A COMPARISON OF PUBLICLY AVAILABLE QTL SSRS FOR MAS WITH TRADITIONAL PLANT BREEDING SELECTION METHODS IN COTTON

A Thesis

by

ALEXANDRA PAIGE ULLRICH

Submitted to the Office of Graduate and Professional Studies of
Texas A&M University
in partial fulfillment of the requirements for the degree of

MASTER OF SCIENCE

Chair of Committee,	Wayne Smith
Committee Members,	Clint Magill
	Eric Hequet
Head of Department,	David Baltensperger

December 2019

Major Subject: Plant Breeding

ABSTRACT

Public breeding programs for upland cotton (*Gossypium hirsutum)* underutilize genotypic selection methods and specifically marker assisted selection for fiber quality selection. A previous study by Kari Hugie sought to analyze publicly available simple sequence repeat (SSR) markers in three diverse populations to quantify those that showed correlative trait stability in different backgrounds. Stable markers identified from that study, 6 for fiber bundle strength (FBS) and 6 for upper half mean length (UHML), as well as two additional markers identified by Dr. Fang of the USDA Louisiana Laboratory were then utilized to make selections in two *G. hirsutum* populations. Population 1 (TAM 11K-13 ELSU/Del Cerro//13P-54 ELSU) was of interspecific background and Population 2 (TAM 11K-13 ELSU/TAM06WE-621 ESU) was of intraspecific background.

In 2016, both populations were grown at the Agrilife Research Center in College Station, Tx and individual plant selections were made separately on the basis of SSR marker number and phenotype for both FBS and UHML to form populations divergent for the trait of interest. Selections were planted in progeny rows in 2017 and fiber samples collected for each row. Fiber data was analyzed to compare selection efficacy within the divergently selected subpopulations and between the marker and phenotypically selected populations.

Divergent selections made on the basis of number of markers in the desired allelic state failed to show significant differences between subpopulations while those made on the basis of high and low value for the phenotypic trait were significant for both traits in both populations. Analyzing individual markers, none were found to be significant ($p < 0.05$) for either trait in both populations for both years. In Population 1 for UHML, two markers were found to be significant only in 2016, $CIR196_{197}$ ($p=0.0416$) and $NAU1369_{247}$ ($p=.0011$), and for FBS in 2016 only $CGR6329_{232}$ was significant ($p= 0.0278$). In Population 2, marker $BNL1604_{98}$ was significant in both years for UHML ($p=0.0270$ and $p=.0448$ in 2016 and 2017 respectively), and in 2016 for FBS ($p=0.0425$). Despite significance, numerical differences between selections with or without the $BNL1604_{98}$ allele were small.

DEDICATION

I would like to dedicate this work to my family who have all continuously supported me.

ACKNOWLEDGEMENTS

I would like to acknowledge all of my committee members as well as Dr. Fang for their contributions to this work.

I would also like to thank my fellow graduate students in the Cotton Improvement Lab and all of the undergraduate workers who helped maintain my plants in the field and helped harvest and process my fiber samples.

CONTRIBUTORS AND FUNDING SOURCES

Contributors

This work was supervised by a thesis committee consisting of Professor C. Wayne Smith [advisor] and Eric Hequet of the Department of Soil and Crop Sciences and Professor Clint Magill of the Department of Genetics.

All plant samples were genotyped by Dr. Fang of the USDA-ARS Cotton Fiber Bioscience Research Unit at the Southern Regional Research Center in New Orleans.

All other work conducted for the thesis was completed by the student independently.

Funding Source

Graduate Study was supported by an assistantship from Texas A&M University.

NOMENCLATURE

ANOVA	Analysis of Variance
ELSU	Extra Long Staple Upland
ESU	Extra Strength Upland
FBS	Fiber Bundle Strength
HVI	High Volume Instrumentation
QTL	Quantitative Trait Locus
SSR	Simple Sequence Repeat
UHML	Upper Half Mean Length

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

CHAPTER I

INTRODUCTION

Upland cotton, *Gossypium hirsutum* L., was introduced to what is now the United States of America in the early 1700s (Smith and Cothren, 1999). From that point, the U.S.A. has become the third largest producer in the world, harvesting approximately 1.8 million hectares in 2017 (National Cotton Council). The majority of this fiber is exported to countries such as China for processing into yarn and fabrics. The product quality that can be produced from any given bale of cotton fiber is highly dependent on the fiber properties within that bale; the most important of which are upper half mean length (UHML), fiber bundle strength (FBS), and fiber length distribution (Simpson et al., 1978). Different spinning methods, of which rotor spinning and ring spinning predominate, may rank these traits at varying levels of importance, however air jet, or vortex spinning, a newer technology that promises to increase spinning speeds without decreasing quality, places even more importance on UHML, length distribution, and FBS. For the U.S.A. to maintain its competitiveness of the global market and take advantage of technological advances, cotton breeders must continue to target these traits.

One tool that has not been taken full advantage of within public cotton breeding programs is marker-assisted selection (MAS). MAS is the process of associating genetic markers with a nearby gene of interest called a quantitative trait locus (QTL) and then using those markers to make plant selections based on the predicted phenotype. The advantages of incorporating MAS include the elimination of poor quality plants in the

seed or seedling stage thereby increasing the frequency of plants with the targeted trait in the field, more easily introgressing or pyramiding beneficial alleles, and trait selection in off-season nurseries. Despite the potential benefits, successful implementation of MAS in other agronomic crops, and numerous mapping studies of *Gossypium* spp. genomes to associate markers with QTLs for different fiber traits, MAS remains underutilized.

Many factors lead into the discrepancy, foremost being that many fiber qualities, including UHML and FBS, are highly quantitative traits controlled by a number of alleles and potentially hundreds of minor effect alleles (Lande et al., 1989). QTLs for these traits tend to be non-robust between populations and may be affected by epistatic interactions. Due to this, a new set of trait-associated QTLs must be generated for each breeding population and possibly for each new environment which would require the planting, harvesting, and genotyping of a training population for each parental combination and maybe each location. In addition to this, UHML and FBS are both moderately to highly heritable traits which can be easily selected given sufficient manpower.

To overcome these disadvantages, Kari Hugie evaluated all publicly available QTLs (Hugie. L, 2015) proposed to be genomically associated with a number of fiber properties, including UHML and FBS. She used three populations, each derived from a different bi-parental background. In her dissertation entitled, "Evaluation of Conventional and Marker-Assisted Breeding Methods for the Improvement of Fiber Quality in *Gossypium* spp.", Dr.Hugie evaluated a large number of QTLs from various mapping studies from which six markers for UHML and six markers for FBS were

identified as robust across all three populations. Grouping plants based on the number of alleles in the desired state for each population, it was then possible to use marker based selections and identify progeny with mean fiber traits not significantly different from those of progeny selected based on traditional phenotype based selection.

CHAPTER II

LITERATURE REVIEW

Overview of Gossypium and Fiber Traits of Interest

Gossypium sp. have been in use as a source of natural fibers since the neolithic era (Christophe et al., 2002) and continues to be the world's most important fiber crop. Four species have been grown at scale for fiber production and are grouped into new and old world species: *G. herbaceum* and *G. arboreum* being diploid old world species denoted by the A genome and *G. hirsutum* and *G. barbadense* being tetraploid, new world species combining the A and D genomes through an interspecific hybridization event (Paterson et al., 2012). Of the cultivated species, *G. hirsutum* is the most widely grown, producing more than 90% of the world's cotton fiber (Wendel et al., 1992). This can be attributed to higher fiber quality than the old world species and agronomic adaptability to a range of growing conditions. While *G. barbadense* has superior fiber qualities compared to *G. hirsutum* cultivation is more difficult and yields tend to be lower in most environments. Thus, *G. barbadense* production is less widespread and often used for specialty and high quality textiles.

There are many traits that dictate good quality such as color, uniformity, and trash content. However, the most important factors dictating the end product quality of woven or knit textiles are length, strength, and length uniformity parameters. UHML and FBS are moderate to highly quantitative traits that are largely controlled by additive gene action (May and Jividen, 1999). The number of genes estimated to control each trait is

variable among mapping studies, but a meta-analysis of fiber trait mapping studies conducted by J. I. Said (2013) showed more than 1,000 trait-associated QTL.

G. hirsutum is indeterminate in habit, causing bolls to mature at different times in the season. Fiber extension from the seed coat begins within 3 days after flowering and continues extension for approximately 30 days (Braden and Smith, 2004). Even though maximum fiber length is reached fairly early in a boll's life span and harvest is timed to maximize fiber maturity, there is still a distribution of fiber lengths created from the bottom to the top of the plant caused by the crops indeterminate habit, and even then fiber lengths vary across the surface of a single seed. Fiber length can also be affected by environmental factors such as day and night temperatures and water stress (Guthrie et al., 1993). There are many different ways to measure fiber length, but the most common measure is upper half mean length (UHML) which is the average length of the longest 50% of fibers as defined by weight. The concept of UHML was introduced based on fibrograph technology as a way to standardize fiber length measurements (Moore, 1996). UHML is measured during High Volume Instrument testing (HVI) by pulling a beard of fiber from the sample with a comb and scanning it with a sensor to produce a fibrogram.

The variation in fiber lengths is the fiber length uniformity, measured as the mean length divided by the UHML and expressed as a percentage (Cotton Incorporated, 2019). Uniformity is affected by the natural variation in fiber lengths and all things which effect it, but is also affected by the ginning process which may cause fiber breakage (Guthrie et al., 1993). Uniformity is complex in that there are many sources that contribute to it, and studies conflict on the genetic basis of the trait. Multiple studies

have found nonadditive gene action for this trait and indicated a high amount of dominant gene action (Khan et al., 2001; Ali et al., 2008). Because UHML is a component of fiber length uniformity and UHML is highly heritable and repeatable, fiber length uniformity was not a focus of this study.

FBS is determined by the amount of carbohydrates deposited into the secondary cell wall as the fibers mature after fiber elongation is completed. FBS is influenced by the environment more than UHML because once bolls open, general weather such as UV exposure and rain tends to degrade the walls of the fiber which weakens them. Carbohydrate deposition is also highly influenced by the environment by such things at light intensity (Pettigrew, 2001). FBS generally has a lower heritability than UHML which makes trait progression in phenotypic selection slower for this trait. FBS is the most common measure of fiber strength in global markets and is measured through HVI using the same fiber beard from the UHML measurement. The beard is held at each end by a pair of clamps or jaws set at a given distance, and the force required to break the fiber bundle is determined.

