ABSTRACT

ANDRES, RYAN JAMES. Genetic Mapping and Molecular Cloning of the Major Leaf Shape Gene (*L-D_1)* in Upland Cotton (*Gossypium hirsutum* L). (Under the direction of Vasu Kuraparthy).

Leaf shape in Upland cotton (*Gossypium hirsutum* L.) is an important trait that influences yield, earliness, flowering rate, disease resistance, and the efficacy of foliar chemical application. A complete understanding of the genetic mechanism controlling leaf shape is essential for its proper manipulation to develop a cotton ideotype that maximizes yield while minimizing inputs. Predominant leaf shapes normal, sub-okra, okra, and super-okra, with varying levels of lobe severity, are controlled by a multiple allelic series at the D-genome locus *L-D_1* (previously the *L* locus*)*. Using a bi-parental F_2 mapping population, the leaf shape gene was localized to a 5.4 cM interval in the distal region of the short arm of chromosome 15. Orthologous mapping of the closely linked markers to the sequenced diploid D genome donor (*Gossypium raimondii*) tentatively resolved the leaf shape locus to a 337kb region containing 34 putative genes. Using shuttle mapping and novel targeted SNP and STS markers in association analysis, the leaf shape locus was further resolved to a 51kb interval containing only four genes, *Gorai.002G243900* through *Gorai.002G244200*. Expression analysis revealed that neither *Gorai.002G243900* nor *Gorai.002G244100* was expressed in young leaf tissue. Two remaining genes, *Gorai.002G244200* (*GhLMI1-D1a*) and *Gorai.002G244000* (*GhLMI1-D1b*), that show similarity to model plant *LMI1*-like genes, thus were identified as possible candidates for the leaf shape trait in cotton. Sanger sequencing of these two candidate genes identified two major polymorphisms in *GhLMI1-D1b*: 1) a 133bp tandem duplication in the promoter region of okra leaf cotton hypothesized to affect the expression pattern of the gene and 2) an 8bp deletion near the beginning of the

third exon of normal leaf cotton that leads to a frameshift mutation and premature stop codon in the putative coding sequence. An allelic intermediate leaf shape (sub-okra) lacks both the promoter duplication and the coding sequence deletion. A fourth extreme allele (super-okra) is identical to okra in regards to these polymorphisms. Association mapping with a large panel of diverse cotton accessions revealed that both polymorphisms remained linked to the leaf shape phenotype despite historical recombination. While *GhLMI1-D1a* was not differentially expressed among leaf shapes, *GhLMI1-D1b* was up-regulated 140-800 fold in the young leaves of okra cotton compared to normal and sub-okra. Finally, Virus-Induced Gene Silencing (VIGS) of *GhLMI1-D1b* in an okra leaf variety of transformation-recalcitrant cotton reduced *GhLMI1-D1b* transcript levels by four fold and led to a pronounced reduction in the degree of leaf lobing. Sequence analysis in diploid progenitor species further established that a diploid with simple leaf shape (*G. raimondii*) possessed a mutated *LMI1-D1b* gene leading to a truncated and presumably reduced function protein. D-genome diploid species with a dissected leaf shape (*G. thurberi* and *G. trilobum*) carry full-length, functional copies of *LMI1-D1b*. Thus, fine mapping using a large F_2 population, an association mapping panel, sequence and gene expression analysis, and complementation using VIGS unequivocally established that *GhLMI1-D1b* is responsible for the major leaf shapes of cotton. The following genetic control of leaf shape in cotton is proposed: The normal leaf shape allele at *GhLMI1-D1b* likely arose from within the *G. raimondii LMI1-D1b* (*Gorai.002G244000*) allele, consistent with the theory that *G. raimondii* is the closest remaining species to the D genome donor. The origin of the sub-okra *GhLMI1-D1b* allele is unknown but may have been transferred into the tetraploid at some point from a lobed D genome diploid. Okra leaf shape arose via the 133bp tandem promoter duplication in sub-okra *GhLMI1-D1b* leading to ectopic expression and enhanced lobing. Super-okra then arose within okra leaf cotton although what separates okra from super-okra leaf shape is not known at this time.

Genetic Mapping and Molecular Cloning of the Major Leaf Shape Gene (*L-*D_1) in Upland Cotton (*Gossypium hirsutum* L.)

by
Ryan James Andres

A dissertation submitted to the Graduate Faculty of
North Carolina State University
in partial fulfillment of the
requirements for the Degree of
Doctor of Philosophy

Crop Science

Raleigh, North Carolina
2015

APPROVED BY:

Vasu Kuraparthy
Committee Chair

Candace Haigler
Co-Chair

Ramsey Lewis

Marcela Rojas-Pierce

DEDICATION

To my wife Ellie for her love and support over the past thirteen years of our lives together. Also, to my parents Tony and Karen for their encouragement and support of my academic and life goals.

BIOGRAPHY

Ryan Andres was born on October 29, 1987 in Poughkeepsie, NY to Anthony and Karen Andres. He relocated along with his family to Cary, NC six years later. After graduating from high school at St. David's School in Raleigh, NC in 2005, Ryan obtained his B.S. in Biology from Elon University in 2009 and his M.S. in Crop Science from N.C State University in 2012. He and his wife Ellie were married on April 14, 2012. Their son, Corey Rode Andres, was born August 26, 2015.

ACKNOWLEDGMENTS

First off, I would like to thank Dr. Vasu Kuraparthy for giving me an opportunity to come to N.C. State and continue my education in the field of crop science. I would also like to thank the Cotton Incorporated Fellowship Program, especially Dr. Don Jones, for providing funding throughout my entire graduate school career. I would like to give special thanks to Dr. Rich Tuttle for his help with the expression and VIGS aspects of this project and Sang-Won Han for performing the co-localization and Arabidopsis transformations. This work would not have been possible without the help of post-doctoral researcher Hui Fang, my fellow graduate students Priyanka Tyagi, Linglong Zhu, and Baljinder Kaur, and the many long-term undergraduate workers, especially Evan Jones, Wess McKnight, Clayton Butcher, and Nathan Kiger. The staffs at the USDA Small Grain Regional Genotyping laboratory, NCSU Genomic Sciences Laboratory, and the NCSU Phytotron contributed immeasurably to this work. I would not be in this position today if not for the loving support from my family including my parents Tony and Karen and my sister Lisa. Finally, I would like to thank my wife Ellie, without whom, none of this would be possible.

TABLE OF CONTENTS

LIST OF TABLES

Chapter 2

Chapter 3

LIST OF FIGURES

Chapter 2

Chapter 3

CHAPTER 1: General Review of Literature

The review below first provides general information about cotton production and breeding (Section 1.1 to 1.7) and then deals with specific information related to leaf shape in cotton (Section 1.8 to 1.10) and leaf shape in general (Section 1.11).

1.1 Cotton Production in the United States

Cotton (*Gossypium* spp.) is the world's most important source of natural fiber as well as one of its leading oilseed crops. In the United States cotton acreage averages 10-12 million acres per year with production ranging from 15-17 million bales annually (7.5-8.5 billion pounds) (NCCA 2013). The production region runs throughout the southern United States ranging from southern Virginia to southern California and is divided into four production zones: the Southeast, the Mid-South along the Mississippi Delta, the Southern High Plains of Texas and Oklahoma, and the Far West. The majority of cotton production in the United States occurs in the Southern High Plains. Cotton planting can begin as early as February in south Texas and as late as June in North Carolina (NCCA 2013). Harvesting occurs between July for the earliest planted crop and November for that in the northern production areas (NCCA 2013). On average, the U.S. exports 80% of this production resulting in a ~$7 billion reduction in the U.S. trade deficit with the remaining ~3.5 million bales consumed by domestic textile mills (NCCA 2013). Only about one third of the harvest by weight consists of lint, of which 75% is used to manufacture apparel, 18% for home furnishings, and 7% for industrial products (NCCA 2013). The remaining two thirds of the crop is cottonseed, which is used to produce 6 billion pounds of animal feed and fertilizer and 90 million gallons of cottonseed oil for use in various food products (NCCA 2013). All told the U.S. cotton industry is estimated to produce $27 billion in revenue and $100 billion in total economic activity (NCCA 2013).

