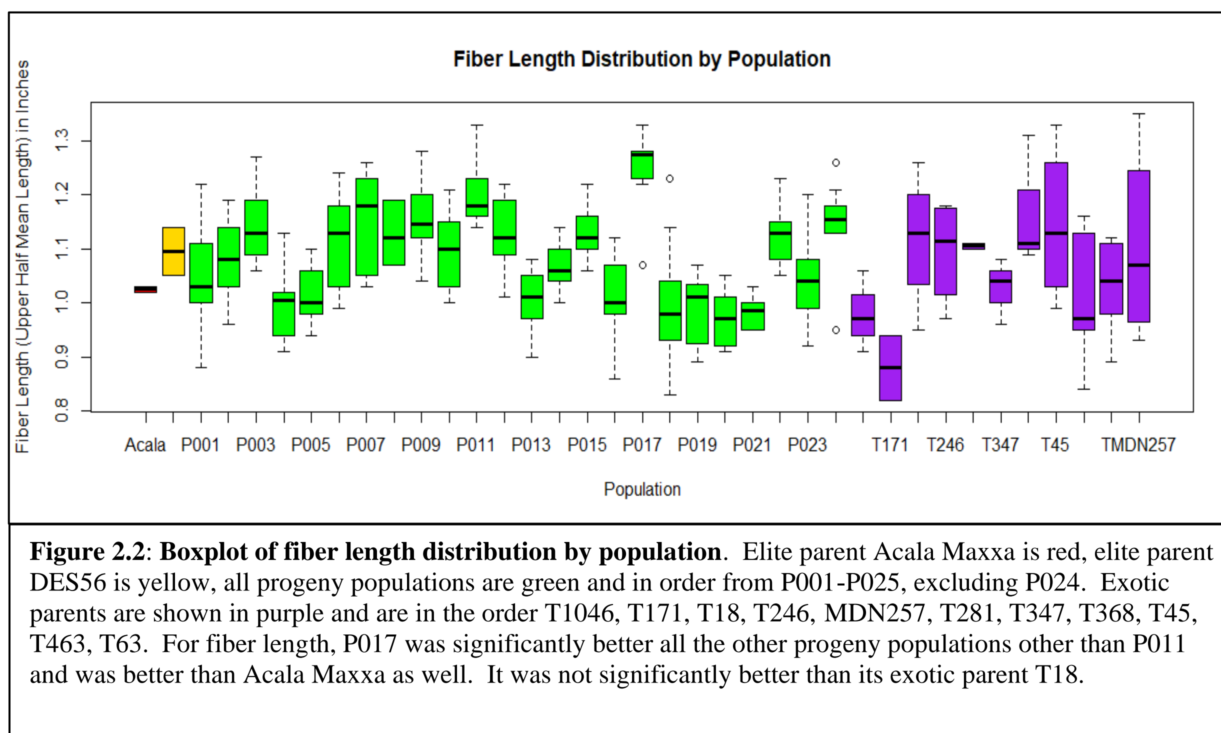


**Figure 2.2**: **Boxplot of fiber length distribution by population**. Elite parent Acala Maxxa is red, elite parent DES56 is yellow, all progeny populations are green and in order from P001-P025, excluding P024. Exotic parents are shown in purple and are in the order T1046, T171, T18, T246, MDN257, T281, T347, T368, T45, T463, T63. For fiber length, P017 was significantly better all the other progeny populations other than P011 and was better than Acala Maxxa as well. It was not significantly better than its exotic parent T18.

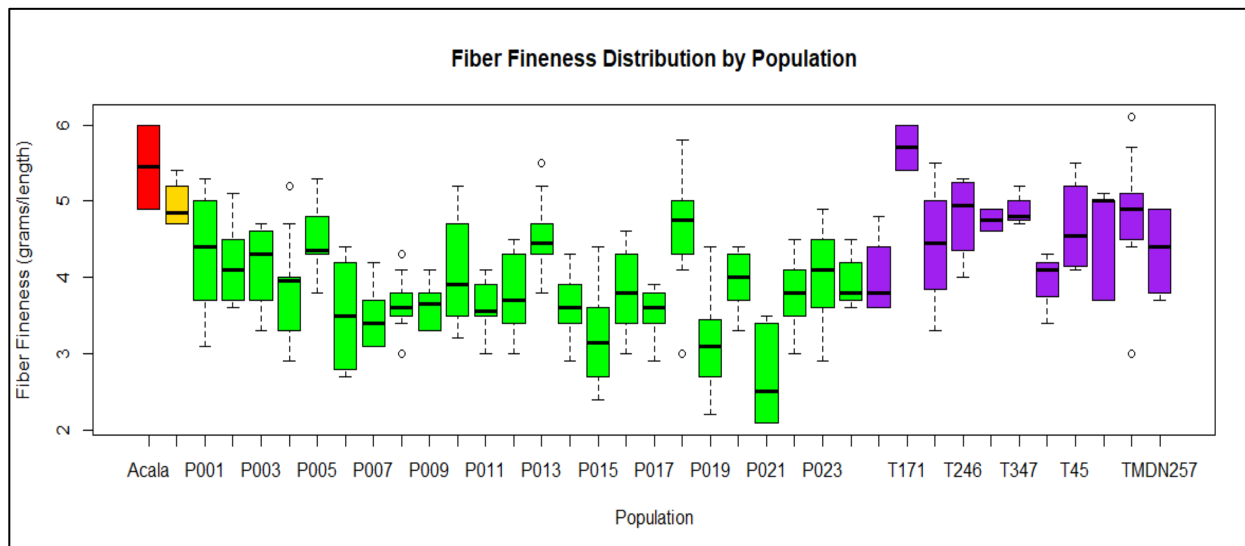


**Figure 2.3**: **Boxplot of fiber fineness distribution by population** (lower values are better for fineness). Elite parent Acala Maxxa is red, elite parent DES56 is yellow, all progeny populations are green and in order from P001-P025, excluding P024. Exotic parents are shown in purple and are in the order T1046, T171, T18, T246, MDN257, T281, T347, T368, T45, T463, T63. Populations P006, P007, P009, P014, P015, P017, and P021 were significantly better than their elite parent Acala Maxxa and populations P008, P011, and P019 were significantly better than their elite parent, DES56. P021 was also better than several other progeny populations.

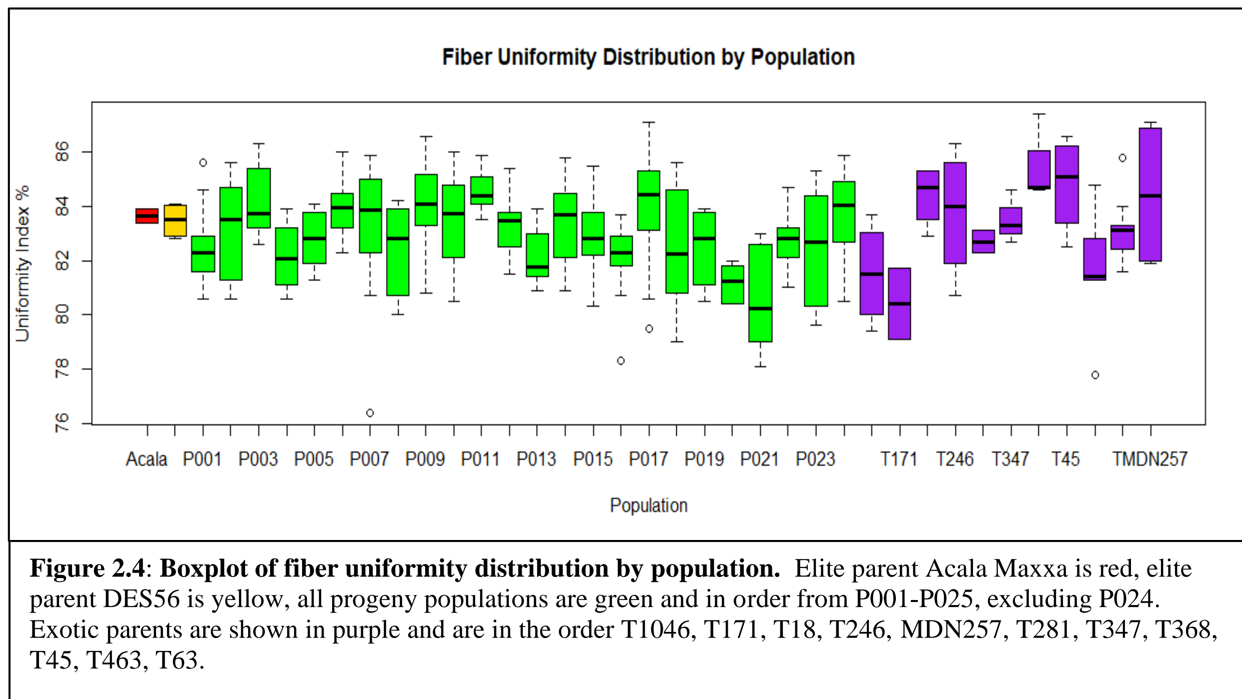


**Figure 2.4**: **Boxplot of fiber uniformity distribution by population.** Elite parent Acala Maxxa is red, elite parent DES56 is yellow, all progeny populations are green and in order from P001-P025, excluding P024. Exotic parents are shown in purple and are in the order T1046, T171, T18, T246, MDN257, T281, T347, T368, T45, T463, T63.

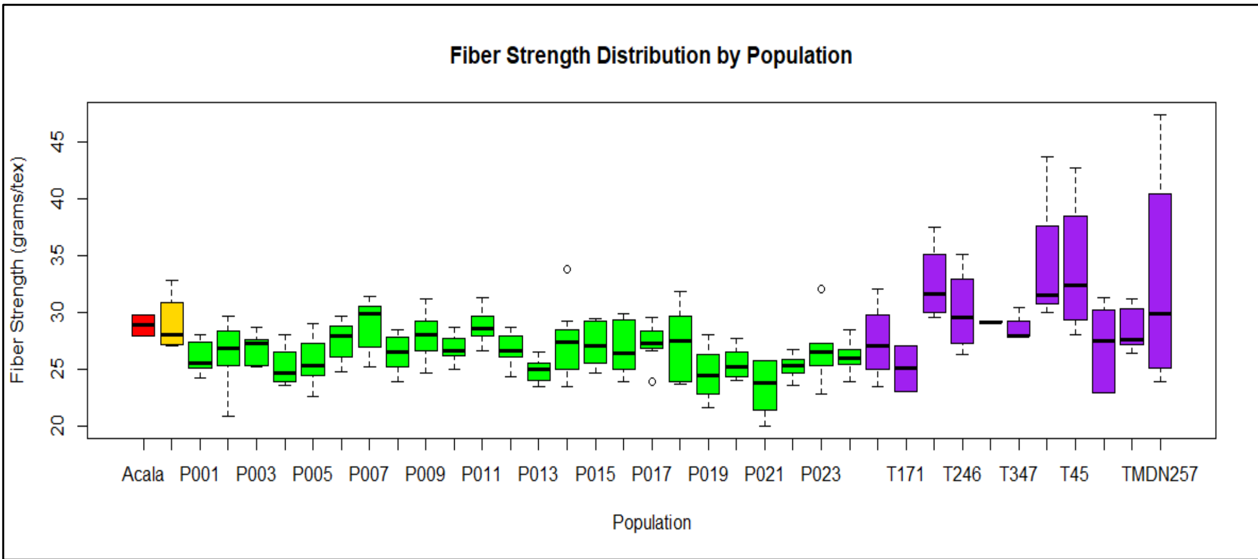


**Figure 2.5**: **Boxplot of fiber strength distribution by population**. Elite parent Acala Maxxa is red, elite parent DES56 is yellow, all progeny populations are green and in order from P001-P025, excluding P024. Exotic parents are shown in purple and are in the order T1046, T171, T18, T246, MDN257, T281, T347, T368, T45, T463, T63.

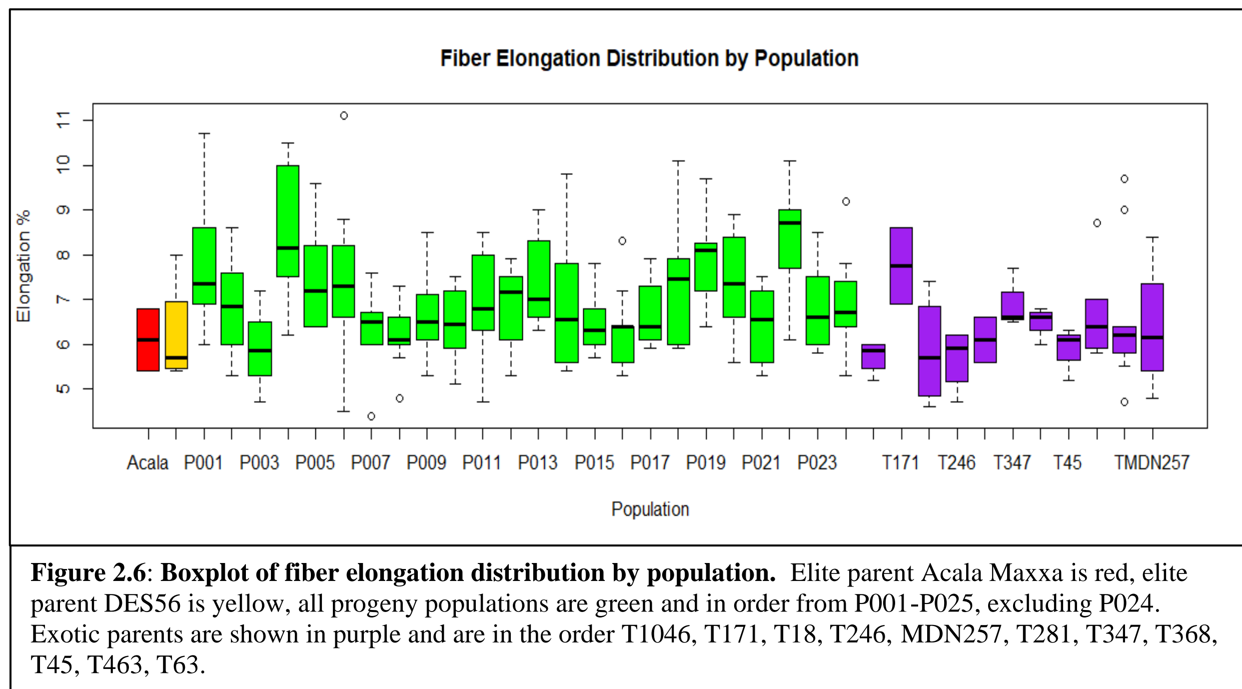


**Figure 2.6**: **Boxplot of fiber elongation distribution by population.** Elite parent Acala Maxxa is red, elite parent DES56 is yellow, all progeny populations are green and in order from P001-P025, excluding P024. Exotic parents are shown in purple and are in the order T1046, T171, T18, T246, MDN257, T281, T347, T368, T45, T463, T63.

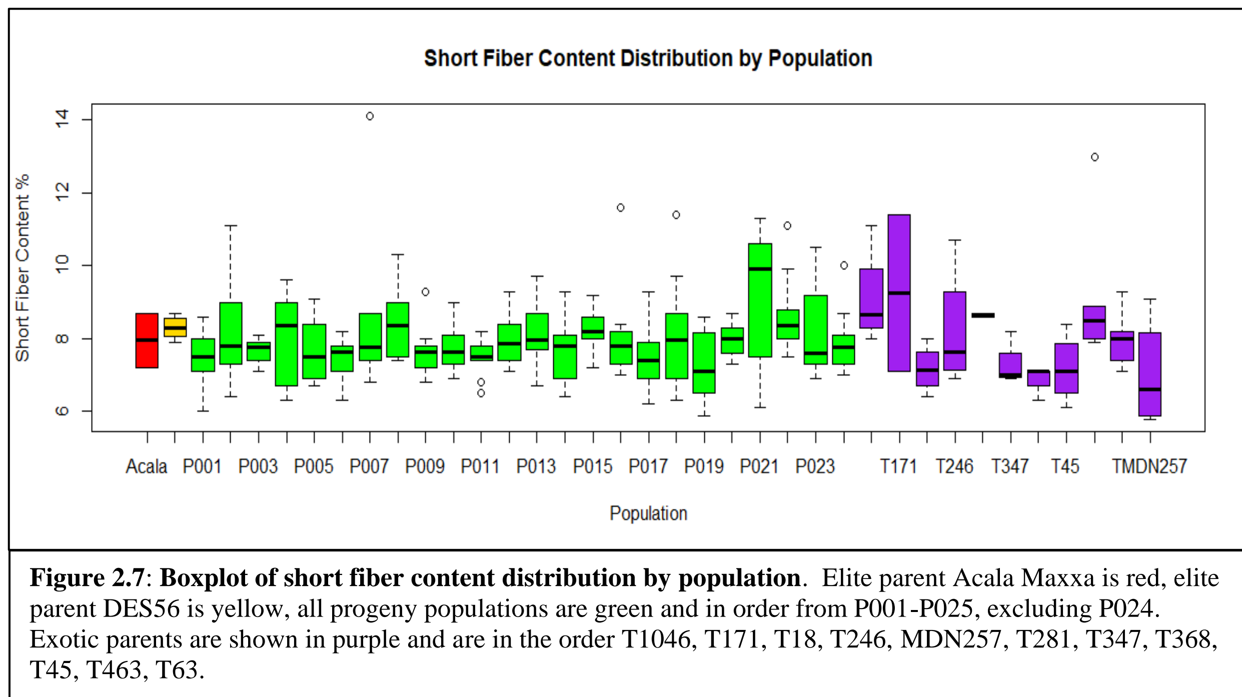


**Figure 2.7**: **Boxplot of short fiber content distribution by population**. Elite parent Acala Maxxa is red, elite parent DES56 is yellow, all progeny populations are green and in order from P001-P025, excluding P024. Exotic parents are shown in purple and are in the order T1046, T171, T18, T246, MDN257, T281, T347, T368, T45, T463, T63.

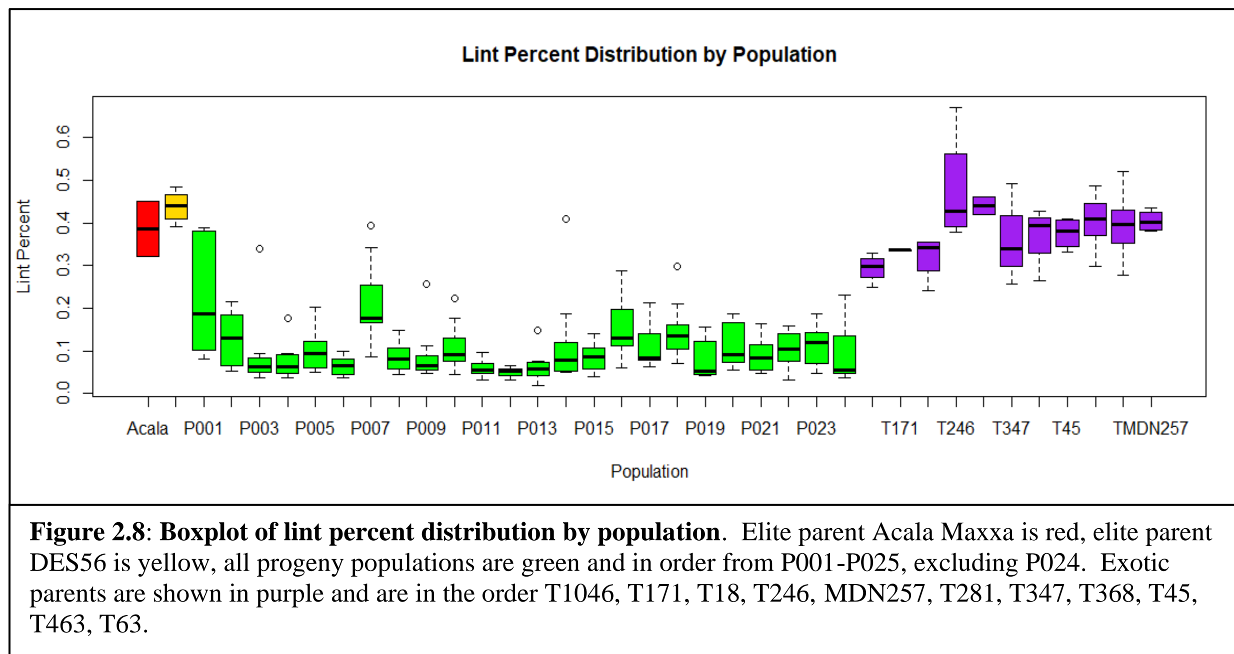


**Figure 2.8**: **Boxplot of lint percent distribution by population**. Elite parent Acala Maxxa is red, elite parent DES56 is yellow, all progeny populations are green and in order from P001-P025, excluding P024. Exotic parents are shown in purple and are in the order T1046, T171, T18, T246, MDN257, T281, T347, T368, T45, T463, T63.

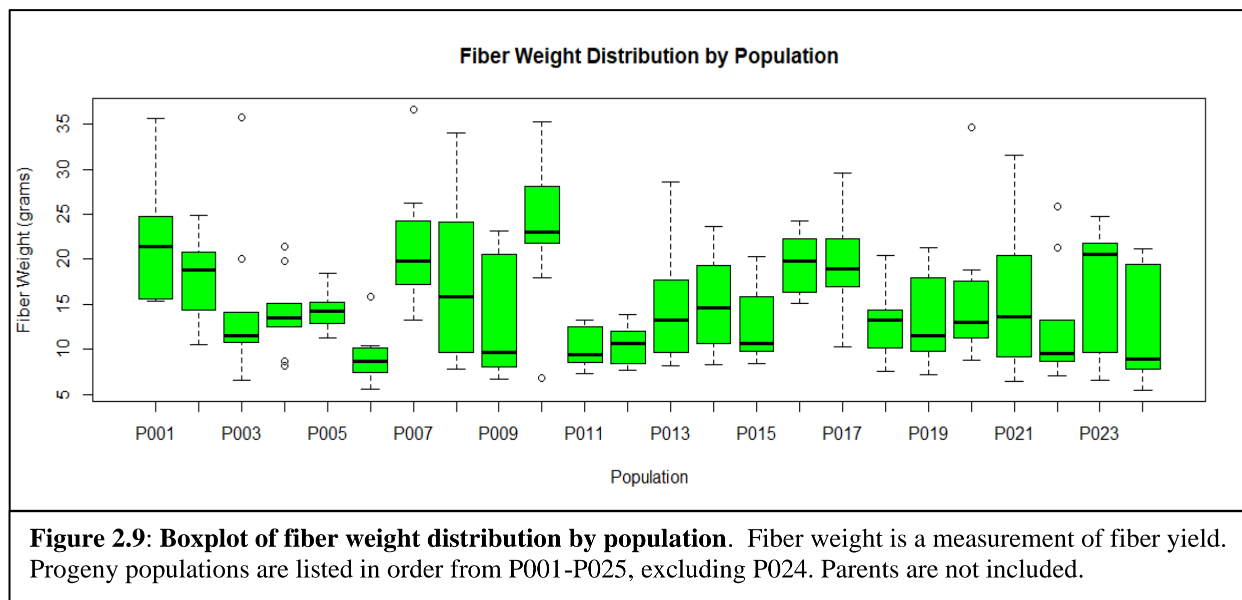


**Figure 2.9**: **Boxplot of fiber weight distribution by population**. Fiber weight is a measurement of fiber yield. Progeny populations are listed in order from P001-P025, excluding P024. Parents are not included.

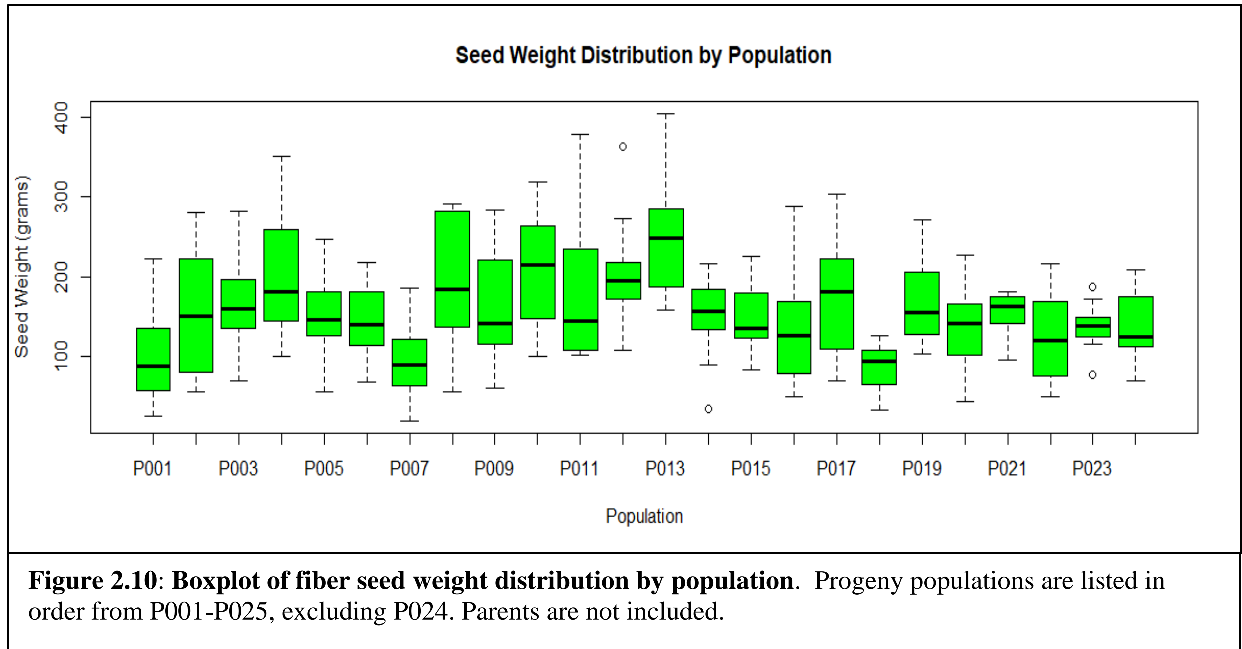


**Figure 2.10**: **Boxplot of fiber seed weight distribution by population**. Progeny populations are listed in order from P001-P025, excluding P024. Parents are not included.

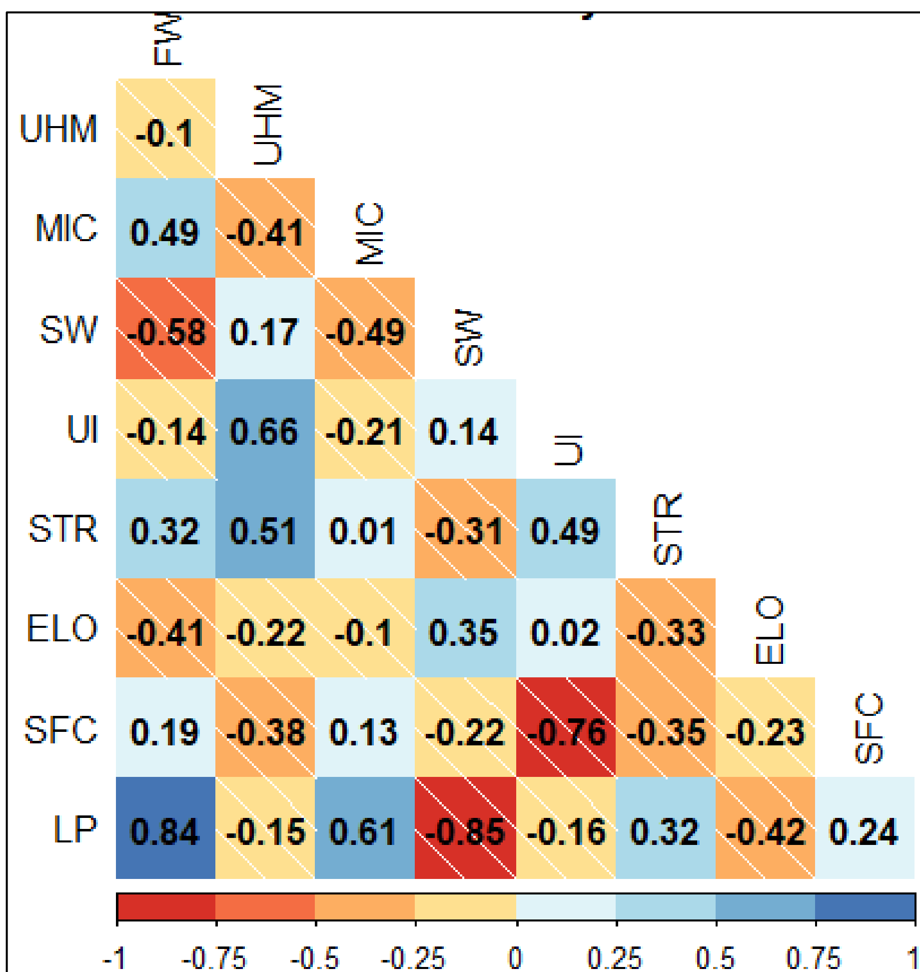


**Figure 2.11**: **Correlation matrix of fiber quality traits from 2015 and 2016**. This is for both the F5 and F6 generations together. Strong positive correlations are shown in darker blue and strong negative correlations are shown in darker red. Abbreviations shown are as follows fiber weight (FW), fiber length (UHM), fiber fineness (MIC), seed weight (SW), uniformity index (UI), fiber strength (STR), fiber elongation (ELO), short fiber content (SFC), and lint percent (LP).