The unique structure of a mature cotton fiber is what gives it its spinnable quality. The fiber is composed of multiple layers which are composed of a crystalline cellulose lattice and which form fibrils that are differently oriented in each layer. As a fiber dries, its cross section takes on a bean-like shape as the hollow inner core flattens and the differences between fibril direction in each layer cause the fiber to twist (Hearle, 2006). This shape allows fibers to adhere to each other when spun into yarn, and yarn

qualities such as strength are directly related to the qualities of the individual fibers that form them and the amount of variation between the fibers.

Spinning Technologies and the Effects of Fiber Traits on Yarn Quality

Fiber spinning is dominated by two major types of technology, ring spinning and rotor spinning, with a new emerging technology in the form of air jet spinning. Each of these processes produces yarns which are different in structure, and while all three place high value on UHML, FBS, and fiber length uniformity in order to produce the highest quality yarns possible, the structure of the yarn dictates how these three traits are ranked. While not discussed as broadly here, fiber fineness also plays a part in yarn strength by increasing the number of fibers in the yarn cross section.

Of the three, ring spinning was the earliest introduction, having been invented in 1832 by John Thorp as an extension and improvement on existing technology. When first introduced, ring spinning speeds far exceeded the other predominate systems of the time and quickly became widespread in the textile industry. Its versatility in producing high quality yarns of various weights and twist densities has led to its continued dominance of the spinning industry despite further advances in speed in newer technologies (Lawrence, 2010).

Ring spinning produces yarns where parallel fibers are twisted throughout the entire yarn (Cotton Incorporated, 2003). The parallel fibers produce yarns that are less bulky and stronger than rotor spun yarns of the same weight, but also means that much of the yarn strength is conveyed by resistance to fiber slippage. While fiber strength plays a direct role in yarn strength as well, the more twists a yarn can be given lends to

more fiber to fiber surface area to impart friction between fibers then the stronger the yarn will be. Longer fibers therefore are more important to overall yarn quality in ring spun yarns in comparison to rotor spun yarns (Rengasamy, 2010).

Rotor spinning is an open spinning method invented in the mid-1900s which works without the need for a spindle. Rotor spinning drafts individual fibers onto the end of a seed yarn and its main advantage over ring spinning is faster production speeds. The addition of individual fibers to the strand means that sliver can be used directly without further processing into roving; eliminating a step and thereby further increase production speeds. The method by which the yarn is formed creates a different yarn structure than ring spun yarns: a dense, heavily twisted core and an outer layer of fibers wrapped around it. Because of the difference in structure, rotor yarns are weaker but less prone to pilling and abrasion, and tend to be more uniform with lower hairiness (Adanur, 2001). Though rotor spinning cannot produce yarns as fine as ring technology, these yarn properties still make it advantageous for products such as wearable fabrics, household products, and various industrial products.

The first air jet machine was introduced in 1981 by Murata, a Japanese company (Hunter, 2006). As a process it is based on open-end technology and uses airflow to form the yarn and impart twist to the fibers. This new technology promises increasingly fast production speeds with current limits being around 400 m min^{-1} but also has the flexibility in yarn weights of a ring spinning system. The produced yarns are somewhat similar in structure to ring spun yarns, except that vortex yarns have a small core with little to no twist. This leads to yarns with the lower hairiness, higher uniformity, and

higher durability to wear than rotor yarns (Basu, 1992). The parallel fibers within the core also leads to increased absorbance and faster drying. The downside of vortex yarns is lower yarn strength because it is the outer wrapper which maintains yarn strength. Air jet spinning ranks fiber strength in higher priority than fiber length, however fiber with increased strength and length is necessary before the full advantage of this technology can be achieved (Günaydin and Soydan, 2016).

Traditional Breeding Methods and Phenotypic Selection of Fiber Traits

'Traditional' breeding methods implies a reliance on visually identifying superior or inferior traits in the field during a time before modern technology allowed phenotyping through instrumentation. Trait advancement relied upon the individual skill of a breeder and the heritability of the trait in question. Breeders currently have access to instruments capable of fast, accurate phenotyping in a combination of traditional, field based selection and technology that still remains distinct from genotypic methodology. Visual selections may be made in the field based purely on traits that the breeder can identify. These selections may be harvested and further tested for phenotype in a lab setting with final selections made based on these data in a two step progression of selecting the best plants or rows to reduce time and money spent harvesting and analyzing poor quality plants.

Up until the 1940s and the application of the fibrogram to testing fiber lengths cotton fiber length in the United States was tested and graded by hand (Hertel, 1940) . Visual grading remains common in other countries and within the U.S.A. in the hobbyist fiber industry. An experienced grader adhering to strict standards as published by the

USDA's *The Classification of Cotton* in a controlled environment and carefully examining and sorting each sample will have good accuracy in measuring the staple length of a sample (USDA, 1980). These conditions become increasingly difficult to meet where environmental conditions cannot be controlled and human error will be higher.

Numerous studies into the heritability of cotton fiber traits have shown that UHML has a high narrow sense heritability. Narrow sense heritability is the proportion of a plant's or variety's phenotype that can be attributed to additive genetic effects. If phenotyping is accurate, selections based on higher UHML than their general population should likewise produce progeny of higher UHML. In consideration to breeder selections of fiber length, even if human error in length measurements are high, as long as selections are not significantly worse than random then trait progress can still continue.

Genotypically Based Trait Selection

Genotypically based selection methods for quantitative traits are based on the association of genetic marker with quantitative trait loci (QTL). QTLs are segments of the genome which, based on statistical analysis, correlate with the genetic variation in a given trait. Genetic markers are specific DNA sequences in linkage disequilibrium to these loci and allow for genomic mapping and identification of allelic differences at these areas that may account for the variation seen in the trait. Linkage disequilibrium is the tendency of loci in close proximity to assort together during genetic recombination rather than randomly (Gupta et al., 2005). Paterson et al. (1988) were the first to use

restriction fragment length polymorphisms (RFLPs) to create a QTL linkage map. Further advances is research brought about amplified fragment length polymorphisms (AFLPs), single nucleotide polymorphisms (SNPs), simple satellite repeats (SSRs), and others. Technological advances have increased the speed and accuracy at which genetic samples can be analyzed while reducing the cost per sample making genotypic selection methods more accessible.

SSRs, or microsatellites, are the marker of the most interest to this study and are characterized by short, repetitive sequences that can be found widely distributed throughout the chromosome of eukaryotic organisms. The structure of these sections leads to DNA slippage during replication causing differing numbers of repeats which are what make them useful in differentiating between alleles at their associated loci (Li et al. 2002). SSRs have numerous advantages, including being widespread through coding and non-coding regions, co-dominant inheritance, having a high amount of polymorphism, and being relatively cheap. The current progress of genetic research also means that numerous mapping studies have been conducted in cotton, thus eliminating the work and time of identifying SSR sequences before application.

Two prominent methods of genetic based selection are the genome-wide association study (GWAS) and marker assisted selection (MAS). In a breeding program and assuming high accuracy in phenotypic prediction, both could save time and increase breeding efficiency by allowing selection at the seed stage and in off season-nurseries, and could help identify unique recombinants and outcrossed plants. For a GWAS, genetic and phenotypic samples are taken from an early generation, recombinant

population and trait associations are analyzed using hundreds to thousands of SNPs across the entire genome (Marees, 2018). The downside of this is that the study must be undertaken for every population in which it is used, and, specifically for moderately to highly heritability traits such as UHML and FBS which are the focus of this study, if phenotypic data must be collected then selecting parents based on phenotype alone will produce acceptable results while requiring significantly less labor. As well, the highly quantitative nature of both FBS, UHML, and other cotton fiber quality traits means that a sample size of at least 500 plants would be required to produce moderately accurate results in accordance the Beavis effect (Xu, 2003). With multiple concurrent populations, this could quickly become a prohibitive number.

MAS may be more beneficial in UHML and FBS, and more applicable in closing the selection gap discussed above. MAS is based on QTL linkage maps and making selections based on a set of specified set of markers already shown to be linked to the trait in order to pyramid genes (Collard and Mackill, 2008). MAS has been successfully applied to select for ear and plant height in maize and in wheat for several quantitative traits including grain weight and yield. However, despite this success in other crops, there are few studies of MAS application to cotton fiber traits in public breeding programs (Zhang et al., 2010.)(Law, 1967). At least part of this is due to the inconsistencies of QTL association found between mapping studies in cotton. One of the mechanisms that could explain this in respect to UHML and FBS is the high number of QTLs associated with both trait. Both are controlled by a few high effect QTL and a large number of QTL with small effects which may affect the consistency or robustness

of QTL estimation and therefore their transferability between populations which preclude the use of MAS without a new mapping study for each population. Similarly controlled traits often have high genotype x environment effects which also reduces transferability between mapping studies from different environments which is difficult to maintain between different studies (Basford and Cooper. 1998). This lack of QTL stability therefore makes application difficult.

Previous Study on Identification of Transferable Markers for UHML and FBS

One of the objective that the Hugie study sought to identify QTL markers in *Gossypium sp.* with stable effects between populations of different heritage (Hugie, 2015). Using three F_3 populations of diverse backgrounds during 2013 and 2014 and using the results of 31 previously published mapping studies for *Gossypium sp.*, the Hugie study attempted to identify SSR markers which had significant trait association for UHML and FBS and which were capable of predicting progeny phenotype on a comparable level to phenotypic selections for each trait. The three populations included in the study included a population with *G. mustelinum* and *G. tomentosum* background. Many mapping studies use interspecific crosses and it is commonly thought that many fiber quality alleles may have originated in non *G. hirsutum* species.