1.2 General Challenges Facing Cotton Production in the United States

The major insect pests of cotton and their effects are described below (Cotton Incorporated 2015). With the successful eradication of the boll weevil (*Anthonomus grandis*), and the effective control of various caterpillar species via the deployment of *Bt* cotton, the major late-season insect pests of cotton are the plant bugs and stink bugs. Both of these insects had previously been adequately controlled by the organophosphate insecticides used to control the weevil and the caterpillars. In the Southeast United States the green stink bug (*Acrosternum hilare*), the southern green stink bug (*Nezara viridula* L.) and the brown stink bug *(Euschitus servus*) are of major concern. These insects feed on seeds during boll development and during feeding inject boll rot pathogens such as *Nematospora coryli*). In the Mid-South the major threat is the tarnished plant bug (*Lygus lineolaris*) with the western tarnished plant bug (*L. hesperus*) in the western cotton belt. Genetic resistance to these pests has not yet been identified in cotton leaving chemical control as the only effective option. In non-*Bt* cotton, caterpillars are still of particular concern, especially the cotton bollworm (*Helicoverpa zea*) and the tobacco budworm (*Heliothis virescens*) in the southeast and the pink bollworm (*Pectinophora gossypiella*) in the west. Other minor caterpillar species include the beet armyworm (*Spodoptera exigua*) and the fall armyworm (*S. frugiperda*).

Thrips, primarily of the genus *Frankliniella* are a sucking insect that attack seedling stage cotton, especially during cool weather. The 2010 EPA ban on Temik 15G, the most widely applied aldicarb insecticide for thrips control, has led to a renewed emphasis for transferring resistance from *G. barbadense* to *G. hirsutum.* Whiteflies are another sucking insect that can stunt growth by feeding on cellular contents, particularly in the west. The major whitefly species is the sweet potato/silver leaf whitefly (*Bemisia tabaci* (Gennadius)*/argentifollii*) along with the banded-wing whitefly (*Trialeurodes abutilonea*) and the greenhouse whitefly (*T. vaporariorum*). Whiteflies also serve as vectors for numerous viral diseases and deposit a sticky honeydew substance that can attract mold and reduce fiber quality. The cotton aphid (*Aphis gossypii*) and spider mites (particularly the two-spotted spider mite *(Tetranychus urticae*) are additional minor insect pests that generally do not cause enough damage to warrant control.

The major diseases facing cotton production in the southeast United States are the fungal seedling disease complex (*Fusarium* spp., *Pythium* spp., *Rhizoctonia solani*, *Ascochyta gossypii*), root knot (*Meloidogyne incognita acrita*) and reniform (*Rotylenchus reniformis*) nematodes, fungal leaf spot (*Cercospora gossypina*, *Alternaria macrospora*), and cotton stem canker (*Ascochyta* spp.) (Koenning 2004). These diseases are mostly combatted by planting resistant varieties, soil fumigation, plowing, chemical control, crop rotations, and maintaining appropriate levels of nutrients and water. The cotton boll rot complex can be caused by over 40 different fungal and bacterial species and is particularly prevalent in coastal cotton regions with high humidity and heavy rainfall. The best way to combat this disease is any measure that creates a more open canopy and discourages excessive vegetative growth such as the planting of a cultivar with mutant leaf shapes. Other important diseases affecting cotton in other areas include cotton root rot (*Phymatotrichum omnivorum*), Verticillium wilt (*Verticillium albo-artum*), rust (*Puccinia cacabata*), Fusarium wilt (*Fusarium oxysporum*), bacterial blight (*Xanthomonas campestris*), and Ascochyta blight (*Ascochyta gossypii*).

Drought, salinity, and extreme temperature are the three major abiotic stresses facing cotton (Allen and Aleman 2011). A growing world population will require greater production from agriculture while pushing farming onto smaller and more marginal pieces of land and reducing the freshwater supply available for irrigation. Global warming will also magnify these abiotic stresses by creating unpredictable weather patterns in terms of both precipitation and temperature. Cotton is susceptible to early frosts that can reduce yield and fiber quality as well as extreme heat during the summer that can stunt growth in the absence of suitable moisture. In addition to drought, excessive rainfall can have adverse effects on a cotton crop, contributing to insect infestations early in the growing season and boll rot and excessive vegetative growth during the boll-filling period. Other threats to U.S. cotton production include competition from uniform synthetic fibers, superior quality of hand-picked foreign cotton, and tariff wars over price supports.

Despite a wealth of genetic variability in the wild cotton germplasm pool, only a relative few lines have been involved in the majority of recently released cultivars (Wallace et al. 2009). It is estimated that only 1% of the available genetic diversity is present in the current Upland cultivars (Wallace et al. 2009). This bottleneck is even more severe in transgenic cotton, which are nearly all derived from the Coker 312 line, resulting in a common linkage block likely present in all GMO cotton (Wallace et al. 2009). This narrow genetic base has made it increasingly difficult to develop improved cotton cultivars in recent years especially in regards to improved yield, fiber quality, abiotic and biotic stress tolerance. More advanced and creative techniques within molecular plant breeding and biotechnology are needed to overcome this genetic bottleneck. These include genome-wide association studies (GWAS), genomic selection (GS), and nested association mapping (NAM) that are powerful technologies better suited to the improvement of complex quantitative traits governed by many loci of minor effect with high precision and efficiency. Another example is the production of gossypol free cottonseed (Sunilkumar et al. 2006). Gossypol is a toxic aldehyde from the *Gossypium* genus that serves as a natural insecticide by inhibiting several dehydrogenase enzymes. However, its presence in cottonseed prevents the use of the seed from being used a food source beyond animals with a ruminant digestive system. Conventional breeding techniques were only able to remove gossypol production from the entire plant, which resulted in varieties overly susceptible to insect damage. Through genetic engineering and RNAi, cotton varieties can now be developed that lack gossypol production only in their seeds, greatly expanding the potential uses of cottonseed (Sunilkumar et al. 2006).

1.3 Evolution and Genetic Diversity of Cotton

As reviewed by Wendel and coworkers (Wendel et al. 2009), roughly 12.5 million years ago (MYA), *Gossypium* diverged from its closest related genera *Gossypioides* and *Kokia*. For the next 5-7 million years, diploid ($2n = 2x = 26$) cotton spread throughout the arid and semi-arid regions of the tropics. This dispersal resulted in eight groups of diploids (A-G and K) broken into three major lineages: the New World diploids (D), the Africa-Asia diploids (A, B, E, and F), and the Australian diploids (C, G, and K). Currently, 43 species of diploid cotton are known to exist across the eight genomes. Spinnable fibers likely evolved only in the A genome diploids and both remaining A genome species, *G. arboreum* and *G. herbaceum*, were domesticated independently.