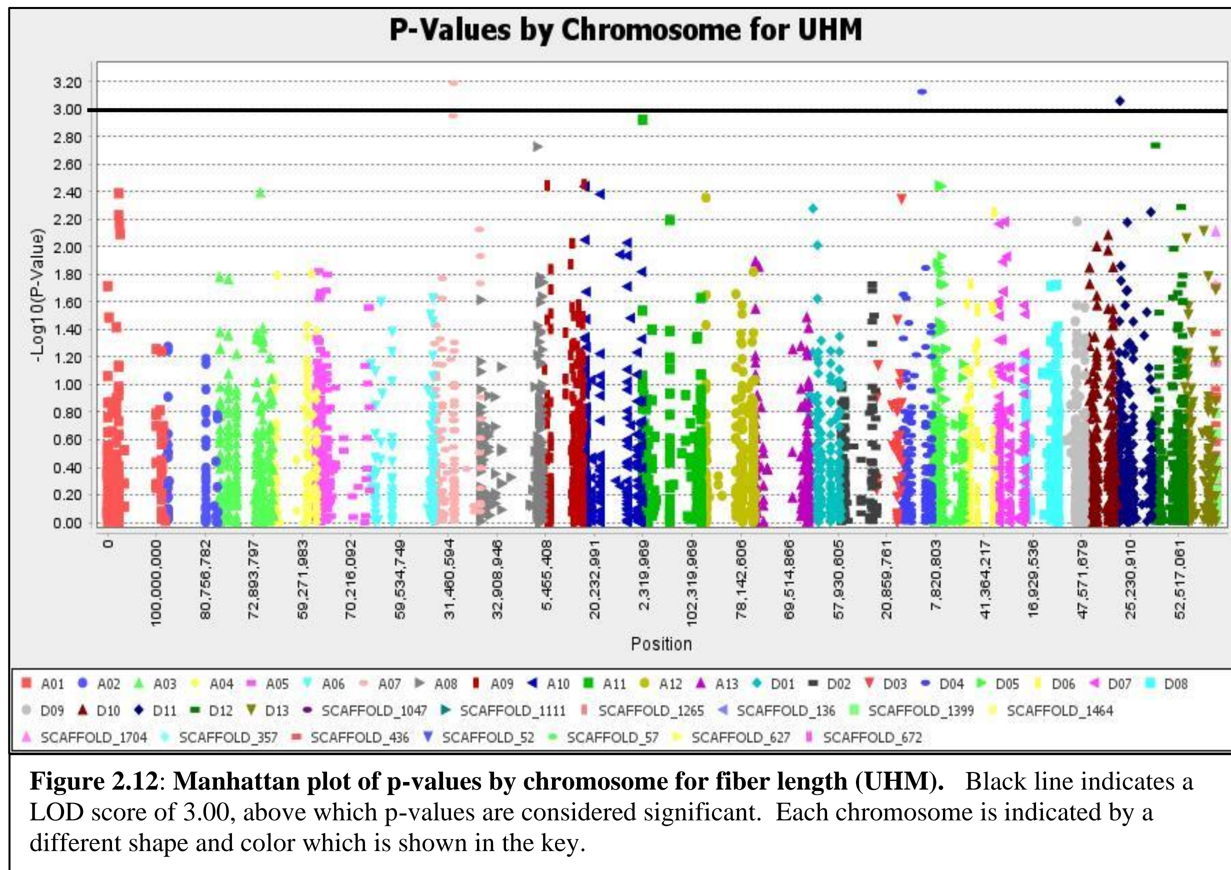


**Figure 2.12**: **Manhattan plot of p-values by chromosome for fiber length (UHM).** Black line indicates a LOD score of 3.00, above which p-values are considered significant. Each chromosome is indicated by a different shape and color which is shown in the key.

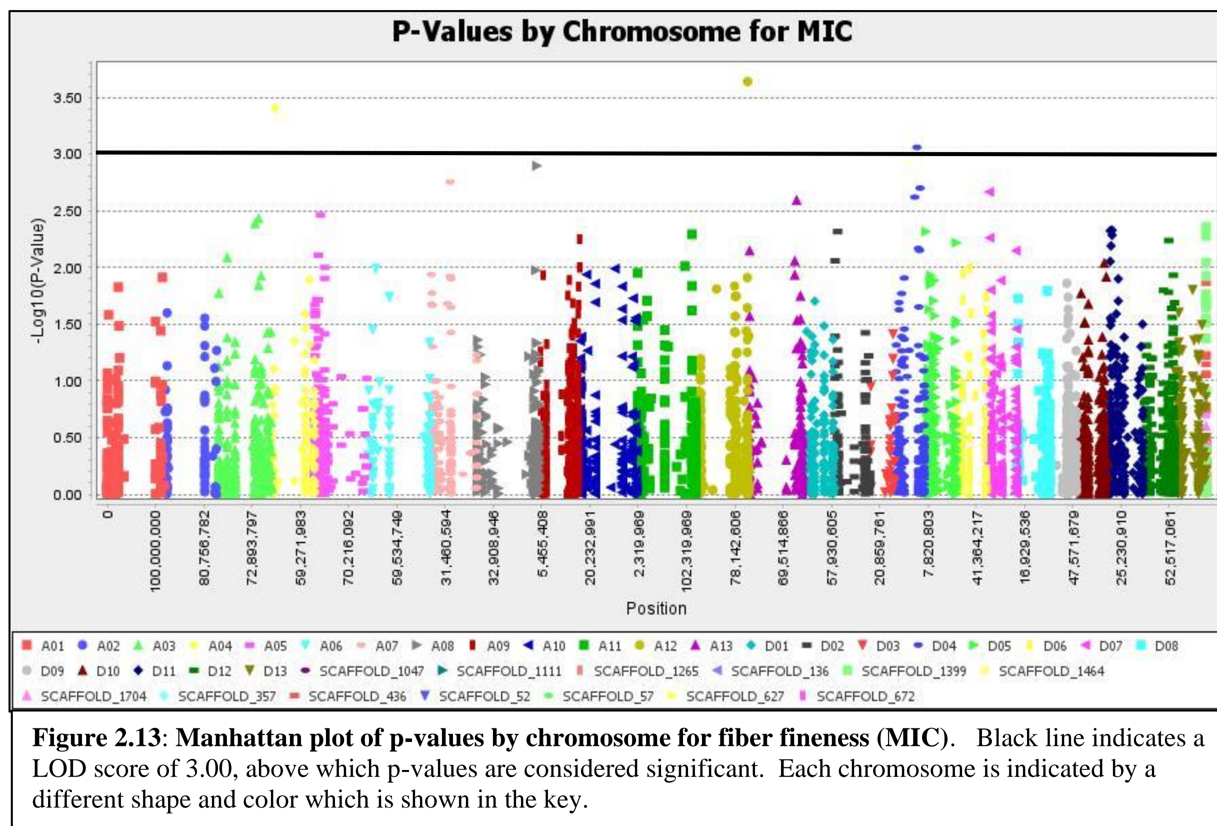


**Figure 2.13**: **Manhattan plot of p-values by chromosome for fiber fineness** (**MIC**). Black line indicates a LOD score of 3.00, above which p-values are considered significant. Each chromosome is indicated by a different shape and color which is shown in the key.

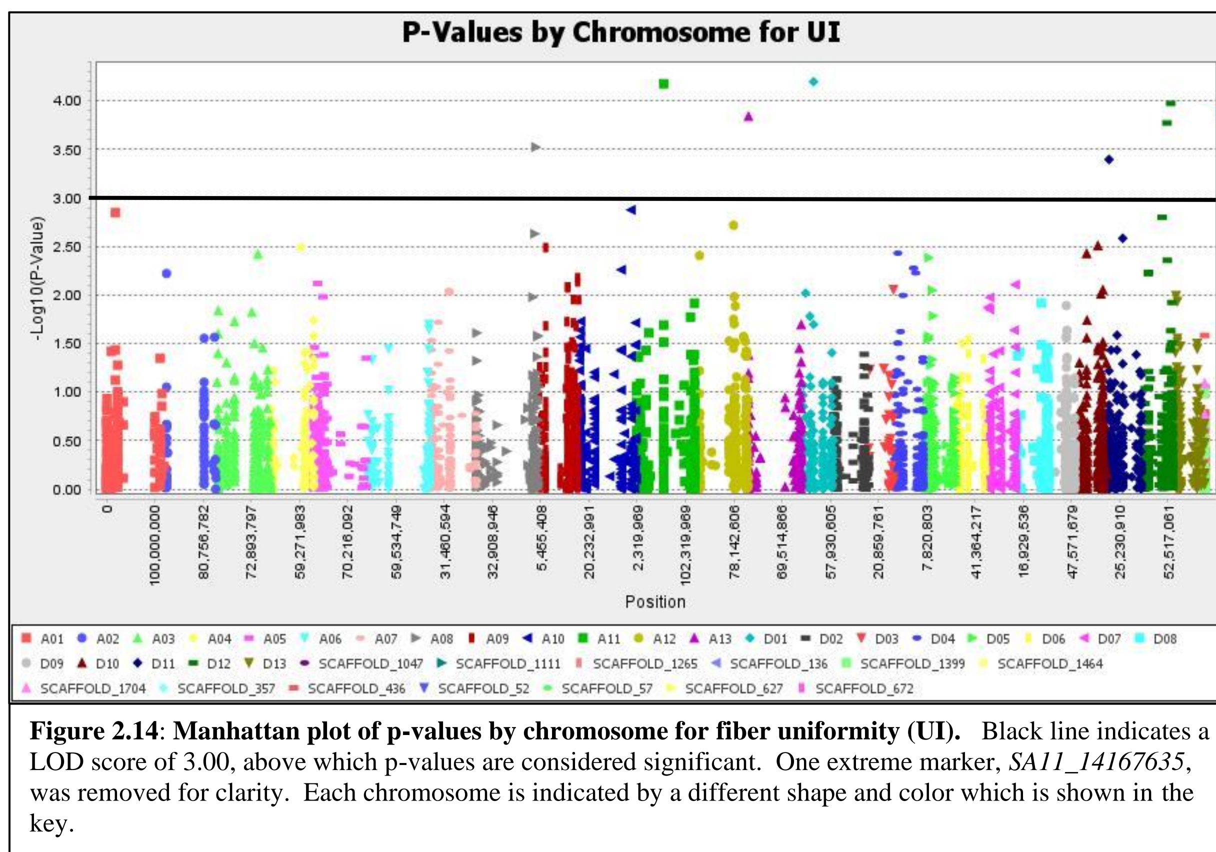


**Figure 2.14**: **Manhattan plot of p-values by chromosome for fiber uniformity (UI).** Black line indicates a LOD score of 3.00, above which p-values are considered significant. One extreme marker, *SA11_14167635*, was removed for clarity. Each chromosome is indicated by a different shape and color which is shown in the key.

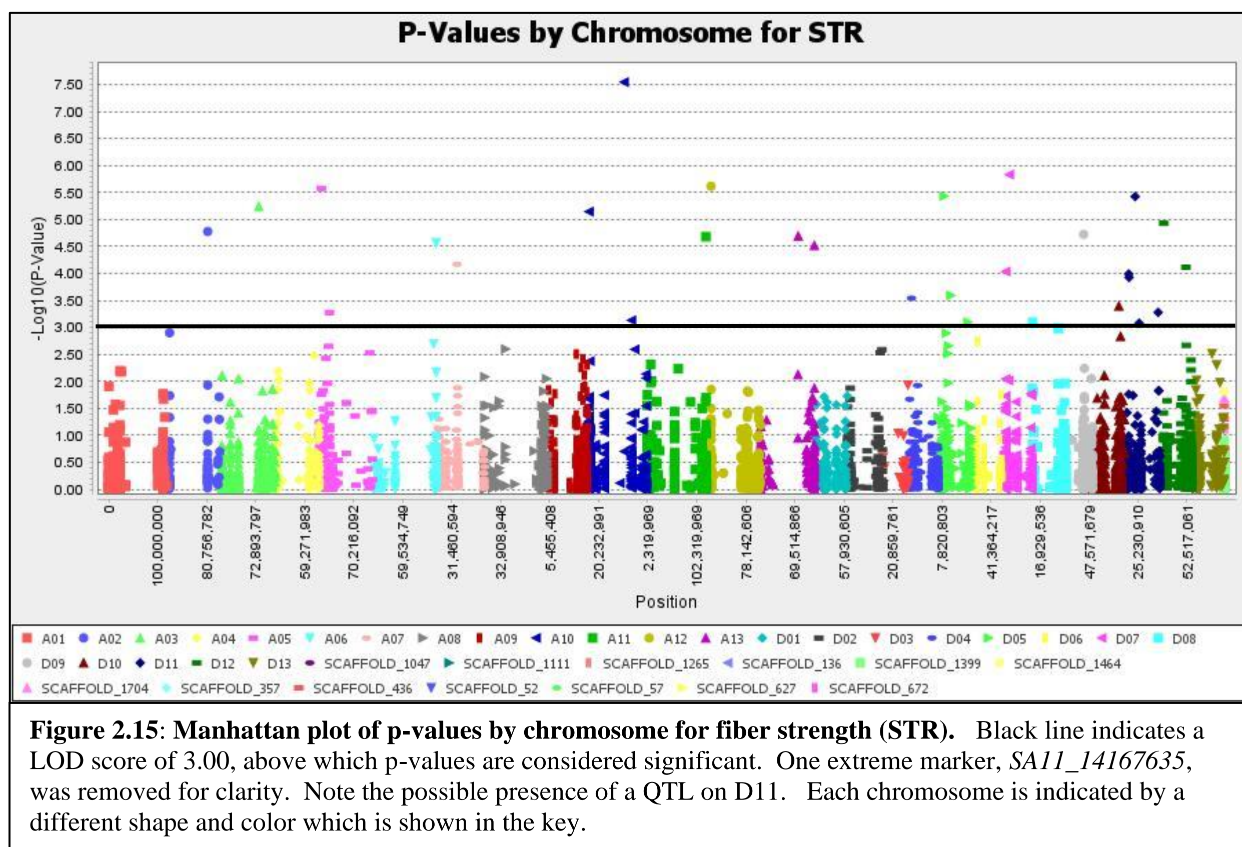


**Figure 2.15**: **Manhattan plot of p-values by chromosome for fiber strength (STR).** Black line indicates a LOD score of 3.00, above which p-values are considered significant. One extreme marker, *SA11_14167635*, was removed for clarity. Note the possible presence of a QTL on D11. Each chromosome is indicated by a different shape and color which is shown in the key.

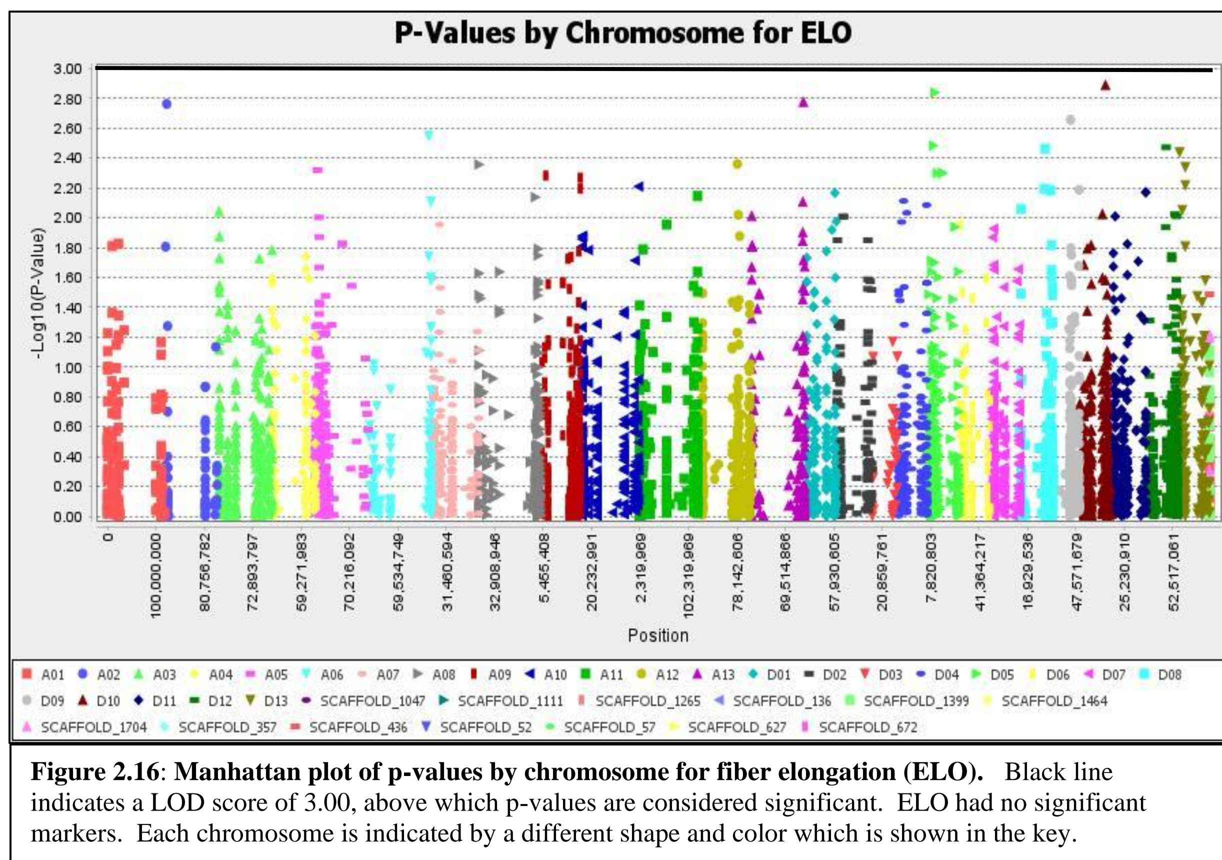


**Figure 2.16**: **Manhattan plot of p-values by chromosome for fiber elongation (ELO).** Black line indicates a LOD score of 3.00, above which p-values are considered significant. ELO had no significant markers. Each chromosome is indicated by a different shape and color which is shown in the key.

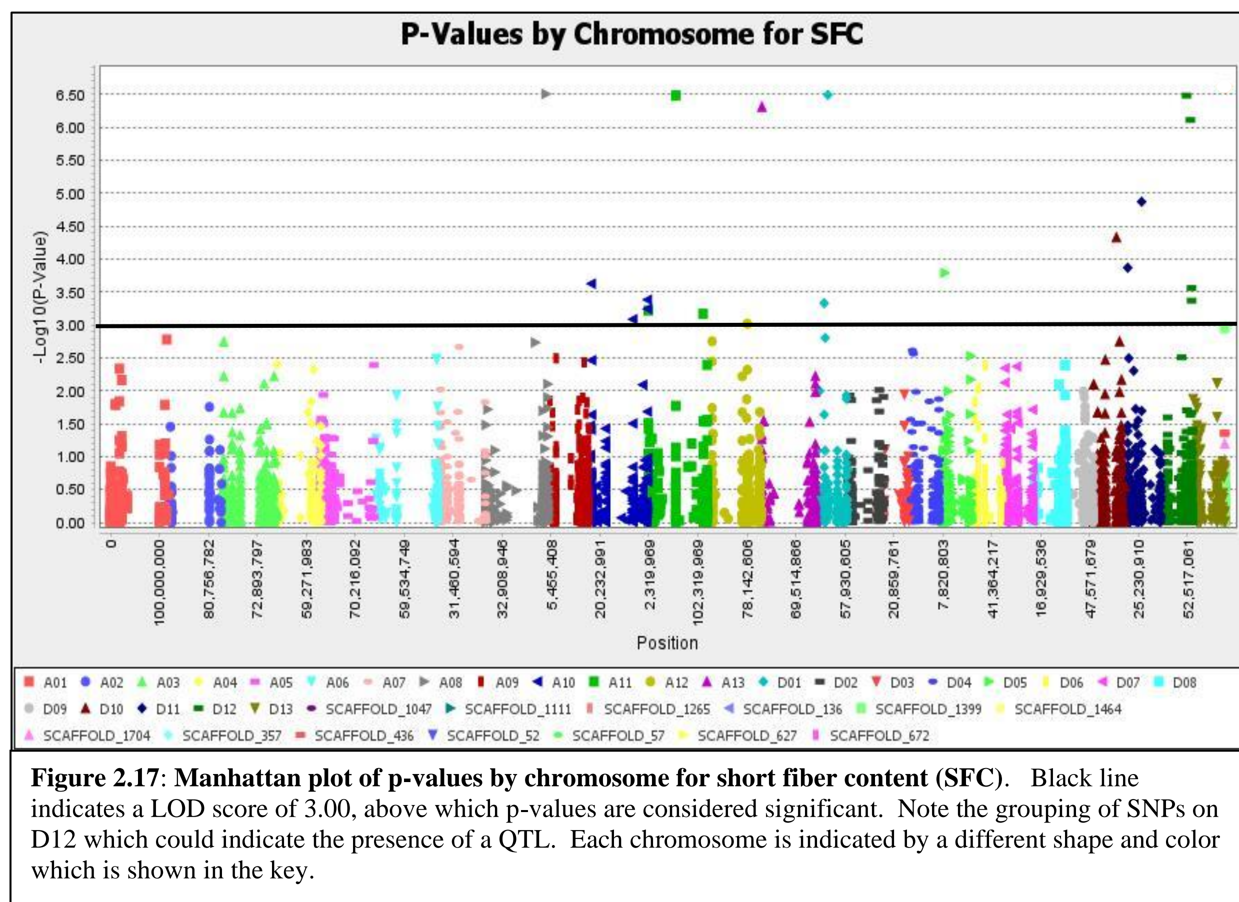


**Figure 2.17**: **Manhattan plot of p-values by chromosome for short fiber content (SFC)**. Black line indicates a LOD score of 3.00, above which p-values are considered significant. Note the grouping of SNPs on D12 which could indicate the presence of a QTL. Each chromosome is indicated by a different shape and color which is shown in the key.

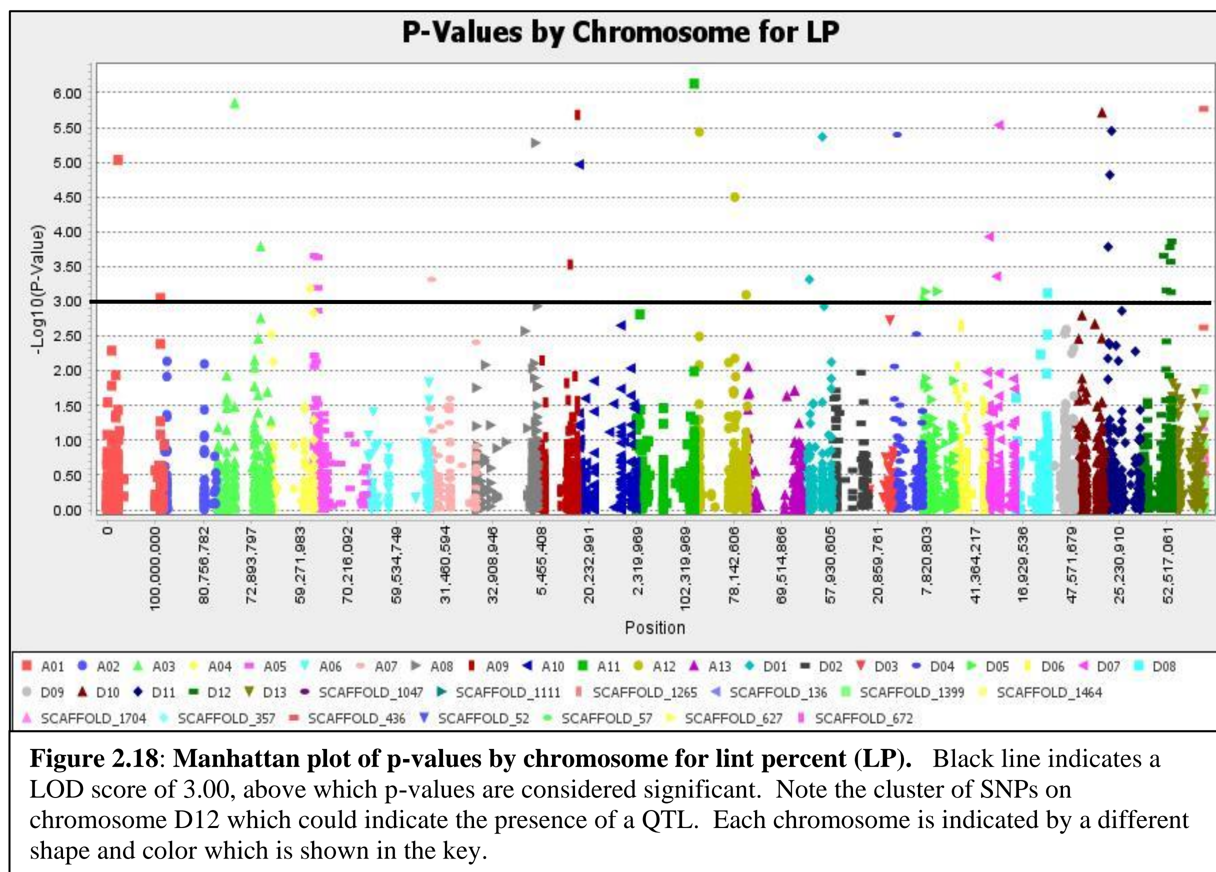


**Figure 2.18**: **Manhattan plot of p-values by chromosome for lint percent (LP).** Black line indicates a LOD score of 3.00, above which p-values are considered significant. Note the cluster of SNPs on chromosome D12 which could indicate the presence of a QTL. Each chromosome is indicated by a different shape and color which is shown in the key.

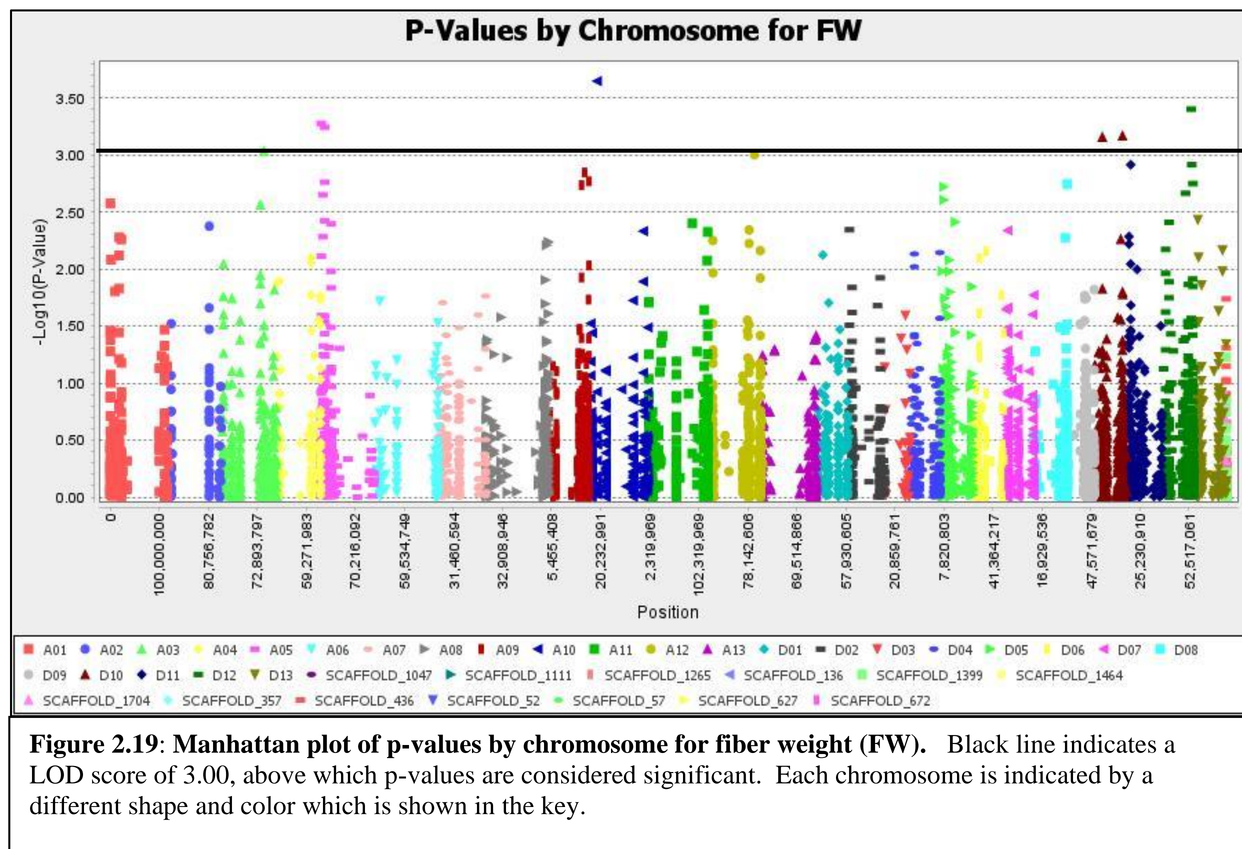


**Figure 2.19**: **Manhattan plot of p-values by chromosome for fiber weight (FW).** Black line indicates a LOD score of 3.00, above which p-values are considered significant. Each chromosome is indicated by a different shape and color which is shown in the key.

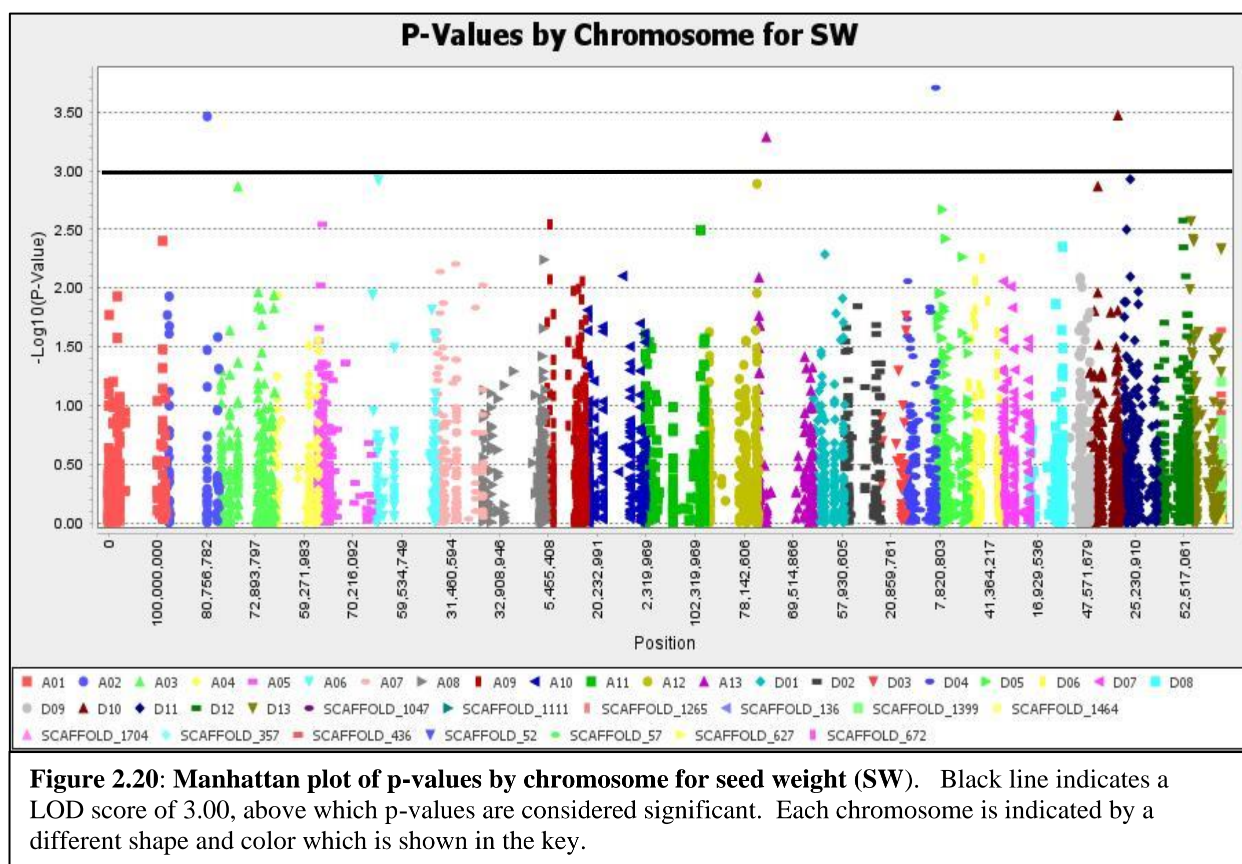


**Figure 2.20**: **Manhattan plot of p-values by chromosome for seed weight (SW)**. Black line indicates a LOD score of 3.00, above which p-values are considered significant. Each chromosome is indicated by a different shape and color which is shown in the key.