From the studies reported in the literature, 536 SSR primer pairs were used to genotype and analyze 713 randomly selected progeny from among the populations with populations individually examined (Hugie, 2015). Within the populations, non-polymorphic markers as well as markers with an extremely low degree of polymorphism were removed from the set and a two-step process of combining single-marker analysis

and then stepwise regression was performed following methodology laid out by Dudley (1993). Markers found nonsignificant at α=.05 were discarded. The remaining set of markers were validated by regressing $F_{3:4}$ row phenotypes onto their respective plant-parent by number of SSR markers and again nonsignificant markers were discarded. Six markers for FBS and six markers for UHML was found to be significant across all populations. A comparison was also made between progeny trait quality based on parental selections of the best 20% for each trait versus progeny trait quality based on 20% of parents with the most markers indicating the desired state of the allele and these two groups were found to not be significantly different for either FBS or UHML with the exception of a single population where phenotypic selections were superior in FBS. Of the identified markers, $BNL1604_{98}$, a marker which was originally mapped in an interspecific population, was found to negatively correlate with high fiber quality for both traits and account for a relatively high amount of trait variation. This marker was therefore suggested as potentially stable through populations and a candidate in possible MAS methodologies.

From this study, the large number of markers that had to be discarded is a clear indication of why MAS is not common for these traits; markers are largely specific to the study from which they originated whether that be due to one of the previously indicated environmental and genetic interactions or due to epistatic interactions between genetic backgrounds as Hugie suggested in the conclusion to her study. The work necessary to validate the markers including the regressions between generation phenotypes is an obvious disincentive when that work could instead be used to either

phenotype more individual plant selections or work with larger or more numerous populations. Identifying a set of markers to work around this would be highly beneficial and this study seeks to validate the conclusions of the Hugie study and observe the identified markers in two new populations.

Objectives

1. Confirm stability of the "Hugie QTLs" and two additional QTLs suggested by Dr.David Fang, USDA-ARS genomist, using SSR (simple sequence repeats) markers in two *G. hirsutum* populations (one F2 and one F3 population).
2. From the same F2 and F3 populations, use traditional phenotypic selection based on high volume instrument (HVI) determined fiber properties.
3. Compare the Hugie and Fang QTLs-based selections with HVI phenotypic selection for effectiveness and efficiency.

CHAPTER III

MATERIALS AND METHODS

Segregating Populations

Two populations were selected for use in the study due to their diverse parentage and the expectation that progeny would segregate for both UHML and FBS. The Hugie study was interested in markers that would be capable of discerning improved fiber qualities associated with interspecific backgrounds which is reflected herein by using one segregating population with 'Del Cerro' (Smith et al., 1999) in its pedigree and thus considered to have some *G. barbadense* introgression and one segregating population with no known interspecific introgression. At the beginning of the study in 2016, both populations were extant at the Texas A&M Cotton Improvement lab (CIL) as early generation populations.

Population one in the present study was an F2 population in 2016 and was derived from a three way cross of TAM 11K-13 ELSU (Smith et al., 2014) / Del Cerro // 13P-54 ELSU. Both 13P-54 ELSU and TAM 11K-13 ELSU are extra-long staple uplands (ELSU) developed at the CIL for improved UHML. 13P-54 ELSU is an unreleased experimental line. Del Cerro is an obsolete cultivar from the early 1900s with a diverse background that included *G. hirsutum*, *G. barbadense*, *G. herbaceum*, and *G. thurberi.* At the time of its original selection, Del Cerro had an UHML and FBS competitive with the pima cultivars of its time, and was later reselected in the 1980s for limited production in Arizona (Smith et al., 1999).

Population two was an F3 population in 2016, derived from a cross between TAM 11K-13 ELSU / TAM 06WE-621 ESU. TAM 06WE-621 ESU is an extra strength upland (ESU) also derived at the CIL. It resulted from a four parent hybridization: DP 491 (PI 618609)/TAM 96WD-18 (Thaxton et al., 2005)//TAM 91C-95Ls (Smith, 2001)/Deltapine Acala90 (PI 564767). Both of the TAM lines are released germplasm lines developed at the CIL for improved fiber quality. DP 491 and Deltapine Acala 90 are Deltapine cultivars developed for general improved quality and performance (Bowman, 2006; USDA, 2006).

Field Trials

In 2016, both populations were planted on May 7 in College Station, Texas at the Texas A&M AgriLife Research Farm. The soil in the area consist of Weswood silt, a very fine Chromic Hapludert, and Ships clay, a fine-silty Udifluventic Haplustept which are both mixed, thermic, and active to superactive. Standard growing procedures were followed including the use of herbicides, pesticides, and furrow irrigation as needed. At approximately 30 days post emergence, plants were thinned to one plant every 20 cm. From each population, both genotypic and phenotypic selections were made in order to examine the efficacy of each selection method.

Genotypic Selections

Two months after planting in 2016, 250 plants were randomly selected from population one and 100 plants selected from population two. Selections were numerically tagged and a 5 g tissue sample was taken from the meristems of each plant. Samples were placed into individually marked test tubes, kept on ice in the field, and

then transferred to storage at -80^0 C before being shipped to the USDA-ARS Cotton Fiber Bioscience Research Unit at the Southern Regional Research Center in New Orleans, LA for genotyping. Non-flanking SSR markers were used following the methodology of Hugie (2015). All markers and their corresponding traits are listed in Table 1.

Table 1. SSR markers previously found to correlate with cotton fiber traits in multiple populations by Hugie and the study in which they were originally selected from.

SSR Marker	Trait	Association†	Publication or Source
$BNL1604_{98}$	UHML‡	Negative	Said et al. 2013
	FBS §	Negative	
$BNL4017_{234}$	UHML	Negative	Zeng et al. 2009
$CGR5548_{162}$	UHML	Negative	Fang et al. 2014
$CIR196_{192}$	UHML	Negative	Zeng et al. 2009
$NAU1369_{247}$	UHML	Positive	Shen et al. 2006
	FBS	Negative	Shen et al. 2007
$NAU5046_{226}$	UHML	Positive	Fang et al. 2014
$CGR6329_{232}$	FBS	Negative	Fang et al. 2014
$DPL0236_{157}$	FBS	Negative	Fang et al 2014
$NAU1102_{231}$	FBS	Positive	Cai et al. 2014
$TMB0382_{179}$	FBS	Positive	Tan et al. 2015
$A07id76_{146}$	FBS	Positive	Fang, personal correspondence, 2016
$C2\text{-}0114_{149}$	FBS	Positive	Fang, personal correspondence, 2016

† Association refers to whether the allele of interest is negatively or positively correlated with the respective trait it has been observed in.
‡Upper half mean length
§ Fiber bundle strength

During the Hugie study (2015), marker sequences had previously been found using the CottonGen database (https://www.cottongen.org). Following the same methodology, fluorescent labelling was applied to the forward primers at the 5' end using either 6-FAM (6-carboxyfluorescein), HEX (4, 7, 2', 4', 5, 7-hexachloro-carboxyfluorescein), or NED (7', 8'-benzo-5-fluoro 2', 4, 7, -trichloro-5-carboxyfluorescein) and multiplex PCR used for each sample. Fang et al. (2010) characterizes the protocol used for the polymerase chain reaction (PCR) for the SSR primer pairs. After amplification, the ABI3730XL (Applied Biosystems Inc., Foster City, CA, USA) automated capillary electrophoresis system was used to separate and size the labeled fragments.

After genotyping, plants were selected based on the total number of alleles in the desired state predicted to maximize their respective trait: present for positively correlating alleles, or absent for negatively correlating alleles. For each trait, UHML and FBS, the goal was to create diverging subpopulations in order to study the efficacy and discriminatory ability of MAS with the given markers. For Population One, the 20% of plants with the largest number of alleles in the desired state and the 20% of plants with the fewest number of alleles in the desired state were advanced to field trials in 2017. The same process was initiated for Population Two and divergent selections were identified, however, due to overlapping research with another graduate student with different objectives, all 250 plants which were genotyped were advanced for observation in 2017.

Phenotypic Selections

Traditional plant breeding methods refer to the process of visually selecting plants based on apparent phenotype. While fiber length can be evaluated to an extent in the field by measuring fibers pulled away from the seed coat, FBS must be measured using objective instrumentation. Therefore, phenotypic selection was a two-step process initially based only on maximum fiber length visually identifiable in the field. At harvest time, from populations one and two, 250 and 100 plants were selected, respectively, on the basis of visual fiber length. Selected plants were marked with ribbon for later sample collection.

Fiber Sample Analysis

For both selection methods, 30 boll samples were harvested by hand from the identified plants and placed into individually labeled sacks. Each sample was ginned on the same 10-saw laboratory gin. The seeds were labeled and stored in a temperature and humidity controlled environment, and the fiber samples were shipped to the Texas Tech University Fiber and Biopolymer Research Institute (FBRI) in Lubbock, Texas for high volume instrument (HVI) testing. The resulting data were used to make the final phenotypic selections; the top and bottom 20% for both UHML and FBS were advanced to 2017 for both populations.

2017 Field Observations

In 2017, individual plants selections were planted on April 26 to progeny rows in the same location as 2016 and with the same management practices. Harvest was delayed due to field flooding caused by a hurricane. Due to the late season and possible

uneven fiber degradation from part of the field retaining more water, 75 boll samples were harvested from each row rather than 30. Bolls were harvested as evenly as possible across the entire row with preference for first node bolls from the middle of the plant. Samples were treated the same as in 2016 and fiber shipped to the FBRI for analysis.

Statistical Analysis

Previous studies have found that UHML and FBS are moderately to highly heritable traits with fluctuations due to environmental and minor genetic effects. Due to the structure of the study as a CRD, genotypic effects cannot be separated into additive and dominant portion preventing calculation of broad sense heritability (the ratio of total genetic variance to phenotypic variance) or narrow sense heritability (the ratio of additive genetic variance divided to phenotypic variance). However, narrow sense heritability can be estimated as the slope of a line drawn by simple linear regression correlating progeny trait values with their midparent value. Here, progeny rows were grown from selfed bolls harvested from individual plant selections meaning that the midparent value is the value of the selfed parent.

Data were first analyzed to compare overall correlations and estimate narrow sense heritability of the traits between 2017 and 2016 observations for both populations. Data for visual and genotypic selections were combined and the averaged trait data from each progeny row from 2017 was regressed onto the individual plant data from 2016 by simple linear regression using JMP (SAS Institute Inc., 2019). These values were compared to the realized heritability which is found using the following formula as described by Fehr (1987).

$$h^2 = \frac{\bar{x}_{high,F3} - \bar{x}_{low,F3}}{\bar{x}_{high,F2} - \bar{x}_{low,F2}}$$

These parameters show preliminary differences in trait progression from one generation to the next between the two selection methods.