At some point approximately 1-2 MYA, an A genome diploid survived a trans-oceanic voyage to the New World where it hybridized with a D genome diploid to form an allopolyploid species ($2n = 4x = 52$, AADD). The cytoplasm of the tetraploid lineage is from the A genome indicating that the A genome diploid was the maternal parent in this cross. The D genome donor diploid was most closely related to *G. raimondii* while the two remaining A genome diploids appear equally closely related to the A genome donor. The A genome diploid parent subsequently died off in the New World following this hybridization. The polyploid then spread throughout the tropics of the New World spawning at least six different species, two of which (*G. hirsutum* and *G. barbadense*) were independently domesticated. *G. hirsutum* is indigenous to Central America and the Caribbean while *G. barbadense* is found naturally in the northern third of South America. Of the four remaining allotetraploids, two are island endemics restricted to the Galapagos Islands (*G. darwinii*) and the Hawaiian Islands (*G. tomentosum*) (Flagel and Wendel 2010). A third, *G. mustelinum*, is found only a small area of northeast Brazil. The sixth allotetraploid species, *G. ekmanianum*¸ was recently identified by DNA analysis and is endemic to the Dominican Republic (Grover et al. 2015).

Upland cotton (*G. hirsutum*) accounts for ~95% of the world cotton production owing to its high yield and adequate fiber quality. Most of the remainder is dedicated to the lower yielding *G. barbadense* that produces premium quality fiber for use in high-end luxury textiles. *G. barbadense* is alternatively known under other names depending on its point of production including Egyptian, Pima/Supima (southwest United States), and Sea-Island (east coast United States). The cultivation of the two domesticated A genome diploids, *G. arboreum* and *G. herbaceum*, is almost non-existent today, having been supplanted by the superior tetraploids. Most cotton species exist naturally as perennial, photoperiodic sprawling bushes or small trees with minimal production of poor quality fiber (Wendel et al. 2009). Today, cotton is cultivated as an annual using photoperiod-insensitive, compact bushes with enhanced yield and fiber quality.

1.4 Polyploidy, Genome Evolution, and Gene Expression

Following a polyploidization event, the new species encounters the difficulty of merging two divergent genomes while surviving any negative interactions the two genomes may have on one another (Flagel and Wendel 2010). These include perturbations to the gene dosage balance, competition between transcription factors, and different levels of various non-coding RNAs (Flagel and Wendel 2010). Methods to compensate for the presence of duplicate genes in the polyploid include loss of one of the duplicates through either genomic fragment loss, mutation, or epigenetic silencing, sub-functionalization in which the duplicate loci evolve to perform only part of their previous function, or neo-functionalization in which a member of a gene pair evolves a new function that it did not previously perform in the parent (Flagel and Wendel 2010). These changes can happen immediately after polyploid formation but also continue to occur throughout the evolution of the polyploid species (Adams and Wendel 2005). These can result in new expression patterns, regulatory networks, and phenotypes, some of which will undoubtedly be favored by natural selection and result in the continued evolution of the polyploid (Adams and Wendel 2005).

As reviewed by Rapp and coworkers (Rapp et al. 2009), despite only a 3% divergence in coding sequence, 53% of genes were differentially expressed between the A genome diploid cotton *G. arboreum* and the D genome diploid *G. thurberi*. When combined in a synthetic allopolyploid, there was a strong genome expression bias exhibited toward the D genome. In other words, the expression profile of the newly synthesized allotetraploid more closely mirrored that of the D genome parent than either the A genome parent or a midparent value. This D genome bias was roughly equally split between genes that were up-regulated vs. down-regulated and the bias was more the result of small biases in many genes rather than drastic differences in a few genes. Interestingly, the D genome bias exhibited by the five allotetraploid species of cotton was smaller than that seen in the newly synthesized synthetic allotetraploid (Flagel and Wendel 2010). This indicates that over the 1-2 million year evolution of the allotetraploid, it evolved to use a more balanced mix of genes from its two genomes as compared to immediately following hybridization (Flagel and Wendel 2010). Consistent with this, sequencing of *G. hirsutum* (Zhang et al. 2015) identified that loss of gene function and other genome modifications were higher in the A genome but there was no genome-wide expression dominance between the sub-genomes. It appears selection in the A genome primarily favored fiber development and quality genes while D genome selection was mostly in the direction of stress response genes (Li et al 2015, Zhang et al. 2015)

1.5 Genetic Mapping

Traditional plant breeding studies attempt to identify molecular markers linked to a trait of interest in one or a few areas of the genome in order to facilitate marker-assisted selection (MAS). The mapping of qualitative traits, controlled by a single gene with little environmental effect is relatively easy. Oftentimes, bi-parental F_2 mapping populations are sufficient, phenotyping can be done rapidly, and markers can be quickly and accurately linked to the causal gene. Once linked markers have been identified and the appropriate genomic resources are available, candidate genes can be quickly identified and validated. Even without the identification of the causal gene, MAS on simply inherited traits is quite effective.

However, the use of MAS to improve complex traits governed by many loci scattered throughout the genome, each with a relatively small effect on the expression of the trait of interest, has been less successful. Many of the most important traits in plant breeding are inherited this way including yield, quality, and biotic and abiotic stress resistance (Laurie et al. 2004, Buckler et al. 2009, Kump et al. 2011, Poland et al. 2011, Tian et al. 2011, Cook et al. 2012, Huang et al. 2012) . These quantitative trait loci (QTL) mapping projects generally use bi-parental mapping populations, which do a poor job of capturing the allelic diversity present in a breeding population. The development of recombinant inbred line (RIL) populations is often required, which requires time and resources. Furthermore, any discovered effects must be validated in subsequent studies, QTL confidence intervals may be large leading to linkage drag, distant markers may disassociate from the trait of interest, and population sizes are insufficient to detect small-effect QTL. Association mapping can be used to increase allele richness and take advantage of historical rates of recombination but suffers from biased effect estimates and population structure and requires substantial genomic knowledge to design a suitable number of markers (Flint-Garcia et al. 2003, Zhu et al. 2008).

With marker development and genotyping costs continuing to fall rapidly, genotyping has become considerably less expensive than phenotyping for many of the major traits of interest. Therefore, genome-wide association studies (GWAS) and genomic selection (GS) have emerged as powerful technologies capable of large improvements in complex quantitative traits. Genomic selection uses a “training population” of lines that have been extensively genotyped and phenotyped in order to estimate the effects of all markers on a given trait. These marker effects are then incorporated into a model that can be used to predict the phenotypic value of any line that has been genotyped. These predicted values are called genomic estimated breeding values (GEBVs). GS then uses a “validation population” where the GEBVs are compared to the actual phenotypic values of lines after they have undergone phenotypic testing. Many studies have shown that GEBVs have been quite accurate at predicting phenotypic values (Jannink et al. 2010).

These models are produced using an array of advanced statistical methods including best linear unbiased predictor (BLUP), ridge regression, Bayesian regression, kernel regression, and machine learning methods (Jannink et al. 2010). The goal is to create the best, most unbiased, and accurate model possible, which is difficult when there are many variables whose relationships are poorly understood and relatively few observations. The best models may actually be those that pool the results from the different methods mentioned above (Jannink et al. 2010).

Markers must be evenly distributed throughout the genome (hence “genome-wide”) and at a very high density. Evenly distributed is based not on physical distance but on genetic distance in centiMorgans (cMs). A higher marker density means a more accurate model, but a point is reached with diminishing returns of adding additional markers (Jannink et al. 2010). SSRs and SNPs are the most common markers. SSRs give more information per marker owing to their multiallelic nature but biallelic SNPs are more common in the genome and easier to analyze and develop.