**Table 2.3: Parental heterozygosity.** A list showing the heterozygosity percentages for all called SNPs and for all significant SNPs for all parents used.

| Parent | **T368(b)** | **T171(1)** | **T326** | **T347** | **T463** | **T246** |
|---|---|---|---|---|---|---|
| Percent heterozygous of all SNPs | 0.80 | 0.95 | 0.78 | 1.03 | 2.24 | 1.22 |
| Percent heterozygous of significant SNPs | 10.58 | 9.62 | 0.96 | 8.65 | 9.62 | 6.73 |
| | | | | | | |
| Parent | **AM2** | **DES56** | **T45** | **T165** | **Mdn257(1)** | **T63** |
| Percent heterozygous of all SNPs | 0.88 | 1.79 | 0.86 | 0.96 | 1.05 | 1.35 |
| Percent heterozygous of significant SNPs | 3.85 | 4.81 | 2.88 | 7.69 | 4.81 | 3.85 |
| | | | | | | |
| Parent | **Dn165** | **T281** | **T18(1)** | **T1046** | **DN117(2)** | |
| Percent heterozygous of all SNPs | 1.94 | 0.32 | 0.77 | 3.14 | 1.50 | |
| Percent heterozygous of significant SNPs | 8.65 | 1.92 | 4.81 | 9.62 | 11.54 | |

**Supplementary Table 2.4: Phenotypic effects of all SNPs and alleles.** This table lists the percent of phenotypic variation explation (%VE) by each SNP for each trait. It also lists the relative effect of each genotype seen at each SNP. The genotype effects are relative to whichever genotype effect is set to 0, and the value of the effect is the difference in phenotypic value for individuals that have that genotype. The genotype codes are as follows: A=A:A, C=C:C, G=G:G, T=T:T, R=A:G, Y=C:T, S=C:G, W=A:T, K=G:T, M=A:C, + =insertion, and - = a deletion.

| Trait | SNP Marker | %VE | Genotype | Genotype Effect |
|---|---|---|---|---|
| FW | SA03_86979875 | 4.0134 | T | -5.20280649 |
| FW | SA03_86979875 | | Y | 0 |
| FW | SA05_901950 | 5.7253 | A | 24.87101723 |
| FW | SA05_901950 | | G | 0.854945709 |
| FW | SA05_901950 | | R | 0 |
| FW | SA05_8155555 | 4.9444 | A | 1.208212445 |
| FW | SA05_8155555 | | M | 24.47692026 |
| FW | SA05_8155555 | | W | 0 |
| FW | SA10_13383304 | 9.2448 | C | -1.363900555 |
| FW | SA10_13383304 | | T | -9.279435923 |
| FW | SA10_13383304 | | Y | 0 |
| FW | SA12_92185502 | 4.3244 | A | -7.580019917 |
| FW | SA12_92185502 | | M | 0 |
| FW | SD10_16414440 | 5.1975 | C | 6.17006104 |
| FW | SD10_16414440 | | T | 22.48686129 |
| FW | SD10_16414440 | | Y | 0 |
| FW | SD10_57514407 | 3.8153 | T | -5.210894105 |
| FW | SD10_57514407 | | Y | 0 |
| FW | SD12_56788408 | 6.7814 | A | 17.44703648 |
| FW | SD12_56788408 | | G | 2.557013633 |
| FW | SD12_56788408 | | R | 0 |
| SW | SA02_83249790 | 5.0245 | T | -81.07677815 |
| SW | SA02_83249790 | | Y | 0 |
| SW | SA13_15083177 | 11.679 | G | 63.78483125 |
| SW | SA13_15083177 | | T | 0 |
| SW | SD04_57862825 | 8.4267 | C | 64.35286652 |
| SW | SD04_57862825 | | T | -75.7508613 |
| SW | SD04_57862825 | | Y | 0 |
| SW | SD10_57506300 | 5.1302 | T | -55.76412698 |
| SW | SD10_57506300 | | Y | 0 |
| MIC | SA04_7831210 | 5.3534 | A | 0.491621852 |
| MIC | SA04_7831210 | | C | 0.011857185 |
| MIC | SA04_7831210 | | M | 0 |
| MIC | SA12_104595544 | 5.7361 | A | -0.626201519 |
| MIC | SA12_104595544 | | R | 0 |

**Supplementary Table 2.4: Phenotypic effects of all SNPs and alleles (cont.)**

| Trait | SNP Marker | %VE | Genotype | Genotype Effect |
|---|---|---|---|---|
| MIC | SD04_42983935 | 6.9223 | A | -0.760955471 |
| MIC | SD04_42983935 | | C | -1.107268379 |
| MIC | SD04_42983935 | | M | 0 |
| UHM | SA07_42400737 | 7.0122 | A | 0.122153066 |
| UHM | SA07_42400737 | | G | 0.06475601 |
| UHM | SA07_42400737 | | R | 0 |
| UHM | SD04_38484354 | 4.4784 | A | -0.006596668 |
| UHM | SD04_38484354 | | G | 0.06995362 |
| UHM | SD04_38484354 | | R | 0 |
| UHM | SD11_4148224 | 5.5353 | C | 0.019373872 |
| UHM | SD11_4148224 | | T | 0.076578701 |
| UHM | SD11_4148224 | | Y | 0 |
| UI | SA08_122918100 | 7.6699 | C | -7.098108608 |
| UI | SA08_122918100 | | T | 0.207303375 |
| UI | SA08_122918100 | | Y | 0 |
| UI | SA11_57094984 | 9.0366 | A | -0.788136197 |
| UI | SA11_57094984 | | G | -8.334158856 |
| UI | SA11_57094984 | | R | 0 |
| UI | SA13_90113 | 9.4526 | C | -7.72133319 |
| UI | SA13_90113 | | T | -0.138798119 |
| UI | SA13_90113 | | Y | 0 |
| UI | SD01_23337801 | 11.213 | C | -7.84313121 |
| UI | SD01_23337801 | | T | 0.397275202 |
| UI | SD01_23337801 | | Y | 0 |
| UI | SD11_2869210 | 6.7185 | A | 1.994185441 |
| UI | SD11_2869210 | | G | 3.413229299 |
| UI | SD11_2869210 | | R | 0 |
| UI | SD12_50099795 | 7.4175 | C | -7.52663521 |
| UI | SD12_50099795 | | T | -0.022587695 |
| UI | SD12_50099795 | | Y | 0 |
| UI | SD12_58107256 | 12.06 | A | 0.679284944 |
| UI | SD12_58107256 | | T | -7.456232455 |
| UI | SD12_58107256 | | W | 0 |
| STR | SA02_83249847 | 9.3854 | A | 0.796921274 |
| STR | SA02_83249847 | | G | 13.65317757 |
| STR | SA02_83249847 | | - | -1.332725188 |
| STR | SA02_83249847 | | R | 1.008384519 |
| STR | SA02_83249847 | | 0 | 0 |
| STR | SA03_79502155 | 8.0209 | A | -12.41112965 |
| STR | SA03_79502155 | | R | -12.78652265 |
| STR | SA03_79502155 | | W | 0 |

**Supplementary Table 2.4: Phenotypic effects of all SNPs and alleles (cont.)**

| Trait | SNP Marker | %VE | Genotype | Genotype Effect |
|---|---|---|---|---|
| STR | SA05_4284946 | 10.028 | A | -0.290417025 |
| STR | SA05_4284946 | | G | 12.92271666 |
| STR | SA05_4284946 | | R | 0 |
| STR | SA05_20746005 | 5.1874 | A | -3.74394292 |
| STR | SA05_20746005 | | T | -2.264976917 |
| STR | SA05_20746005 | | W | 0 |
| STR | SA06_128074102 | 8.9477 | A | 12.42696535 |
| STR | SA06_128074102 | | T | -0.429483337 |
| STR | SA06_128074102 | | W | -0.211974836 |
| STR | SA06_128074102 | | Y | 0 |
| STR | SA07_42400737 | 9.8643 | A | 5.867639017 |
| STR | SA07_42400737 | | G | 4.644610406 |
| STR | SA07_42400737 | | R | 0 |
| STR | SA10_254377 | 7.9718 | C | -0.087095447 |
| STR | SA10_254377 | | T | 12.34505108 |
| STR | SA10_254377 | | Y | 0 |
| STR | SA10_72077971 | 11.022 | C | -15.11921041 |
| STR | SA10_72077971 | | G | -12.74686314 |
| STR | SA10_72077971 | | S | 0 |
| STR | SA10_86862262 | 5.4053 | C | 6.93805821 |
| STR | SA10_86862262 | | T | 0.403779571 |
| STR | SA10_86862262 | | Y | 0 |
| STR | SA11_120873612 | 7.5463 | A | 11.03810012 |
| STR | SA11_120873612 | | G | -0.747486663 |
| STR | SA11_120873612 | | R | 0 |
| STR | SA12_7162372 | 9.6555 | A | -0.069541284 |
| STR | SA12_7162372 | | T | 12.94210055 |
| STR | SA12_7162372 | | W | 0 |
| STR | SA13_76462449 | 7.611 | A | -10.08569059 |
| STR | SA13_76462449 | | C | -7.541219053 |
| STR | SA13_76462449 | | M | 0 |
| STR | SA13_109335847 | 7.7004 | C | 11.64732632 |
| STR | SA13_109335847 | | T | -0.315042528 |
| STR | SA13_109335847 | | Y | 0 |
| STR | SD04_3856727 | 5.0821 | A | -5.198354375 |
| STR | SD04_3856727 | | G | -1.890212727 |
| STR | SD04_3856727 | | R | 0 |
| STR | SD05_12607139 | 9.2516 | A | 13.0509247 |
| STR | SD05_12607139 | | C | 0.160466013 |
| STR | SD05_12607139 | | M | 0 |
| STR | SD05_26301075 | 5.3491 | A | -1.67215326 |

**Supplementary Table 2.4: Phenotypic effects of all SNPs and alleles (cont.)**

| Trait | SNP Marker | %VE | Genotype | Genotype Effect |
|---|---|---|---|---|
| STR | SD05_26301075 | | G | -5.332755156 |
| STR | SD05_26301075 | | R | 0 |
| STR | SD07_5193330 | 6.8161 | A | 11.44707236 |
| STR | SD07_5193330 | | G | 0.381956916 |
| STR | SD07_5193330 | | R | 0 |
| STR | SD07_12801944 | 8.7801 | C | -14.48803347 |
| STR | SD07_12801944 | | T | -12.5521386 |
| STR | SD07_12801944 | | Y | 0 |
| STR | SD08_1998600 | 4.5818 | C | 4.797261312 |
| STR | SD08_1998600 | | T | -1.50893899 |
| STR | SD08_1998600 | | Y | 0 |
| STR | SD09_36693375 | 9.0471 | A | 0.625853789 |
| STR | SD09_36693375 | | G | 12.71455253 |
| STR | SD09_36693375 | | R | 0 |
| STR | SD10_54031653 | 4.2698 | T | -3.338281031 |
| STR | SD10_54031653 | | K | 0 |
| STR | SD11_6282948 | 9.0791 | A | 13.56489818 |
| STR | SD11_6282948 | | G | 1.389842877 |
| STR | SD11_6282948 | | T | 1.015981002 |
| STR | SD11_6282948 | | R | 0.689806539 |
| STR | SD11_6282948 | | K | 0 |
| STR | SD11_6943459 | 7.5918 | C | 11.04029647 |
| STR | SD11_6943459 | | T | 0.66366639 |
| STR | SD11_6943459 | | Y | -0.201384078 |
| STR | SD11_6943459 | | 0 | 0 |
| STR | SD11_19403607 | 8.0834 | C | -12.58741777 |
| STR | SD11_19403607 | | T | -11.98613718 |
| STR | SD11_19403607 | | Y | 0 |
| STR | SD11_66257876 | 4.8165 | C | 5.79874869 |
| STR | SD11_66257876 | | T | 6.457952542 |
| STR | SD11_66257876 | | Y | 0 |
| STR | SD12_4848862 | 10.109 | A | -0.050509988 |
| STR | SD12_4848862 | | G | 12.63619969 |
| STR | SD12_4848862 | | R | 0 |
| STR | SD12_50099837 | 6.7412 | A | 0.029976351 |
| STR | SD12_50099837 | | G | 6.269649851 |
| STR | SD12_50099837 | | R | 0 |
| SFC | SA08_122918100 | 14.58 | C | 6.86045688 |
| SFC | SA08_122918100 | | T | 0.42916945 |
| SFC | SA08_122918100 | | Y | 0 |
| SFC | SA10_2214083 | 5.7626 | T | -1.771254745 |

**Supplementary Table 2.4: Phenotypic effects of all SNPs and alleles (cont.)**

| Trait | SNP Marker | %VE | Genotype | Genotype Effect |
|---|---|---|---|---|
| SFC | SA10_2214083 | | K | 0 |
| SFC | SA10_86862295 | 6.9277 | C | -1.35918801 |
| SFC | SA10_86862295 | | T | -2.87801981 |
| SFC | SA10_86862295 | | Y | 0 |
| SFC | SA10_116893612 | 5.1681 | A | -1.548061619 |
| SFC | SA10_116893612 | | R | 0 |
| SFC | SA10_116893638 | 4.9295 | T | -1.555252434 |
| SFC | SA10_116893638 | | Y | 0 |
| SFC | SA11_1109913 | 9.5087 | C | -0.162141531 |
| SFC | SA11_1109913 | | T | 0.669510427 |
| SFC | SA11_1109913 | | W | 2.10542928 |
| SFC | SA11_1109913 | | Y | 0 |
| SFC | SA11_57094984 | 16.181 | A | 0.089152348 |
| SFC | SA11_57094984 | | G | 6.592475433 |
| SFC | SA11_57094984 | | R | 0 |
| SFC | SA11_112523478 | 17.857 | C | 4.318460686 |
| SFC | SA11_112523478 | | T | 0.468002826 |
| SFC | SA11_112523478 | | Y | 0 |
| SFC | SA12_79044294 | 8.5611 | A | 3.778228925 |
| SFC | SA12_79044294 | | C | 1.338984432 |
| SFC | SA12_79044294 | | T | -0.552627747 |
| SFC | SA12_79044294 | | W | -0.925193105 |
| SFC | SA12_79044294 | | Y | 0 |
| SFC | SA13_90113 | 18.447 | C | 6.28298612 |
| SFC | SA13_90113 | | T | -0.102057543 |
| SFC | SA13_90113 | | Y | 0 |
| SFC | SD01_15768704 | 7.3074 | G | -1.555980778 |
| SFC | SD01_15768704 | | T | -1.84489858 |
| SFC | SD01_15768704 | | K | 0 |
| SFC | SD01_23337801 | 21.405 | C | 6.810993949 |
| SFC | SD01_23337801 | | T | 0.057844355 |
| SFC | SD01_23337801 | | Y | 0 |
| SFC | SD05_12607083 | 7.9834 | A | 0.059182543 |
| SFC | SD05_12607083 | | C | 3.461876802 |
| SFC | SD05_12607083 | | M | 0 |
| SFC | SD10_47645825 | 11.846 | A | -3.55390768 |
| SFC | SD10_47645825 | | C | -2.146119403 |
| SFC | SD10_47645825 | | M | -3.28597551 |
| SFC | SD10_47645825 | | W | 0 |
| SFC | SD11_2869210 | 8.6002 | A | -1.383063359 |
| SFC | SD11_2869210 | | G | -2.126379747 |

**Supplementary Table 2.4: Phenotypic effects of all SNPs and alleles (cont.)**

| Trait | SNP Marker | %VE | Genotype | Genotype Effect |
|---|---|---|---|---|
| SFC | SD11_2869210 | | R | 0 |
| SFC | SD11_31298915 | 14.313 | C | -0.719366831 |
| SFC | SD11_31298915 | | T | -3.539698093 |
| SFC | SD11_31298915 | | M | 0.54179077 |
| SFC | SD11_31298915 | | W | -1.52580787 |
| SFC | SD11_31298915 | | Y | 0 |
| SFC | SD12_50099795 | 14.689 | C | 6.221165782 |
| SFC | SD12_50099795 | | T | -0.227955907 |
| SFC | SD12_50099795 | | Y | 0 |
| SFC | SD12_58107256 | 18.807 | A | -0.620552879 |
| SFC | SD12_58107256 | | T | 5.987148797 |
| SFC | SD12_58107256 | | W | 0 |
| SFC | SD12_60563503 | 6.4941 | A | -1.276147931 |
| SFC | SD12_60563503 | | W | 0 |
| SFC | SD12_60658786 | 7.1835 | A | -0.55301552 |
| SFC | SD12_60658786 | | G | 3.734371885 |
| SFC | SD12_60658786 | | R | 0 |
| LP | SA01_23151361 | 5.9992 | C | 0.345339566 |
| LP | SA01_23151361 | | T | 0.013360029 |
| LP | SA01_23151361 | | Y | 0 |
| LP | SA01_111579988 | 2.8887 | T | -0.068625494 |
| LP | SA01_111579988 | | Y | 0 |
| LP | SA03_38202138 | 6.4289 | A | -0.033685532 |
| LP | SA03_38202138 | | G | 0.320533324 |
| LP | SA03_38202138 | | R | -0.026881166 |
| LP | SA03_38202138 | | W | 0 |
| LP | SA03_92087644 | 5.8432 | A | 0.011097781 |
| LP | SA03_92087644 | | T | 0.0364735 |
| LP | SA03_92087644 | | M | 0.388609782 |
| LP | SA03_92087644 | | W | 0.040134604 |
| LP | SA03_92087644 | | S | 0.037846566 |
| LP | SA03_92087644 | | Y | 0.028302121 |
| LP | SA03_92087644 | | K | 0 |
| LP | SA04_81021342 | 3.0043 | G | -9.86E-04 |
| LP | SA04_81021342 | | S | 0.158667489 |
| LP | SA04_81021342 | | K | 0 |
| LP | SA05_901950 | 3.9992 | A | 0.285009707 |
| LP | SA05_901950 | | G | 0.0061456 |
| LP | SA05_901950 | | R | 0 |
| LP | SA05_8155555 | 3.2747 | A | 0.013930434 |
| LP | SA05_8155555 | | M | 0.282241976 |

**Supplementary Table 2.4: Phenotypic effects of all SNPs and alleles (cont.)**

| Trait | SNP Marker | %VE | Genotype | Genotype Effect |
|---|---|---|---|---|
| LP | SA05_8155555 | | W | 0 |
| LP | SA05_8155557 | 3.3658 | A | 0.001510022 |
| LP | SA05_8155557 | | T | -0.076565618 |
| LP | SA05_8155557 | | M | 0.270567617 |
| LP | SA05_8155557 | | W | 0 |
| LP | SA07_5551984 | 3.5612 | A | -0.007793504 |
| LP | SA07_5551984 | | G | 0.273072543 |
| LP | SA07_5551984 | | R | 0 |
| LP | SA08_122918080 | 6.8882 | C | 0.320489514 |
| LP | SA08_122918080 | | T | -0.024802877 |
| LP | SA08_122918080 | | Y | 0 |
| LP | SA09_66750666 | 4.3095 | C | 0.265566915 |
| LP | SA09_66750666 | | T | -0.010234518 |
| LP | SA09_66750666 | | Y | 0 |
| LP | SA09_82364060 | 6.9022 | A | 0.368853387 |
| LP | SA09_82364060 | | T | 0.011721724 |
| LP | SA09_82364060 | | W | 0 |
| LP | SA10_254461 | 5.3832 | A | 0.372288964 |
| LP | SA10_254461 | | T | 0.033769339 |
| LP | SA10_254461 | | W | 0.012821164 |
| LP | SA10_254461 | | Y | 0 |
| LP | SA11_120873572 | 6.6444 | A | -0.036907801 |
| LP | SA11_120873572 | | G | 0.309049519 |
| LP | SA11_120873572 | | R | 0 |
| LP | SA12_7564233 | 5.5028 | A | -0.00760859 |
| LP | SA12_7564233 | | C | 0.34066413 |
| LP | SA12_7564233 | | M | 0 |
| LP | SA12_81233850 | 5.3728 | A | 0.324061949 |
| LP | SA12_81233850 | | C | -0.087575425 |
| LP | SA12_81233850 | | G | -0.013693973 |
| LP | SA12_81233850 | | R | 0.011742858 |
| LP | SA12_81233850 | | S | 0 |
| LP | SA12_104595550 | 3.6011 | A | -0.088886213 |
| LP | SA12_104595550 | | G | -0.10566467 |
| LP | SA12_104595550 | | R | 0 |
| LP | SD01_15768689 | 4.9368 | A | -0.003270831 |
| LP | SD01_15768689 | | C | 0.184179612 |
| LP | SD01_15768689 | | T | -0.010252456 |
| LP | SD01_15768689 | | M | 0.011514222 |
| LP | SD01_15768689 | | W | 4.17E-04 |
| LP | SD01_15768689 | | Y | 0 |

**Supplementary Table 2.4: Phenotypic effects of all SNPs and alleles (cont.)**

| Trait | SNP Marker | %VE | Genotype | Genotype Effect |
|---|---|---|---|---|
| LP | SD01_42750508 | 5.092 | C | 0.362772805 |
| LP | SD01_42750508 | | G | 0.02535812 |
| LP | SD01_42750508 | | S | 0 |
| LP | SD04_6012369 | 5.9238 | C | 0.345821826 |
| LP | SD04_6012369 | | T | -0.004445156 |
| LP | SD04_6012369 | | Y | 0 |
| LP | SD05_4702874 | 4.1204 | C | 0.161208686 |
| LP | SD05_4702874 | | T | -0.027419717 |
| LP | SD05_4702874 | | Y | 0 |
| LP | SD05_7014280 | 2.1262 | A | 0.050955009 |
| LP | SD05_7014280 | | R | 0 |
| LP | SD05_31739215 | 4.3152 | G | 0.213864278 |
| LP | SD05_31739215 | | T | 0.025378223 |
| LP | SD05_31739215 | | K | 0 |
| LP | SD07_5846236 | 4.4435 | G | -0.003626068 |
| LP | SD07_5846236 | | T | 0.167572117 |
| LP | SD07_5846236 | | K | 0 |
| LP | SD07_20984540 | 2.557 | A | 0.091070038 |
| LP | SD07_20984540 | | C | 0 |
| LP | SD07_26289135 | 5.7357 | A | -0.013842099 |
| LP | SD07_26289135 | | G | 0.332643725 |
| LP | SD07_26289135 | | R | 0 |
| LP | SD08_68897859 | 4.2168 | C | 0.182162396 |
| LP | SD08_68897859 | | T | -0.009687924 |
| LP | SD08_68897859 | | Y | 0 |
| LP | SD10_58019010 | 5.6906 | A | 0.031960873 |
| LP | SD10_58019010 | | G | 0.381023948 |
| LP | SD10_58019010 | | R | 0 |
| LP | SD11_2869179 | 3.5697 | C | 0.009431064 |
| LP | SD11_2869179 | | G | 0.215833118 |
| LP | SD11_2869179 | | S | 0 |
| LP | SD11_6282973 | 5.2663 | A | -0.050680026 |
| LP | SD11_6282973 | | G | 0.102488749 |
| LP | SD11_6282973 | | R | 0 |
| LP | SD11_9980089 | 7.5804 | C | 0.371703497 |
| LP | SD11_9980089 | | T | 0.009059323 |
| LP | SD11_9980089 | | Y | 0 |
| LP | SD12_45219434 | 5.2056 | C | 0.299954713 |
| LP | SD12_45219434 | | T | 0.011759911 |
| LP | SD12_45219434 | | Y | 0 |
| LP | SD12_50746497 | 3.1471 | A | -0.025814805 |

**Supplementary Table 2.4: Phenotypic effects of all SNPs and alleles (cont.)**

| Trait | SNP Marker | %VE | Genotype | Genotype Effect |
|---|---|---|---|---|
| LP | SD12_50746497 | | G | 0.220473621 |
| LP | SD12_50746497 | | R | 0 |
| LP | SD12_58107221 | 6.4429 | A | -0.032110764 |
| LP | SD12_58107221 | | T | 0.246469157 |
| LP | SD12_58107221 | | W | 0 |
| LP | SD12_60070368 | 3.6281 | C | 0.199656957 |
| LP | SD12_60070368 | | T | 0.001413645 |
| LP | SD12_60070368 | | Y | 0 |
| LP | SD12_60658745 | 3.057 | C | 0.269004841 |
| LP | SD12_60658745 | | T | -0.003795243 |
| LP | SD12_60658745 | | Y | 0 |
| LP | SD12_61811284 | 3.4761 | C | -0.10926617 |
| LP | SD12_61811284 | | M | -0.113230348 |
| LP | SD12_61811284 | | S | 0 |

**References**

Abbas A., Iqbal M.A., Rahman M.U., Paterson A.H. (2015) Estimating genetic diversity among selected cotton genotypes and the identification of DNA markers associated with resistance to cotton leaf curl disease. Turkish Journal of Botany 39:1033-1041. DOI: 10.3906/bot-1505-22.

Abdurakhmonov I.Y., Kohel R.J., Yu J.Z., Pepper A.E., Abdullaev A.A., Kushanov F.N., Salakhutdinov I.B., Buriev Z.T., Saha S., Scheffler B.E., Jenkins J.N., Abdukarimov A. (2008) Molecular diversity and association mapping of fiber quality traits in exotic G. hirsutum L. germplasm. Genomics 92:478-487. DOI: 10.1016/j.ygeno.2008.07.013.

Adhikari J., Das S., Wang Z.N., Khanal S., Chandnani R., Patel J.D., Goff V.H., Auckland S., Rainville L.K., Jones D., Paterson A.H. (2017) Targeted identification of association between cotton fiber quality traits and microsatellite markers. Euphytica 213. DOI: 10.1007/s10681-017-1853-0.

Ali M.J., Xing G.N., He J.B., Zhao T.J., Gai J.Y. (2020) Detecting the QTL-allele system controlling seed-flooding tolerance in a nested association mapping population of soybean. Crop Journal 8:781-792. DOI: 10.1016/j.cj.2020.06.008.

Bajgain P., Sallam A.H., Annor G., Conley E., Steffenson B.J., Muehlbauer G.J., Anderson J.A. (2021) Genetic characterization of flour quality and bread-making traits in a spring wheat nested association mapping population. Crop Science 61:1168-1183. DOI: 10.1002/csc2.20432.

Bradbury P.J., Zhang Z., Kroon D.E., Casstevens T.M., Ramdoss Y., Buckler E.S. (2007) TASSEL: software for association mapping of complex traits in diverse samples. Bioinformatics 23:2633-5. DOI: 10.1093/bioinformatics/btm308.

Brock M.T., Rubin M.J., DellaPenna D., Weinig C. (2020) A Nested Association Mapping Panel inArabidopsis thalianafor Mapping and Characterizing Genetic Architecture. G3-Genes Genomes Genetics 10:3701-3708. DOI: 10.1534/g3.120.401239.

Brown N., Kumar P., Singh R., Lubbers E., Campbell B.T., Myers G.O., Wright R.J., Subramani J., Jones D., Paterson A.H., Chee P.W. (2019) Evaluation of a Chromosome Segment from Gossypium barbadense Harboring the Fiber Length QTL qFL-Chr.25 in Four Diverse Upland Cotton Genetic Backgrounds. Crop Science 59:2621-2633. DOI: 10.2135/cropsci2019.05.0321.

Bu S.H., Wu W.R., Zhang Y.M. (2021) A Multi-Locus Association Model Framework for Nested Association Mapping With Discriminating QTL Effects in Various Subpopulations. Frontiers in Genetics 11. DOI: 10.3389/fgene.2020.590012.

Buttner B., Draba V., Pillen K., Schweizer G., Maurer A. (2020) Identification of QTLs conferring resistance to scald (Rhynchosporium commune) in the barley nested association mapping population HEB-25. Bmc Genomics 21. DOI: 10.1186/s12864-020-07258-7.

Chee P., Draye X., Jiang C.X., Decanini L., Delmonte T.A., Bredhauer R., Smith C.W., Paterson A.H. (2005) Molecular dissection of interspecific variation between Gossypium hirsutum and Gossypium barbadense (cotton) by a backcross-self approach: I. Fiber elongation. Theoretical and Applied Genetics 111:757-763. DOI: 10.1007/s00122-005-2063-z.

Chen Z.J., Scheffler B.E., Dennis E., Triplett B.A., Zhang T.Z., Guo W.Z., Chen X.Y., Stelly D.M., Rabinowicz P.D., Town C.D., Arioli T., Brubaker C., Cantrell R.G., Lacape J.M., Ulloa M., Chee P., Gingle A.R., Haigler C.H., Percy R., Saha S., Wilkins T., Wright R.J., Van Deynze A., Zhu Y.X., Yu S.X., Abdurakhmonov I., Katageri I., Kumar P.A.,

Mehboob ur R., Zafar Y., Yu J.Z., Kohel R.J., Wendel J.F., Paterson A.H. (2007) Toward Sequencing cotton (Gossypium) Genomes. Plant Physiology 145:1303-1310. DOI: 10.1104/pp.107.107672.