Within Each population and trait, visually selected progeny and genotypically selected progeny were then separately analyzed by analysis of variance (ANOVA) in JMP to observe statistical differences between means of the divergent subgroups. Each was analyzed as a completely random design (CRD) with no random effects such that the linear modelling statement is:

$$Y_{ij} = \mu + \tau_i + \varepsilon_{ij}$$

where Y_{ij} is the j^{th} observation of the i^{th} subgroup, μ is the population mean, τ_i is the effect of the i^{th} subgroup, and ε_{ij} is the random error. τ_i has three levels of I, i.e., 0 for the divergent subgroup formed by the individual plants selected for predicted longest or strongest fibers, 1 for the subgroup formed by the plants selected for predicted shortest or weakest fibers, and 2 for plants that were selected based on the converse fiber trait which serve as a control. For the ANOVAs that indicated significant differences between subgroup means, the subgroups were further examined using Tukey's HSD to determine which subgroups differed.

An ANOVA was then used to compare the subgroups from the visual selections to those of the genotypic selections with the modelling statement:

$$y_{ijk} = \mu + \alpha_i + \beta_j + \alpha\beta_{ij} + \varepsilon_{ijk},$$

where y_{ijk} is the kth observation of the i^{th} subgroup of the j^{th} selection method, μ is the population mean, α_i is the effect of the i^{th} subgroup, β_j is the effect of the j^{th} selection method, and $\alpha\beta_{ij}$ is the interaction between the selection method and the subgroup. α_i has the same levels defined above for τ_i. β_j has two levels of j, 0 for visual selections and 1 for genotypic selections. Tukey's HSD was again used to separate the subgroups. By comparing the subgroup means directly, we summarize which method had the greatest efficacy in selecting plants whose progeny have improved trait characteristics.

Finally to examine the effects and group effects of the individual SSR alleles included in the study, a stepwise multiple linear regression was used following the same methodology as Hugie (2015). The regression was calculated in mixed direction and the probability-to-enter and probability-to-leave were both set as $\alpha > .05$. Because the ANOVAs only show the effects of the alleles as a group, it is possible that alleles with a significant trait effect could be hidden by non-significant alleles; this regression ensures that all possibilities were considered.

CHAPTER IV

RESULTS AND DISCUSSION

Trait Heritability

Fiber length and strength are two of the most important traits controlling final product quality in textile mills; as such, both traits and their genetic basis have been thoroughly researched. Previous studies have found that UHML and FBS are moderately to highly heritable traits with fluctuations due to environmental and minor genetic effects. Due to the structure of the study as a CRD, genotypic effects cannot be separated into additive and dominant portions, preventing calculation of broad sense heritability (the ratio of total genetic variance to phenotypic variance) or narrow sense heritability (the ratio of additive genetic variance divided to phenotypic variance). However, narrow sense heritability can be estimated as the slope of a line drawn by simple linear regression correlating progeny trait values with their midparent value. Here, progeny rows were grown from selfed bolls harvested from individual plant selections meaning that the midparent value is the value of the selfed parent.

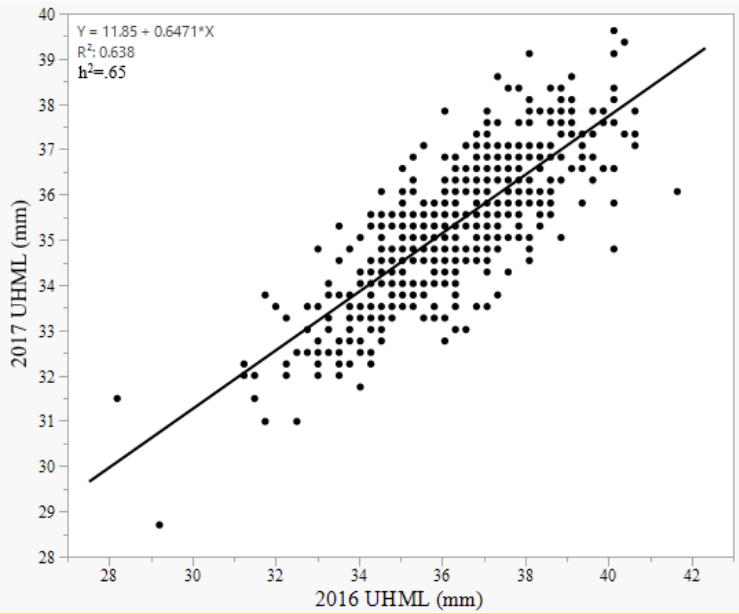

Figure 1. Estimation of narrow sense heritability of upper half mean length (UHML) using a simple linear regression of 2017 progeny UHML on 2016 parent UHML. The slope of the line of best fit estimates the narrow sense heritability for UHML. The Pearson correlation coefficient indicates the amount of variation explained by the model.

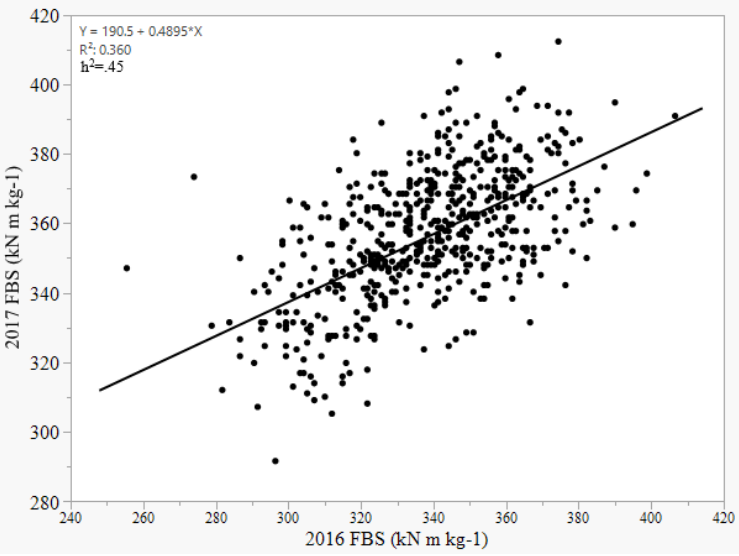

Figure 2. Estimation of narrow sense heritability of fiber bundle strength (FBS) using a simple linear regression of 2017 progeny FBS on 2016 parent FBS with combined populations. The slope of the line of best fit estimates the narrow sense heritability for FBS. The Pearson correlation coefficient indicates the amount of variation explained by the model.

Using combined phenotypic data from both populations and selection methods, overall trait heritability for UHML was estimated as 0.65 (Figure 1) and for FBS as 0.49 (Figure 2). Estimating heritability for the two populations separately (Figure 3; Figure 4) shows a lower correlation for each which is expected due to the smaller population size and differing fiber qualities within each population. Realized heritabilities calculated following Fehr's procedure (1987) for the phenotypic selections were 0.59 for Population 1 for UHML and 0.38 for FBS. Population 2 exhibited similar values at 0.60 and 0.27 for UHML and FBS, respectively. It has been shown that heritability can differ depending on the genetic background of the population and these estimates conform to that standard (El-Hashash, 2017).

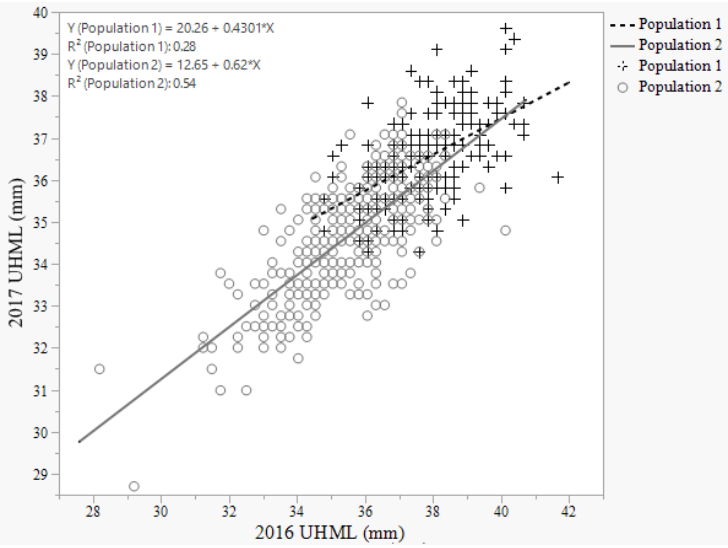

Figure 3. Estimation of narrow sense heritability of upper half mean length (UHML) using a simple linear regression of 2017 progeny UHML on 2016 parent UHML with two segregating populations. The slope of the line of best fit estimates the narrow sense heritability for UHML. The Pearson correlation coefficient indicates the amount of variation explained by the model. Population 1 was an F_2 from TAM 11K-13 ELSU/ Del Cerro//13P-54 ELSU and Population 2 was an F_3 from TAM 11K-13/ TAM 06WE-621 ESU.

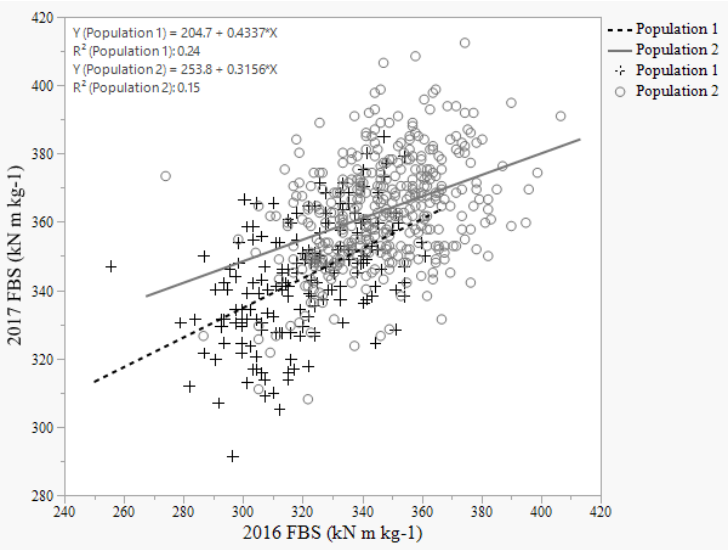

Figure 4. Estimation of narrow sense heritability of fiber bundle strength (FBS) using a simple linear regression of 2017 progeny FBS on 2016 parent FBS with two segregating populations. The slope of the line of best fit estimates the narrow sense heritability for FBS. The Pearson correlation coefficient indicates the amount of variation explained by the model. Population 1 was an F_2 from TAM 11K-13 ELSU/ Del Cerro//13P-54 ELSU and Population 2 was an F_3 from TAM 11K-13/ TAM 06WE-621 ESU.