Larger population sizes also increase the accuracy of the model. However, the effective population size must be considered where adding more distantly related individuals is more valuable than adding closely related ones (Jannink et al. 2010). Increasing population size will lead to a more accurate model than increasing marker density if resources are constrained (Jannink et al. 2010). When constructing a training population with restricted space, it is better to include more lines with a single replication than to perform multiple replications (Jannink et a. 2010). The ultimate goal of GS is to make predictions and selections before advancing lines to expensive, replicated phenotypic testing. Using GS should result in a greater proportion of the selected lines performing at a desirable level. GS with markers can also be done in off-season nurseries that can maximize the gain per unit time. Traditional phenotypic selection can only evaluate one generation per year.

There are a few major drawbacks to GS/GWAS. First, the technique itself does not tell you anything about how the underlying genes are altering the trait of interest. However, associated markers can be used in subsequent studies to ascertain these mechanisms. Also, it is unclear how much, if any, GS will reduce genetic diversity over time. However, it appears that this may not be as drastic as previously thought (Jannink et al. 2010). Models developed for one training population also have a poor track record of being able to accurately predict performance in other populations (Albrecht et al. 2011, Massman et al. 2013, Wallace et al. 2014). Another major area of future research is how best to design training populations in order to maximize marker variance and reduce resources spent on phenotyping (Jannink et al. 2010).

Nested association mapping (NAM) is another technique that combines the advantages of association and bi-parental linkage mapping while minimizing their disadvantages. This potentially allows large gains to be made in the improvement of complex traits. NAM involves the mating of a diverse panel of lines to a common parent, usually the genetic standard of a particular crop species. From each cross a RIL population is then developed. Each RIL is then genotyped with the same panel of markers in order facilitate comparison among RIL populations. Selection of markers for which the common parent has a rare allele can facilitate this process and increase the number of markers available for genotyping.

Then all of the parents lines are either sequenced or high-density genotyped. Overlaying this parental data on the RIL genotyping data allows the breeder to essentially sequence or high-density genotype every RIL line. Phenotypic analysis of the RIL population then allows high resolution and statistical power mapping of markers associated with QTLs for traits of interest.

These advanced techniques could be best deployed in cotton to improve complex quantitative traits, particularly yield and quality. Gains in selection of both of these traits may be plateauing under traditional phenotypic and/or marker-assisted selection (Paterson et al. 2004). Furthermore, it may be able to finally break the well-documented negative correlation between fiber yield and quality, which may be due to the very tight repulsion phase linkage of genes underlying these traits (Clement et al. 2012).

1.6 Genomic Resources in Cotton

The first restriction fragment length polymorphism (RFLP) map, consisting of 705 markers, was published by Reinisch et al. (1994) using an interspecific *G. barbadense* x *G. hirsutum* mapping population. The first intraspecific *G. hirsutum* RFLP map was developed in 1998 by Shappley et al. The interspecific RFLP map was later expanded to 2,584 loci (Rong et al. 2004). Numerous issues with RFLP markers, especially the time and cost-intensive nature of the technique, prompted the development of better marker systems including amplified fragment length polymorphism (AFLP) and simple sequence repeat (SSR) markers.

Numerous studies developed genetic maps using varying combinations of marker types in both interspecific *G. hirsutum* x *G. barbadense* (Lacape et al. 2003, Nguyen et al. 2004, Park et al. 2005 Yu et al. 2007, Guo et al. 2007, Lacape et al. 2009, Yu et al. 2011, Yu et al. 2012) and intraspecific *G. hirsutum* (Ulloa et al. 2002, Lin et al. 2009, Zhang et al. 2012) populations. SSRs quickly became the favored marker in cotton genetics during this time, mainly due to their abundance, diversity, repeatability, and amenability to high-throughput genotyping. However, SSR markers in cotton were developed independently by at least 17 different groups leading to a high degree of marker redundancy that complicated comparisons between different studies. Blenda et al. (2012) investigated 17,343 SSR markers in cotton and found that 20% (3,410) were redundant and could be grouped into clusters. The resulting high-density consensus (HDC) genetic map in cotton of AFLP, RFLP, and SSR markers (Blenda et al. 2012) facilitated the comparison of genetic mapping studies in cotton.

The use of next-generation, single nucleotide polymorphisms (SNPs) to develop genetic maps in cotton has also increased in recent years (Yu et al. 2012, Byers et al. 2012, Zhu et al. 2014). The development of the Cotton SNP Chip, based on the Illumina BeadArray™ technology, should further facilitate the movement towards high-throughput SNP genotyping in cotton. This assay, developed by the International Cotton SNP Genotyping Panel, contains 70,000 publicly available SNPs as its “Base Content” with room for an additional 20,000 private SNPs to meet the specific needs of the user. Of the “Base Content” SNPs, 50,000 are intraspecific to *G. hirsutum* while 16,000 are interspecific tetraploid SNPs and the remaining 4,000 are reserved for the diploid cottons.

A diversity array technology (DArT) marker platform has also been developed for use in cotton (Reddy et al. 2011). DArT markers were genetically mapped with a much greater frequency than AFLP markers (Reddy et al. 2011). Furthermore, the majority of DArT marker sequences were homologous to sequenced genes in both cotton and Arabidopsis. This indicates that this marker type may be especially functionally relevant for marker-assisted selection and map-based cloning if they can be linked to traits of economic importance (Reddy et al. 2011).

Efforts to improve the fiber quality of *G. hirsutum* through introgression of elite alleles from *G. barbadense* is often hindered by hybrid breakdown (Stephens 1950) and preferential elimination of *G. barbadense* alleles in segregating populations derived from this cross (Stephens 1949, Gore et al. 2014). In an effort to break this genetic bottleneck, a RIL population was developed from TM-1 crossed to NM24016, a *G. hirsutum* line stably carrying introgressions from *G. barbadense* (Cantrell and Davis 2000, Gore et al. 2009). Genetic mapping with SSRs and genotype-by-sequencing (GBS) SNPs identified 28 QTLs associated with 10 agronomic and fiber quality traits (Gore et al. 2014). However, these QTLs explained only a fraction of the phenotypic variation (Gore et al. 2014). A small population size likely missed small-effect QTLs (Gore et al. 2014). Furthermore, due to large identical-by-descent (IBD) blocks between the parents, marker coverage of the genome was only 50% (Gore et al. 2014). Optimized enzyme cocktails, a GBS system expanded to InDels and dominant markers, and a NAM population structure were all proposed to better identify QTLs underlying important agronomic and fiber quality traits in cotton (Gore et al. 2014).

The genome sequence of allotetraploid cotton has been recently completed (Zhang et al. 2015, Li et al. 2015). However, while this sequence can be downloaded and BLAST searched against, it has not yet been published in a user-friendly format. Additionally, two independent draft sequences of the presumed D-genome donor diploid, *G. raimondii*, have recently been released (Paterson et al. 2012, Wang et al. 2012). The A genome diploid *G. arboreum* has also been recently sequenced (Li et al. 2014). The availability of these draft genome sequences should help facilitate the fine mapping and cloning of genes in cotton by providing a physical map to integrate with constructed genetic maps. The gene annotation of these draft genomes, as well as the available protein homology and gene ancestry tools of the Paterson et al. (2012) genome, may allow the combination of physical and genetic maps to identify highly plausible candidate genes underlying phenotypes of interest.

1.7 Genetic Mapping in Cotton

The majority of genetic mapping studies in cotton have focused on fiber quality traits (strength, length, uniformity, micronaire, color, etc.) with yield components being a distant second (Rong et al. 2007, Lacape et al. 2010, Said et al. 2013). However, other QTL mapping studies have investigated disease and pest resistance as well as seed and morphological characteristics (Said et al. 2013).