Chen Z.J., Sreedasyam A., Ando A., Song Q., De Santiago L.M., Hulse-Kemp A.M., Ding M., Ye W., Kirkbride R.C., Jenkins J., Plott C., Lovell J., Lin Y.-M., Vaughn R., Liu B., Simpson S., Scheffler B.E., Wen L., Saski C.A., Grover C.E., Hu G., Conover J.L., Carlson J.W., Shu S., Boston L.B., Williams M., Peterson D.G., McGee K., Jones D.C., Wendel J.F., Stelly D.M., Grimwood J., Schmutz J. (2020) Genomic diversifications of five Gossypium allopolyploid species and their impact on cotton improvement. Nature Genetics 52:525-533. DOI: 10.1038/s41588-020-0614-5.

Cortes A.J., Lopez-Hernandez F. (2021) Harnessing Crop Wild Diversity for Climate Change Adaptation. Genes 12. DOI: 10.3390/genes12050783.

d'Eeckenbrugge G.C., Lacape J.M. (2014) Distribution and Differentiation of Wild, Feral, and Cultivated Populations of Perennial Upland Cotton (Gossypium hirsutum L.) in Mesoamerica and the Caribbean. Plos One 9. DOI: 10.1371/journal.pone.0107458.

de Carvalho L.P., Farias F.J.C., Rodrigues J.I.S., Suassuna N.D., Teodoro P.E. (2017) Genetic diversity among exotic cotton accessions as for qualitative and quantitative traits. Genetics and Molecular Research 16. DOI: 10.4238/gmr16019590.

Feng L.C., Zhang S.W., Xing L.T., Yang B., Gao X., Xie X.M., Zhou B.L. (2019) QTL analysis for yield and fibre quality traits using three sets of introgression lines developed from three Gossypium hirsutum race stocks. Molecular Genetics and Genomics 294:789-810. DOI: 10.1007/s00438-019-01548-w.

Gage J.L., Monier B., Giri A., Buckler E.S. (2020) Ten Years of the Maize Nested Association Mapping Population: Impact, Limitations, and Future Directions. Plant Cell 32:2083-2093. DOI: 10.1105/tpc.19.00951.

Guo W.Z., Cai C.P., Wang C.B., Han Z.G., Song X.L., Wang K., Niu X.W., Wang C., Lu K.Y., Shi B., Zhang T.Z. (2007) A microsatellite-based, gene-rich linkage map reveals genome structure, function and evolution in gossypium. Genetics 176:527-541. DOI: 10.1534/genetics.107.070375.

He S.P., Sun J.L., Zhang C., Du X.M. (2011) Identification of exotic genetic components and DNA methylation pattern analysis of three cotton introgression lines from Gossypium bickii. Molecular Biology 45:204-210. DOI: 10.1134/s002689331102018x.

Jan S.F., Khan M.R., Iqbal A., Khan F.U., Ali S. (2020) Genetic diversity in exotic oat germplasm & resistance against barley yellow dwarf virus. Saudi Journal of Biological Sciences 27:2622-2631. DOI: 10.1016/j.sjbs.2020.05.042.

Mace E.S., Cruickshank A.W., Tao Y.F., Hunt C.H., Jordan D.R. (2021) A global resource for exploring and exploiting genetic variation in sorghum crop wild relatives. Crop Science 61:150-162. DOI: 10.1002/csc2.20332.

Mansoor S., Paterson A.H. (2012) Genomes for jeans: cotton genomics for engineering superior fiber. Trends in Biotechnology 30:521-527. DOI: 10.1016/j.tibtech.2012.06.003.

Olatoye M.O., Hu Z.B., Morris G.P. (2020) Genome-wide mapping and prediction of plant architecture in a sorghum nested association mapping population. Plant Genome 13. DOI: 10.1002/tpg2.20038.

Palomeque L., Liu L.J., Li W.B., Hedges B., Cober E.R., Rajcan I. (2009) QTL in mega-environments: II. Agronomic trait QTL co-localized with seed yield QTL detected in a

population derived from a cross of high-yielding adapted x high-yielding exotic soybean lines. Theoretical and Applied Genetics 119:429-436. DOI: 10.1007/s00122-009-1048-8.

Paterson A.H., Boman R.K., Brown S.M., Chee P.W., Gannaway J.R., Gingle A.R., May L., Smith C.W. (2004) Reducing the genetic vulnerability of cotton. Crop Science 44:1900-1901. DOI: 10.2135/cropsci2004.1900.

Paterson A.H., C. L. Brubaker, and J.F. Wendel. (1993) A rapid method for extraction of cotton (Gossypium spp.) genomic DNA suitable for RFLP or PCR analysis. Plant Molecular Biology 11:122-127. DOI: 10.1007/ BF02670470.

Perumal R., Tesso T.T., Morris G.P., Jagadish S.V.K., Little C.R., Bean S.R., Yu J.M., Prasad P.V.V., Tuinstra M.R. Registration of the sorghum nested association mapping (NAM) population in RTx430 background. Journal of Plant Registrations. DOI: 10.1002/plr2.20110.

Poland J.A., Rife T.W. (2012) Genotyping-by-Sequencing for Plant Breeding and Genetics. Plant Genome 5:92-102. DOI: 10.3835/plantgenome2012.05.0005.

Quisenberry J.E., Jordan W.R., Roark B.A., Fryrear D.W. (1981) EXOTIC COTTONS AS GENETIC SOURCES FOR DROUGHT RESISTANCE. Crop Science 21:889-895. DOI: 10.2135/cropsci1981.0011183X002100060022x.

Saha S., Bellaloui N., Jenkins J.N., McCarty J.C., Stelly D.M. (2020) Effect of chromosome substitutions fromGossypium barbadenseL.,G. tomentosumNutt. Ex Seem andG. mustelinumWatt intoG. hirsutumL. on cottonseed protein and oil content. Euphytica 216. DOI: 10.1007/s10681-020-02644-4.

Said J.I., Lin Z.X., Zhang X.L., Song M.Z., Zhang J.F. (2013) A comprehensive meta QTL analysis for fiber quality, yield, yield related and morphological traits, drought tolerance,

and disease resistance in tetraploid cotton. Bmc Genomics 14. DOI: 10.1186/1471-2164-14-776.

Said J.I., Song M.Z., Wang H.T., Lin Z.X., Zhang X.L., Fang D.D., Zhang J.F. (2015) A comparative meta-analysis of QTL between intraspecific Gossypium hirsutum and interspecific G. hirsutum x G. barbadense populations. Molecular Genetics and Genomics 290:1003-1025. DOI: 10.1007/s00438-014-0963-9.

Seyoum M., Du X.M., He S.P., Jia Y.H., Pan Z.E., Sun J.L. (2018) Analysis of genetic diversity and population structure in upland cotton (Gossypium hirsutum L.) germplasm using simple sequence repeats. Journal of Genetics 97:513-522. DOI: 10.1007/s12041-018-0943-7.

Wang B.H., Zhuang Z.M., Zhang Z.S., Draye X., Shuang L.S., Shehzad T., Lubbers E.L., Jones D., May O.L., Paterson A.H., Chee P.W. (2017) Advanced Backcross QTL Analysis of Fiber Strength and Fineness in a Cross between Gossypium hirsutum and G. mustelinum. Frontiers in Plant Science 8. DOI: 10.3389/fpls.2017.01848.

Warburton M.L., Rauf S., Marek L., Hussain M., Ogunola O., Gonzalez J.D.S. (2017) The Use of Crop Wild Relatives in Maize and Sunflower Breeding. Crop Science 57:1227-1240. DOI: 10.2135/cropsci2016.10.0855.

Yu J., Jung S., Cheng C.-H., Ficklin S.P., Lee T., Zheng P., Jones D., Percy R.G., Main D. (2014) CottonGen: a genomics, genetics and breeding database for cotton research. Nucleic Acids Research 42:D1229-D1236. DOI: 10.1093/nar/gkt1064.

Zaidi S., Naqvi R.Z., Asif M., Strickler S., Shakir S., Shafiq M., Khan A.M., Amin I., Mishra B., Mukhtar M.S., Scheffler B.E., Scheffler J.A., Mueller L.A., Mansoor S. (2020) Molecular insight into cotton leaf curl geminivirus disease resistance in cultivated cotton

(Gossypium hirsutum). Plant Biotechnology Journal 18:691-706. DOI: 10.1111/pbi.13236.

Zeng L., Meredith W.R. (2009) Registration of Five Exotic Germplasm Lines of Cotton Derived from Multiple Crosses among Gossypium Tetraploid Species. Journal of Plant Registrations 3:77-80. DOI: 10.3198/jpr2008.06.0337crg.

Zhang Z., Liu A.Y., Huang Z., Fan S.M., Zou X.Y., Deng X.Y., Ge Q., Gong J.W., Li J.W., Gong W.K., Shi Y.Z., Fan L.Q., Zhang Z.B., Jiang X., Lei K., Yuan Y.L., Xu A.X., Shang H.H. (2019) Quantitative trait locus mapping for plant height and branch number in an upland cotton recombinant inbred line with an SNP-based high-density genetic map. Euphytica 215. DOI: 10.1007/s10681-019-2428-z.

Zhang Z.S., Rong J.K., Waghmare V.N., Chee P.W., May O.L., Wright R.J., Gannaway J.R., Paterson A.H. (2011) QTL alleles for improved fiber quality from a wild Hawaiian cotton, Gossypium tomentosum. Theoretical and Applied Genetics 123:1075-1088. DOI: 10.1007/s00122-011-1649-x.

# CHAPTER 3

# MAPPING MUTATIONS IN *GOSSYPIUM HIRSUTUM* CAUSED BY ETHYL METHANESULFONATE

## Introduction

Cotton is an extremely important crop for economies around the world, including providing approximately 50% of the raw material for the textile industry (Rong et al., 2007). Globally, there is approximately 30 million hectares dedicated to cotton production (Ulloa et al., 2020). In the 2018/2019 growing season alone, cotton was grown in about 30 countries and approximately 116 million bales were produced (Zhang et al., 2019b). In the 2019/2020 growing season this estimation increased to approximately 125.3 million bales produced (Saha et al., 2020). In addition to cotton production, 150 countries in total are involved in cotton export and import, with the United States being the top global cotton exporter (Chen et al., 2007). In 2016 alone, 2 million metric tons of cotton were exported which equates to a multi-billion dollar industry for the US alone (Chandnani et al., 2017). Zhang et al. (2019a) states that the US exported approximately 80.7% of the cotton that was produced here. In 2017, this equated to $74 billion contributed to the gross domestic product (GDP) and is directly responsible for 1.3 million jobs (Lopes et al., 2020).

The most common species of cotton grown is *Gossypium hirsutum*, or upland cotton, which makes up 95% of cotton yield around the world. This is mainly because it has a higher yield, although lower quality, than the second most commonly produced species of cotton, *G. barbadense*. The United States has been able to continue to be competitive in the international

cotton market due to superior fiber quality, therefore it is extremely important that quality continues to improve without sacrificing yield (Zhang et al. 2019). However, a narrow genetic base in "genetically impoverished" upland cotton is the major obstacle to cotton improvement (Paterson et al., 2004). This low genetic diversity is likely from a series of bottlenecks from polyploidization, domestication, and migration, as well as an overexploitation of a few genetic backgrounds in recent breeding programs (Adhikari et al., 2017). Abbas et al. (2015) used SSR markers to find that most genotypes of domesticated cotton have high levels of genetics similarity ranging from 81.1% to 98.7%. However, Lopes et. al 2020 suggested that it is possible that less than 1% of total cotton genetic variability resides in US commercial cultivars, so it is likely possible to regain some genetic diversity through external methods.

The two largest sources to increase genetic diversity in upland cotton are breeding with wild or exotic relatives or creating diversity by creating new alleles, such as through mutagenesis (Mansoor and Paterson, 2012). It is possible to cause mutagenesis in cotton with a number of techniques. Luckett (1989) used colchicine, which is a chemical that causes chromosome loss or rearrangement as well as helps to isolate homozygous genotypes, to increase heritable variation in cotton. Another technique is to use gamma radiation on cotton pollen and subsequent hybridization to create new genetic variability (Haidar et al., 2016). However, radiation tends to result in higher mortality of individuals exposed (Haidar et. al 2016). Recently, the CRISPR/Cas9 genome editing system has been used to increase genetic diversity in cotton as well. This technique is relatively new and the scope of its uses are still being discovered, but it has been demonstrated to be effective at targeted mutagenesis of specific genes in cotton, as well as single base mutations from C-G to T-A (Janga et al., 2017; Qin et al., 2020). While these are

all effective mutagenesis techniques, one mutagenesis agent that has been successfully used in a variety of crops is ethyl methanesulfonate or EMS.

Chemical mutagens, such as ethyl methanesulfonate are able to confer stable single nucleotide changes, rather than deletions that usually have some negative or even lethal results (Parry et al., 2009). EMS causes modifications of individual nucleotides in the DNA, which causes guanine (G) to pair with thymine (T) instead of cytosine (C) (Greene et al., 2003). The cell will attempt to repair the mismatch but will replace the guanine with adenine (A) instead of replacing the thymine. The result is an overall change from a GC nucleotide pair to an AT nucleotide pair, which is a point mutation (Greene et. al 2003). EMS is a frequently used mutagen, because it is more efficient than irradiation and has a lower mortality rate (Favret, 1960). EMS is efficient because has a high mutation rate that allows for smaller sample sizes to still be saturated with mutations (Talebi, 2012). This method has been used to cause mutations with economic benefits in a variety of crops including peas, soybeans, barley, tomatoes, wheat, and others (Blixt et al., 1966; Espina et al., 2018; Gichner and Ehrenberg, 1966; Hildering and Vanderve.Jh, 1966; Rao and Sears, 1962; Vanderve.Jh, 1967). Specifically, this method of mutagenesis has led to the release of numerous commercially improved cotton lines (Auld et al., 2007; Bechere et al., 2007; Bechere et al., 2011; Brown et al., 2012; Herring et al., 2004). Mutagenesis, and specifically EMS, is a simple, useful way to generate novel variation in crops where inbreeding and heavy selection has dramatically decreased the natural genetic variation present. The present study will map mutations caused by EMS to candidate QTLs in a population that has shown improvement in fiber quality characteristics.

**Mutant S Population Background**

Previously, two varieties of *G. hirsutum*, TAM94L-25 (Smith, 2003) and Acala 1517-99 (Cantrell et al., 2000), were treated with EMS according to Auld et al. (1992) and advanced to the M5 generation by single boll descent (Patel et al., 2014). Around 3000 of these mutant lines were tested for improved fiber quality and yield in Texas and Georgia. A total of 157 lines were found to have at least one improved fiber property. Additionally, some mutant lines were crossed to each other to stack the mutations in the greenhouse in the Summer of 2012. The parents of these crosses were from different variety backgrounds, so one parent was a mutant of the Acala variety, and one parent was a mutant of the TAM94-L variety. The F1 hybrids of these crosses were then crossed together to continue to stack mutations and hopefully improve fiber quality. The resulting hybrid from the second cross was then selfed to create an F2 population. This F2 generation was evaluated in the same two locations as above. All the mutant lines were developed by Jinesh Patel at the Plant Genome Mapping Laboratory at the University of Georgia in Athens, GA. The population being mapped in this study, which will now be referred to as the "S population," was formed by crossing four mutant lines in total together. This population had lines 926 (TAM94L-25) X 3028 (Acala 1517-99) as one parent and lines 2888(TAM94L-25) X 1903 (Acala 1517-99) as the second parent. These crosses were also done by Jinesh Patel at the University of Georgia Plant Genome Mapping Laboratory.

**Materials and Methods**

One hundred lines of the S population were grown in Watkinsville, GA in 2018, with two replicates of each line. Fiber weight and seed weight were collected in the field after harvesting. Fiber quality parameters were measured by Cotton Inc in Cary, NC on a high-volume instrument (HVI). Additionally, seed surface area was analyzed as well. Phenotypic analysis was

completed in RStudio using ANOVA and Tukey's Honest Significant difference tests. Correlations between fiber quality characteristics were also found using R. In addition to samples grown in 2018, phenotypic analysis was completed for samples in four total environments, Watkinsville/Athens in 2013, 2014, and 2018 and Tifton in 2014, to analyze any genotype by environment effects.

Additionally, young leaf tissue samples were taken for DNA extraction in 2018. DNA was extracted according to the methods described in (Paterson, 1993) which uses a modified CTAB method and a combination of buffers to extract DNA of a size and quality that is suitable for analysis. DNA concentration was checked on a nanodrop to make sure it was of high enough quality for sequencing library creation.

Libraries were created according to a modified version of the methods in Elshire et al. (2011) for the ApeKI enzyme. DNA samples were digested with CutSmart® Buffer NEB3.1 and the ApeKI enzyme and run in a thermocycler at 75°C for three hours. Barcode adapters were then added to the DNA samples with T4 DNA Ligase buffer and T4 DNA Ligase (Enzymatics). To ligate the barcodes to the samples the plate was run in the thermocycler at 25°C for two hours and 65°C for twenty minutes. The DNA samples were then pooled together and cleaned using a Qiagen PCR clean-up kit. The DNA was then run on a 2% TAE size selection gel at 80V for one and a half hours. The fragment between 200-500bp was cut out and used for gel elution. Gel elution was done using a Qiagen gel elution kit and then the DNA concentration was checked on Qubit. Afterward, the DNA samples were amplified using 2X GoTaq Master mix and forward and reverse primers. The plate was run for 18 cycles of PCR, then pooled and the concentration taken again. Another size selection gel and gel elution step were completed to remove any DNA

strands that were shorter than 200bp. Samples were then sent to the Georgia Genome Facility for fragment analysis and sequencing using Illumina NextSeq.

The sequences from Illumina NextSeq were processed in the TASSEL 5.0 standalone pipeline (Bradbury et al., 2007). This pipeline calls the distinct barcodes associated with each sample and aligns each tag with a reference genome. The reference genome used was the *Gossypium hirsutum* (AD1) 'TM-1' genome UTX_v2.1 genome alignment (Chen et al., 2020; Yu et al., 2014). The individual sequences were aligned to the reference genome using the bwa alignment tool. SNPs were then called from the sequences using the ProductionSNPCallerPlugin in TASSEL. The resulting SNPs were output to a HapMap file for use in analysis.

A total of 13,346 SNPs were called by TASSEL. After filtering so that each site had a minor allele frequency (MAF) of at least 0.01 and all sites had at least 20 (out of 96) taxa with non-missing data, there were 7,451 SNPs left. This was done to remove low coverage and monomorphic sites. The ABH genotype function was then used in TASSEL to find markers in the progeny that can be attributed to either parent. This resulted in SNP markers to be used for QTL mapping in r/qtl. Single QTL analysis used the Extended Haley-Knott regression which is an approximation of standard interval mapping which considers both the mean and variances of phenotypes and is more robust than standard interval mapping. This method requires genotype probabilities to be calculated first, using the Hidden Markov Model, with a density of 1cM, to calculate for QTL in between available markers (Broman, 2009). Markers with LOD scores above 3.00 were considered significant. Two-dimensional QTL analysis was done to look for any linked loci, epistatic effects, or interactions between two QTLs. This analysis was also done using the Haley-Knott regression method and calculation of genotype probabilities with a density of 1cM. The two-dimensional model uses a string of thresholds for LOD scores for comparing

the full vs. additive model, comparing the full model vs. the single QTL model, the interaction between QTLs, the additive model, and comparing the additive model to the single QTL model, which are 9.1, 7.1, 6.3, 6.3, and 3.3, respectively. These threshold scores have been previously estimated by permutation tests for the 95 percentiles of the null distributions for an intercross (Broman, 2009). The high thresholds help to correction for multiple testing across the genome.

Multiple QTL mapping looks at overall model testing of how all QTLs found in either the single-QTL analysis or the two-dimensional QTL analysis fits the data. The Haley-Knott regression method was also used for multiple QTL mapping. Multiple QTL mapping in R uses the locations of a set of QTLs to be considered, in this case the results from the single or two-dimensional QTL analysis and fits them into a defined model to see the effects of all QTLs and whether they interact. It also drops one term at a time from the model to assess support for individual terms and interactions by comparing between the full model and the model with the term omitted. (Broman, 2009). First, models with no interactions present were considered. Then, the command "addint" was used to add in each interaction one at a time. If interactions were significant, they were then added to the model. Multiple QTL analysis also gives the effect of each QTL, which is estimated as the difference between the phenotypic averages for the heterozygotes and the homozygotes. A positive effect indicates that the heterozygote had the greater phenotypic value than the homozygote, which may or may not be preferred depending on the trait as lower values are preferred for fineness and short fiber content.

Two separate QTL analyses were completed, one in which the parents in the analysis mimicked what occurred in the cross (two TAMxAcala hybrids crossed together), and one in which one parent was both Acala mutants and the other parent was both TAM mutants. These

two analyses were considered to see how they might differ and to find additional sites for candidates of a fiber length mutation.

## Results

### Phenotypic Analysis

For the 2018 samples only, correlations between all traits can be seen in Figure 3.1. There were six pairs of strong negatively correlated traits: UHM and MIC, SFC and UHM, UI, and STR, and LP and UHM and UI. There were two pairs of strong positively correlated traits fiber weight and seed weight, and UHM and UI. The S population was significantly lower than parent 1903 in length and uniformity. The S population was also significantly lower in uniformity than parent 926. The S population was significantly better than its elite ancestral parents, Acala Maxxa and TAM94L-25 in fiber fineness, length, strength, and short fiber content. The S population has a significantly higher lint percent than 926, 1903, 2888, and 3028 and was not significantly different from Acala or TAM. Boxplots showing the distribution of all traits that showed differences is shown in Figure 3.2.

Environment had a strong effect on the phenotypes of all fiber quality traits in the S population. The environments are referred to as A13, A14, and A18 for samples grown in Athens/Watkinsville in 2013, 2014, and 2018, and T14 for samples grown in Tifton in 2014. For MIC, samples performed significantly better (lower is better) in A13 and T14 than in A14 and A18. For UHM, samples in A13 performed significantly better than all other environments, and A18 performed worse than other environments. For uniformity index, all environments were significantly different from each other and A13 performed the best and A18 performed the worst. For fiber strength, A18 and A14 were significantly higher than T14. For elongation, all

environments performed differently from one another with A18 performing the best and T14 performing the worst. For short fiber content, A13 performed better than all other environments. For lint percent, all environments were different from each other and A18 was highest overall and T14 was the lowest overall. The distribution of all traits across all environments are shown in Figure 3.3.

**QTL Analysis of (Acala X TAM) crossed to (Acala X TAM)**

**General Data Checking for QTL analysis**

The ABH genotype function from TASSEL gave 131 SNP markers to be analyzed in QTL mapping. Parent A was 926 X 3028, and Parent B was 2888 X 1903. It is important to note that chromosomes A09 and D06 did not have any markers associated with them. The chromosomes with the most markers were A08 and D10 with 11 and 10 markers each, respectively. In addition, there was a lot of missing data, with only 46.8% of all sites across all individuals being genotyped. Markers on different chromosomes did not appear to be linked, which can be seen in the plot of pairwise recombination fractions and LOD scores in Figure 3.4.

**Single QTL Analysis**

Single QTL analysis found five potential QTLs for MIC, four potential QTLs for UI, seven for STR, eight for ELO, eight for SFC, seven for seed surface area, six for fiber weight, five for seed weight and none for fiber length or lint percent. These were across sixteen chromosomes. The chromosomes that had QTLs for the most traits were eight and thirteen, both had potential QTLs for seven of eight fiber quality traits. Chromosome eight did not have a QTL for fiber weight and chromosome thirteen did not have a QTL for seed weight. Chromosome seven also had QTLs for six out of eight fiber quality traits, all but ELO and UI. The QTLs found in this single QTL analysis will be used to fit into a model for multiple QTL analysis.

**Two-Dimensional QTL analysis**

For the Two-dimensional QTL analysis, only one trait showed a possible interaction between multiple QTLs, elongation. There were two pairs of QTLs found for elongation, on chromosome two at 30.1cM and chromosome 24 at 61.5cM, and on chromosome five at 102.8cM and chromosome 17 at 20.5cM. Both pairs show evidence of the existence of a pair of QTLS when an interaction is allowed, and there is evidence for an interaction as well. The single QTL analysis for elongation also found potential QTLs on chromosomes 2, 5, and 17, but did not find evidence of one on chromosome 24, potentially due to its interaction with chromosome 2. A table with LOD scores for this analysis can be seen in Table 3.1.

**Multiple QTL Analysis**

For all fiber traits except for elongation, the potential QTLs that were found using the single QTL analysis were used in modeling for multiple QTL analysis. Fiber elongation used both the single and two-dimensional scan results in separate multiple QTL analyses because it was the only trait that had pairs over the thresholds given.

For fiber fineness, MIC, five QTL in total on chromosomes 4, 7, 8, 11, and 13 were considered in the model. The overall model had a LOD score of 3.26 which means that it is a better fit than the null model, containing no QTLs. This model explains approximately 15.08% of the phenotypic variation seen. However, no individual term, when dropped from the model had a LOD score above 3, which does not indicate strong evidence for any of these loci. When interaction terms were added one by one, none had a LOD score >2, so it is possible that none of them are significant. Three QTLs, on chromosomes 4, 7, and 13, had negative effects and two, on chromosomes 8 and 11, had positive effects. It is important to note that in MIC, lower values

are better, so here a "negative effect" is beneficial. None of the interaction terms were significant and therefore not included in the model.

For fiber uniformity, UI, four QTL in total on chromosomes 8, 10, 13, and 20 were considered in the model. The overall model had a LOD score of 3.66 which means that at a threshold of 3, it is significantly better than the null model containing no QTLs. This model explained approximately 16.7% of phenotypic variation observed. However, when each locus is dropped individually, only one locus had a LOD score >1, which does not indicate strong evidence for the presence of these QTLs. When interaction terms were added one by one, none had a LOD score >2, so it is possible that none of them are significant. Three QTLs, on chromosomes 8, 13, and 20, had negative effects and one QTL, on chromosome 10, had positive effects.

For fiber strength, STR, seven QTL in total on chromosomes 3, 7, 8, 11, 13, 15, and 18 were considered in the model. The overall model had a LOD score of 5.58, which means that it is significantly better than the null model containing no QTLs. This model explained approximately 24.4% of the phenotypic variation observed. When one locus at a time was dropped, one QTL had a LOD score >3, which indicates strong evidence for its presence. The QTL on chromosome 13 at 15.9cM had a LOD score of 3.22 and explained 13.2% of phenotypic variation. When interaction terms were added one by one, one interaction, between chromosome seven and chromosome 15 was significant with a LOD score of 3.5. This term was then added back into the original model. The resulting model had a LOD score of 9.12 and explained approximately 36.6% of the phenotypic variation. In this model, three potential QTLs and the interaction had LOD scores above 3 and each explained between 10-16% of the variation observed. When interaction terms were added to this model one by one, an additional

interaction, between chromosomes 7 and 18, had a LOD score of 3.9. When these two interaction terms were considered, the model had a LOD score of 13.07 and explained 48% of the observed variation. There is strong evidence for the presence of QTLs on chromosomes 7, 13, 15, and 18, and both interactions as they all have LOD scores >3. Five QTLs, on chromosomes 7, 8, 11, 13, and 18, contributed negative effects and two, on chromosomes 3 and 15, contributed positive effects.

For fiber elongation, eight total QTL on chromosomes 2, 5, 8, 11, 12, 13, 17, and 18 from the single-QTL mapping were considered in this model. The overall model had a LOD score of 5.04, therefore it is significantly better than the null model. This model explained 22.3% of the overall phenotypic variation. However, when each locus is dropped individually, only one locus had a LOD score >1, which does not indicate strong evidence for the presence of these QTLs. When interaction terms were added to this model individually, two interaction terms were significant, between chromosome 2 and 11, and chromosomes 2 and 17. The resulting model with these two terms included, had a LOD score of 12.22 and explained 45% of the variation. Chromosomes 2, 11, 17 and both interaction terms were significant. When interaction terms were added individually to this model, two more interaction terms, between chromosome 2 and 5 and between chromosomes 11 and 13 were also significant. This model had a LOD score of 18 and explained approximately 59.5% of the variation. The QTL on chromosome 2 at 19.1cM explained the most phenotypic variation at 36.7% and contributed a small negative effect of -0.68. The model with all interactions included had a LOD score of 81.52 and explained 98% of the variation observed. When evaluated individually, all main effects and interactions were significant.

For the multiple QTL analysis of the two-dimensional QTL scan for fiber elongation the model, which contained the two interactions previously stated, had a LOD score of 13.4 and explained approximately 48.8% of the phenotypic variation. When model terms were dropped individually, only one QTL, on chromosome 2 at 30.1cM had a LOD score higher than 3. This QTL had a small negative effect of -0.023, but the interaction between chromosomes 2 and 24 had a positive effect of 0.83.

For short fiber content, SFC, eight total QTL on chromosomes 3, 7, 8, 10, 11, 13, 15, and 18 were included in the model. The overall model had a LOD score of 3.55 and explained approximately 16.3% of the phenotypic variation. When removed from the model individually, no QTL had a LOD score higher than 1, which does not support evidence for their presence. When interaction terms were added individually, none of them had a LOD score greater than three and therefore did not appear to be significant.

For seed surface area, seven total QTL on chromosomes 4, 5, 7, 8, 12, 13, and 25 were included in the model. The model had an LOD score of 7.4 and explained approximately 33% of the variation observed. When each term was dropped individually from the model, none had a LOD score higher than three which does not support evidence for their presence. When adding interactions to the model one by one, none appeared to be significant (none had a LOD score higher than 3) and were therefore not included.

For fiber weight six total QTLs on chromosomes 3, 7, 10, 12, 13, and 22 were included in the model. The model had an LOD score of 4.5 and explained approximately 20.4% of the variation observed. Model terms when dropped individually from the model, did not have any that had LOD scores >3. Four QTLs, on chromosomes 3, 10, 13, and 22, had positive effects and two, on chromosomes 7 and 12, had negative effects. When interactions were individually

looked at, one interaction, between chromosome 7 and 13 was significant and added into the model. This new model had a LOD score of 7.62 and explained approximately 31.7% of the phenotypic variation observed. When dropped from the model individually, chromosomes 7, 13, and the interaction were shown to be significant. No other interaction terms were significant.

For seed weight, five total QTL on chromosomes 3, 4, 7, 8, and 10 were considered. The model had a LOD score of 4.5 and explained approximately 20.3% of the phenotypic variation. When individual QTLs were dropped, none had a LOD score >3, which does not indicate evidence for their presence. Three QTLs, on chromosomes 3, 4, and 10, had positive effects and two, on chromosomes 7 and 8, had negative effects. None of the interaction terms were significant and therefore not included in the model.