These values help define potential advances that can be made in each trait solely through phenotypic selection. In consideration of visually based selection and testing of plants in the field, a comparison of estimated heritability and realized heritability shows how effective a given plant breeder was at identifying the trait of interest. If the goal is to make fiber quality advances as quickly and efficiently as possible, then new methods of selection could reduce the differences between realized heritability and narrow sense heritability.

In smaller breeding programs where space, money, and time are limited, the main benefits of field screening, especially in early generations where genetic diversity is highest, is as a resource saving measure. Field screening for visually obvious traits,

mean fewer plants to hand harvest and fewer samples to be objectively or machine phenotyped, which, in the case of cotton, means HVI or the newer Advanced Fiber Information System (AFIS) which provides more information but is subsequently more expensive. In a perfect program, where every single plant in an early generation field could be accurately phenotyped, trait progression between generations would be maximized for highly heritable traits such as FBS or UHML. Introducing human deficiencies in selection creates a quality gap between the actual selections and the 'perfect' selections. Large, commercial programs can minimize the difference by harvesting and phenotyping with instrumentation a larger number of samples, and even this does not account for error caused by genotypic and environmental interactions. This gap is an area where newer technologies such as high throughput phenotyping and genotypic methods have the most potential in advancing trait progress.

In 2016, initial field selections for the phenotypic groups were made based on visual observations of fiber length. Figure 5 shows the UHML distributions of these selections compared with the random plants selected for QTL marker analysis. In population 1, the mean UHML of the random plants was 37.9 mm and 38.1 in the visual selections was 38.1 mm were not found to be different ($p=.2943$) using a student's t-test with $\alpha=.05$. In population 2, the random plant mean was 35.3 mm and the visual mean was 35.7 mm which were different ($p=.0013$) using the same test.

The visual and marker assisted selections for Population 1 do not differ significantly, and while the means for Population 2 are significantly different; numerically they are within a millimeter of each other. Even though realized heritability

was high, this distribution supports that it is likely a large number of plants with good fiber traits within the visual group remained unselected due to human bias caused by skill level in selections, outer appearance of the plant regardless of fiber traits, and time restrictions. Even with highly heritable traits, genetic screening may be competitive versus phenotypic screenings the fiber traits of all plants in the population are estimated.

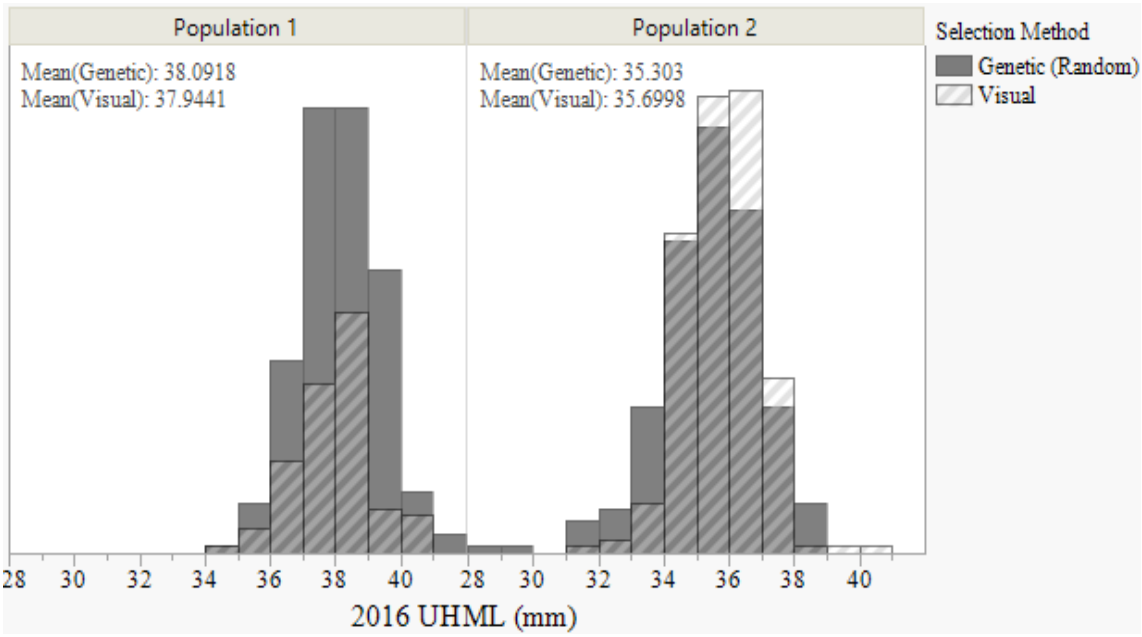

Figure 5. Comparison of upper half mean length (UHML) distribution from visually selected plants to plants selected for simple sequence repeat (SSR) based marker assisted selection (MAS) in populations 1 and 2. The visually selected plants were identified as independent plant selections on the basis of fiber length, while the genetic selections were random. Population 1 was an F_2 from TAM 11K-13 ELSU/ Del Cerro//13P-54 ELSU and Population 2 was an F_3 from TAM 11K-13/ TAM 06WE-621 ESU.

Analysis of Divergent Subgroups for Each Trait and Population

By selecting for divergent populations, the ability of each selection method to accurately discriminate between high and low value plants was evaluated. Across both populations, ANOVA for phenotypic selections showed significant divergence (p=.001) in trait means for both UHML and FBS (Table 2).

Table 2. ANOVA of 2017 UHML and FBS for divergent subgroups selected based on phenotype of parents grown in 2016 in College Station.

			Mean Square	
	Source	DF	UHML[†]	FBS[‡]
Population 1	subgroup[§]	2	22.23***	1701.57***
	error	61	.84	221.92
Population 2	subgroup	2	49.00***	2072.31***
	error	152	.70	206.78

*** Significant at the .001 probability level.

[†]Upper half mean length

[‡] Fiber bundle strength

[§]Three subgroups are included. Two divergent subgroups were formed from the 20% of plants with the highest UHML and FBS, the 20% of plants with the lowest UHML and FBS. The third subgroup is composed of plants selected for inclusion in 2017 based on the other fiber trait. For UHML this would be plants chosen based on either high or low FBS and vice versa.

Table 3. UHML and FBS 2017 means for divergent subgroups selected based on the phenotype of parents grown in 2016 in College Station.

Population 1	Mean UHML† (mm)	Mean FBS‡ (kN m kg^{-1})
Top 20% §	37.7 a ¶	357 a
No selection	37.2 a	347 b
Bottom 20%	35.9 b	339 b
Population 2		
Top 20%	35.8a	368 a
No selection	35.2 b	362 a
Bottom 20%	33.9 c	355 b

†Upper half mean length

‡Fiber bundle strength

§ The top 20% subgroup is composed of 2017 progeny selected based on independent plant selections which had the highest UHML and FBS in 2016. The bottom 20% were based on the lowest UHML and FBS. The third, no selection, subgroup is composed of plants selected for inclusion in 2017 based on the other fiber trait. For UHML this would be plants chosen based on either high or low FBS and vice versa.

¶ Subgroup Means within columns and populations that are followed by the same letter are not significantly different according to Tukey's HSD at $\alpha = 0.05$.

The random selection subgroup was significantly different from the divergent subgroup for UHML in Population 2, but not for Population 1 (Table 3). Though not significantly different, the random selection subgroup was numerically intermediate to the other two which does agree with expectations based on the realized heritability for the population.

Conversely, subgroups formed by genotypic MAS showed no significant differences in either population or for either trait (Table 4). Numerical differences between subgroups also failed to follow any trend showing that the MAS failed to discriminate at even a minor level (Table 5). As well, realized heritabilities could not be accurately calculated for the genotypic selections due to inflation of the heritability value caused by lack of differentiation in the parental subgroups from 2016.

Table 4. ANOVA of 2017 UHML and FBS for divergent subgroups selected using SSR based MAS of genotyped parents grown in 2016 in College Station.

			Mean Square	
	source	DF	UHML†	FBS‡
Population 1	supgroup§	2	.28	706.70
	error	106	.86	272.81
Population 2	supgroup	2	0.42	383.43
	error	241	1.70	265.37

* Significant at the .05 probability level.
† Upper half mean length
‡ Fiber bundle strength
§ Three subgroups are included. Two divergent subgroups were formed from the 20% of plants with the largest number of alleles in the desired state for UHML and FBS in 2016, and the 20% of plants with the fewest number of alleles in the desired state for UHML and FBS in 2016. The third subgroup is composed of plants selected for inclusion in 2017 based on the other fiber trait. For UHML this would be plants chosen based on FBS QTLs.

Table 5. UHML and FBS 2017 means for divergent subgroups selected using SSR based MAS of genotyped parents grown in 2016 in College Station.

Population 1	Mean UHML† (mm)	Mean FBS‡ (kN m kg^{-1})
Top MAS-predicted 20%§	36.2 a¶	346 a
No selection	36.4 a	343 a
Bottom MAS-predicted 20%	36.3 a	337 a
Population 2		
Top MAS-predicted 20%	34.5 a	364 a
No selection	34.5 a	362 a
Bottom MAS-predicted 20%	34.4 a	366 a

†Upper half mean length
‡ Fiber bundle strength
§The top MAS-predicted 20% subgroup was derived from the 20% of plants with the largest number of alleles in the desired state for UHML and FBS in 2016. The bottom MAS-predicted 20% subgroup was derived from the 20% of plants with the fewest number of alleles in the desired state for UHML and FBS in 2016. The third, no selection, subgroup is composed of plants selected for inclusion in 2017 based on the other fiber trait. For UHML this would be plants chosen based on FBS QTLs.
¶ Means within columns and populations followed by the same letter are not significantly different according to Tukey's HSD at $\alpha = 0.05$.