Comparison among QTL mapping studies is complicated by the use of different marker types/panels, parents, population types (F_2, BC, and RIL, interspecific vs. intraspecific), environments, amount of phenotypic variance explained, and LOD scores used to declare significance (Rong et al. 2007, Lacape et al. 2010, Said et al. 2013). Lack of suitable reference maps can also hinder comparisons among studies (Rong et al. 2007). Earlier QTL studies struggled due to a lack of sufficient markers, small population sizes, and failure to analyze multiple locations and years (Lacape et al. 2010, Said et al. 2013). This resulted in large QTL confidence intervals and inflation of QTLs with low heritability and high genotype x environment effects (Lacape et al. 2010, Said et al. 2013). Methods to correct these problems included larger population sizes, testing in more environments, and comparisons across multiple studies (Lacape et al. 2010). Rapid decreases in genotyping costs solved the issues with marker density and the publication of consensus genetic maps allowed for the undertaking of meta-analyses across QTL mapping studies in cotton.

The overall goal of these meta-analyses was to identify consistently detectable QTLs in order to promote marker-assisted selection (Said et al. 2013). In theory, these are likely to also be some of the largest effect QTLs and comparing multiple studies may assist in narrowing down confidence intervals. Meta-analyses revealed that QTLs were not distributed randomly throughout the cotton genome but were instead grouped in clusters and hotspots (Said et al. 2013). Congruence between QTL studies was usually low to moderate although consistent QTLs affecting fiber quality, bacterial resistance, nematode resistance, and leaf morphology have been identified (Said et al. 2013). Since most QTL mapping studies involved fiber quality, the greatest insights have come in regard to this trait. Positive effect QTLs on fiber quality were roughly evenly distributed across the A and D sub-genomes (Rong et al. 2007, Lacape et al. 2010). This was surprising given that the evolution of spinnable fibers occurred only in the A genome and indicates evolution and selection of D genome loci to improve fiber characteristics. Furthermore, only rarely were such loci likely to be homeologous (Rong et al. 2007, Lacape et al. 2010, Said et al. 2013). Also positive effect fiber quality QTLs were equally prevalent in *G. hirsutum* as *G. barbadense* despite the fact that *G. barbadense* is widely known to produce superior quality fiber (Lacape et al.

2010). This result may have been inflated by the preferential elimination of *G. barbadense* alleles in the segregating progeny of interspecific crosses (Lacape et al. 2010). This indicated that transgressive segregates for fiber quality were possible and they were readily obtained in an interspecific RIL population (Lacape et al. 2010).

1.8 Description of Leaf Shape in Cotton

Nearly all Upland cotton (*G. hirsutum*) cultivars possess the normal or broad leaf type. Along with normal (*l*), numerous mutant leaf types make up an allelic series at the leaf shape locus (L or L-D_1) including okra (L^O), sub-okra (L^U), sea-island (L^E), and super-okra (L^S) (Stephens 1945, Green 1953). Pictures of leaf shapes are available in Figure 1 of Chapters 2 and 3. The normal leaf shape is broad and palmate with five readily observable, yet insipid lobes. The okra leaf shape is characterized by reduced photosynthetic area per leaf (Karami and Weaver 1972, Pegelow et al. 1977), a more pronounced lobing pattern with deeper sinuses, and ectopic outgrowths leading to abnormal leaf margins. The appearance of ectopic outgrowths along the leaf margins is common but their position, shape, and number appear random. Sub-okra and sea-island leaf shapes are very similar to one another. Both show a degree of lobing that is intermediate between that of normal and okra, similar to the heterozygote between these two types. Sea-island is predominantly found in cultivars of extra-long staple cotton (*G. barbadense*) while sub-okra is typically found in Upland cotton. Ectopic outgrowths in both of these leaf shapes are rare and manifestation of these two mutant leaf shapes is generally not clear until later in development compared to okra.

However, it has not been definitively proven that sea-island is a separate allele from sub-okra (Meredith 1984). Instead, the relatively minor differences between the two leaf shapes may be the result of different alleles at loci of minor effect on leaf shape in the genetic background of the two species. Super-okra is an even more severe leaf shape mutant than okra with lobing stretching all the way to the petiole and approaching classification as a compound leaf. At maturity, the leaf has been further reduced to a single strip emanating from the petiole. Ectopic outgrowths are occasionally present but less numerous and noticeable than in okra. Across all the alleles, heterozygotes are intermediate between the two respective parental types and all three classes can usually be categorized with brief visual observation.

Allometric analysis has indicated that the factor underlying okra leaf shape in cotton acts very early during leaf development, during the initiation of the primordium (Hammond 1941a, Dolan and Poethig 1991). After the shape has been determined, it remains constant through the remaining growth and development of the leaf (Dolan and Poethig 1991). The okra leaf factor delays the appearance of lateral lobes, increases the rate of lobe growth relative to sinus growth, and increases the rate of lobe elongation compared to lobe widening (Hammond 1941a, Dolan and Poethig 1991). Okra may do this by substantially extending the early growth pattern of the primordia (Dolan and Poethig 1991). The use of genetic mosaics has also shown that the factor underlying okra leaf shape is cell autonomous and not diffusible across cells (Dolan and Poethig 1991). The alleles at the major leaf shape locus in the A genome diploids also act very early in leaf development, similar to the okra gene in the tetraploid (Hammond 1941b).

The reduced leaf area of the leaf shape mutants can delay the attainment of canopy closure when compared to normal, especially under adverse early-season growing conditions. This can lead to reduced yields in mutant leaf shape varieties and exacerbate their issues in regards to weed pressure and increased boll shed. It is likely that an ideal cotton cultivar would produce a lower canopy of normal leaves before switching over to an upper canopy of okra leaves roughly about the time of canopy closure. While such a phenotype does not appear to be present in the natural variation of cotton, it is plausible that such plant architecture would combine the advantages of both leaf shapes while minimizing their disadvantages. In order to develop a plant with this transforming leaf phenotype, a more complete understanding of the genetic mechanism controlling leaf shape in cotton is needed.

1.9 Inheritance of Leaf Shape in Cotton

Okra leaf has been known to exist in cotton since before the 19th century (Mell 1890). Shoemaker (1909) first showed that okra was controlled by a single gene. Sea-island was also shown to be simply inherited a short time later (McClendon 1912). Super-okra was first reported as a spontaneous mutant in a population of Acala Okra grown in Trinidad and subsequently observed to be simply inherited (Harland 1932). The allele symbols for normal (*l*), okra (L^{O}), and super-okra (L^{S}) were given by Hutchinson and Silow (1939). In 1945, Stephens demonstrated that super-okra and sea-island (L^E) were alleles at the same locus as normal and okra (Stephens 1945).

Green (1953) reported on a new leaf shape observed in the progeny of a synthetic tetraploid derived from a cross between the A genome diploid *G. arboreum* and the D genome diploid *G. thurberi*. This novel leaf shape was named sub-okra (L^U) and shown to be a fifth allele at the leaf shape locus (Green 1953). Citing only “independent evidence” Green claimed that the sub-okra allele originated in *G. thurberi*, thus initiating the belief that the locus resides in the D genome. It is likely that the leaf shapes of the parents played some role in this decision as *G. thurberi* possesses a highly lobed leaf similar to okra while most *G. arboreum* leaves are similar to normal or sub-okra. In 1955, Stephens showed that the leaf shape locus (*L*) was genetically linked to the crinkle leaf (*cr*) and green lint (*Lg*) loci. Stephens then cited Green (1953) as saying that since sub-okra came from *G. thurberi*, the whole linkage group must reside in the D genome. Unpublished data from Hutchinson saying that the green lint gene came from the D genome diploid *G. armourianum* was also cited to confirm the placement of the linkage group in the D genome (Stephens 1955). The chromosome carrying the *L* locus was named chromosome 15 (Chr15) in cytogenetic work utilizing monosomes (Endrizzi and Brown 1964). Later telosomic stocks were used to place the *L* locus on the short arm of Chr15, opposite *cr* and *Lg* on the long arm (Endrizzi and Kohel 1966).