**QTL Analysis of (Acala X Acala) crossed to (TAM X TAM) (experimental analysis)**

**General Data Checking for QTL analysis**

This analysis used the two TAM parents as parent A and the two Acala parents as parent B. The ABH genotype function from TASSEL gave 208 SNP markers to be analyzed in QTL mapping. Parent A was 926 and 2888, and Parent B was 3028 and 1903. All chromosomes had markers associated with them. The chromosome with the most markers were A05 with 33 markers. The two chromosomes with the next highest numbers of markers were A08 and D11 each with 12 markers. In addition, there was a lot of missing data, with only 50.3% of all sites across all individuals being genotyped.

**Single QTL Analysis**

Single QTL analysis found five potential QTLs for MIC, one potential QTL for UHM, ten potential QTLs for UI, seven for STR, sixteen for ELO, nine for SFC, five for seed surface area, eight for fiber weight, four for seed weight and none for lint percent. These were across

sixteen chromosomes. The chromosomes that had QTLs for the most traits were three and twenty-four, both had potential QTLs for six of nine fiber quality traits. Chromosome three did not have a QTL for fiber weight, length, and seed weight, and chromosome twenty-four did not have a QTL for seed weight, MIC, or length. Chromosomes two, five, six and eleven also had QTLs for five out of nine fiber quality traits. The QTLs found in this single QTL analysis will be used to fit into a model for multiple QTL analysis.

**Two-Dimensional QTL analysis**

For the Two-dimensional QTL analysis, only one trait showed a possible interaction between multiple QTLs, fiber elongation. There were eighteen pairs of QTLs found for elongation. A list of all chromosome pairs found and their locations can be seen in Table 3.2. All pairs show evidence of the existence of a pair of QTLS when interactions are allowed, and there is evidence for interactions as well. The single QTL analysis for elongation also found potential QTLs on all chromosomes found in the two-dimensional scan except for the QTL on chromosome 21, potentially due to its interaction with chromosome 11.

**Multiple QTL Analysis**

For all fiber traits except for elongation, the potential QTLs that were found using the single QTL analysis were used in modeling for multiple QTL analysis. Fiber elongation used both the single and two-dimensional scan results in separate multiple QTL analyses because it was the only trait that had pairs over the thresholds given.

For fiber fineness, MIC, five QTL in total on chromosomes 2, 3, 5, 11, and 16 were considered in the model. The overall model had a LOD score of 2.9 which does not exceed the threshold value of 3, so it is not a better fit than the null model, containing no QTLs. This model explains approximately 13.52% of the phenotypic variation seen. In addition, no individual

term, when dropped from the model had a LOD score above 3, which does not indicate strong evidence for any of these loci. When interaction terms were added one by one, none had a LOD score >2.5, so it is possible that none of them are significant. Three QTLs, on chromosomes 3, 11, and 16, had negative effects and two, on chromosomes 2 and 5, had positive effects. It is important to note that in MIC, lower values are better, so here a "negative effect" is beneficial. None of the interaction terms were significant and therefore not included in the model.

For fiber uniformity, UI, ten QTL in total on chromosomes 2, 3, 5, 6, 8, 9, 18, 22, 24, and 26 were considered in the model. The overall model had a LOD score of 6.68, which means that at a threshold of 3, it is significantly better than the null model containing no QTLs. This model explained approximately 28.4% of phenotypic variation observed. However, when each locus is dropped individually, only two loci had a LOD score >1.5, which does not indicate strong evidence for the presence of these QTLs. When interaction terms were added one by one, one interaction, between chromosome three and chromosome 26 was significant with a LOD score of 3.33. This term was then added back into the original model. The resulting model had a LOD score of 10.02 and explained 39.4% of the phenotypic variation. In this model, two potential QTLs and the interaction had LOD scores above 3 and explained between 11-17% of the variation observed. When interaction terms were added one by one, three additional interactions, between chromosome 5 and 22, between 18 and 26, and between 22 and 26, had LOD scores above three and were added back into the model. The resulting had a LOD score of 19.5 and explained 62.3% of the phenotypic variation. In this model four potential QTLs and one interaction had LOD scores above 3. Seven QTLs, on chromosomes 2, 3, 5, 6, 8, 22, and 26, had negative effects and three QTLs, on chromosomes 9, 18, and 24, had positive effects.

For fiber strength, STR, seven QTL in total on chromosomes 2, 3, 8, 11, 15, 20, and 24 were considered in the model. The overall model had a LOD score of 4.84, which means that it is significantly better than the null model containing no QTLs. This model explained approximately 21.5% of the phenotypic variation observed. When one locus at a time was dropped, no QTLs had a LOD score >3, which does not indicate strong evidence for the presence of these QTLs. When interaction terms were added one by one, none had a LOD score >3, so it is possible that none of them are significant. Four QTLs, on chromosomes 3, 11, 15, and 20, contributed negative effects and three, on chromosomes 2, 8, and 24, contributed positive effects.

For fiber elongation, sixteen total QTL on chromosomes 1, 2, 3, 5, 6, 9, 10, 11, 12, 13, 15, 16, 17, 23, 24, and 25 from the single-QTL mapping were considered in this model. The overall model had a LOD score of 13.34, therefore it is significantly better than the null model. This model explained 48.7% of the overall phenotypic variation. When each locus was dropped individually, three QTLs, on chromosomes 10, 13, and 16, had LOD scores above 3, which does indicate strong evidence for the presence of these QTLs. When interaction terms were added to this model individually, 26 interaction terms were significant. This is too many interactions to add in at once, so the interactions which had a LOD score >5 were included in the second interaction of the model. The resulting model with these ten terms included, had a LOD score of 51.2 and explained 92.3% of the variation. All chromosomes except 5, 6, and 10 had LOD scores greater than 3. Eight out of ten interaction terms were also significant. The resulting model had 34 more interaction terms that were significant, 10 of which had LOD scores greater than 10.0. Because there are so many significant interactions and all individual QTLs have significant interactions, there is strong evidence for the presence of these QTLs. Ten QTLs, on

chromosomes 1, 2, 3, 6, 10, 13, 15, 16, 17, and 23, had negative effects and six QTLs, on chromosomes 5, 9, 11, 12, 24, and 25, had positive effects.

For the multiple QTL analysis of the two-dimensional QTL scan for fiber elongation the model, which contained all interactions previously stated, had a LOD score of 51.17 and explained approximately 92.28% of the phenotypic variation. When model terms were dropped individually all QTL and interactions had a LOD score higher than 3, which is significant evidence for their presence.

For short fiber content, SFC, ten total QTL on chromosomes 1, 2, 3, 6, 8, 9, 15, 18, 22, and 24 were included in the model. The overall model had a LOD score of 5.17 and explained approximately 22.8% of the phenotypic variation. When removed from the model individually, no QTL had a LOD score higher than 1.5, which does not support evidence for their presence. When interaction terms were added individually, four of them, between 1 and 24, between 2 and 3, between 8 and 24, and between 15 and 22, had a LOD score greater than three and were added back into the model. The resulting model had a LOD score of 19.11 and explained 61.59% of the phenotypic variation. Seven QTL and all the interactions had LOD scores greater than three which is good evidence that they are actually significant. Three additional interactions were significant and added back into the model. The resulting model had a LOD score of 31.15 and explained 78.97% of the variation. Nine of the QTLs present and all of the interactions were significant and therefore had evidence for the presence of the QTLs. The only QTL that did not have a LOD score greater than three was the QTL on chromosome 6. Five QTLs, on chromosomes 1, 2, 9, 22, and 24, had negative effects and five QTLs, on chromosomes 3, 6, 8, 15, and 18, had a positive effect.

For seed surface area, five total QTL on chromosomes 3, 5, 11, 23, and 24 were included in the model. The model had an LOD score of 3.7 and explained approximately 18.2% of the variation observed. When each term was dropped individually from the model, none had a LOD score higher than three, which does not support evidence for their presence. When adding interactions to the model one by one, none appeared to be significant (none had a LOD score higher than 3) and were therefore not included. Four QTLs, on chromosomes 3, 5, 11, and 24, had negative effects and one, on chromosome 23, had positive effects.

For fiber weight eight total QTLs on chromosomes 1, 5, 6, 8, 12, 19, 22, and 24 were included in the model. The model had an LOD score of 8.4 and explained approximately 34.3% of the variation observed. Model terms when dropped individually from the model, had one QTL that had a LOD score >3, which was on chromosome 8. Four QTLs had positive effects and two had negative effects. When interactions were individually looked at, one interaction, between chromosome 19 and 24 was significant and added into the model. This new model had a LOD score of 11.46 and explained approximately 43.65% of the phenotypic variation observed. When dropped from the model individually, chromosomes 19, 24, and the interaction were shown to be significant. Six QTLs, on chromosomes 1, 5, 8, 12, 19, and 24, were shown to have negative effects and two, on chromosomes 6 and 22, to have positive effects. No other interaction terms were significant.

For seed weight, four total QTL on chromosomes 1, 6, 11, and 26 were considered. The model had a LOD score of 3.4 and explained approximately 15.6% of the phenotypic variation. When individual QTLs were dropped, none had a LOD score >3, which does not indicate evidence for their presence. Two QTLs, on chromosomes 3 and 26, had positive effects and two,

on chromosomes 1 and 11, had negative effects. None of the interaction terms were significant and therefore not included in the model.

**Discussion**

Overall, 50 QTLs were found in single QTL mapping and an additional two pairs were found for elongation in a two-dimensional QTL scan. These QTLs spanned 17 chromosomes, however, four of these chromosomes only had one QTL each (chromosomes 20, 22, 24, and 25) and four more only had two QTL each (chromosomes 2, 10, 15, and 17). Three chromosomes in particular hosted many overall QTLs. Chromosomes 7, 8, and 13 had 6, 7, and 7 QTLs each, respectively, which makes up 37% of all QTLs discovered. This supports previous evidence that fiber quality QTLs are non-randomly distributed throughout the genome (Chandnani et al., 2018). These three chromosomes also show evidence of QTL clustering. QTL clusters are genomic regions that harbor more than two QTL for different traits within 20cM (Chandnani et al., 2018). For chromosome 7, all six QTLs were clustered together ranging from 69cM to 91cM, approximately 22cM. For chromosome 8, four out the seven QTLs were clustered within 8cM, from 105cM to 113cM, which may also be evidence that a single QTL affects multiple traits, as opposed to multiple QTLs being clustered together. For chromosome 13, two clusters were observed. The first cluster contained four QTLs and ranged from 17.9cM to 37.9cM and the second cluster contained three QTLs and ranged from 50cM to 58cM. Chromosomes 3 and 10, which each contained four QTL, also had QTL clusters. Chromosome 3 had all four QTL found within 13cM, ranging from 72cM to 85cM. Chromosome 10 also had all four QTL clustered within 14cM, ranging from 65cM to 79cM. Again, this supports previous findings that QTLs are nonrandomly clustered throughout the cotton genome. Knowledge of these QTL clusters can greatly help with marker assisted selection but can also increase the linkage drag of

unfavorable traits, since a single genomic region influences so many traits (Chandnani et al., 2018).

In addition, there were three pairs of QTLs in different traits, on chromosomes 3, 4, and 12, that were very close together and therefore may be part of a single QTL for multiple traits. Chromosome 4 had QTLs for MIC and seed weight at 59.7 and 58.7cM, respectively. Chromosome 3 had QTLs for fiber weight and seed weight at 72.2 and 74.2cM, respectively. These traits were also very highly correlated, so it is highly likely that this QTL affects both traits. Chromosome 12 had QTL for seed surface area and fiber weight at 59.6 and 60.6cM, respectively. This is further evidence towards the non-random distribution of QTLs in the genome, as even chromosomes without QTL clusters (4 and 12) had more than one trait associated with very close positions in the genome.

Some QTLs found in this study were within the range of other QTLs previously found according to the CottonQTLdB (Said et al., 2013; Said et al., 2015). Two out of four QTLs for fiber uniformity were within previously discovered QTLs, three out of seven QTLs for fiber strength were within previously known QTLs, and five out of twelve QTLs found for elongation were within known QTLs. The CottonQTLdB does not have many known QTLs listed for SFC or seed surface area. A table listing all QTLs found and whether they are in a known QTL can be seen in Table 3.3.

In this study, 80% of the QTLs that were found, were in the A genome, and all chromosomes that were found to be QTL rich were also within the A genome. This contrasts with a number of previous studies which found more QTLs in the D genome (Chandnani et al., 2018; Jiang et al., 1998; Rong et al., 2007). The D subgenome's closest ancestor, *Gossypium raimondii*, does not produce spinnable fiber, but the D subgenome in tetraploids have been

shown to make an important contribution to fiber production (Adhikari et al., 2017). While the majority of QTLs were not found in the D genome, almost all fiber traits studied had at least one QTL that was found in the D genome. Only MIC and seed weight did not have associated QTLs within the D subgenome. This does still suggest importance of the D genome in fiber quality in this population. However, there are other studies that have found no difference in the number of QTL between the subgenomes (Chee et al., 2005; Grover et al., 2020) and studies that have found more QTLs in the A subgenome as well (Ijaz et al., 2019; Mei et al., 2004). Therefore, it is likely that these results could depend on specific population and background. Also, QTLs detected in both subgenomes suggest that fiber related traits result from gene expression and interaction between the homologous subgenomes (Wendel, 2000). While interactions between QTL were only found in one trait, fiber elongation, each pair of interacting loci had one QTL in the A subgenome and one in the D subgenome. Studying the potential interactions between chromosomes and QTL, especially between the subgenomes, could provide a lot of insight to improvement of beneficial traits.

One interesting finding is that there were no QTLs found for fiber length, even though this population was found to have superior fiber length, both in this study and in previous evaluations (Patel et al., 2014). The four mutant parents selected for this study all had fiber length superior to that of the ancestral populations, Acala and TAM. One of these parental lines was also superior to the progeny lines, but three were not. The progeny lines were still superior to the ancestral populations however, so it is possible that there was not enough variation in the population to map for the four segregating mutants. While it is common that only QTLs above a specified threshold are considered significant, since fiber length was of particular interest in this population, QTLs below the significance threshold were also looked at to see if there were any

near misses close to the significance threshold. In the single QTL model, this threshold is an LOD score of 3.0. The highest LOD score of any QTL found for fiber length was 1.98, which was associated with marker *SA11_10627656* on chromosome 11. For the double QTL model, the significance threshold for the full model is higher (9.1) in order to correct for multiple testing. The QTL pair that had the highest LOD score was chromosome four and seven at positions 87.73cM and 91.95cM respectively, with a LOD score of 6.85. The markers on chromosomes A04 and A11 are within reported QTL from the CottonQTLdB. The possible QTL on A04 was within *qFL-C4-2* and the possible QTL on chromosome A11 is within four reported QTLs: *qFL-A11-1*, *qFL11.1*, *qFL-c11*, and *FL2.u*. These QTLs are likely too far below the threshold levels to be considered near misses in this study, but they could be areas to consider more in depth. If we assume that each of the four parents had different mutations (which is likely), then it is possible that the nature of the crossing scheme (where a TAM and Acala mutant were crossed together and then crossed to another TAM x Acala F1 hybrid) resulted in one Acala mutant masking the mutation of the other in the progeny, and therefore reducing the ability of the mutation to be mapped in a QTL. This scenario would also be likely for the TAM mutants. It is likely that this map in general picked up differences that exist between the two elite lines but are not a result of the mutagenesis.

The experimental analysis had similar results to the initial analysis, although there were more significant interactions present in most traits. However, there was one QTL for fiber length that was significant in the experimental analysis on chromosome D05 and was also associated with a specific SNP marker *SD05_2219717*. This QTL could also be further, more finely mapped as a potential site of EMS mutations. While this potential QTL for fiber length could be worth studying, as well as the other QTLs found, overall this analysis is a less accurate

representation of the S population, as it was created by crossing two Acala x Tam F1 hybrids, and it is likely that this mapping scheme mainly picked up differences in the two elite lines, rather than differences due to EMS caused mutations.

Overall, this study found numerous potential QTLs associated with eight fiber quality characteristics. The locations of QTL from analyses in which the designated 'parents' mimicked what occurred in the cross (two TAM x Acala hybrids crossed together) could represent places where EMS generated mutations occur and provide economic benefits and could be candidates for more precise fine mapping in the future.

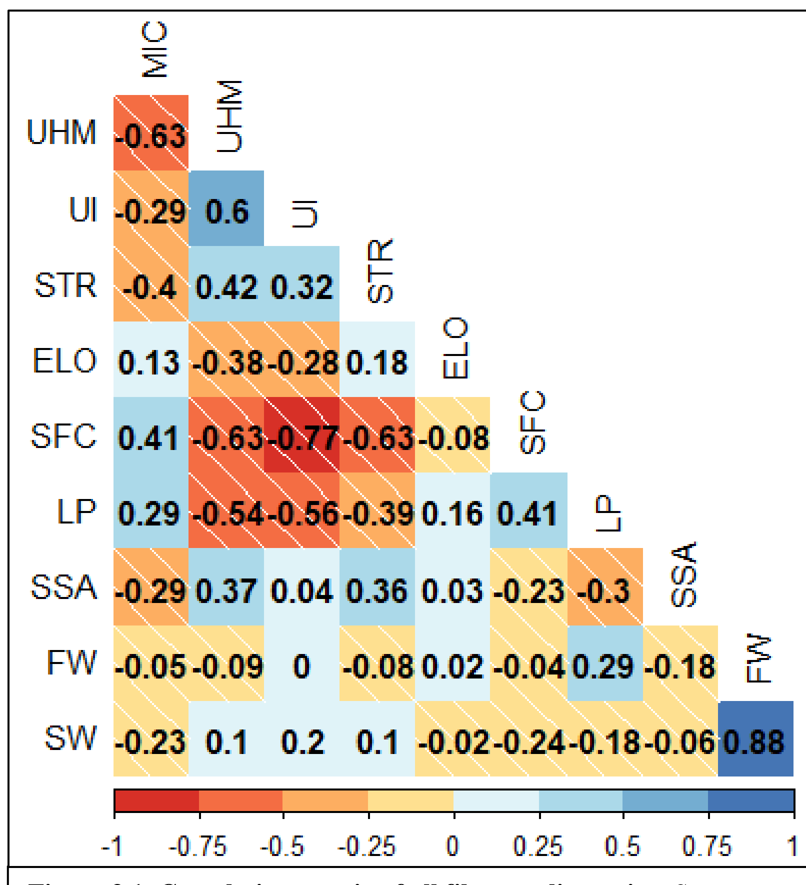


**Figure 3.1**: **Correlation matrix of all fiber quality traits.** Strong positive correlations are shown in darker blue and strong negative correlations are shown in darker red. Abbreviations shown are as follows fiber weight (FW), fiber length (UHM), fiber fineness (MIC), seed weight (SW), uniformity index (UI), fiber strength (STR), fiber elongation (ELO), short fiber content (SFC), lint percent (LP), and seed surface area (SSA).

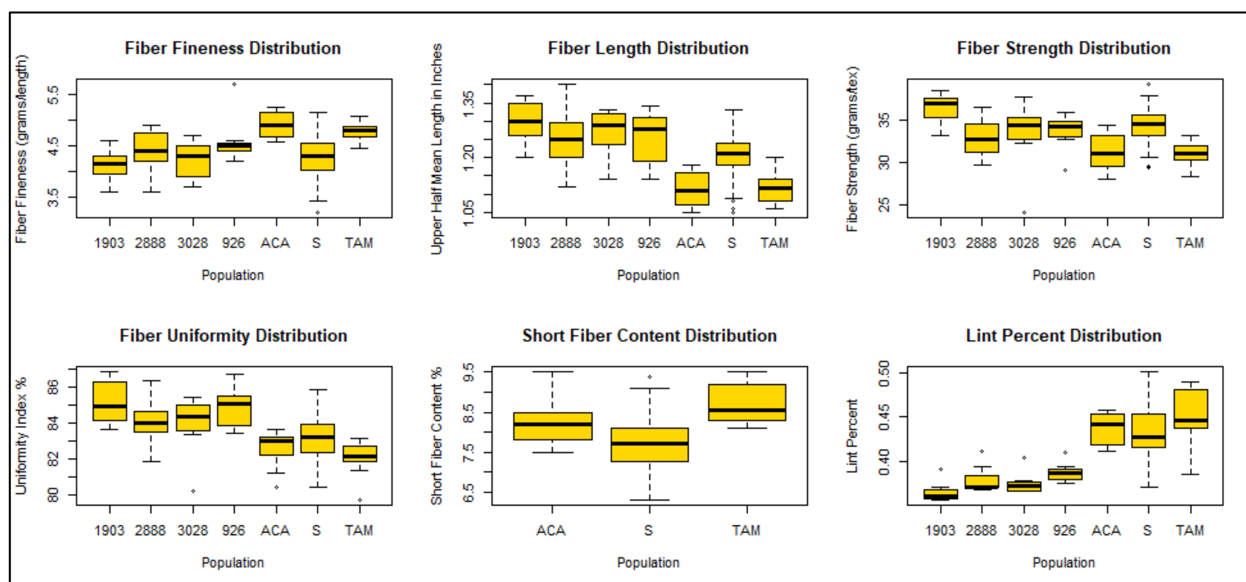


**Figure 3.2**: **Boxplots showing the distribution of six fiber traits**. The traits are fineness, length, strength, uniformity, short fiber content (SFC), and lint percent among the four mutant parental lines, the two elite ancestral lines and the S population. Note that SFC does not have data for the four mutant parental lines and only shows the differences between the S population and Acala and TAM. The four mutant parents are 1903, 2888, 3028, and 926.

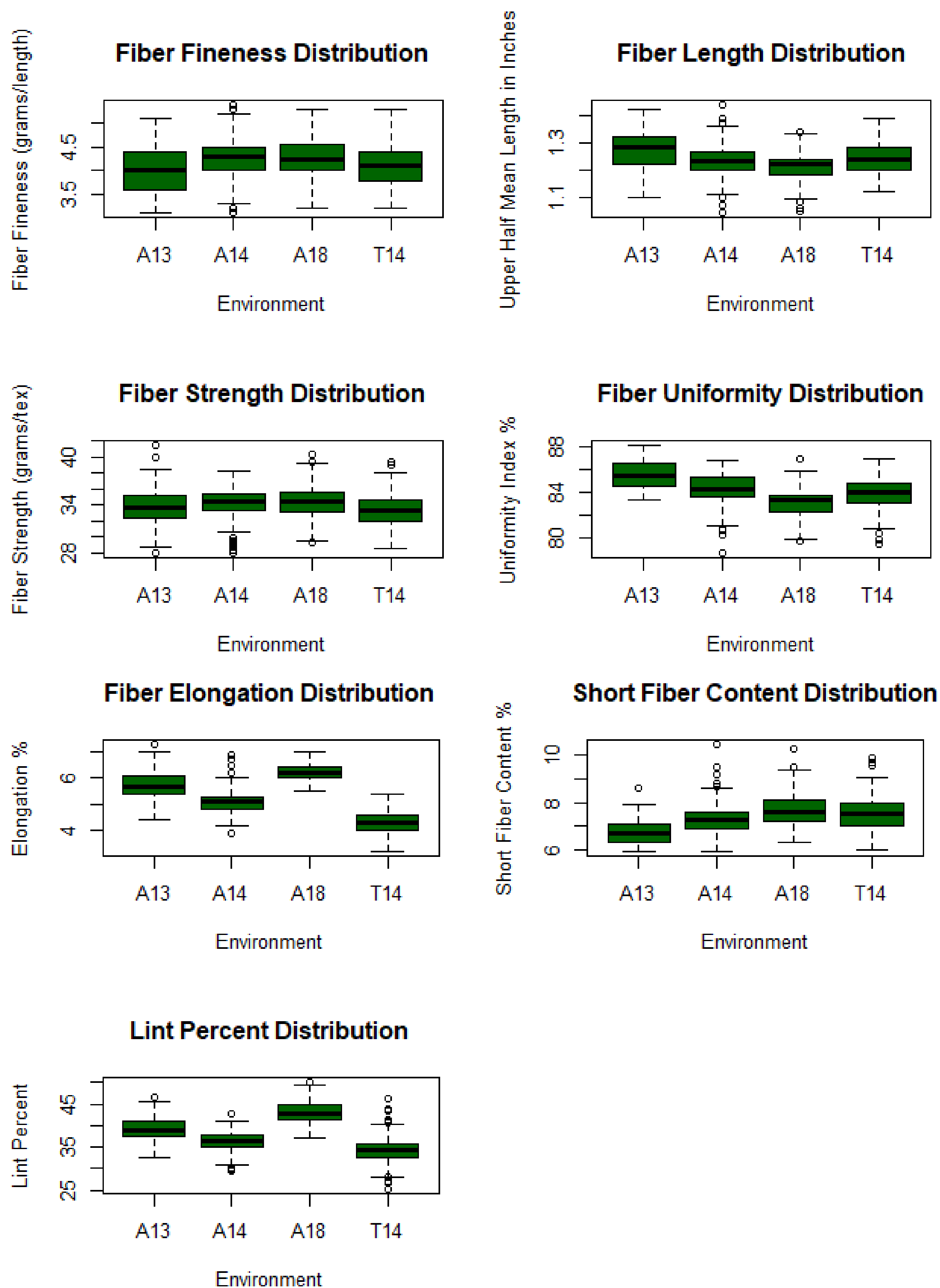


**Figure 3.3**: **Boxplots showing the distribution fiber quality traits across four environments.** The seven traits are fineness, length, strength, uniformity, elongation, short fiber content, and lint percent for the S population. The four environments are Athens in 2013 (A13), Athens in 2014 (A14), Athens in 2018 (A18), and Tifton in 2014 (T14).

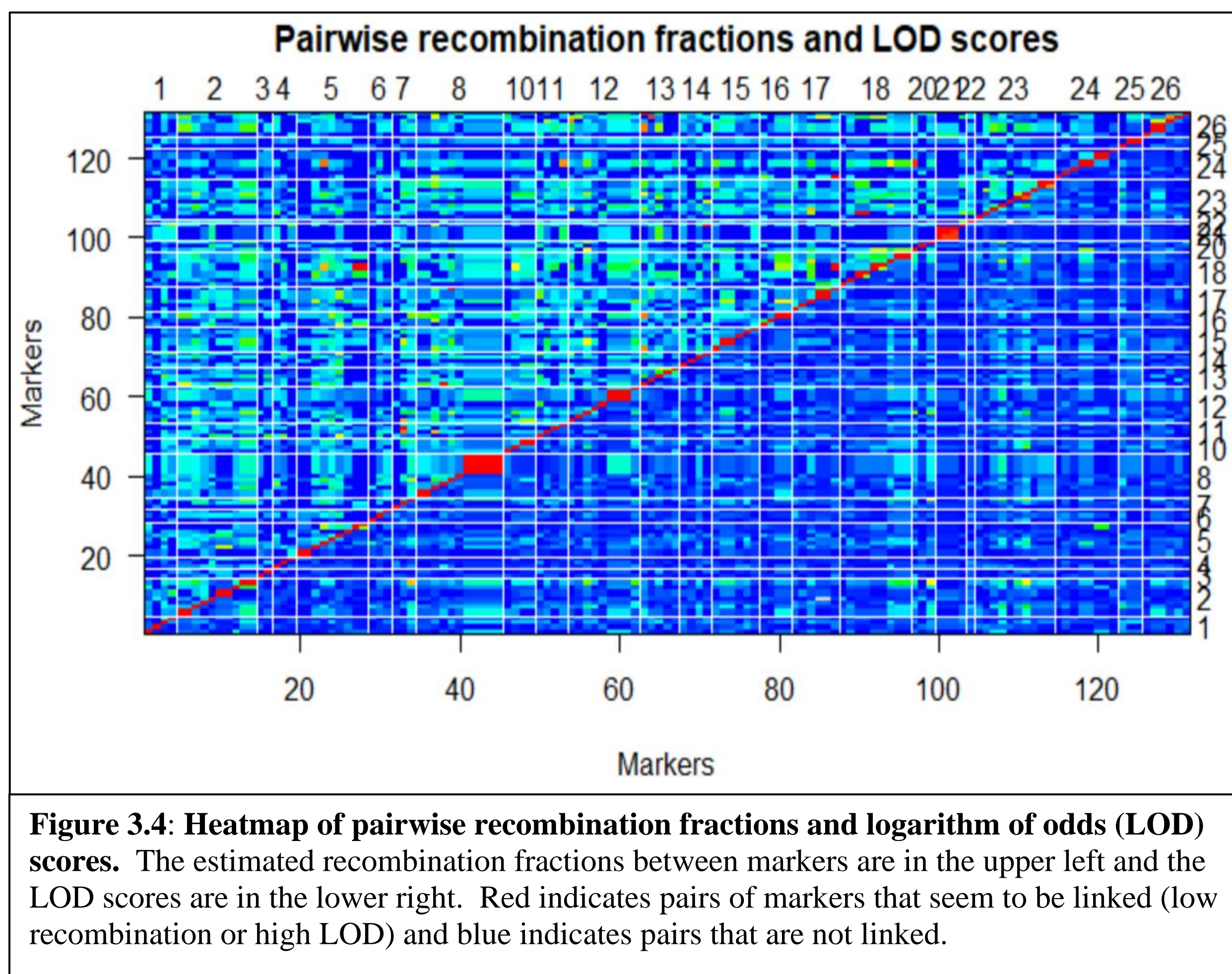


**Figure 3.4**: **Heatmap of pairwise recombination fractions and logarithm of odds (LOD) scores.** The estimated recombination fractions between markers are in the upper left and the LOD scores are in the lower right. Red indicates pairs of markers that seem to be linked (low recombination or high LOD) and blue indicates pairs that are not linked.