Direct Comparison of Selection Methods

The two populations utilized in the study had different parental origins and therefore ANOVAs were run separately for each population rather than include a three-way-interaction in order to make results as clear as possible. Of the resulting data tables, the UHML analyses showed two way interactions of selection method and divergent subgroup that were significant at α=.01 and the FBS analyses showed this interaction significant at α=.05 (Table 6; Table 7).

Table 6. ANOVA for Population 1, TAM 11K-13 ELSU/Del Cerro//13P-54 ELSU, comparing divergent subgroups formed by phenotypic and SSR† marker assisted selection methods for UHML and FBS.

		Mean Square	
Source‡	df	UHML§ (mm)	FBS¶ ($kNmkg^{-1}$)
Subgroup	2	20.48***	1600.22**
Selection Method	1	13.90***	1157.49*
Selection Method x Subgroup	2	29.33***	820.43*
error	167	0.86	254.22

* Significant at the .05 probability level, ** Significant at the .01 probability level, *** Significant at the .001 probability level.

†Simple sequence repeats

‡Divergent subgroups were selected from 2016 parents. For the phenotypic selection method, subgroups were chosen based on 20% highest and 20% lowest for UHML and FBS. Genotypic subgroups were chosen based on 20% with the most and 20% with the least number of alleles in the desired state for UHML and FBS. For both phenotypic and genotypic methods, a third subgroup is included based on plants selected for inclusion in 2017 trials based on the other fiber trait. For UHML this would be plants included on the basis of FBS and vice versa.

§Upper half mean length

¶Fiber bundle strength

Table 7. ANOVA for Population 2, TAM 11K-13 ELSU/TAM06WE-621 ESU, comparing divergent subgroups formed by phenotypic and SSR† MAS‡ methods.

		Mean Square	
Source	df	UHML§ (mm)	FBS¶ ($kNmkg^{-1}$)
Subgroup#	2	29.95***	919.41*
Selection Method	1	22.59***	679.55
Selection Method x Subgroup	2	41.78***	1915.48***
error	393	1.31	242.71

* Significant at the .05 probability level, ** Significant at the .01 probability level, *** Significant at the .001 probability level.
† Simple sequence repeat
‡ Marker assisted selection
§ Upper half mean length
¶ Fiber bundle strength
Divergent subgroups were selected from 2016 parents. For the phenotypic selection method, subgroups were chosen based on 20% highest and 20% lowest for UHML and FBS. Genotypic subgroups were chosen based on 20% with the most and 20% with the least number of alleles in the desired state for UHML and FBS. For both phenotypic and genotypic methods, a third subgroup is included based on plants selected for inclusion in 2017 trials based on the other fiber trait. For UHML this would be plants included on the basis of FBS and vice versa.

For fiber length in Population 1, the phenotypic selections based on longest UHML and those which were randomly selected statistically had the longest (p=.05) mean UHML while the other subgroups were undifferentiated (Table 8). Population 2 was similar (Table 9). It is interesting to note that the genotypic subgroups with no selection had a lower (p=.05) mean than the phenotypic subgroup with no selection even though data showed that the distribution between the parents of both groups had similar means.

Table 8. Mean 2017 UHML in Population 1, TAM 11K-13 ELSU/Del Cerro//13P-54 ELSU, comparing divergent subgroups selected using phenotypic selection to those selected using SSR† MAS‡.

Method by Subgroup§	Mean UHML¶ (mm)
Visual - Longest 20%	37.7 a#
Visual - No Selection	37.2 a
Genotypic - Shortest MAS-predicted	36.4 b
Genotypic - No Selection	36.3 b
Genotypic - Longest MAS- predicted	36.2 b
Visual - Shortest 20%	35.9 b

†Simple sequence repeats

‡Marker assisted selection

§Divergent subgroups were selected from 2016 parents. For the phenotypic selection method, subgroups were chosen based on 20% longest and 20% shortest UHML. Genotypic subgroups were chosen based on 20% with the most and 20% with the least number of alleles in the desired state for UHML. For both phenotypic and genotypic methods, a third subgroup is included based on plants selected for inclusion in 2017 trials based on fiber bundle strength.

¶Upper half mean length

#Means followed by the same letter are not significantly different according to Tukey's HSD at $\alpha = 0.05$.

Table 9. Mean 2017 UHML in Population 2, TAM 11K-13 ELSU/TAM06WE-621 ESU, comparing divergent subgroups selected using phenotypic selection to those selected using SSR† MAS‡.

Method by Subgroup§	Mean UHML¶ (mm)
Visual - Longest 20%	35.8 a#
Visual – No Selection	35.2 a
Genotypic - Longest MAS-predicted	34.5 bc
Genotypic – No Selection	34.5 b
Genotypic – Shortest MAS-predicted	34.4 bc
Visual – Shortest 20%	33.9 c

† Simple sequence repeats

‡ Marker assisted selection

§ Divergent subgroups were selected from 2016 parents. For the phenotypic selection method, subgroups were chosen based on 20% longest and 20% shortest UHML. Genotypic subgroups were chosen based on 20% with the most and 20% with the least number of alleles in the desired state for UHML. For both phenotypic and genotypic methods, a third subgroup is included based on plants selected for inclusion in 2017 trials based on fiber bundle strength.

¶ Upper half mean length

Means followed by the same letter are not significantly different according to Tukey's HSD at $\alpha = 0.05$.

The analyses for FBS overall showed less separation between groups, likely due to overall lower heritability but most likely because the parents were less variable for FBS in each population. In both populations, while the phenotypically selected subgroup for strongest fibers had the numerically highest mean, it was not different (p=.05) from the mean FBS of the group genotypically predicted to have the strongest fiber (Table 10; Table 11). Literature commonly cites that genotypic selection methods are often more efficient than phenotypic methods for lower heritability traits (Hospital, 1997). Trait progression with FBS may therefore benefit more through genotypic selection than UHML. Table 3 shows that the visually selected subgroups for FBS do show a level of discrimination for phenotype. The population would need to be further followed through generations to observe whether phenotypic and genotypic subgroups segregate to a definitive level.

Table 10. Mean 2017 FBS in Population 1, TAM 11K-13 ELSU/Del Cerro//13P-54 ELSU, comparing divergent subgroups selected using phenotypic selection to those selected using SSR† MAS‡.

Method by Subgroup§	Mean FBS¶ ($kNmkg^{-1}$)
Visual - Strongest 20%	357 a#
Visual - No Selection	347 ab
Genotypic - Strongest MAS-predicted	346 ab
Genotypic - Weakest MAS-predicted	343 b
Visual - Weakest 20%	339 b
Genotypic- No Selection	337 b

†Simple sequence repeats
‡Marker assisted selection
§Divergent subgroups were selected from 2016 parents. For the phenotypic selection method, subgroups were chosen based on 20% strongest and 20% weakest FBS. Genotypic subgroups were chosen based on 20% with the most and 20% with the least number of alleles in the desired state for FBS. For both phenotypic and genotypic methods, a third subgroup is included based on plants selected for inclusion in 2017 trials based on upper half mean length.
¶Fiber bundle strength
#Means followed by the same letter are not significantly different according to Tukey's HSD at $\alpha = 0.05$.

Table 11. Mean 2017 FBS in Population 2, TAM 11K-13 ELSU/TAM06WE-621 ESU, comparing divergent subgroups selected using phenotypic selection to those selected using SSR† MAS‡.

Method by Subgroup§	Mean FBS¶ ($kNmkg^{-1}$)
Visual - Strongest 20%	368 a#
Genotypic - Weakest MAS-predicted	366 a
Genotypic - Strongest MAS-predicted	364 a
Visual - No Selection	362 ab
Genotypic - No Selection	362 ab
Visual - Weakest 20%	354 b

† Simple sequence repeat
‡ Marker assisted selection
§ Divergent subgroups were selected from 2016 parents. For the phenotypic selection method, subgroups were chosen based on 20% strongest and 20% weakest FBS. Genotypic subgroups were chosen based on 20% with the most and 20% with the least number of alleles in the desired state for FBS. For both phenotypic and genotypic methods, a third subgroup is included based on plants selected for inclusion in 2017 trials based on upper half mean length.
¶Fiber bundle strength
#Means followed by the same letter are not significantly different according to Tukey's HSD at $\alpha = 0.05$.

Individual Marker Analysis

Only markers with sufficient segregation across all three populations were considered by (2015). Utilizing the final set of Hugie markers to evaluate the two populations used in this study showed that several markers were homogenous in one or both populations or did not have enough variation for non-biased analysis. These markers were therefore precluded from the individual marker ANOVAs and the stepwise regressions.

Table 12. ANOVA for Individual Markers previously correlated with UHML in Populations 1, TAM 11K-13 ELSU/Del Cerro//13P-54 ELSU, and 2, TAM 11K-13 ELSU/TAM06WE-621 ESU. Prob>F

UHML†	Population 1		Population 2	
Marker ‡	2016	2017	2016	2017
$BNL1604_{98}$	0.3679	0.9595	0.027*	0.0448*
$CIR196_{197}$	0.0416*	0.2128	.	.
$CGR5548_{162}$	.	.	.	.
$NAU5046_{226}$	.	.	0.3092	0.4442
$NAU1369_{247}$	0.0011**	0.2958	0.3380	0.6537
$BNL4017_{234}$	0.3250	0.9903	.	.

* Significant at the .05 probability level, ** Significant at the .01 probability level.

† Upper half mean length

‡ Each marker that was previously identified to correlate with UHML by Kari Hugie was individually analyzed using an ANOVA to determine whether there were significant effects on UHML between plants with QTL which were homozygous for the desired allele, heterozygous, or homozygous for an undesired allele.

Table 13. ANOVA for Individual Markers previously correlated with FBS in Populations 1, TAM 11K-13 ELSU/Del Cerro//13P-54 ELSU, and 2, TAM 11K-13 ELSU/TAM06WE-621 ESU. Prob>F

FBS†	Population 1		Population 2	
Marker‡	2016	2017	2016	2017
$BNL1604_{98}$	0.8301	0.3080	0.0425*	0.1447
$CGR6329_{232}$	0.1712	0.0278*	0.1852	0.2332
$NAU1369_{247}$	0.1064	0.3082	0.1487	0.5540
$TMB0382_{179}$	0.3278	0.2464	.	.
$A07id76_{146}$	0.5674	0.1070	.	.
$C2\text{-}0114_{149}$	.	.	.	.
$NAU1102_{231}$	.	.	0.6705	0.1356
$DPL0236_{157}$	.	.	.	.