Jiang et al. (2000) used restriction fragment length polymorphism (RFLP) markers to map leaf shape quantitative trait loci (QTL) in an interspecific *G. hirsutum* (okra) x *G.*

barbadense (normal) F_2 population. They identified a large, multiple effect QTL on Chr15, presumably the *L* locus, mapped between the RFLP markers pAR019 and pAR1001 (Jiang et al. 2000). It was also proposed that the QTL may be a cluster of tightly linked leaf shape modifiers, rather than a single locus (Jiang et al. 2000). Song et al. (2005) also mapped leaf shape QTL using an interspecific *G. hirsutum* (normal) x *G. barbadense* (okra) BC_1 population and simple sequence repeat (SSR) markers. A large, multiple effect QTL was not found in this population on Chr15, but a QTL of minor effect on leaf lobe width was mapped to a ~30cM region bounded by SSRs BNL2440 and JESPR152 (Song et al. 2005). The authors comment that the disparity of their findings to Jiang et al. (2000) may be the result of differences in population size, population type, marker type, or parents (Song et al. 2005). However, based on statements made and data presented on the parental lobe lengths, the *G. barbadense* parent may have been sub-okra or sea-island rather than okra. Therefore, the lack of the detection of the classical *L* locus in this study may be better explained by the greater similarity in shape of sub-okra to normal as well as the exclusion of sub-okra homozygotes in the BC_1 mapping population backcrossed to the normal leaf parent. Lacape et al. (2013) used amplified fragment length polymorphism (AFLP) and SSR markers to map leaf shape and other traits in an interspecific *G. hirsutum* (normal) x *G. barbadense* (sub-okra) recombinant inbred line (RIL) population. The major leaf shape QTL was placed in a 26.9 cM interval bounded by the SSR BNL2440 and the AFLP E43M52-M326.0 on Chr15 (Lacape et al. 2013). Three markers within this region appeared to be highly associated with the major leaf shape QTL, SSRs BNL1693 and MGHES32 and the AFLP 151u. In summary, although classical genetic studies showed that okra leaf shape was inherited as a single gene, almost all molecular mapping efforts reported that leaf shape in cotton was controlled by QTLs of varying effect sizes.

An alternate leaf shape locus is also known to exist in the A genome of cotton. In a series of crosses among and between the diploid A genome species *G. arboreum* and *G. herbaceum* Hutchinson (1934) demonstrated that there existed five leaf shapes in Asiatic cotton, all of which are allelomorphic: laciniate (L^L), arboreum (L), recessive broad (l), mutant broad (L^B) and mutant intermediate (L^I). Only laciniate, which is phenotypically similar to okra, was transferred from *G. arboreum* to *G. hirsutum* by a Dr. C. Rhyne ~1960 (Endrizzi and Stein 1975, Jones 1982). The laciniate locus was placed on Chr01 of the A genome in cytogenetic work using monosomes (White and Endrizzi 1965). Since okra and laciniate alleles have similar effects on leaf shape, they were considered to be genes at "duplicate" loci in the two genomes (White and Endrizzi 1965). This served as the basis for establishing that Chr01 and Chr15 are homeologous chromosomes in tetraploid cotton (White and Endrizzi 1965). There exists no mention of any of the other leaf shape alleles being transferred to *G. hirsutum* from Asiatic cotton. However, since they are considered an allelic series in Asiatic cotton, it is assumed that they could also make up an allelic series at the Chr01 A genome homeolog in *G. hirsutum* (Jones 1982, Meredith 1984). Some *G. hirsutum* lines still exist today that purportedly carry the laciniate allele. However, the term laciniate is occasionally used incorrectly and interchangeably with okra and super-okra in the literature. Therefore, which of these lines truly carry the laciniate allele is unknown.

1.10 Advantages and Disadvantages of the Major Leaf Shapes in Cotton Production

1.10.1 Boll Rot Resistance

Okra cultivars reduce losses due to boll rot by 7-11% compared to normal when conditions are low to moderate (Jones and Andries 1967, Rao and Weaver 1976) and 43-45% when environments are more severe (Andries et al. 1969, Karami and Weaver 1972). Super-okra shows an even greater reduction in boll rot (55%) than normal under severe boll rot conditions (Andries et al. 1970). The reduction in boll rot of these mutants is generally attributed to microclimatic differences. Okra and super-okra develop a more open canopy allowing for greater air circulation and light penetration, creating a less favorable environment for microbial growth.

1.10.2 Earliness to Maturity and Flowering Rates

Okra cultivars are earlier to flowering and maturity than normal with effects reported from two to twelve days earlier (Jones and Andries 1967, Andries et al. 1969, Karami and Weaver 1972, Rao and Weaver 1976, Heitholt and Meredith 1998). Super-okra matures nine to twelve days earlier than normal (Andries et al. 1970). Both mutant leaf shapes also display drastically increased flowering rates with okra flowering 25-50% and super-okra 85-100% more often than normal (Andries et al. 1969, 1970, Heitholt 1993, Gonias et al. 2011). This increase in flowering is generally offset by a marked increase in boll abortion (Kerby and Buxton 1976, Heitholt 1993, Heitholt 1995). Flowering rates and boll abscission of sub-okra cotton are not different than normal (Heitholt 1993). The increase in flowering rate and decrease in time to maturity is thought to be physiological in nature (Landivar et al. 1983). Since okra and super-okra need less photosynthate to both develop new leaves and maintain existing ones, they can contribute more photosynthate to reproductive development and do so earlier (Landivar et al. 1983). However, these varieties initiate more bolls than they can support, leading to higher rates of abortion (Landivar et al. 1983).

10.1.3 Insect Resistance

Two studies using okra vs. normal near-isolines in different genetic backgrounds associated okra with a reduction in damage by the pink bollworm (PBW) *Pectinophora gossypiella* (Wilson et al.1979, Wilson and George 1982). This reduction in PBW damage was at least partially attributed to the earliness of okra leaf (Wilson et al.1979). However, a later study utilizing seven pairs of leaf shape near-isolines found a reduction in PBW damage by okra in only three of the genetic backgrounds (Wilson 1986). In a follow-up study Stoneville 7A Okra, which had twice shown less damage to PBW than its normal leaf near-isoline (Wilson et al.1979, Wilson 1986), was shown to have decreased boll penetration by PBW larvae (Wilson et al.1986). The authors hypothesized the okra leaf character could result in increased thickness of the carpel (boll) walls, but only in certain genetic backgrounds (Wilson et al.1986). Wilson et al. (1986) also demonstrated that increased canopy and soil temperatures of Stoneville 7A Okra had no effect on the growth and development of PBW at multiple life stages. Later, under laboratory conditions, Stoneville 7A Okra did not have fewer larval entrance holes than Stoneville 7A Normal but did have fewer larvae per entrance hole in the first 50 minutes after infestation (Wilson et al.1992). There was no difference in either the number of entrance holes or larvae per hole between Stoneville 213 Okra and Stoneville 213 Normal (Wilson et al.1992). Furthermore, Stoneville 7A Okra had thinner carpel walls than Stoneville 7A Normal while the reverse was true in the Stoneville 213 near-isolines (Wilson et al.1992). This indicates that carpel wall thickness has no effect on PBW resistance (Wilson et al.1992). Therefore, there does not appear to be anything intrinsic to okra leaf shape that reduces PBW damage. In any case, the more complete control of PBW provided by the deployment of *Bt* cotton in the mid-1990s appears to have obviated the need for further research into the minor level of PBW resistance occasionally seen in okra leaf shape lines.