**Table 3.1: The LOD from the two-dimensional QTL scan for fiber elongation**. Two pair of interacting QTLs were found on the chromosomes listed. LOD.full is the LOD score of the full model and LOD.int is the LOD score of the interaction.

| Chromosome Pair | Position 1 | Position 2 | LOD.full | LOD.int |
|---|---|---|---|---|
| c2:c24 | 30.1 | 61.5 | 10.76 | 6.3 |
| c5:c17 | 102.8 | 20.5 | 9.37 | 8.16 |

**Table 3.2: A list of pairs of chromosomes found for fiber elongation in the two-dimensional QTL analysis**. Pos1 and pos2 indicate the positions of each interacting pair along the chromosome, lod.full indicates the LOD value for those chromosomes, and lod.int indicates the LOD score for the interaction.

| | **pos1** | **pos2** | **lod.full** | **lod.int** |
|---|---|---|---|---|
| **c1 :c2** | 118.16 | 33.262 | 10.57 | 5.72 |
| **c1 :c10** | 118.16 | 0.17 | 11.13 | 7.29 |
| **c1 :c12** | 104.16 | 28.15 | 10.2 | 6.97 |
| **c1 :c16** | 118.16 | 25.454 | 10.68 | 7.06 |
| **c2 :c10** | 33.262 | 0.17 | 10.03 | 5.97 |
| **c2 :c13** | 35.262 | 0.916 | 10.44 | 5.43 |
| **c3 :c5** | 113.32 | 6.424 | 13.76 | 10.96 |
| **c3 :c13** | 113.32 | 1.916 | 12.81 | 8.89 |
| **c3 :c25** | 0.32 | 6.934 | 9.29 | 7.74 |
| **c9 :c16** | 83.763 | 0.454 | 9.85 | 7.8 |
| **c9 :c23** | 80.763 | 45.111 | 9.44 | 6.42 |
| **c10:c16** | 21.17 | 0.454 | 10.7 | 8.14 |
| **c11:c12** | 17.047 | 103.15 | 11.85 | 8.45 |
| **c11:c21** | 18.047 | 26.912 | 10.13 | 6.51 |
| **c12:c13** | 102.15 | 4.916 | 12.53 | 8.75 |
| **c13:c16** | 0.916 | 0.454 | 11.26 | 6.91 |
| **c13:c21** | 16.916 | 25.912 | 10.11 | 5.45 |
| **c16:c17** | 6.454 | 47.514 | 9.27 | 7.54 |

**Table 3.3**: **List of all QTLs found**. The table lists each fiber quality characteristic, the chromosome, its position in cM on the chromosome, whether or not a QTL has been previously identified at that location, the phenotypic effect of each QTL from the additive model, and the percent of phenotypic variation explained (%VE) by each QTL. ELO2 refers to the QTL found in the two-dimensional scan.

| Trait | Chrom. | Pos (cM) | Known QTL | Closest SNP | Effect | %VE |
|---|---|---|---|---|---|---|
| Mic | 4 | 59.7 | N/A | SA04_88479487 | -0.2623 | 0.9793 |
| Mic | 7 | 71 | N/A | SA07_94730177 | -0.2634 | 0.7455 |
| Mic | 8 | 105.5 | N/A | SA08_95567902 | 0.2576 | 1.8234 |
| Mic | 11 | 45.3 | N/A | SA11_19747786 | 0.26606 | 0.7388 |
| Mic | 13 | 50.9 | N/A | SA13_93133823 | -0.0026 | 1.4466 |
| UI | 8 | 109.5 | qFU08.2 | SA08_117743844 | -0.9427 | 4.266 |
| UI | 10 | 77.6 | N/A | SA10_114885133 | 0.297 | 1.859 |
| UI | 13 | 23.9 | qFU-C13-1 | SA13_915703 | -0.337 | 3.75 |
| UI | 20 | 25.3 | N/A | SD07_43059834 | -0.3558 | 5.192 |
| STR | 3 | 80.2 | qFBS-c3 | SA03_109415594 | 0.57099 | 0.7098 |
| STR | 7 | 91 | N/A | SA07_94730177 | -1.9048 | 3.8628 |
| STR | 8 | 111.5 | N/A | SA08_117743844 | -0.4943 | 0.91 |
| STR | 11 | 90.3 | qFS11.1 | SA11_121748774 | -1.5391 | 0.9645 |
| STR | 13 | 17.9 | N/A | SA13_915703 | -0.0786 | 13.246 |
| STR | 15 | 36.7 | qFS-D1-1 | SD02_55224952 | 0.32258 | 4.4629 |
| STR | 18 | 41.6 | N/A | SD05_26000948 | -0.6846 | 0.4953 |
| ELO | 2 | 16.12 | qFE-A2-1 | SA02_12220430 | -0.3221 | 5.8656 |
| ELO | 5 | 57.76 | N/A | SA05_89129732 | 0.10133 | 1.9987 |
| ELO | 8 | 56.46 | N/A | SA08_85966778 | 0.2601 | 3.3554 |
| ELO | 11 | 69.33 | N/A | SA11_19747786 | 0.56885 | 2.5581 |
| ELO | 12 | 9.64 | qFE-A12-1 | SA12_7041864 | -0.1957 | 2.7183 |
| ELO | 13 | 58.92 | qFE-C13-1 | SA13_93133823 | 0.05705 | 0.0259 |
| ELO | 17 | 15.55 | N/A | SD04_7484036 | -0.1542 | 0.9584 |
| ELO | 18 | 9.63 | qFE-C18-1 | SD05_12605755 | -0.1245 | 1.655 |
| ELO2 | 2 | 30.1 | qFE-A2-1 | SA02_32206877 | -0.6225 | 15.261 |
| ELO2 | 5 | 102.8 | N/A | SA05_105477515 | -0.0931 | 1.1445 |
| ELO2 | 17 | 20.5 | N/A | SD04_7484036 | -0.0188 | 0.3737 |
| ELO2 | 24 | 61.5 | N/A | SD11_62849237 | -0.205 | 2.0279 |

**Table 3.3: List of all QTLs found (cont.)**

| Trait | Chrom. | Pos (cM) | Known QTL | Closest SNP | Effect | %VE |
|---|---|---|---|---|---|---|
| SFC | 3 | 85.2 | N/A | SA03_109415594 | -0.06093 | 0.2069 |
| SFC | 7 | 73 | N/A | SA07_94730177 | 0.22564 | 1.0655 |
| SFC | 8 | 113.5 | N/A | SA08_117743844 | 0.36213 | 2.4943 |
| SFC | 10 | 71.6 | N/A | SA10_39628043 | -0.12109 | 0.6435 |
| SFC | 11 | 56.3 | N/A | SA11_19747786 | 0.38071 | 2.9587 |
| SFC | 13 | 31.9 | N/A | SA13_915703 | 0.67888 | 3.2611 |
| SFC | 15 | 25.7 | N/A | SD02_10715660 | -0.09796 | 0.7628 |
| SFC | 18 | 17.6 | N/A | SD05_12605797 | -0.09562 | 1.2406 |
| SSA | 4 | 34.7 | N/A | SA04_14119735 | 3.525 | 3.811 |
| SSA | 5 | 52.8 | N/A | SA05_89129732 | 3.382 | 2.5564 |
| SSA | 7 | 86 | N/A | SA07_94730177 | -1.5777 | 9.2113 |
| SSA | 8 | 38.5 | N/A | SA08_9460644 | -0.2756 | 0.0276 |
| SSA | 12 | 59.6 | N/A | SA12_30016746 | -2.2693 | 8.786 |
| SSA | 13 | 37.9 | N/A | SA13_915703 | -7.8306 | 4.8619 |
| SSA | 25 | 40.7 | N/A | SD12_58998289 | 1.9197 | 7.1284 |
| FW | 3 | 72.2 | N/A | SA03_41186322 | 1.5827 | 0.5041 |
| FW | 7 | 69 | N/A | SA07_45482778 | -3.5307 | 3.3093 |
| FW | 10 | 65.6 | N/A | SA10_39628043 | 3.0833 | 9.2319 |
| FW | 12 | 60.6 | N/A | SA12_30016746 | -3.7918 | 2.8359 |
| FW | 13 | 57.9 | N/A | SA13_93133823 | 3.8861 | 1.5292 |
| FW | 22 | 48 | N/A | SD09_44966934 | 1.2737 | 0.37 |
| SW | 3 | 74.2 | N/A | SA03_41186322 | 5.0179 | 3.3307 |
| SW | 4 | 58.7 | N/A | SA04_88479487 | 6.9319 | 3.8207 |
| SW | 7 | 79 | N/A | SA07_94730177 | -0.5048 | 2.5755 |
| SW | 8 | 61.5 | N/A | SA08_85966778 | -3.2639 | 0.7172 |
| SW | 10 | 79.6 | N/A | SA10_114885133 | 4.5481 | 10.2421 |

**References**

Abbas A., Iqbal M.A., Rahman M.U., Paterson A.H. (2015) Estimating genetic diversity among selected cotton genotypes and the identification of DNA markers associated with resistance to cotton leaf curl disease. Turkish Journal of Botany 39:1033-1041. DOI: 10.3906/bot-1505-22.

Adhikari J., Das S., Wang Z.N., Khanal S., Chandnani R., Patel J.D., Goff V.H., Auckland S., Rainville L.K., Jones D., Paterson A.H. (2017) Targeted identification of association between cotton fiber quality traits and microsatellite markers. Euphytica 213. DOI: 10.1007/s10681-017-1853-0.

Auld D.L., Bechere E., Krifa M., Kebede H., Hequet E., Wright R., Misra S. (2007) Registration of 'Raider 276', a High-Yielding, Improved-Quality Upland Mutant Cotton Cultivar. Journal of Plant Registrations 1:115-116. DOI: 10.3198/jpr2007.01.0059crc.

Auld D.L., Heikkinen M.K., Erickson D.A., Sernyk J.L., Romero J.E. (1992) RAPESEED MUTANTS WITH REDUCED LEVELS OF POLYUNSATURATED FATTY-ACIDS AND INCREASED LEVELS OF OLEIC-ACID. Crop Science 32:657-662. DOI: 10.2135/cropsci1992.0011183X003200030016x.

Bechere E., Auld D.L., Cantrell R.G., Hequet E., Krifa M., Misra S., Smith C.W. (2007) Registration of TTU 0774-3-3 and TTU 0808-1-6-1 Upland Cotton Germplasm Lines with Improved Fiber Length and Strength. Journal of Plant Registrations 1:58-59. DOI: 10.3198/jpr2006.09.0587crg.

Bechere E., Auld D.L., Krifa M., Smith C.W., Cantrell R.G. (2011) Registration of TTU 0782, an Upland Cotton Germplasm Line with Superior Fiber Quality. Journal of Plant Registrations 5:207-210. DOI: 10.3198/jpr2010.07.0404crg.

Blixt S., Ehrenberg L., Gelin O. (1966) STUDIES OF INDUCED MUTATIONS IN PEAS .18. EMS-TREATMENT OF DIFFERENT DURATION OF VARIETIES WEITOR AND WITHAM WONDER. Agri Hortique Genetica 24:138-+.

Bradbury P.J., Zhang Z., Kroon D.E., Casstevens T.M., Ramdoss Y., Buckler E.S. (2007) TASSEL: software for association mapping of complex traits in diverse samples. Bioinformatics 23:2633-5. DOI: 10.1093/bioinformatics/btm308.

Broman K.a.S., Saunak. (2009) A Guide to QTL Mapping with R/qtl Springer Science+Business Media, LLC.

Brown A., Paterson A.H., Li L. (2012) Genomics and Breeding in Food Crops.

Cantrell R.G., Roberts C.L., Waddell C. (2000) Registration of 'Acala 1517-99' cotton. Crop Science 40:1200-1201. DOI: 10.2135/cropsci2000.0049rcv.

Chandnani R., Kim C., Guo H., Shehzad T., Wallace J.G., He D.H., Zhang Z.S., Patel J.D., Adhikari J., Khanal S., Paterson A.H. (2018) Genetic Analysis of Gossypium Fiber Quality Traits in Reciprocal Advanced Backcross Populations. Plant Genome 11:15. DOI: 10.3835/plantgenome2017.06.0057.

Chandnani R., Zhang Z.S., Patel J.D., Adhikari J., Khanal S., He D.H., Brown N., Chee P.W., Paterson A.H. (2017) Comparative genetic variation of fiber quality traits in reciprocal advanced backcross populations. Euphytica 213. DOI: 10.1007/s10681-017-2029-7.

Chee P.W., Draye X., Jiang C.X., Decanini L., Delmonte T.A., Bredhauer R., Smith C.W., Paterson A.H. (2005) Molecular dissection of phenotypic variation between Gossypium hirsutum and Gossypium barbadense (cotton) by a backcross-self approach: III. Fiber length. Theoretical and Applied Genetics 111:772-781. DOI: 10.1007/s00122-005-2062-0.

Chen Z.J., Scheffler B.E., Dennis E., Triplett B.A., Zhang T.Z., Guo W.Z., Chen X.Y., Stelly D.M., Rabinowicz P.D., Town C.D., Arioli T., Brubaker C., Cantrell R.G., Lacape J.M., Ulloa M., Chee P., Gingle A.R., Haigler C.H., Percy R., Saha S., Wilkins T., Wright R.J., Van Deynze A., Zhu Y.X., Yu S.X., Abdurakhmonov I., Katageri I., Kumar P.A., Mehboob ur R., Zafar Y., Yu J.Z., Kohel R.J., Wendel J.F., Paterson A.H. (2007) Toward Sequencing cotton (Gossypium) Genomes. Plant Physiology 145:1303-1310. DOI: 10.1104/pp.107.107672.

Chen Z.J., Sreedasyam A., Ando A., Song Q., De Santiago L.M., Hulse-Kemp A.M., Ding M., Ye W., Kirkbride R.C., Jenkins J., Plott C., Lovell J., Lin Y.-M., Vaughn R., Liu B., Simpson S., Scheffler B.E., Wen L., Saski C.A., Grover C.E., Hu G., Conover J.L., Carlson J.W., Shu S., Boston L.B., Williams M., Peterson D.G., McGee K., Jones D.C., Wendel J.F., Stelly D.M., Grimwood J., Schmutz J. (2020) Genomic diversifications of five Gossypium allopolyploid species and their impact on cotton improvement. Nature Genetics 52:525-533. DOI: 10.1038/s41588-020-0614-5.

Elshire R.J., Glaubitz J.C., Sun Q., Poland J.A., Kawamoto K., Buckler E.S., Mitchell S.E. (2011) A Robust, Simple Genotyping-by-Sequencing (GBS) Approach for High Diversity Species. Plos One 6. DOI: 10.1371/journal.pone.0019379.

Espina M.J., Ahmed C.M.S., Bernardini A., Adeleke E., Yadegari Z., Arelli P., Pantalone V., Taheri A. (2018) Development and Phenotypic Screening of an Ethyl Methane Sulfonate Mutant Population in Soybean. Frontiers in Plant Science 9. DOI: 10.3389/fpls.2018.00394.

Favret E.A. (1960) Somatic mutations of four genes for albinism in barley induced by X-rays and ethyl methane-sulphonate. Hereditas 46:622-634.

Gichner T., Ehrenberg L. (1966) INFLUENCE OF POST-TREATMENT STORAGE ON FREQUENCY OF EMS-INDUCED CHROMOSOMAL ABERRATIONS IN BARLEY. Biologia Plantarum 8:256-+. DOI: 10.1007/bf02930787.

Greene E.A., Codomo C.A., Taylor N.E., Henikoff J.G., Till B.J., Reynolds S.H., Enns L.C., Burtner C., Johnson J.E., Odden A.R., Comai L., Henikoff S. (2003) Spectrum of chemically induced mutations from a large-scale reverse-genetic screen in arabidopsis. Genetics 164:731-740.

Grover C.E., Yoo M.J., Lin M., Murphy M.D., Harker D.B., Byers R.L., Lipka A.E., Hu G.J., Yuan D.J., Conover J.L., Udall J.A., Paterson A.H., Gore M.A., Wendel J.F. (2020) Genetic Analysis of the Transition from Wild to Domesticated Cotton (Gossypium hirsutum L.). G3-Genes Genomes Genetics 10:731-754. DOI: 10.1534/g3.119.400909.

Haidar S., Aslam M., Ul Haq M.A. (2016) NIAB-852: A NEW HIGH YIELDING AND BETTER FIBRE QUALITY COTTON MUTANT DEVELOPED THROUGH POLLEN IRRADIATION TECHNIQUE. Pakistan Journal of Botany 48:2297-2305.

Herring A.D., Auld D.L., Ethridge M.D., Hequet E.F., Bechere E., Green C.J., Cantrell R.G. (2004) Inheritance of fiber quality and lint yield in a chemically mutated population of cotton. Euphytica 136:333-339. DOI: 10.1023/b:euph.0000032747.97343.54.

Hildering G.J., Vanderve.Jh. (1966) MUTUAL INDEPENDENCE OF M1-FERTILITY AND MUTANT YIELD IN EMS TREATED TOMATOES. Euphytica 15:412-+. DOI: 10.1007/bf00022188.

Ijaz B., Zhao N., Kong J., Hua J.P. (2019) Fiber Quality Improvement in Upland Cotton (Gossypium hirsutum L.): Quantitative Trait Loci Mapping and Marker Assisted Selection Application. Frontiers in Plant Science 10. DOI: 10.3389/fpls.2019.01585.

Janga M.R., Campbell L.M., Rathore K.S. (2017) CRISPR/Cas9-mediated targeted mutagenesis in upland cotton (Gossypium hirsutum L.). Plant Molecular Biology 94:349-360. DOI: 10.1007/s11103-017-0599-3.

Jiang C.X., Wright R.J., El-Zik K.M., Paterson A.H. (1998) Polyploid formation created unique avenues for response to selection in Gossypium (cotton). Proceedings of the National Academy of Sciences of the United States of America 95:4419-4424. DOI: 10.1073/pnas.95.8.4419.

Lopes C.M.L., Suassuna N.D., Cares J.E., Gomes A., Perina F.J., Nascimento G.F., Mendonca J.S.F., Moita A.W., Carneiro R. (2020) Marker-assisted selection in Gossypium spp. for Meloidogyne incognita resistance and histopathological characterization of a near immune line. Euphytica 216. DOI: 10.1007/s10681-020-2554-7.

Luckett D.J. (1989) COLCHICINE MUTAGENESIS IS ASSOCIATED WITH SUBSTANTIAL HERITABLE VARIATION IN COTTON. Euphytica 42:177-182. DOI: 10.1007/bf00042630.

Mansoor S., Paterson A.H. (2012) Genomes for jeans: cotton genomics for engineering superior fiber. Trends in Biotechnology 30:521-527. DOI: 10.1016/j.tibtech.2012.06.003.

Mei M., Syed N.H., Gao W., Thaxton P.M., Smith C.W., Stelly D.M., Chen Z.J. (2004) Genetic mapping and QTL analysis of fiber-related traits in cotton (Gossypium). Theoretical and Applied Genetics 108:280-291. DOI: 10.1007/s00122-003-1433-7.

Parry M.A.J., Madgwick P.J., Bayon C., Tearall K., Hernandez-Lopez A., Baudo M., Rakszegi M., Hamada W., Al-Yassin A., Ouabbou H., Labhilili M., Phillips A.L. (2009) Mutation discovery for crop improvement. Journal of Experimental Botany 60:2817-2825. DOI: 10.1093/jxb/erp189.

Patel J.D., Wright R.J., Auld D., Chandnani R., Goff V.H., Ingles J., Pierce G.J., Torres M.J., Paterson A.H. (2014) Alleles conferring improved fiber quality from EMS mutagenesis of elite cotton genotypes. Theoretical and Applied Genetics 127:821-830. DOI: 10.1007/s00122-013-2259-6.

Paterson A.H., Boman R.K., Brown S.M., Chee P.W., Gannaway J.R., Gingle A.R., May L., Smith C.W. (2004) Reducing the genetic vulnerability of cotton. Crop Science 44:1900-1901. DOI: 10.2135/cropsci2004.1900.

Paterson A.H., C. L. Brubaker, and J.F. Wendel. (1993) A rapid method for extraction of cotton (Gossypium spp.) genomic DNA suitable for RFLP or PCR analysis. Plant Molecular Biology 11:122-127. DOI: 10.1007/ BF02670470.

Qin L., Li J.Y., Wang Q.Q., Xu Z.P., Sun L., Alariqi M., Manghwar H., Wang G.Y., Li B., Ding X., Rui H.P., Huang H.M., Lu T.L., Lindsey K., Daniell H., Zhang X.L., Jin S.X. (2020) High-efficient and precise base editing of C center dot G to T center dot A in the allotetraploid cotton (Gossypium hirsutum) genome using a modified CRISPR/Cas9 system. Plant Biotechnology Journal 18:45-56. DOI: 10.1111/pbi.13168.

Rao S.H.K., Sears E.R. (1962) EMS-INDUCED MUTATIONS IN HEXAPLOID WHEAT. Genetics 47:983-&.

Rong J., Feltus E.A., Waghmare V.N., Pierce G.J., Chee P.W., Draye X., Saranga Y., Wright R.J., Wilkins T.A., May O.L., Smith C.W., Gannaway J.R., Wendel J.R., Paterson A.H. (2007) Meta-analysis of polyploid cotton QTL shows unequal contributions of subgenomes to a complex network of genes and gene clusters implicated in lint fiber development. Genetics 176:2577-2588. DOI: 10.1534/genetics.107.074518.

Saha S., Bellaloui N., Jenkins J.N., McCarty J.C., Stelly D.M. (2020) Effect of chromosome substitutions fromGossypium barbadenseL.,G. tomentosumNutt. Ex Seem andG. mustelinumWatt intoG. hirsutumL. on cottonseed protein and oil content. Euphytica 216. DOI: 10.1007/s10681-020-02644-4.

Said J.I., Lin Z.X., Zhang X.L., Song M.Z., Zhang J.F. (2013) A comprehensive meta QTL analysis for fiber quality, yield, yield related and morphological traits, drought tolerance, and disease resistance in tetraploid cotton. Bmc Genomics 14. DOI: 10.1186/1471-2164-14-776.

Said J.I., Song M.Z., Wang H.T., Lin Z.X., Zhang X.L., Fang D.D., Zhang J.F. (2015) A comparative meta-analysis of QTL between intraspecific Gossypium hirsutum and interspecific G. hirsutum x G. barbadense populations. Molecular Genetics and Genomics 290:1003-1025. DOI: 10.1007/s00438-014-0963-9.

Smith C.W. (2003) Registration of TAM 94L-25 and TAM 94J-3 germplasm lines of upland cotton with improved fiber length. Crop Science 43:742-743. DOI: 10.2135/cropsci2003.0742.

Talebi A.B., Shahrokhifar, B. (2012) Ethyl methane sulphonate (EMS) induced mutagenesis in Malaysian rice (cv. MR219) for lethal dose determination. American Journal of Plant Sciences.

Ulloa M., De Santiago L.M., Hulse-Kemp A.M., Stelly D.M., Burke J.J. (2020) Enhancing Upland cotton for drought resilience, productivity, and fiber quality: comparative evaluation and genetic dissection. Molecular Genetics and Genomics 295:155-176. DOI: 10.1007/s00438-019-01611-6.

Vanderve.Jh. (1967) EMS INDUCED NON-RECESSIVE EMBRYONIC LETHALS IN ARABIDOPSIS THALIANA. Genetica 38:61-&.

Wendel J.F. (2000) Genome evolution in polyploids. Plant Molecular Biology 42:225-249. DOI: 10.1023/a:1006392424384.

Yu J., Jung S., Cheng C.-H., Ficklin S.P., Lee T., Zheng P., Jones D., Percy R.G., Main D. (2014) CottonGen: a genomics, genetics and breeding database for cotton research. Nucleic Acids Research 42:D1229-D1236. DOI: 10.1093/nar/gkt1064.

Zhang K., Kuraparthy V., Fang H., Zhu L.L., Sood S., Jones D.C. (2019a) High-density linkage map construction and QTL analyses for fiber quality, yield and morphological traits using CottonSNP63K array in upland cotton (Gossypium hirsutum L.). Bmc Genomics 20. DOI: 10.1186/s12864-019-6214-z.

Zhang Z., Liu A.Y., Huang Z., Fan S.M., Zou X.Y., Deng X.Y., Ge Q., Gong J.W., Li J.W., Gong W.K., Shi Y.Z., Fan L.Q., Zhang Z.B., Jiang X., Lei K., Yuan Y.L., Xu A.X., Shang H.H. (2019b) Quantitative trait locus mapping for plant height and branch number in an upland cotton recombinant inbred line with an SNP-based high-density genetic map. Euphytica 215. DOI: 10.1007/s10681-019-2428-z.

# CHAPTER 4

# HYBRIDIZATION OF COTTON LINES DEVELOPED BY MUTAGENESIS AND COTTON LINES DEVELOPED BY QTL STACKING

## Introduction

Cotton is an extremely important global crop which provides approximately 50% of the raw material for the textile industry as well as contributing to the medical industry and the agricultural industry, with cottonseed being an important feed source for livestock (Bi et al., 2016; Rong et al., 2007). While four *Gossypium* species are cultivated, 95% of the global output is a single species, *Gossypium hirsutum,* which is native to Central America. Cotton is an important source of income for numerous countries around the world, including the US. As of 2018, the US is the third largest cotton fiber producing country and the largest cotton fiber exporting country, with 18.59 million bales produced and 15 million bales exported (Chandnani et al., 2017b). This has resulted in a multi-billion dollar industry in the US alone (Chandnani et al., 2017a). However, the US has only been able to remain competitive in the international market because of superior fiber quality, which has also been continuing to improve in other cotton producing countries (Zhang et al., 2019). Therefore, it is extremely pertinent that the US market can continue to improve and increase cotton fiber quality and yield.

Previously, breeders primarily focused on improved yield with some resources dedicated to improving length and fineness (Perkins, 1984). Recently, however, there have been several reasons for the necessity of increased of fiber quality in in addition to elevated yield. The improvement towards more efficient spinning technologies as well as increased competition

from synthetic fibers have caused the improvement of other fiber quality characteristic to be desirable. The most important fiber quality metrics are length, strength and fineness, although uniformity and short fiber content have also been heavily selected for as well, because these traits allow for more efficiency in the process of spinning fiber into yarn while also producing less waste (Chee et al., 2005; Patel et al., 2014). However, the previous narrow focus on improving yield, may have contributed to a stagnation in fiber quality improvement, especially since it has been shown that many yield and quality characteristics are often negatively correlated (Brown et al., 2019b). This stagnation is likely due to a lack of genetic diversity resulting from a long history of inbreeding few elite cotton lines in order to increase the overall yield.

Upland cotton, *Gossypium hirsutum*, is one of the most genetically impoverished crops, likely due to a series of bottlenecks from polyploidization, migration, domestication, and inbreeding (Paterson et al., 2004). In fact, Abbas et al. (2015) found through SSR analysis that most genotypes of domesticated cotton have between 81.1% and 98.7% genetic similarity. Genetic variation is important because it can help increase resistance to pests, drought and other biotic and abiotic stresses. Also, the obstacle of this narrow genetic base must be overcome in order to further cotton fiber improvement. However, it is possible that only a fraction of the cotton genetic base has been exploited and much of the useful gene pool may be unknown (Bowman et al., 2003; de Menezes et al., 2015). In addition, to discovering genetic variation within commercial cultivars, there is also a lot of variation to be exploited wild relative. Esbroeck (1998) estimate that that less than 1% of all genetic variability in cotton resides in US commercial cultivars. Therefore, there is a lot of potential for increasing fiber yield and quality.

There are numerous ways to introduce or increase genetic diversity into a cultivar, but one of the most common methods is through hybridization, either from other cultivars of the same species, wild relatives of the same species or from different species. Hybridization is the main component of all plant breeding programs and the more common way to improve desirable traits in all crops. In present breeding programs, hybridization helps with the introgression of specific traits and elite cultivars of all cultivated species have histories of hybridization (Chandnani et al., 2018). Most commonly, *Gossypium barbadense* and *Gossypium hirsutum* are hybridized (interspecific hybridization) and are often used for mapping. These hybrids are selfed and/or backcrossed for multiple generations to achieve improved quality in the targeted traits. Then the progeny of these crosses between elite cultivars is selected by pedigree for further fiber quality improvement (Chandnani et al., 2017b). Intraspecific hybridization occurs as well as interspecific hybridization, however intraspecific hybridization has occurred more commonly in *G. barbadense* rather than *G. hirsutum* (Chandnani et al., 2017b). Overall, hybridization has been and continues to be a major source of selective variation for important traits in cotton and all crops (Chandnani et al., 2017a).

In combination with hybridization, through complex crossing schemes, QTL stacking or QTL pyramiding, can bring variation into a cultivar from multiple different sources. Gene or QTL pyramiding is the assembly of multiple desirable genes or QTLs into a single genotype (Parhe et al., 2017). This technique is commonly used to introgress different types of resistance genes to gain maximum resistance (to pests, stress, environment), but it has also been used to improve economically important characteristics as well (Bokore et al., 2017; Gennaro et al., 2012; Miedaner et al., 2006; Parhe et al., 2017). QTL pyramiding can be used to circumvent breakdown of single gene resistance and instability as well as having a minimal chance of

linkage drag or detrimental genes when crossing parents with similar backgrounds (Bokore et al., 2017; Parhe et al., 2017). This method is also useful when working with QTLs that have too small of an individual effect to be useful in breeding but could be beneficial when combined with other traits (den Boer et al., 2014). This is especially useful in cotton where genetic background has a very strong influence on fiber quality traits, even stronger than environmental effects (Shi et al., 2019). When stacking QTLs, the effects are often additive, however these are often less than the estimated additive effects, which may result from epistasis (Miedaner et al., 2006). This method also helps to understand the interactions between QTLs (den Boer et al., 2014). However, some of the negative components of this method are that it is possible that genes in the same pathway do not show an increased effect when combined, due to epistasis, and there is evidence of negative epistatic interactions as well (den Boer et al., 2014). Overall, QTL stacking has a strong potential to improve economically important traits in cotton.

While hybridization and QTL stacking can introduce variation into a specific cultivar by taking existing variation from somewhere else, chemical mutagens such as ethyl methanesulfonate, are able to create new alleles with point mutations in key genes. Chemical mutagens are able to confer stable single nucleotide changes, as opposed to deletions that have at least some negative or even lethal results (Parry et al., 2009). Ethyl methanesulfonate, or EMS, is a frequently favored chemical mutagen, as is more efficient than irradiation and has a higher survival rate (Favret, 1960). EMS causes a modification of the nucleotides, which then causes guanine to bond with thymine instead of cytosine. The cell will attempt to repair these mismatches but replaces the guanine with adenine instead of replacing the thymine. This results in an overall change from a GC nucleotide pair to an AT nucleotide pair (Greene et al., 2003). This is what is called a point mutation. EMS is also a very simple method of mutagenesis, as it

has a high mutation rate that allows for smaller sample sizes to still be saturated with mutations (Talebi, 2012). This method has been used to cause beneficial mutations in a variety of crops including peas, soybeans, barley, tomatoes, wheat, and other (Blixt et al., 1966; Espina et al., 2018; Gichner and Ehrenberg, 1966; Hildering and Vanderve.Jh, 1966; Rao and Sears, 1962; Vanderve.Jh, 1967). Specifically, this method of mutagenesis has led to the release of numerous improved cotton lines (Auld et al., 2007; Bechere et al., 2007; Bechere et al., 2011; Brown et al., 2012; Herring et al., 2004). Mutagenesis, and specifically EMS, is a simple, useful way to generate novel variation in crops where inbreeding and heavy selection has dramatically decreased the natural genetic variation present.