* Significant at the .05 probability level.

† Fiber bundle strength

‡ Each marker that was previously identified to correlate with FBS by Kari Hugie (2015) was individually analyzed using an ANOVA to determine whether there were significant effects on FBS between plants with SSRs which were homozygous for the desired allele, heterozygous, or homozygous for an undesired allele.

For Population 1, none of the SSR markers evaluated were significant for either trait across both 2016 and 2017 (Table 12; Table 13). $CIR196_{197}$ identified plants with improved UHML (p=.04), however the trait association was negative for the desired state while positive in the Hugie (2015) study. Similarly, heterozygotes of $NAU1369_{247}$ exhibited the highest UHML means. The stepwise regressions reflects the significance of these two markers by including them in the models for UHML for both years, however, $CIR196_{197}$ showed opposite effects between the models for 2016 and 2017 (Table 14).

Table 14. Stepwise Regression of UHML† for individual plants in 2016 and progeny rows in 2017 for Populations 1, TAM 11K-13 ELSU/Del Cerro//13P-54 ELSU, and 2, TAM 11K-13 ELSU /TAM06WE-621 ESU, based on SSR‡ single marker analysis.

Population 1			Population 2		
2016	Estimate	Publication	2016	Estimate	Publication
Intercept	37.9604		Intercept	35.4252	
$CIR196_{192}$ [2-1&0] §	-0.1581*	Zeng et al.2009	$BNL1604_{98}$ [0&1-2]	-0.2618*	Said et al. 2013
$NAU1369_{247}$ [1-0&2]	-0.298**	Shen et al. 2006			
adj. R^2 = 0.0649			adj. R^2 = .0243		
2017			2017		
Intercept	36.4588		Intercept	34.4337	
$CIR196_{192}$ [1&2-0]	-0.2537*	Zeng et al. 2009	$BNL1604_{98}$ [0&1-2]	-0.1905*	Said et al. 2013
$NAU1369_{247}$ [1-0&2]	-0.2385*	Shen et al. 2006			
adj. R^2 = 0.0817			adj. R^2 = .0215		

* Significant at the .05 probability level, ** Significant at the .01 probability level.

† Upper half mean length

‡ Simple sequence repeat

§ Markers were classified based on the desired state of the allele. 2 signifies homozygous for the desired state. 1 signifies heterozygous. 0 signifies homozygous in an undesirable state. Number in brackets following the marker identify the allele state which the estimates are based on. Only markers with a significant effect are included in the table.

The stepwise models have adjusted R^2 values of 0.0649 and 0.0817 for 2016 and 2017, respectively, meaning that useful applicability of the models is doubtful. The stepwise regression models for FBS did not have any markers which remained significant across both years (Table 15).

Table 15. Stepwise Regression of FBS† for individual plants in 2016 and progeny rows in 2017 for Populations 1, TAM 11K-13 ELSU/Del Cerro//13P-54 ELSU, and 2, TAM 11K-13 ELSU /TAM06WE-621 ESU, based on SSR‡ single marker analysis.

Population 1			Population 2		
2016	Estimate	Publication	2016	Estimate	Publication
Intercept	320.6338		Intercept	352.415	
$NAU1369_{247}$ [0-1&2] §	-2.5537*	Shen et al. 2007	$BNL1604_{98}$ [0&1-2]	-3.6587**	Said et al. 2013
			$CGR6329_{232}$ [1&2-0]	-2.9433*	Fang et al. 2014
adj. R^2 = .0156			adj. R^2 = .0385		
2017			2017		
Intercept	343.6081		Intercept	365.6213	
$CGR6329_{232}$ [0&1-2]	-5.2367*	Fang et al. 2014	$BNL1604_{98}$ [0&1-2]	-2.3188*	Said et al. 2013
			$CGR6329_{232}$ [1&2-0]	-3.264*	Fang et al. 2014
adj R^2= .0492			adj R^2 = .0299		

* Significant at the .05 probability level, ** Significant at the .01 probability level.

† Fiber bundle strength

‡ Simple sequence repeat

§ Markers were classified based on the desired state of the allele. 2 signifies homozygous for the desired state. 1 signifies heterozygous. 0 signifies homozygous in an undesirable state. Number in brackets following the marker identify the allele state which the estimates are based on. Only markers with a significant effect are included in the table.

For Population 2, marker $BNL1604_{98}$ was significant in at least one year for both traits. $BNL1604_{98}$ maintained significance during both years for UHML and in 2016 for FBS (Table 12; Table 13). In each case, the marker showed the correct trend in that the homozygous desired state had a higher (p=.05) mean. The stepwise models for both fiber traits included $BNL1604_{98}$ and included $CGR6329_{232}$ in both years for FBS (Table 14; Table 15). Like Population 1, the models all had R^2 values less than 0.1 making them poor predictors for fiber quality. In both populations, the stepwise models included

markers that were not significant on an individual level due to differences in error partitioning between the single and multivariate models.

Hugie (2015) indicated that marker $BNL1604_{98}$ had potential for portability between populations. The significance within Population 2 may show that this marker is potentially transferable. However, data in Table 16 indicate that numerical differences between plants homozygous for the undesirable state and for the desirable state were biologically small. Even if the marker can be transferred between populations its effects may remain limited.

Table 16. Mean UHML and FBS in 2016 and 2017 to show influence of presence or absence of $BNL1604_{98}$ for Population 2, TAM 11K-13 ELSU/TAM06WE-621 ESU, in 2016 and 2017.

Year	mean UHML† (mm)		mean FBS‡ (kN m kg-1)	
	absent§	*present*	*absent*	*present*
2016	35.08 b¶	35.70 a	345.8 b	352.8 a
2017	34.23 b	34.70 a	361.9 a	364.8 a

† Upper half mean length
‡ Fiber bundle strength
§ The absent state is homozygous for the undesirable allele while the present state indicates homozygosity for the desirable allele state.
¶ Means followed by the same letter within each row and header are not significantly different according to Tukey's HSD at $\alpha = 0.05$.

CHAPTER V

CONCLUSION

One of the main goals of this study was to compare the efficacy of phenotypic selection with MAS using markers identified in a previous study undertaken by Kari Hugie (2015). In both populations used in the present study, trait selection based solely on phenotypic selection proved capable of selecting plants which would produce progeny with desirable UHML and FBS. Divergent populations selected from length data showed more differentiation than those selected for FBS which agrees with estimates of narrow sense heritability from this study and more advanced calculations from genetic mapping studies (Paterson et al., 2003). The FBS of the parents used to develop the segregating populations for this study also were apparently not sufficiently diverse to create the appropriate variability.

The markers identified by Hugie (2015) used for the genetic based portion of this study failed to predict divergent populations or populations that differed significantly from their means for either trait in either population. The phenotypic based selections therefore predicted progeny phenotype more accurately than the genotypic selections. As stated in the literature review, one of the issues facing widespread use of MAS for improvement of fiber quality in cotton is the lack of robust and portable genetic markers. The Hugie study identified QTL markers that predicted strain phenotype on a level equal or near equal to phenotypic selection within three diverse populations, but this study does not support her conclusions when those same QTL markers were applied to F_2 or F_3

individual plants used herein. It is likely that by starting with a large number of markers and narrowing down to a subset applicable to all three populations as with the Hugie (2015) results effectively mirrored a genome-wide association study which is already known to be effective but only in the initial population in which the study is conducted, and which therefore lacks many of the benefits of MAS.

On a single marker basis, it is more difficult to draw conclusions because several of the markers of interest were already homogenous in one or both of the populations. It could be possible that the homogenous markers are robust and could be used concomitantly in making selections or in choosing parents and that the qualities that make them so lead to them being quickly fixed in any given breeding program. By using at least one interspecific population, such alleles may have shown their merit during analysis, but in populations from material that has been cultured for an extensive time those alleles and their respective markers no longer segregate and are therefore not useful in their new background.

For the markers which were not homogenous, few showed significance and in all but one case, even when they did show significance the data either showed an opposite correlation to the published information or did not show additive inheritance of the trait. Hugie (2015) also observed and disqualified a number of markers showing opposite effects within its populations and it could be that genotype by environment (G x E) interactions are the cause. G x E interactions are one of the main issues in assigning allelic effects and these interactions may even fluctuate between different years at the same location (Davidonis et al., 2004). Studying and accounting for these interactions

would require analysis and separation of genetic effects which would necessitate a more complex experimental design than the simplistic design of this study.

There was one marker that was significant in Population 2 (TAM 11K-13 ELSU / TAM 06WE-621 ESU; true upland / upland population) for UHML and FBS in 2016 and for length in 2017. $BNL1604_{98,}$ which had previously been found to positively correlate with both traits had the same trend in all cases here. Although mean separation of positive and negative homozygotes was relatively small, this does suggest that there are markers with potential robustness in multiple populations (Table 16).

The results of this study suggest that the QTL markers included cannot be applied to individual plants of segregating populations if those QTL markers were derived from different parental material. Using MAS in this manner should allow prediction of UHML or FBS in cotton. However, looking at the length distributions of visually selected plants from the field in comparison to the randomly selected plants for the genotypically selected portion (Figure 5) and the lack of difference between the two, it is clear that there must be a level of skill involved in field based selections, and that level of care when observing any given plant is negatively correlated with how many plants a single person can observe. The lack of difference means that, depending on the trait of interest, time spent making selections by persons with low skill in identifying superior genotypes is better spent taking plant samples for phenotyping by instrumentation, and this leaves a gap that superior phenotyping or genotyping methods could fill. While field based selections may allow even inexperienced breeders to exclude poor quality plants not related to the trait of interest, genotypic based methods

still have the potential to offer improved screening of quantitative traits under the correct circumstances.

REFERENCES

Ali, M. A., I. A Khan, S. I. Awan, S. Ali, and S. Niaz. 2008. Genetics of fibre quality traits in cotton (*Gossypium hirsutum L.*). Aust. J. Crop Sci. 2: 10-17.