The silverleaf/sweet potato whitefly (*Bemisia tabaci/argentifolli*) and the banded-winged whitefly (*Trialeurodes abutilonea*) are serious insect pests of cotton. Yield is lost due to direct insect feeding and the transmission of viral diseases while the deposition of sticky honeydew on lint can reduce fiber quality (Centintas and McAuslane 2009). In numerous germplasm screens okra leaf shape has been associated with lower numbers of whitefly adults, nymphs, or eggs (Jones et al. 1975, Ozgur and Sekeroglu 1986, Sippell et al. 1987, Butler et al. 1988, Butler et al. 1991, Chu et al. 1999, Chu et al. 2000, Chu et al. 2002, Centintas and McAuslane 2009). One germplasm screen reported that okra did not have a consistent effect on nymph numbers (Flint and Parks 1990). In studies using isogenic lines, results where more mixed with results varying from okra being more resistant (Jones et al. 1975) to mixed (Butler et al. 1986, Butler and Wilson 1986) to okra having no or a negative effect on whitefly resistance (Butler and Wilson 1984). The decrease in whitefly numbers was largely attributed to microclimatic differences (Ozgur and Sekeroglu 1986, Sippell et al. 1987, Chu et al. 1999, Chu et al. 2002). A more open canopy may result in a reduced number of protected sites for feeding and oviposition and therefore create greater exposure to sunlight, high temperature, and low humidity. The majority of these studies do not investigate the impact of these reduced whitefly populations on yield and those that do provide mixed results (Butler et al. 1988, Chu et al. 1999). In conclusion, there may be some ability of mutant leaf shapes to reduce whitefly populations but the effect may also depend on other characters such as leaf hairiness (Walker and Natwick 2006) and the depth of vascular bundles from the leaf surface (Chu et al. 1999).

In a germplasm screen for resistance to the twospotted spider mite *Tetranychus urticae*, Bailey et al. (1978) noted that okra varieties tended to support lower mite population numbers than normal under insecticide treated fields. This difference was not seen in the untreated fields and was therefore attributed to greater pesticide penetration in okra (Bailey et al. 1978). Bailey and Meredith (1983) later showed that an okra isoline had fewer mites than frego, smooth, normal, nectariless, and glandless isolines. Comparing only Siokra to Deltapine 90 (normal), Wilson (1993) demonstrated that mite populations developed slower and reached lower peak levels on Siokra than Deltapine 90. This resulted in smaller yield reductions due to mite infestations in Siokra than Deltapine 90 (Wilson 1993). The mechanism of resistance of okra leaf cotton to twospotted spider mite was later attributed to reduced protected area for oviposition and feeding in a study utilizing isolines (Wilson 1994).

Okra leaf has been mentioned as providing some resistance to the boll weevil *Anthonomus grandis* (Pieters and Bird 1977, Jones 1982). The mechanisms of this resistance have been attributed to earliness (Pieters and Bird 1977) and a hotter and drier microclimate (Jones 1982). However, more detailed studies on the effect of the okra leaf character on the boll weevil were likely obviated by the successful eradication of the boll weevil through the Boll Weevil Eradication Program.

10.1.4 Application of Foliar Chemicals

Okra leaf shape has been credited with increased spray penetration of insecticides under field conditions (Jones et al. 1987). No effort was made to gauge the effect of this increased penetration but the theory is that the delivery of more insecticide to middle and lower parts of the canopy, where bolls and insects are more plentiful, can reduce damage and/or the number of insecticide applications (Jones et al. 1987). Super-okra would likely have the same or greater effect on spray penetration but this has not been documented.

10.1.5 Photosynthesis and Water Use Efficiency

Based on models and air-flow experiments, it was hypothesized that an okra leaf would have a thinner boundary layer than normal (Baker and Myhre 1969). This would make it easier for CO_2 to move into an okra leaf, thereby allowing a higher photosynthetic rate but would also increase water lost through transpiration (Baker and Myhre 1969). However, single okra leaves failed to fix more CO_2 than normal under experimental conditions and it was hypothesized that the boundary layer does not have a large enough effect on diffusion resistance to have an impact (Baker and Myhre 1969). Based on modeling studies, it was proposed that on a single leaf basis okra fixes 5% more CO_2 than normal but loses 7.5% more water through transpiration under conditions of low relative humidity (Buxton and Stapleton 1970). However, there was no difference in transpiration rates under high relative humidity, indicating that okra would have a photosynthetic and water use efficiency advantage under high humidity conditions (Buxton and Stapleton 1970). This advantage might be even larger when expanded to a full canopy effect rather than just a single leaf (Buxton and Stapleton 1970). However, later field studies showed no difference in CO_2 fixation rates between normal and okra canopies while super-okra fixed 29% less CO_2 (Pegelow et al. 1977). There was also no difference in transpiration rates or photosynthesis-to-transpiration ratios among the three leaf shapes (Pegelow et al. 1977). Karami et al. (1980) showed that there was no difference in water relations or photosynthetic rates among single leaves of normal, laciniate, and super-okra under well-watered greenhouse conditions. However, under drought-stress, super-okra had statistically improved water relations and carbon fixation with laciniate an intermediate between the two (Karami et al. 1980). This was attributed to the mutant leaves requiring less water to maintain themselves and thus being less susceptible to water stress (Karami et al. 1980). Two studies showed that there was no consistent difference in CO_2 canopy fixation rates among leaf shape isolines (Kerby et al. 1980, Peng and Krieg 1991). On a single leaf basis, one study found higher per unit leaf area fixation in okra compared to normal (Peng and Krieg 1991) while another found no consistent differences among normal, okra and super-okra (Perry et al. 1993). Wells et al. (1986) using leaf shape isolines in a MD 65-11 background found that okra and super-okra canopies had reduced canopy photosynthetic rates compared to normal and sub-okra. This was attributed to insufficient leaf area failing to maximize light interception in okra and super-okra (Wells et al. 1986). However, using the same leaf shape isolines but single leaf measurements Pettigrew et al. (1993) found that okra and super-okra fixed more CO_2 than normal. Combined with lower stomatal conductance measurements this lead to higher water use efficiency for the mutants compared to normal (Pettigrew et al. 1993). The effects were largely attributed to increases in leaf thickness and chlorophyll concentration of the mutants (Pettigrew et al. 1993). Contradictions to Wells et al. (1986) were attributed to differences in the time of day measurements were taken (Pettigrew et al. 1993) but no mention was made about single leaf vs. canopy measurements. The advantage of okra in carbon fixation and chlorophyll concentration, but not stomatal conductance, was confirmed in a later study using a different genetic background, but again on single leaves (Pettigrew 2004). The superiority of okra leaf in photosynthesis was attributed to improved light adapted photosystem II quantum efficiency, electron transport rate, and non-photochemical quenching (Pettigrew 2004). Gonias et al. (2011) used single leaf measurements on chamber-grown plants to show no difference in leaf photosynthesis, PSII quantum yield, or membrane leakage between normal and okra in the FM 832 background. Okra did display higher chlorophyll content while normal had thicker leaves (Gonias et al. 2011). Therefore, the reported effects of leaf shape on photosynthesis and transpiration indicate that individual leaves of the mutants may have certain advantages. However, when examined over an entire canopy these effects are generally negligible.