All these methods can be used to increase genetic diversity as well as improve economically desirable traits. The purpose of this research was to take lines developed from QTL pyramiding and lines developed from mutagenesis and cross them together to potentially increase diversity and fiber quality.

**Germplasm Background**

**QTL Pyramided Lines**

Previously, Waghmare et al. (2005) created a cross between *Gossypium hirsutum* and *Gossypium tomentosum*. *G. tomentosum* is a wild tetraploid cotton species that is endemic to the Hawaiian Islands (Waghmare et al., 2005). The main purpose for this cross was to create a genetic map and compare it to *G. hirsutum* and *G. barbadense* crosses to increase our understanding about the cotton genome. This cross used the *G. hirsutum acc*. TMS-22 and *G. tomentosum acc*. WT936 (Waghmare et al., 2005).

Then Zhang et al. (2011) backcrossed the F2 hybrids to *G. hirsutum* cv. CA3093 and selfed until the BC3F3 generation. This population was used for identifying QTLs in *G.*

*tomentosum* that may provide improved fiber quality to *G. hirsutum*. The BC3F3 generation was crossed to elite varieties, such as GA2004230 and R01-40-08. The F1 hybrids were genotyped to find which individuals had desirable QTLs, mainly for fiber elongation and fineness. In addition, four *G. mustelinum* QTL NIILs (near isogenic introgression lines) in *G. hirsutum* backgrounds were obtained from Dr. Peng-Wah Chee and the Molecular Cotton Breeding Laboratory in Tifton, GA. These lines were selected based on specific target QTLs for improvement of fiber elongation and strength. The *G. mustelinum* lines were in the same backgrounds as the *G. tomentosum* lines. In the fall of 2012, the *Gt* and *Gm* lines were crossed together within the two backgrounds mentioned.

The hybrid *G. tomentosum* and *G. mustelinum* progeny were selfed to the F2 generation and the F2:3 generation was used to validate QTL affects in the two different genetic backgrounds of *G. hirsutum*. The final QTL pyramided lines that were used in creating the hybrids in this research were developed by Sameer Khanal at the University of Georgia Plant Genome Mapping Laboratory in Athens, GA.

**Mutant Lines**

Previously, two varieties of *G. hirsutum*, TAM94L-25 (Smith, 2003) and Acala 1517-99 (Cantrell et al., 2000), were treated with EMS according to Auld et al. (1992) and advanced to the M5 generation by single boll descent (Patel et al., 2014). Around 3000 of these mutant lines were tested for improved fiber quality and yield in Texas and Georgia. 157 lines were found to have at least one improved fiber property. Nine of these 157 lines were selected to be crossed with an elite variety of the Eastern Cotton Belt, GA2004230 (GA230) in the greenhouse in Summer 2012. The F1 hybrids of the mutants and GA230 were then crossed together to stack the individual mutations. The F2 progenies were grown in Watkinsville, GA in 2013. In 2014,

progenies with sufficient seeds were evaluated for improved fiber characteristics in two locations, Watkinsville and Tifton, GA, with two replicates in each location. This is how the lines J29, K75, O037, Q12, K68, P058, and P059 were developed.

In addition to the crosses made above, some of the mutant lines were also crossed to each other to stack the mutations. The parents of these crosses were from different variety backgrounds, so one parent was a mutant of the Acala variety, and one parent was a mutant of the TAM94-L variety. The F1 hybrids of these crosses were then also crossed together to continue to stack mutations and hopefully improve fiber quality. This F2 generation was evaluated in the same two locations as above. This is how the lines R043, R076, S9, S49, S50, S67, S32, and S45 were developed. All the mutant lines were developed by Jinesh Patel at the Plant Genome Mapping Laboratory at the University of Georgia in Athens, GA.

**Materials and Methods**

Seeds from both the QTL stacked lines and the mutant lines were obtained from the Plant Genome Mapping Laboratory in Athens, GA and planted on June 7, 2018, at the UGA Iron Horse Farm in Watkinsville, GA. Crosses between these two populations were made from the beginning of August to the end of September in 2018. Crossing of two *G. hirsutum* parents is done by first emasculating the female parent in the afternoon. Emasculation is done by removing the petals and the anthers of the flower before it is open, so that only the stigma is left. The stigma and the rest of the flower is covered by a wax bag and secured with a twist tie for protection from the elements and insects. The next morning, an open or partially open flower is taken from the male parent and used to lightly brush pollen onto the stigma of the female parent. The cross was then recovered with the wax bag and labeled.

Four lines from the QTL stacked population, 241-036, 241-039, 239-145, and 242-020 were used in these crosses. 241-036 and 241-039 were selected for superior fiber strength and 239-145 and 242-020 were selected for superior fiber fineness. Ten of the mutant lines, J29, K75, O037, Q12, R043, R076, S9, S49, S50 and S67 were used in these crosses. Lines 241-036 and 241-039 and lines S67, S50, K75, R043, and Q12 were crossed to each other and were selected for increased fiber strength. Lines 239-145 and 242-020 and lines O037, S49, J29, R076, and S9 were crossed to each other and were selected for improved fiber fineness. In these crosses, the lines from the QTL stacked population were always the male parent and the mutant lines were always the female parent. Approximately ten crosses were made for each combination, however, some bolls died during the season, therefore some combinations had as few as 7 individual bolls. Bolls were harvested in early December. During harvesting, individual bolls from individual crosses were kept separate.

In 2019 five seeds from two individual bolls from each of the twenty combinations were planted in May at the UGA Iron Horse Farm in Watkinsville, GA, for a total of 40 plots. These seeds represent the $F_1$ generation of these hybrids. These plants were self-pollinated from late July to September of 2019. Self-pollination of cotton plants is done by securing the tips of unopened flowers with a twist-tie, so that they cannot open and be pollinated by insects. These bolls were harvested in December of 2019 and bolls from different plants were kept separate. Fiber samples of 25 bolls were taken from the whole plot, weighed, and ten grams were sent to Cotton Incorporated in Cary, NC for fiber quality analysis. Seeds from self-pollinated bolls were saved to create the F2 generation.

Out of the five seeds that were planted of the F1 generation in 2019, three samples were taken from different plants for DNA extraction and analysis. DNA was extracted according the

methods detailed in Paterson (1993). This method uses fresh, young leaves and a combination of buffers to extract DNA of a size and quality that is suitable for analysis. The DNA was then amplified using a polymerase chain reaction (PCR) and analyzed for confirmation of hybridization on acrylamide gels with silver stain. Hybrids were confirmed with simple sequence repeat markers (SSRs) to identify if paternal alleles were present.

In addition to crossing lines created from EMS and created from QTL stacking, different lines created using EMS were crossed together to attempt to stack mutations and improve multiple fiber quality metrics at once. There were three sets of five lines each, for a total of fifteen mutant lines that were used. The first set contained the lines Q12, K75, R43, S67, and S50 which were crossed to each other. The second set contained the lines J29, O037, R076, S9, and S49 which were all crossed to each other. The third set contained the lines K68, P058, P059, S45, and S32 which were all crossed to each other. Lines in different sets were not crossed to each other. The crosses were made in the same season and location as the mutant/QTL crosses and the F1 progeny were also grown in the same location and time. The analysis was the same for both kinds of crosses, except that the mutant x mutant lines were not confirmed with SSR markers.

To see if there were already differences between the parents and the progeny F1 generation, fiber quality and yield data was analyzed using R. Analysis of variance (ANOVA) and Tukey's honest significant difference tests were used to find if there were any differences between the generations.

**Results**

**Hybrid Confirmation**

Of the 40 Mutant x QTL stacked crosses that were planted in 2019, 26 of them (65%) had all three DNA samples confirm the presence of paternal alleles. Nine other crosses had at least one of the three DNA samples that confirmed the presence of paternal alleles and five crosses did not show any presence of paternal alleles. The crosses that did not show any evidence of paternal alleles were neither replicate from Q12 x 241-039, neither replicate from K75 x 241-039, and one replicate from S9 x 239-145. There were six crosses that only had one of the three DNA samples show evidence of paternal alleles and they were both replicates of S67 x 241-039, one replicate of S50 x 241-036, one replicate of S49 x 242-020, one replicate of O037 x 242-020, and one replicate of S9 x 242-020. There were three crosses where two of the three DNA samples had evidence of paternal alleles which were one replicate of K75 x 241-036, one replicate of J29 x 242-020, and one replicate of Q12 x 241-036. All other crosses had all three DNA samples show the presence of paternal alleles.

**Statistical Analysis- QTL Stacked crossed with Mutant lines**

The statistical tests were done to evaluate if there were any significant differences between the parental generation and the F1 progeny in any fiber characteristics. The averages and significance values for all traits with the parents K75, Q12, R43, S50, S67, 241-036 or 241-039 can be found in Table 4.1. Specific p-values for all crosses can be found in supplemental table 4.6. Q12 x 241-036 had a significantly higher lint percentage than the paternal parent, 241-036. The two parents were also significantly different from each other, but not the progeny, in fineness and length. Q12 x 241-039 had significant differences from the parents in fineness (MIC), length (UHM), uniformity (UI), and lint percent. The cross had a significantly higher

fineness and lint percent than 241-039 (the paternal parent) and was not significantly different from the maternal parent (Q12). The cross was significantly shorter than both parents and had a significantly lower uniformity than the paternal parent.

K75 x 241-039 was significantly higher than the paternal parent, 241-039, in both MIC and lint percent. K75 x 241-036 was also significantly higher than its paternal parent, 241-036, in MIC and lint percent. It is important to note that for fiber fineness, a lower value is considered more desirable. In addition, the two parents were significantly different in UHM, but the progeny did not differ from either parent.

R043 x 241-039 had a significantly higher lint percent than 241-039. In addition, the two parents were significantly different in uniformity index (UI), but the progeny was not significantly different from either parent. R043 x 241-036 also had a significantly higher lint percent than its paternal parent, 241-036.

Neither S50 x 241-039 or S50 x 241-036 had any significant differences between the progeny and the parents. However, S50 and 241-036 were significantly different in UHM but the cross was not significantly different from either parent.

S67 x 241-039 had significant differences from the parents in UI, ELO, and lint percent. The cross had a significantly lower UI than the paternal parent (241-039), had a higher ELO than the maternal parent (S67), and had a higher lint percent than both parents. S67 x 241-036 had significant differences from the parents in UHM and lint percent. The progeny was significantly longer than the paternal parent (241-036) and had a significantly higher lint percent than both parents.

The averages for all traits containing the parents J29, O037, R076, S9, S49, 242-020 or 239-145 can be found in Table 4.2. Specific p-values for all crosses can also be found in

supplemental table 1. J29 x 242-020 had no significant differences from either parent in any fiber quality characteristics, but the two parents were significantly different in UHM. J29 x 239-145 had significant differences from the parents in UI and SFI. The progeny had a significantly higher uniformity than the maternal parent, J29, but had a lower SFI than J29 as well.

O037 x 242-020 did not have any significant differences from either parent in any fiber quality characteristics. O037 x 239-145 had a significantly higher MIC than the paternal parent, 239-145.

S9 x 239-145 had significantly higher strength than its paternal parent, 239-145. S9 x 242-020 had significant differences from the parents in STR and lint %. The cross was higher than the paternal parent (242-020) in fiber strength and had significantly higher lint percent than the maternal parent (S9). In addition, the two parents were significantly different in UHM and SFI but were not significantly different from the progeny.

S49 x 242-020 did not have significant differences between the parents and the progeny, however the two parents were significantly different in UHM, UI, STR, and SFI. The cross did not differ from either parent in any of these metrics. S49 x 239-145 had significantly lower SFI than both parents. Lower SFI, short fiber index, is also preferred.

R076 x 239-145 had significant differences from the parents in MIC, STR, and lint percent. The progeny was significantly higher than the maternal parent, R076, in STR and lint percent. The progeny was also significantly higher than both parents for MIC. In addition, the two parents were significantly different from each other in uniformity index (UI). R076 x 242-020 had significant differences from the parents in MIC, UHM, and lint percent. The progeny had significantly higher MIC than both parents, was significantly longer than the paternal parent (242-020) and had a significantly higher lint percent than the maternal parent.

**Statistical Analysis- Mutants lines crossed to each other**

The first set of mutants contains lines Q12, K75, R43, S50, and S67 which were all crossed to each other. The averages of all traits can be found in Table 4.3 and p-values for any differences stated can be found in supplemental table 4.6. Q12 X K75 did not have any significant differences between the progeny and the parents. However, in lint percent, the two parents were significantly different from each other, but the progeny was not significantly different from either parent. Q12 X S67 had significant differences between the parents and the progeny in lint percent. The progeny had a significantly higher lint percent than the paternal parent S67 but was significantly lower than the maternal parent, Q12.

R43 X Q12 had significant differences between the progeny and the parents in MIC and lint percent. The progeny had a higher fiber fineness than the maternal parent, R43, and had a lower lint percent than the paternal parent, Q12. R43 X S50 had significantly higher lint percent than the paternal parent, S50.

K75 X R43 had a significantly lower SFI than both parents. In addition, for fiber strength, the parents were significantly different from each other, but the progeny was not significantly different from either parent. K75 X S67 had significant difference between the parents and progeny in SFI and lint percent. The progeny had a lower SFI than both parents and had a higher lint percent than the paternal parent, S67. In addition, for elongation (ELO) the parents were significantly different from each other, but the progeny was not significantly different from either parent.

S50 X K75 had a significantly lower SFI than both parents. In addition, the two parents were significantly different from each other in MIC, STR, and lint percent, but the progeny was not significantly different from either parent. S50 X Q12 had a significantly higher lint percent

than the maternal parent, S50, but was significantly lower than the maternal parent, Q12. The two parents significantly differed in MIC, but the progeny did not differ from either parent.

S67 X R43 had significant differences in SFI and lint percent. The cross had significantly lower SFI than both parents and a significantly higher lint percent than the maternal parent, S67. S67 X S50 was significantly lower than both parents for SFI. The parents were also significantly different from each other in ELO, but the progeny was not significantly different from either.

The second set of mutant hybrids contained the lines J29, R76, S9, O037, and S49 which were all crossed to each other. The averages of all traits can be seen in Table 4.4 and p-values for any differences stated can be found in supplemental table 4.6. J29 X R76 had significant differences in lint percent and SFI. The progeny had a significantly lower SFI than both parents and a significantly higher lint percent than the paternal parent, R76. J29 X S9 had no significant differences between the parents and the progeny in any fiber characteristics. However, the parents were significantly different from each other in uniformity index (UI) and STR while the progeny was not different from either parent.

O037 X J29 had no significant differences between the progeny and the parents or between the two parents. O037 X S49 had a significantly lower SFI than the maternal parent, O037. In addition, the parents were significantly different from each other in STR, UI, and UHM, but the progeny was not significantly different from either parent.

R76 X O037 had significant differences between the parents and the progeny in MIC, UHM, STR, SFI, and lint percent. The progeny had a higher MIC and lint percent than the maternal parent, R076. The progeny was higher than the paternal parent, O037, in UHM and STR. The progeny had significantly lower SFI than both parents. R76 X S49 had significant differences between the progeny and the parents in STR, SFI, and lint percent. The progeny had

significantly higher STR and LP than the maternal parent, R76, and had a significantly lower SFI than the maternal parent as well. In addition, the two parents were significantly different from each other in UI, but the progeny was not significantly different from either parent.

S9 X O037 had significant differences between the progeny and the parents in UHM, STR, and SFI. For UHM and STR, the progeny was significantly higher than the paternal parent, O037. For SFI, the progeny was significantly lower than both parents. S9 X R76 was significantly lower than the paternal parent in SFI. In addition, the two parents were significantly different in STR, and the progeny did not differ from either parent.

S49 X J29 had a significantly lower SFI than the paternal parent, J29. The two parents were also significantly different in STR, but the progeny was not significantly different from either parent. S49 X S9 had no significant differences in any fiber quality characteristics.

The third set of mutant hybrids crossed the lines K68, S45, P058, P059, and S32 to each other. The averages of all traits can be seen in Table 4.5 and p-values for any differences stated can be found in supplemental table 4.6. K68 X S45 had a significantly higher MIC than both parents. The two parents differed in lint percent, but the progeny did not significantly differ from either parent. K68 X P058 did not have any significant differences in any fiber quality characteristics.

P058 X S32 did not have any significant differences in any fiber quality characteristics. P058 X S45 had significant differences between the parents and the progeny in STR and SFI. The progeny had a higher STR than both parents and had a lower SFI than the paternal parent. The parents were significantly different from each other in MIC, but the progeny did not differ from either parent.

P059 X K68 had significant differences between the parents and the progeny in UI and SFI. The progeny had a significantly higher UI than the paternal parent, K68, and a significantly lower SFI than the paternal parent. P059 X P058 had no significant differences in any of the fiber quality characteristics.

S32 X K68 had no significant differences between the parents and the progeny. However, for UI the two parents were significantly different, but the progeny did not differ from either parent. S32 X P059 had no significant differences in any fiber quality characteristics.

S45 X P059 had a significantly lower SFI than the maternal parent, S45. The two parents were significantly different in MIC, but the progeny did not differ from either parent. S45 X S32 did not have any significant differences in any fiber quality characteristics.

**Discussion**

**Mutant x QTL stacked lines**

For the Mutant x QTL stacked hybrids, in general, fiber fineness was higher in the hybrid than in the paternal parent (lower is better) which is to be expected since the paternal parent was selected for superior fiber fineness. Also, generally, lint percent of the progeny was significantly higher than the paternal parents 241-039 and 241-036. When SFI was significantly different it was lower than both parents. Strength showed significant differences in three crosses, where the paternal parents were 242-020 and 239-145, and it was always higher in the progeny. These were not the lines that were selected for superior fiber strength. In crosses where the paternal parents were either 242-020 or 239-145, lint percent was significantly different in three crosses and the progeny was always higher than the maternal parent. Additionally, in the first set of mutant and QTL crosses, the best cross for UHM was from the mutant S67, which was also the best UHM mutant, even though it did not exceed the parental UHM. The QTL-stacked line 241-

036 consistently conferred higher strength than 241-039 and the highest overall strength came from 241-036 X K75. For the second set, of mutant and QTL crosses, QTL line 239-145 consistently conferred higher UHM than 242-020 and the highest lengths were S9 X 239-145 and S49 X 239-145 at 1.26in. S9 X 239-145 was also an overall very good cross as it was not only superior in fiber length, but also had the lowest MIC and highest strength. However, lint percent was low in this cross.

Approximately two-thirds (65%) of the crosses were able to be fully confirmed. This could be due to several reasons, such as contamination from insect pollination or self-pollination or a failed cross. Another reason for this could be that it was not necessarily a failed cross, but the cross was just not able to be confirmed with the SSR markers used. There was only a small number of SSRs used for confirmation, eight different markers for each set. These markers were also specific to the introgressions that were previously crossed into the paternal parent. Therefore, hybridization could have occurred but if the specific introgressed QTLs did not get passed down, then the hybridization would not have been able to be confirmed. However, the introgressions are responsible for improved fiber quality of the paternal parents, so it would likely be the most beneficial if they were present. It is also possible that recombination occurred between the markers and the specific QTLs. Therefore, the beneficial paternal QTLs would still be present, but not able to be found with the SSR markers used because the markers themselves are not present.

**Mutant x Mutant lines**

In the first set of mutant crosses, containing the lines Q12, K75, R43, S67, and S50, there were no differences in length or uniformity in any cross. The most common differences between parents and progeny occurred in short fiber index (SFI) and lint percent. SFI had differences in

six out of ten crosses and five of those were significantly less than both parents. The hybrids consistently had lower SFI, which is preferred, than all parents other than Q12. For lint percent, the hybrids consistently had a lower lint percent than the parent Q12, but a higher lint percent than the parents S50 and S67. When there were differences in MIC, ELO, and STR, they were generally only between the two parents and the progeny were in between the two parents. Additionally, crosses where the maternal parent was S67, had the best MIC. S67 also had the highest length and all its crosses had even higher length, including the two-best overall (S67 X R43, S67 X S50). Mutant K75 had the best overall strength and consistently produced crosses with the best strength as well.

In the second set of mutant crosses, containing the lines J29, O037, R076, S9, and S49, there were no differences in elongation. The most common differences occurred in the traits of strength and short fiber index. In short fiber index, the progeny was always lower when there were differences. In addition, the progeny was always significantly lower than the parent R076, and generally lower than the parent O037. For differences in strength, generally the parents were significantly different from each other, and the progeny did not differ from either parent. However, when the progeny did differ from a parent, it was always higher than one parent, usually O037. When there were differences in length, the progeny was consistently longer than the parent O037. For uniformity index, the progeny was consistently between the values of the two parents when the two parents were significantly different. When there were differences in lint percent, the progeny was always higher than the parent R76. Overall, crosses involving mutant J29 had the lowest MIC values and the highest UHM values. Crosses where at least one of the parents were S9 or S49 generally had high UHM and also had the highest STR. The highest STR was the cross between these two lines as well.

In the third set of mutant crosses, containing the lines K68, P058, P059, S32, and S45, there were less differences overall, and no differences seen in length or elongation. Short fiber index was the trait that had the most total differences, with three out of ten of the crosses having progeny that was lower than one parent. Crosses with P058 and P059 generally had high UHM, especially when they were crossed to each other, S45, or S32. This group of crosses also had an overall high STR, especially cross P058 X S45, which had a STR value of 40.3 grams/tex, which is rather exceptional. More research and replication will be needed to find higher quality progeny from these parents.

Hybridization is the most commonly used method, and often combined with other methods, for improvement of a variety of crop species, including cotton. Elite cultivars of the two most heavily cultivated species of cotton, *G. hirsutum* and *G. barbadense*, have histories of hybridization (Chandnani et al., 2018). Prior to becoming a new variety, cotton lines generally go through multiple rounds of self-pollination or backcrossing to make sure the particular gene, QTL, or chromosomal segment of interest is fixed. For example, the mutant cotton that was used as the basis for the mutant lines in this experiment was self-pollinated and advanced by single boll descent to the M5 generation before it was able to be registered (Auld et al., 2000). This practice is common because, while there has been evidence of heterosis in cotton hybrids for yield characteristics, there has not been evidence shown for heterosis in fiber characteristics (Tyagi et al., 2014). In addition, producing hybrid seed in cotton is much more intensive and impractical than a crop such as corn to be commercially exploited (Tyagi et al., 2014). Heterosis in cotton is also extremely dependent on different environments, especially with regards to water availability, and is more obvious in generally low yielding environments (Tyagi et al., 2014). In addition, it has been shown that heterosis is more apparent when two lower yielding inbreds are

crossed and over time, heterosis has decreased in cotton due to faster improvement in parental inbreds (Troyer and Wellin, 2009). Therefore, while the results of the F1 hybrids studied in this experiment are promising, it is not surprising that they do not show many significant differences from the parents since it is still the early stages of development of lines.

While there are minimal signs of heterosis in cotton fiber quality there have been some instances of incomplete dominance demonstrated in cotton. Many of these have been simple traits, such as the okra leaf trait or trichomes, but have been associated with important economic traits as well (Nawab et al., 2011a; Nawab et al., 2011b). Incomplete dominance is where the progeny of the cross demonstrates a phenotype that is in between or a mixture of the two parental phenotypic extremes, e.g. a blue flower and a red flower crossed results in a purple flowered hybrid (Nawab et al., 2011a). In the hybrids presented here, there are a few instances where the progeny did not differ from either parent, however, the 2 parents were significantly different from each other. This is consistent with patterns of incomplete dominance. There were results that followed this pattern in all sets of crosses. This suggests at least successful hybridization since the progeny had a value that was between the parents. Therefore, there is potential that the traits of interest from each of the parents could eventually become fixed in the population if the progeny were continued to be self-pollinated or potentially backcrossed to one or both parents.

**Conclusion**

In conclusion, while these initial results show promise for improved fiber quality from the hybridization of two unique lines, more work needs to be done to fully evaluate their potential. The rate of hybridization success, the patterns suggesting incomplete dominance in certain traits, and the improvement of some traits in the progeny over the parents all suggest future success of the hybridization of these two lines. In addition, only a subset of the crosses

that were made in 2018 have been evaluated, so it is possible that additional crosses may have even better results than these hybrids. Overall, these F1 hybrids have shown potential for improved fiber quality and yield but need to be further advanced via self-pollination or backcrossing in order to be able to significantly incorporate the benefits from both parents.

**Table 4.1**: **Average values for all traits in the first set of crosses between the mutant lines and the QTL stacked lines.** This contains the parents (241-036, 241-039, Q12, K75, R43, S50, and S67). Traits are listed in the order fiber fineness (Mic), fiber length (UHM), uniformity index (UI), fiber strength (Str), fiber elongation (Elo), short fiber index (SFI%) and lint percent (Lint %). Superscripts denote significant differences. "p" means that the progeny was different than the paternal parent, "m" means that the progeny was different from the maternal parent, and "B" means that the progeny was different from both parents. Crosses are always listed as maternal X paternal.

| Population | Mic | UHM | UI | Str | Elo | SFI (%) | Lint % |
|---|---|---|---|---|---|---|---|
| 241-036 | 3.83 | 1.13 | 84.91 | 34.35 | 5.38 | 7.11 | 0.32 |
| 241-039 | 3.89 | 1.19 | 85.51 | 33.61 | 5.50 | 7.10 | 0.33 |
| Q12 | 4.53 | 1.23 | 84.80 | 35.65 | 4.93 | 7.38 | 0.48 |
| K75 | 4.33 | 1.21 | 84.46 | 36.71 | 5.81 | 6.83 | 0.40 |
| R43 | 4.06 | 1.21 | 83.74 | 33.27 | 5.24 | 6.88 | 0.39 |
| S50 | 3.88 | 1.23 | 83.83 | 33.47 | 5.87 | 7.48 | 0.34 |
| S67 | 4.33 | 1.24 | 83.88 | 35.50 | 4.43 | 7.75 | 0.35 |
| Q12 X 241-036 | 4.54 | 1.18 | 84.65 | 36.20 | 5.40 | 6.00 | **0.41$^{p}$** |
| Q12 X 241-039 | **4.62$^{p}$** | **1.11$^{B}$** | **83.05$^{p}$** | 33.55 | 5.75 | 7.80 | **0.43$^{p}$** |
| K75 X 241-036 | **4.83$^{p}$** | 1.15 | 85.20 | 38.25 | 5.35 | 6.00 | **0.42$^{p}$** |
| K75 X 241-039 | **4.6$^{p}$** | 1.20 | 85.25 | 34.65 | 6.05 | 5.45 | **0.4$^{p}$** |
| R43 X 241-036 | 4.37 | 1.11 | 83.95 | 36.50 | 5.10 | 7.05 | **0.41$^{p}$** |
| R43 X 241-039 | 4.03 | 1.21 | 83.75 | 33.75 | 5.85 | 5.80 | **0.41$^{p}$** |
| S50 X 241-036 | 4.23 | 1.20 | 84.80 | 36.60 | 5.85 | 5.65 | 0.38 |
| S50 X 241-039 | 4.26 | 1.16 | 83.50 | 34.40 | 5.20 | 7.05 | 0.38 |
| S67 X 241-036 | 4.69 | **1.23$^{p}$** | 83.15 | 35.20 | 4.85 | 6.85 | **0.41$^{B}$** |
| S67 X 241-039 | 4.45 | 1.23 | **83.15$^{p}$** | 32.80 | **5.8$^{m}$** | 6.70 | **0.41$^{B}$** |

**Table 4.2**: **Average values for all traits in the second set of crosses between the mutant lines and the QTL stacked lines.** This set contains the parents (239-145, 242-020, J29, O037, R76, S9, and S49). Traits are listed in the order fiber fineness (Mic), fiber length (UHM), uniformity index (UI), fiber strength (Str), fiber elongation (Elo), short fiber index (SFI%) and lint percent (Lint %). Superscripts denote significant differences. "p" means that the progeny was different than the paternal parent, "m" means that the progeny was different from the maternal parent, and "B" means that the progeny was different from both parents. Crosses are always listed as maternal X paternal.

| Population | Mic | UHM | UI | Str | Elo | SFI (%) | Lint % |
|---|---|---|---|---|---|---|---|
| 239-145 | 3.38 | 1.20 | 84.87 | 31.39 | 5.45 | 7.28 | 0.38 |
| 242-020 | 3.52 | 1.12 | 81.59 | 29.58 | 4.89 | 8.88 | 0.38 |
| J29 | 3.45 | 1.23 | 81.05 | 30.15 | 5.50 | 8.35 | 0.41 |
| O037 | 3.43 | 1.14 | 81.40 | 28.35 | 5.35 | 8.75 | 0.38 |
| R76 | 3.34 | 1.25 | 82.12 | 30.54 | 5.32 | 8.05 | 0.32 |
| S9 | 3.61 | 1.26 | 83.44 | 34.78 | 5.53 | 6.86 | 0.33 |
| S49 | 3.70 | 1.23 | 84.48 | 34.64 | 5.29 | 7.08 | 0.36 |
| J29 X 239-145 | 3.89 | 1.21 | **83.85**$^{m}$ | 31.65 | 5.50 | **6.55**$^{m}$ | 0.41 |
| J29 X 242-020 | 3.70 | 1.15 | 81.75 | 31.50 | 5.30 | 9.45 | 0.44 |
| O037 X 239-145 | **4.15**$^{p}$ | 1.19 | 83.00 | 32.00 | 5.45 | 8.00 | 0.41 |
| O037 X242-020 | 4.17 | 1.16 | 82.80 | 30.35 | 6.05 | 8.30 | 0.43 |
| R76 X 239-145 | **4.72**$^{B}$ | 1.24 | 84.90 | **34.75**$^{m}$ | 5.75 | 5.95 | **0.40**$^{m}$ |
| R76 X 242-020 | **4.73**$^{B}$ | **1.23**$^{p}$ | 84.20 | 34.50 | 5.35 | 6.15 | **0.43**$^{m}$ |
| S9 X 239-145 | 3.47 | 1.26 | 83.65 | **37.3**$^{p}$ | 5.45 | 5.70 | 0.34 |
| S9 X 242-020 | 4.06 | 1.20 | 84.10 | **34.9**$^{p}$ | 5.55 | 6.65 | **0.39**$^{m}$ |
| S49 X 239-145 | 4.03 | 1.26 | 84.90 | 35.10 | 5.85 | **5.25**$^{B}$ | 0.38 |
| S49 X 242-020 | 4.17 | 1.18 | 82.65 | 31.75 | 5.35 | 7.95 | 0.42 |