Adanur, S. 2001. Handbook of Weaving. CRC Press. p. 14-16.

Basu, A. 1992. Structure and properties of air-jet spun yarn. The Indian Text. J. March p. 38-41.

Basford K., and M. Cooper. 1998. Genotype x environment interactions and some considerations of their implications for wheat breeding in Australia. Aust. J. Agric. Res. 49: 153 – 174.

Bowman, D. T., et al. 2006. Pedigrees of Upland Cotton Cultivars Released Between 1970 and 1990. Mississippi Agricultural & Forestry Experiment Station. Handb. 1155. Mississippi State University.

Braden, C., and C. W. Smith. 2004. Fiber length development in nearlong staple upland cotton. Crop Sci. 44(5): 1553-1559.

Cai, C., W. Ye, T. Zhang, and W. Guo. 2014. Association analysis of fiber quality traits and exploration of elite alleles in Upland cotton cultivars/accessions (*Gossypium hirsutum L.*). J. Int. Plant Bi. 56: 51—62.

Christophe, M., T. Margareta, J. F. Haquet, M.Benoît. 2002. First Evidence of Cotton at Neolithic Mehrgarh, Pakistan: Analysis of Mineralized Fibres from a Copper Bead. J. Archaeological Sci. 29(12): 1393-1401.

Collard, B. C., and D. J. Mackill. 2008. Marker-assisted selection: an approach for precision plant breeding in the twenty-first century. Philosophical Transactions of the Royal Society of London. Series B, Biological Sciences. 363(1491): 557–572.

Cotton Incorporated. 2003. Introduction to open-end spinning. Technical Bulletin TRI 1004.

Cotton Incorporated. 2019. Classification of Upland Cotton. Cotton Incorporated Classification of Cotton. https://www.cottoninc.com/cotton-production/quality/classification-of-cotton/classification-of-upland-cotton. (accessed May 25, 2019).

Davidonis, G. H., A. S.Johnson, J. A. Landivar, and C. J. Fernandez. 2004. Cotton fiber quality is related to boll location and planting date. Agron. J. 96:42-47.

El-Hashash, E. 2017. Comparison of various methods for estimating of heritability and number of genes affecting the quantitative traits in cotton. Egypt. J. Plant Breed. 21(5): 612-628.

Fang, D. D., J. N. Jenkins, D. D. Deng, J. C. McCarty, P. Li, and J. Wu. 2014. Quantitative trait loci analysis of fiber quality traits using a random-mated recombinant inbred population in Upland cotton (Gossypium hirsutum L.). BMC Genomics 15: 397-411.

Fang, D.D., J. Xiao, P.C. Canci, and R.G. Cantrell. 2010. A new SNP haplotype associated with blue disease resistance gene in cotton (Gossypium hirsutum L.). Theor. Appl. Genet. 120: 943-953.

Günaydin, G.K. and A. S. Soydan. 2016. Vortex Spinning system and Vortex Yarn Structure. IntechOpen. DOI: 10.5772/67076

Gupta P. K., S. Rustgi and P. L. Kulwal. 2005 Linkage disequilibrium and association studies in higher plants: present status and future prospects. Plant Mol. Biol. 57: 461–485.

Guthrie, D., M. Watson, K. Hake. 1993. The 1993 Cotton Crop – Quality Trends. Cotton Phys. Today. Newsletter of the Cotton Phys. Edu. Program. Vol. 4(10).

Hearle, J. W. S. 2006. Physical structure and properties of cotton. *In* Gordon, S., Y.L. Hsieh, (eds). Cotton: Science and Technology. Woodhead Publishing Limited. Cambridge, England. p. 35-63.

Hertel, K. L. 1940. A Method of Fibre-Length Analysis Using the Fibrograph. Textile Research, 10(12): 510–520.

Hospital, F., L. Moreau, F. Lacoudre, A. Charcosset, and A. Gallais. 1997. More on the efficiency of marker-assisted selection. Theor. Appl. Genet. 95: 1181.

Hugie, L. 2015. Evaluation of Conventional and Marker-Assisted Breeding Methods for the Improvement of Fiber Quality in *Gossypium* spp. Ph.D. diss. Texas A&M Univ., College Station.

Hunter, L. 2006. Cotton Spinning Technology. *In* Gordon, S., Y.L. Hsieh. (eds). Cotton: Science and Technology. Woodhead Publishing Limited. Cambridge, England. p. 35-63.

Khan, I. A., A. Shakeel, and F. M. Azhar. 2001. Genetic analysis of fiber quality traits in upland cotton. Sci Int (Lahore). 2: 165-169.

Lande, R., and R. Thompson. 1989. Efficiency of Marker-Assisted Selection in the Improvement of Quantitative Traits. The Genetics Society of America. 124(3): 743-756.

Law, C.N. 1967. The location of genetic factors controlling a number of quantitative characters in wheat. Genetics 56(3): 445-461.

Lawrence, C.A. 2010. Overview of developments in yarn spinning technology. *In* Lawrence, C.A. (ed) Advances in yarn spinning technology. The Textile Institute. pp 3-40.

Li et al. 2002. Microsatellites: genomic distribution, putative functions and mutational mechanisms: a review. Mol. Ecol. 11(12): 2453-2465

National Cotton Council of America. 2018. National & State Cotton Area, Yield and Production. NCCA. Memphis, TN. url: http://www.cotton.org/econ/cropinfo/cropdata/state_data.cfm. (accessed May 25, 2019).

Marees, A.T. 2018. A tutorial on conducting genome-wide association studies: Quality control and statistical analysis. Int. J. Methods Psychiatr. Res. e1608. doi: 10.1002/mpr.1608.

May, O.L., and G.M. Jividen. 1999. Genetic modification of cotton fiber properties as measured by single- and high-volume instruments. Crop Sci. 39:328-333.

Moore, J.F. 1996. Cotton classification and quality. p. 51-57. In R.J. Kohel and C.F. Lewis. (eds.). The cotton industry in the United States. ASA, CSSA, and SSSA, Madison, WI.

Paterson, A. H., E. S. Lander, J. D. Hewitt, S. Peterson, S. E. Lincoln, and S. D. Tanksley. 1988. Resolution of quantitative traits into Mendelian factors by using a complete linkage map of restriction fragment length polymorphisms. Nature 335: 721-726.

Paterson, A. H., Y. Saranga, M. Menz, C-X Jiang, and R. Wright. 2003. QTL analysis of genotype × environment interactions affecting cotton fiber quality. TAG. Theoretical and applied genetics. Theoretische und angewandte Genetik. 106: 384-96. doi: 10.1007/s00122-002-1025-y.

Paterson, A. H. et al. 2012. Repeated polyploidization of Gossypium genomes and the evolution of spinnable cotton fibres. Nature. 492:423–427.

Pettigrew, W. T. 2001. Environmental Effects on Cotton Fiber Carbohydrate Concentration and Quality. Crop Science. 41. doi: 10.2135/cropsci2001.4141108x.

Rengasamy, R. S. 2010. Fundamental principles of ring spinning in yarns. *In* Lawrence, C.A. (ed) Advances in yarn spinning technology. The Textile Institute. p 3-40.

Said, J. I., Z. Lin, X. Zhang, M. Song, and J. Zhang. 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 14:776.

Shen, X., W. Guo, Z. Lu, Z. Zhu, Y. Yuan, and T. Zhang. 2007. Genetic mapping of quantitative trait loci for fiber quality and yield trait by RIL approach in upland cotton. Euphytica. 155: 371-380.

Shen, X., T. Zhang, W. Guo, X. Zhu, and X. Zhang. 2006. Mapping fiber and yield QTLs with main, epistatic, and QTL x environment interaction effects in recombinant inbred lines of upland cotton. Crop Sci. 46: 61-66.

Simpson, J., and M. F. Murray. 1978. Effect of Cotton Fiber Fineness and Strength on Mechanical Processing and Open-End Spinning and Yarn Properties. Textile Research Journal. 48: 247-251.

Smith, C. W. 2001. Registration of nine germplasm lines of upland cotton with improved fiber strength. Crop Sci. 41: 1370-1371.

Smith, C. W., et al. 2008. Development of extra-long staple upland cotton. Crop Sci. 48(5): 1823-1831.

Smith, C. W., Roy G. Cantrell, Hal S. Moser, and Stephen R. Oakley. 1999. History of Cultivar Development in the United States. *In* Smith, C. W. and J. T. Cothren (eds) Cotton: Origin, History, Technology and Production. J. Wiley & Sons, Inc. Oxford, UK. p 150-154.

Smith, C. W., E. Hequet, , S. Hague, and D. Jones. 2014. Registration of TAM 06WE-621 Upland Cotton with Improved Fiber Strength and Yarn Performance. J. Plant Registrations. 8: 308-312.

Tan, Z. et al. 2015. Genetic map and QTL controlling fiber quality traits in upland cotton (*Gossypium hirsutum L.).* Euphytica. 203: 615-628.

Thaxton, P., W. Smith, and R. Cantrell. 2005. Registration of TAM 96WD-18 upland cotton germplasm line with improved fiber length and strength. Crop Sci. 45: 1172.

USDA. 1980. The Classification of Cotton. Agriculture Handb. 566: 28-34.

USDA. 2006. National Cotton Variety Test. Crop Genetics & Production Research Unit. https://www.ars.usda.gov/ARSUserFiles/60661000/2006NCVTIndividualRegions/central.pdf.

Wendel, J., C. Brubaker, A. Percival. 1992. Genetic diversity in *Gossypium hirsutum* and the origin of upland cotton. Am J Bot. 79:1291–1310.

Xu, S. 2003. Theoretical basis of the Beavis effect. *Genetics* vol. 165(4): 2259-68.

Zeng, L., W. R. Meredith Jr., O.A. Gutierrez, and D.L. Boykin. 2009. Identification of associations between SSR markers and fiber traits in an exotic germplasm derived from multiple crosses among Gossypium tetraploid species. Theoretical and Applied Genetics. 119: 93-103.

Zhang et al. 2010. Stability of QTL across environments and QTL-by-environment interactions for plant and ear height in maize. Agricultural Sciences in China. 9: 1400-1412.