**Table 4.3**: **Average values for all traits in the first set of crosses between different mutant lines.** This set contains the parents (Q12, K75, R43, S50, and S67). Traits are listed in the order fiber fineness (Mic), fiber length (UHM), uniformity index (UI), fiber strength (Str), fiber elongation (Elo), short fiber index (SFI%) and lint percent (Lint %). Superscripts denote significant differences. "p" means that the progeny was different than the paternal parent, "m" means that the progeny was different from the maternal parent, and "B" means that the progeny was different from both parents. Crosses are always listed as maternal X paternal.

| Population | Mic | UHM | UI | Str | Elo | SFI (%) | Lint % |
|---|---|---|---|---|---|---|---|
| Q12 | 4.53 | 1.23 | 84.80 | 35.65 | 4.93 | 7.38 | 0.48 |
| K75 | 4.33 | 1.21 | 84.46 | 36.71 | 5.81 | 6.83 | 0.40 |
| R43 | 4.06 | 1.21 | 83.74 | 33.27 | 5.24 | 6.88 | 0.39 |
| S50 | 3.88 | 1.23 | 83.83 | 33.47 | 5.87 | 7.48 | 0.34 |
| S67 | 4.33 | 1.24 | 83.88 | 35.50 | 4.43 | 7.75 | 0.35 |
| Q12 X K75 | 4.63 | 1.29 | 83.80 | 34.45 | 5.40 | 5.85 | 0.44 |
| Q12 X S67 | 4.65 | 1.26 | 84.05 | 35.40 | 5.00 | 5.80 | **0.45$^{B}$** |
| R43 X Q12 | **5.02$^{m}$** | 1.21 | 84.80 | 33.00 | 5.25 | 5.85 | **0.41$^{p}$** |
| R43 X S50 | 4.41 | 1.20 | 84.10 | 33.05 | 5.85 | 6.75 | **0.43$^{p}$** |
| K75 X R43 | 4.56 | 1.26 | 85.60 | 35.55 | 5.70 | **4.65$^{B}$** | 0.40 |
| K75 X S67 | 4.55 | 1.30 | 85.25 | 36.35 | 5.45 | **4.55$^{B}$** | **0.42$^{p}$** |
| S50 X K75 | 4.24 | 1.29 | 84.85 | 36.50 | 5.50 | **4.6$^{B}$** | 0.38 |
| S50 X Q12 | 4.25 | 1.23 | 85.15 | 35.55 | 5.00 | 5.45 | **0.40$^{B}$** |
| S67 X R43 | 3.91 | 1.32 | 85.30 | 34.35 | 5.75 | **4.3$^{B}$** | **0.39$^{m}$** |
| S67 X S50 | 3.65 | 1.33 | 83.70 | 33.25 | 5.25 | **4.0$^{B}$** | 0.35 |

**Table 4.4**: **Average values for all traits in the second set of crosses between different mutant lines.** This set contains the parents (J29, O037, R76, S9, and S49). Traits are listed in the order fiber fineness (Mic), fiber length (UHM), uniformity index (UI), fiber strength (Str), fiber elongation (Elo), short fiber index (SFI%) and lint percent (Lint %). Superscripts denote significant differences. "p" means that the progeny was different than the paternal parent, "m" means that the progeny was different from the maternal parent, and "B" means that the progeny was different from both parents. Crosses are always listed as maternal X paternal.

| Population | Mic | UHM | UI | Str | Elo | SFI (%) | Lint % |
|---|---|---|---|---|---|---|---|
| J29 | 3.45 | 1.23 | 81.05 | 30.15 | 5.50 | 8.35 | 0.41 |
| O037 | 3.43 | 1.14 | 81.40 | 28.35 | 5.35 | 8.75 | 0.38 |
| R76 | 3.34 | 1.25 | 82.12 | 30.54 | 5.32 | 8.05 | 0.32 |
| S9 | 3.61 | 1.26 | 83.44 | 34.78 | 5.53 | 6.86 | 0.33 |
| S49 | 3.70 | 1.23 | 84.48 | 34.64 | 5.29 | 7.08 | 0.36 |
| J29 X R76 | 3.08 | 1.31 | 84.30 | 31.85 | 5.35 | $\mathbf{4.55^{B}}$ | $\mathbf{0.38^{p}}$ |
| J29 X S9 | 3.61 | 1.23 | 83.15 | 33.03 | 5.63 | 7.48 | 0.40 |
| O037 X J29 | 3.17 | 1.21 | 81.30 | 31.45 | 5.65 | 7.90 | 0.39 |
| O037 X S49 | 3.50 | 1.15 | 84.40 | 31.55 | 6.95 | $\mathbf{6.25^{m}}$ | 0.41 |
| R76 X O037 | $\mathbf{4.12^{m}}$ | $\mathbf{1.29^{p}}$ | 84.28 | $\mathbf{32.43^{p}}$ | 5.90 | $\mathbf{4.83^{B}}$ | $\mathbf{0.42^{m}}$ |
| R76 X S49 | 3.96 | 1.29 | 85.10 | $\mathbf{36.55^{m}}$ | 5.65 | $\mathbf{4.55^{m}}$ | $\mathbf{0.38^{m}}$ |
| S9 X O037 | 3.48 | $\mathbf{1.31^{p}}$ | 84.08 | $\mathbf{34.00^{p}}$ | 5.45 | $\mathbf{4.70^{B}}$ | 0.38 |
| S9 X R76 | 3.82 | 1.30 | 85.90 | 34.35 | 5.70 | $\mathbf{4.25^{p}}$ | 0.33 |
| S49 X J29 | 3.90 | 1.28 | 84.80 | 34.15 | 5.75 | $\mathbf{5.55^{p}}$ | 0.41 |
| S49 X S9 | 3.63 | 1.27 | 84.48 | 36.68 | 5.48 | 5.25 | 0.33 |

**Table 4.5**: **Average values for all traits in the third set of crosses between different mutant lines.** This set contains the parents (K68, P058, P059, S32, and S45). Traits are listed in the order fiber fineness (Mic), fiber length (UHM), uniformity index (UI), fiber strength (Str), fiber elongation (Elo), short fiber index (SFI%) and lint percent (Lint %). Superscripts denote significant differences. "p" means that the progeny was different than the paternal parent, "m" means that the progeny was different from the maternal parent, and "B" means that the progeny was different from both parents. Crosses are always listed as maternal X paternal.

| Population | Mic | UHM | UI | Str | Elo | SFI (%) | Lint % |
|---|---|---|---|---|---|---|---|
| K68 | 3.97 | 1.25 | 82.96 | 34.20 | 5.36 | 7.09 | 0.42 |
| | | | | | | | |
| P058 | 4.16 | 1.27 | 83.89 | 33.60 | 5.20 | 6.31 | 0.41 |
| P059 | 4.24 | 1.27 | 84.84 | 34.74 | 5.44 | 6.02 | 0.40 |
| S32 | 3.93 | 1.28 | 84.44 | 33.95 | 5.62 | 6.71 | 0.39 |
| S45 | 3.72 | 1.28 | 83.83 | 33.92 | 5.26 | 7.06 | 0.37 |
| K68 X P058 | 4.05 | 1.27 | 85.65 | 37.25 | 5.50 | 4.80 | 0.40 |
| K68 X S45 | **4.89$^{B}$** | 1.18 | 82.90 | 33.95 | 5.50 | 7.45 | 0.41 |
| P058 X S32 | 4.29 | 1.37 | 85.45 | 37.55 | 5.35 | 3.85 | 0.43 |
| P058 X S45 | 3.90 | 1.35 | 85.45 | **40.30$^{B}$** | 5.40 | **3.75$^{p}$** | 0.41 |
| P059 X K68 | 4.49 | 1.29 | **85.75$^{p}$** | 35.75 | 5.60 | **4.20$^{p}$** | 0.42 |
| P059 X P058 | 4.50 | 1.34 | 85.95 | 37.30 | 5.25 | 3.90 | 0.43 |
| S32 X K68 | 4.58 | 1.27 | 85.10 | 38.40 | 5.70 | 4.90 | 0.42 |
| S32 X P059 | 4.54 | 1.28 | 84.85 | 37.15 | 5.70 | 4.75 | 0.37 |
| S45 X P059 | 4.19 | 1.29 | 85.95 | 36.30 | 5.30 | **4.65$^{m}$** | 0.36 |
| S45 X S32 | 4.26 | 1.20 | 84.85 | 35.30 | 6.20 | 5.90 | 0.37 |

**Supplemental Table 4.6**: **List of all pairs of progeny and parents that were significantly different from each other, and parents that were significantly different from each other**. The p-values shown are from the Tukey's Honest Significant Difference test. Traits listed are fiber fineness (MIC), fiber length (UHM), uniformity index (UI), fiber strength (STR), fiber elongation (ELO), short fiber index (SFI) and lint percent (Lint %).

| Pair | Trait | p-value | | Pair | Trait | p-value |
|---|---|---|---|---|---|---|
| Q12X241-036 and 241-036 | Lint % | 0.012 | | J29 and 242-020 | UHM | 0.07 |
| Q12 and 241-036 | MIC | 0.07 | | J29X239-145 and J29 | UI | 0.005 |
| Q12 and 241-036 | UHM | 0.03 | | J29X239-145 and J29 | SFI | 0.029 |
| Q12X241-039 and 241-039 | MIC | 0.022 | | O037X239-145 and 239-145 | MIC | 0.07 |
| Q12X241-039 and 241-039 | UHM | 0.05 | | S9X239-145 and 239-145 | STR | 0.023 |
| Q12X241-039 and Q12 | UHM | 0.016 | | S9X242-020 and 242-020 | STR | 0.034 |
| Q12X241-039 and 241-039 | UI | 0.035 | | S9X242-020 and S9 | Lint % | 0.033 |
| Q12X241-039 and 241-039 | Lint % | 0.005 | | S9 and 242-020 | UHM | 0.021 |
| K75X241-039 and 241-039 | MIC | 0.007 | | S9 and 242-020 | SFI | 0.043 |
| K75X241-039 and 241-039 | Lint % | 0.018 | | S49 and 242-020 | UHM | 0.021 |
| K75X241-036 and 241-036 | MIC | 0.006 | | S49 and 242-020 | UI | 0.008 |
| K75X241-036 and 241-036 | Lint % | 0.0006 | | S49 and 242-020 | STR | 0.016 |
| K75 and 241-036 | UHM | 0.0274 | | S49 and 242-020 | SFI | 0.017 |
| R043X241-039 and 241-039 | Lint % | 0.009 | | R076X239-145 and 239-145 | MIC | 0.04 |
| R043 and 241-039 | UI | 0.034 | | R076X239-145 and R076 | MIC | 0.022 |
| R043X241-036 and 241-036 | Lint % | 0.0006 | | R076X239-145 and R076 | STR | 0.0645 |
| S50 and 241-036 | UHM | 0.008 | | R076X239-145 and R076 | Lint % | 0.0355 |
| S67X241-039 and 241-039 | UI | 0.08 | | R076X242-020 and 242-020 | MIC | 0.06 |
| S67X241-039 and S67 | ELO | 0.02 | | R076X242-020 and R076 | MIC | 0.019 |
| S67X241-039 and 241-039 | Lint % | 0.004 | | R076X242-020 and 242-020 | UHM | 0.021 |
| S67X241-039 and S67 | Lint % | 0.012 | | R076X242-020 and R076 | Lint % | 0.0006 |
| S67X241-036 and 241-036 | UHM | 0.016 | | J29XR076 and J29 | SFI | 0.08 |
| S67X241-036 and 241-036 | Lint % | 0.003 | | J29XR076 and R076 | SFI | 0.07 |
| S67X241-036 and S67 | Lint % | 0.016 | | J29XR076 and R076 | Lint % | 0.026 |
| Q12 and K75 | Lint % | 0.016 | | J29 X S9 | UI | 0.059 |
| Q12XS67 and S67 | Lint % | 0.00001 | | J29 X S9 | STR | 0.022 |
| Q12XS67 and Q12 | Lint % | 0.016 | | O037XS49 and O037 | SFI | 0.043 |
| R43XQ12 and R43 | MIC | 0.009 | | O037 and S49 | STR | 0.008 |
| R43XQ12 and Q12 | Lint % | 0.02 | | O037 and S49 | UI | 0.008 |
| R43XS50 and S50 | Lint % | 0.007 | | O037 and S49 | UHM | 0.08 |
| K75XR43 and K75 | SFI | 0.057 | | R076XO037 and R76 | MIC | 0.047 |
| K75XR43 and R43 | SFI | 0.051 | | R076XO037 and O037 | UHM | 0.0002 |

| K75 and R43 | STR | 0.0326 | | R076XO037 and O037 | STR | 0.042 |
|---|---|---|---|---|---|---|
| K75XS67 and K75 | SFI | 0.014 | | R076XO037 and O037 | SFI | 0.013 |
| K75XS67 and S67 | SFI | 0.002 | | R076XO037 and R076 | SFI | 0.015 |
| K75XS67 and S67 | Lint % | 0.0079 | | R076XO037 and R076 | Lint % | 0.00005 |
| K75 and S67 | ELO | 0.011 | | R076XS49 and R076 | STR | 0.003 |
| S50XK75 and S50 | SFI | 0.016 | | R076XS49 and R076 | SFI | 0.003 |
| S50XK75 and K75 | SFI | 0.068 | | R076XS49 and R076 | Lint % | 0.08 |
| S50 and K75 | MIC | 0.0079 | | R076 and S49 | UI | 0.056 |
| S50 and K75 | STR | 0.029 | | S9XO037 and O037 | UHM | 0.02 |
| S50 and K75 | Lint % | 0.0013 | | S9XO037 and O037 | STR | 0.007 |
| S50XQ12 and S50 | Lint % | 0.052 | | S9XO037 and O037 | SFI | 0.004 |
| S50XQ12 and Q12 | Lint % | 0.081 | | S9XO037 and S9 | SFI | 0.06 |
| S50 and Q12 | MIC | 0.019 | | S9XR076 and R076 | SFI | 0.041 |
| S67XR43 and S67 | SFI | 0.011 | | S9 and R076 | STR | 0.0018 |
| S67XR43 and R43 | SFI | 0.03 | | S49XJ29 and J29 | SFI | 0.053 |
| S67XR43 and S67 | Lint % | 0.04 | | S49 and J29 | STR | 0.049 |
| S67XS50 and S67 | SFI | 0.012 | | K68XS45 and K68 | MIC | 0.07 |
| S67XS50 and S50 | SFI | 0.009 | | K68XS45 and S45 | MIC | 0.014 |
| S67 and S50 | ELO | 0.0037 | | K68 and S45 | Lint % | 0.003 |
| P059XK68 and K68 | SFI | 0.026 | | P058XS45 and P058 | STR | 0.007 |
| P059XK68 and K68 | UI | 0.079 | | P058XS45 and S45 | STR | 0.008 |
| S32 and K68 | UI | 0.064 | | P058XS45 and S45 | SFI | 0.041 |
| S45XP059 and S45 | SFI | 0.037 | | P058 and S45 | MIC | 0.083 |
| S45 and P059 | MIC | 0.021 | | | | |

**References**

Abbas A., Iqbal M.A., Rahman M.U., Paterson A.H. (2015) Estimating genetic diversity among selected cotton genotypes and the identification of DNA markers associated with resistance to cotton leaf curl disease. Turkish Journal of Botany 39:1033-1041. DOI: 10.3906/bot-1505-22.

Auld D.L., Bechere E., Ethridge M.D., Becker W.D., Hequet E., Cantrell R.G. (2000) Registration of TTU 202-1107-B and TTU 271-2155-C mutant germplasm lines of upland cotton with improved fiber quality. Crop Science 40:1835-1836.

Auld D.L., Bechere E., Krifa M., Kebede H., Hequet E., Wright R., Misra S. (2007) Registration of 'Raider 276', a High-Yielding, Improved-Quality Upland Mutant Cotton Cultivar. Journal of Plant Registrations 1:115-116. DOI: 10.3198/jpr2007.01.0059crc.

Auld D.L., Heikkinen M.K., Erickson D.A., Sernyk J.L., Romero J.E. (1992) RAPESEED MUTANTS WITH REDUCED LEVELS OF POLYUNSATURATED FATTY-ACIDS AND INCREASED LEVELS OF OLEIC-ACID. Crop Science 32:657-662. DOI: 10.2135/cropsci1992.0011183X003200030016x.

Bechere E., Auld D.L., Cantrell R.G., Hequet E., Krifa M., Misra S., Smith C.W. (2007) Registration of TTU 0774-3-3 and TTU 0808-1-6-1 Upland Cotton Germplasm Lines with Improved Fiber Length and Strength. Journal of Plant Registrations 1:58-59. DOI: 10.3198/jpr2006.09.0587crg.

Bechere E., Auld D.L., Krifa M., Smith C.W., Cantrell R.G. (2011) Registration of TTU 0782, an Upland Cotton Germplasm Line with Superior Fiber Quality. Journal of Plant Registrations 5:207-210. DOI: 10.3198/jpr2010.07.0404crg.

Bi C.W., Paterson A.H., Wang X.L., Xu Y.Q., Wu D.Y., Qu Y.S., Jiang A.N., Ye Q.L., Ye N. (2016) Analysis of the Complete Mitochondrial Genome Sequence of the Diploid Cotton Gossypium raimondii by Comparative Genomics Approaches. Biomed Research International. DOI: 10.1155/2016/5040598.

Blixt S., Ehrenberg L., Gelin O. (1966) STUDIES OF INDUCED MUTATIONS IN PEAS .18. EMS-TREATMENT OF DIFFERENT DURATION OF VARIETIES WEITOR AND WITHAM WONDER. Agri Hortique Genetica 24:138-+.

Bokore F.E., Cuthbert R.D., Knox R.E., Randhawa H.S., Hiebert C.W., DePauw R.M., Singh A.K., Singh A., Sharpe A.G., N'Diaye A., Pozniak C.J., McCartney C., Ruan Y.F., Berraies S., Meyer B., Munro C., Hay A., Ammar K., Huerta-Espino J., Bhavani S. (2017) Quantitative trait loci for resistance to stripe rust of wheat revealed using global field nurseries and opportunities for stacking resistance genes. Theoretical and Applied Genetics 130:2617-2635. DOI: 10.1007/s00122-017-2980-7.

Bowman D.T., May O.L., Creech J.B. (2003) Genetic uniformity of the US upland cotton crop since the introduction of transgenic cottons. Crop Science 43:515-518. DOI: 10.2135/cropsci2003.0515.

Brown A., Paterson A.H., Li L. (2012) Genomics and Breeding in Food Crops.

Brown N., Kumar P., Singh R., Lubbers E., Campbell B.T., Myers G.O., Wright R.J., Subramani J., Jones D., Paterson A.H., Chee P.W. (2019a) Evaluation of a Chromosome Segment from Gossypium barbadense Harboring the Fiber Length QTL qFL-Chr.25 in Four Diverse Upland Cotton Genetic Backgrounds. Crop Science 59:2621-2633. DOI: 10.2135/cropsci2019.05.0321.

Brown N., Shen X.L., Lubbers E.L., Kumar P., McBlanchett J., Smith C.W., Jones D., Paterson A.H., Chee P.W. (2019b) Registration of GA R01-40-08, a Gossypium hirsutum Upland Cotton Germplasm Line with qFL-Chr.1 Introgressed from Gossypium barbadense Conferring Improved Fiber Length. Journal of Plant Registrations 13:406-410. DOI: 10.3198/jpr2019.03.0014crg.

Cantrell R.G., Roberts C.L., Waddell C. (2000) Registration of 'Acala 1517-99' cotton. Crop Science 40:1200-1201. DOI: 10.2135/cropsci2000.0049rcv.

Chandnani R., Kim C., Guo H., Shehzad T., Wallace J.G., He D.H., Zhang Z.S., Patel J.D., Adhikari J., Khanal S., Paterson A.H. (2018) Genetic Analysis of Gossypium Fiber Quality Traits in Reciprocal Advanced Backcross Populations. Plant Genome 11:15. DOI: 10.3835/plantgenome2017.06.0057.

Chandnani R., Wang B.H., Draye X., Rainville L.K., Auckland S., Zhuang Z.M., Lubbers E.L., May O.L., Chee P.W., Paterson A.H. (2017a) Segregation distortion and genome-wide digenic interactions affect transmission of introgressed chromatin from wild cotton species. Theoretical and Applied Genetics 130:2219-2230. DOI: 10.1007/s00122-017-2952-y.

Chandnani R., Zhang Z.S., Patel J.D., Adhikari J., Khanal S., He D.H., Brown N., Chee P.W., Paterson A.H. (2017b) Comparative genetic variation of fiber quality traits in reciprocal advanced backcross populations. Euphytica 213. DOI: 10.1007/s10681-017-2029-7.

Chee P.W., Draye X., Jiang C.X., Decanini L., Delmonte T.A., Bredhauer R., Smith C.W., Paterson A.H. (2005) Molecular dissection of phenotypic variation between Gossypium hirsutum and Gossypium barbadense (cotton) by a backcross-self approach: III. Fiber

length. Theoretical and Applied Genetics 111:772-781. DOI: 10.1007/s00122-005-2062-0.

de Menezes I.P.P., da Silva J.O., Malafaia G., Silveira R.D.D., Barroso P.A.V. (2015) Natural hybridization between Gossypium mustelinum and exotic allotetraploid cotton species. Genetics and Molecular Research 14:14177-14180. DOI: 10.4238/2015.October.29.40.

den Boer E., Pelgrom K.T.B., Zhang N.W.W., Visser R.G.F., Niks R.E., Jeuken M.J.W. (2014) Effects of stacked quantitative resistances to downy mildew in lettuce do not simply add up. Theoretical and Applied Genetics 127:1805-1816. DOI: 10.1007/s00122-014-2342-7.

Esbroeck G.B., DT (1998) Cotton germplasm diversity and its importance to cultivar development. Journal of Cotton Science 2:121-129.

Espina M.J., Ahmed C.M.S., Bernardini A., Adeleke E., Yadegari Z., Arelli P., Pantalone V., Taheri A. (2018) Development and Phenotypic Screening of an Ethyl Methane Sulfonate Mutant Population in Soybean. Frontiers in Plant Science 9. DOI: 10.3389/fpls.2018.00394.

Favret E.A. (1960) Somatic mutations of four genes for albinism in barley induced by X-rays and ethyl methane-sulphonate. Hereditas 46:622-634.

Gennaro A., Forte P., Panichi D., Lafiandra D., Pagnotta M.A., D'Egidio M.G., Ceoloni C. (2012) Stacking small segments of the 1D chromosome of bread wheat containing major gluten quality genes into durum wheat: transfer strategy and breeding prospects. Molecular Breeding 30:149-167. DOI: 10.1007/s11032-011-9606-6.

Gichner T., Ehrenberg L. (1966) INFLUENCE OF POST-TREATMENT STORAGE ON FREQUENCY OF EMS-INDUCED CHROMOSOMAL ABERRATIONS IN BARLEY. Biologia Plantarum 8:256-+. DOI: 10.1007/bf02930787.

Greene E.A., Codomo C.A., Taylor N.E., Henikoff J.G., Till B.J., Reynolds S.H., Enns L.C., Burtner C., Johnson J.E., Odden A.R., Comai L., Henikoff S. (2003) Spectrum of chemically induced mutations from a large-scale reverse-genetic screen in arabidopsis. Genetics 164:731-740.

Herring A.D., Auld D.L., Ethridge M.D., Hequet E.F., Bechere E., Green C.J., Cantrell R.G. (2004) Inheritance of fiber quality and lint yield in a chemically mutated population of cotton. Euphytica 136:333-339. DOI: 10.1023/b:euph.0000032747.97343.54.

Hildering G.J., Vanderve.Jh. (1966) MUTUAL INDEPENDENCE OF M1-FERTILITY AND MUTANT YIELD IN EMS TREATED TOMATOES. Euphytica 15:412-+. DOI: 10.1007/bf00022188.

Miedaner T., Wilde F., Steiner B., Buerstmayr H., Korzun V., Ebmeyer E. (2006) Stacking quantitative trait loci (QTL) for Fusarium head blight resistance from non-adapted sources in an European elite spring wheat background and assessing their effects on deoxynivalenol (DON) content and disease severity. Theoretical and Applied Genetics 112:562-569. DOI: 10.1007/s00122-005-0163-4.

Nawab N.N., Saeed A., Tariq M.S., Nadeem K., Mahmood K., Ul-Hassan M., Shakil Q., Alam M.S., Hussain S.I., Khan A.A. (2011a) Inheritance of okra leaf type in different genetic backgrounds and its effects on fibre and agronomic traits in cotton. African Journal of Biotechnology 10:16484-16490. DOI: 10.5897/ajb11.1445.

Nawab N.N., Shakil Q., Niaz S., Iqbal M.M., Asif M.A., Khan I.A. (2011b) Genetics of trichomes and its association with fibre and agronomic traits in cotton. African Journal of Biotechnology 10:3964-3972.

Parhe S.D., Chimote V.P., Deshmukh M.P., Chandra K., Akash M. (2017) Marker-assisted pyramiding of four QTL/genes for Asian rust (Phakopsora pachyrhizi) resistance in soybean. Journal of Crop Improvement 31:689-711. DOI: 10.1080/15427528.2017.1347593.

Parry M.A.J., Madgwick P.J., Bayon C., Tearall K., Hernandez-Lopez A., Baudo M., Rakszegi M., Hamada W., Al-Yassin A., Ouabbou H., Labhilili M., Phillips A.L. (2009) Mutation discovery for crop improvement. Journal of Experimental Botany 60:2817-2825. DOI: 10.1093/jxb/erp189.

Patel J.D., Wright R.J., Auld D., Chandnani R., Goff V.H., Ingles J., Pierce G.J., Torres M.J., Paterson A.H. (2014) Alleles conferring improved fiber quality from EMS mutagenesis of elite cotton genotypes. Theoretical and Applied Genetics 127:821-830. DOI: 10.1007/s00122-013-2259-6.

Paterson A.H., Boman R.K., Brown S.M., Chee P.W., Gannaway J.R., Gingle A.R., May L., Smith C.W. (2004) Reducing the genetic vulnerability of cotton. Crop Science 44:1900-1901. DOI: 10.2135/cropsci2004.1900.

Paterson A.H., C. L. Brubaker, and J.F. Wendel. (1993) A rapid method for extraction of cotton (Gossypium spp.) genomic DNA suitable for RFLP or PCR analysis. Plant Molecular Biology 11:122-127. DOI: 10.1007/ BF02670470.

Perkins H.E., DE; Bragg, CK. (1984) Fiber American Society of Agronomy:438–508.

Rao S.H.K., Sears E.R. (1962) EMS-INDUCED MUTATIONS IN HEXAPLOID WHEAT. Genetics 47:983-&.

Rong J., Feltus E.A., Waghmare V.N., Pierce G.J., Chee P.W., Draye X., Saranga Y., Wright R.J., Wilkins T.A., May O.L., Smith C.W., Gannaway J.R., Wendel J.R., Paterson A.H.

(2007) Meta-analysis of polyploid cotton QTL shows unequal contributions of subgenomes to a complex network of genes and gene clusters implicated in lint fiber development. Genetics 176:2577-2588. DOI: 10.1534/genetics.107.074518.

Shi Y.Z., Liu A.Y., Li J.W., Zhang J.F., Zhang B.C., Ge Q., Jamshed M., Lu Q.W., Li S.Q., Xiang X.H., Gong J.W., Gong W.K., Shang H.H., Deng X.Y., Pan J.T., Yuan Y.L. (2019) Dissecting the genetic basis of fiber quality and yield traits in interspecific backcross populations of Gossypium hirsutum x Gossypium barbadense. Molecular Genetics and Genomics 294:1385-1402. DOI: 10.1007/s00438-019-01582-8.

Smith C.W. (2003) Registration of TAM 94L-25 and TAM 94J-3 germplasm lines of upland cotton with improved fiber length. Crop Science 43:742-743. DOI: 10.2135/cropsci2003.0742.

Talebi A.B., Shahrokhifar, B. (2012) Ethyl methane sulphonate (EMS) induced mutagenesis in Malaysian rice (cv. MR219) for lethal dose determination. American Journal of Plant Sciences.

Troyer A.F., Wellin E.J. (2009) Heterosis Decreasing in Hybrids: Yield Test Inbreds. Crop Science 49:1969-1976. DOI: 10.2135/cropsci2009.04.0170.

Tyagi P., Bowman D.T., Bourland F.M., Edmisten K., Campbell B.T., Fraser D.E., Wallace T., Kuraparthy V. (2014) Components of hybrid vigor in upland cotton (Gossypium hirsutum L.) and their relationship with environment. Euphytica 195:117-127. DOI: 10.1007/s10681-013-0987-y.

Vanderve.Jh. (1967) EMS INDUCED NON-RECESSIVE EMBRYONIC LETHALS IN ARABIDOPSIS THALIANA. Genetica 38:61-&.

Waghmare V.N., Rong J.K., Rogers C.J., Pierce G.J., Wendel J.F., Paterson A.H. (2005) Genetic mapping of a cross between Gossypium hirsutum (cotton) and the Hawaiian endemic, Gossypium tomentosum. Theoretical and Applied Genetics 111:665-676. DOI: 10.1007/s00122-005-2032-6.

Zhang Z., Zou X.Y., Huang Z., Fan S.M., Qun G., Liu A.Y., Gong J.W., Li J.W., Gong W.K., Shi Y.Z., Fan L.Q., Zhang Z.B., Liu R.X., Jiang X., Lei K., Shang H.H., Xu A.X., Yuan Y.L. (2019) Genome-wide identification and analysis of the evolution and expression patterns of the GATA transcription factors in three species of Gossypium genus. Gene 680:72-83. DOI: 10.1016/j.gene.2018.09.039.

Zhang Z.S., Rong J.K., Waghmare V.N., Chee P.W., May O.L., Wright R.J., Gannaway J.R., Paterson A.H. (2011) QTL alleles for improved fiber quality from a wild Hawaiian cotton, Gossypium tomentosum. Theoretical and Applied Genetics 123:1075-1088. DOI: 10.1007/s00122-011-1649-x.