10.1.6 Lint Trash

Novick et al. (1991) investigated the effect of leaf shape on lint trash in two genetic backgrounds. Super-okra was found to reduce motes by 15% and small leaf trash by 20% (Novick et al. 1991). The reduction in motes was attributed to higher pollination rates due to increased pollen viability in the less humid and unshaded super-okra canopy (Novick et al. 1991). Both okra and super-okra were associated with increased cleanability due to a greater percentage of the leaf trash being harder materials such as veins and petioles that are less prone to shattering (Novick et al. 1991). Okra leaf can actually increase lint trash due to its propensity to hang on branches following defoliation (Novick et al. 1991). Overall, super-okra and sub-okra showed improved grades compared to the normal check following cleaning (Novick et al. 1991).

10.1.7 Yield and Fiber Quality

Studies comparing lint yield on a per plot basis in leaf shape isolines have reported effects ranging from +20% (Karami and Weaver 1972) to -8% (Wilson and George 1982).

Using single plant harvesting or hybrid cultivars have led to reports of the okra leaf yield penalty being as high as -17% and -38.5% respectively (Wilson 1986, Zhu et al. 2008). However, the majority of the isogenic studies found that leaf shape had no effect on yield (Jones and Andries 1967, Meredith 1983, 1984, 1985, Heitholt 1993, Heitholt et al. 1996, Gonias et al. 2011, Riar et al. 2013). Increases in yield of okra leaf isolines have been attributed to reductions in boll rot losses (Andries et al. 1969), earliness to escape late-season stress (Andries et al. 1969), improved harvest index (Karami and Weaver 1972), positive response to reduced insecticide application (Thomson et al. 1987), improved suitability to increased planting density (Heitholt et al. 1992, 1994), and higher water use efficiency (Stiller et al. 2004). However, increased planting density can be offset by difficulties in adaptation to mechanized production (Heitholt et al. 1996) and increased seed and seed treatment costs (Riar et al. 2013). Decreases in yield of okra leaf isolines have been attributed to their construction via backcrossing in elite backgrounds for normal leaf (Meredith and Wells 1986), insufficient light interception and photosynthesis (Wells et al. 1986), and increased evaporation of soil moisture (Zhu et al. 2008).

In conclusion, the effect of okra leaf shape on yield is likely heavily influenced by the environmental conditions in a given year. Early season stresses are more likely to negatively affect okra leaf lines by delaying the attainment of canopy closure (Heitholt and Meredith 1998). As long as okra does not reach canopy closure substantially later than normal leaf, it should be in a position to out-yield normal due to its greater relative partitioning of assimilate to reproductive growth throughout the season. On the other hand, late season stresses are more likely to depress yields in normal leaf lines because they produce fewer flowers and mature later (Heitholt and Meredith 1998). Therefore, an ideal cotton cultivar would produce normal leaves up until the point canopy closure is obtained. This would protect against adverse early season growing conditions while minimizing inputs and seeding rates. Then, once canopy closure was obtained, the cultivar would switch over to an open canopy effect of okra or even super-okra leaves. Subsequent wasteful vegetative growth through shading would be minimized and more photosynthate would be directed towards reproduction. Lower bolls and leaves would remain relatively unshaded leading to lower rates of low position boll shedding. The upper open canopy would render the crop less susceptible to late season stresses. Therefore, consistently higher yields may be obtained than with either okra or normal leaf types alone.

The effect of super-okra isolines on yield is either none (Thomson 1971, Heitholt 1993) or slightly negative (-6-8.8% (Andries et al. 1970, Meredith 1984). Zhu et al. (2008) did report that super-okra hybrids yielded only 42% of a normal leaf check hybrid. The inferiority of super-okra leaf in terms of yield is likely attributable to insufficient light interception and canopy photosynthesis as well increased weed competition (Thomson 1971, Zhu et al. 2008). However, these may be offset by increasing planting density (Thomson 1971). The effect of sub-okra isolines on yield ranges from none (Heitholt 1993, Meredith et al. 1996) to a slight increase of 3-4.8% (Meredith 1984, Meredith and Wells 1987). The yield advantage of sub-okra may result from the combination of its ability to quickly reach canopy closure along with the benefits of a slightly more open canopy structure later in the season (Meredith 1984).

While individual studies occasionally report small statistically differences in either boll or fiber quality characters due to leaf shape, these differences are usually not substantial and consistent across studies. Therefore, it can be concluded that leaf shape has no impact on either boll characteristics or fiber quality.

1.11 General Growth and Development of Leaves in Model Species

1.11.1 Maintenance of the Shoot Apical Meristem (SAM)

The shoot apical meristem (SAM) is a mass of pluripotent stem cells that forms at the tip of the main stem and gives rise to all of the aboveground organs of the plant, including the leaves. The SAM is divided into three zones: the central zone (CZ) at the very tip, a rib zone (RZ) just below the CZ, and a peripheral zone (PZ) on the flanks of the CZ and RZ (Bar and Ori 2014, Kalve et al. 2014, Sluis and Hake 2015). Slow division in the CZ eventually pushes clumps of stem cells into the PZ where they are recruited to initiate the next leaf primordia (Townsley and Sinha 2012, Kalve et al. 2014, Sluis and Hake 2015). The pluripotent stem cell pool must be maintained in order to provide fresh stem cells for additional organs, but also cannot grow uncontrollably. Maintenance of the size of the stem cell pool is primarily controlled by a negative feedback loop between *WUSCHEL* (*WUS*) and *CLAVATA* (*CLV*)*1*, *2*, and *3* but also involving *LEAF CURLING RESPONSIVENESS* (*LCR*), microRNA (miR)394, *ARABIDOPSIS RESPONSE REGULATOR* (*ARR*) and the phytohormone cytokinin (Kalve et al. 2014, Sluis and Hake 2015). Loss-of-function of *WUS* leads to abolishment of the SAM and plants that abort early, after developing no or a few above-ground organs (Sluis and Hake 2015). In *clv1* and *clv3* mutants, mis-expression of *WUS* leads to an enlarged or fasciated SAM (Kalve et al. 2014, Sluis and Hake 2015).

CLV3 produces a mobile protein at the top of the CZ in the SAM (Kalve et al. 2014, Sluis and Hake 2015). CLV3 moves into the lower layers of the SAM where it interacts with CLV1 and CLV2 to repress *WUS* (Kalve et al. 2014, Sluis and Hake 2015). *WUS* is active in the RZ where it produces a signal that moves into the CZ to activate *CLV3* expression (Kalve et al. 2014). LCR can interfere with this process but is itself negatively regulated by miR394 (Kalve et al. 2014). *WUS* expression is thought to be activated by cytokinin but cytokinins can be negatively regulated by *ARR* (Kalve et al. 2014). *WUS* can inhibit *ARR*, completing another loop (Kalve et al. 2014).

Class I *KNOTTED-LIKE HOMEOBOX* (*KNOX1*) genes are also required for both the formation and maintenance of the SAM. *KNOX1* genes work together to maintain the SAM by activating cytokinins through *ISOPENTENYL TRANSFERASE 7* (*IPT7*) and repressing gibberellins (Jasinski et al. 2005, Yanai et al. 2005, Kalve et al. 2014). Repression of giberrellins occurs through the both the activation of *GA2ox* involved in gibberellin catabolism and the repression of *GA20ox* involved in gibberellin biosynthesis (Jasinski et al. 2005, Kalve et al. 2014, Bar and Ori 2014). Cytokinins are thought to help maintain stem cell indeterminacy while giberrellins promote cellular maturation and differentiation. Data in rice suggests that *KNOX1* genes up-regulate brassinosteroid (BR) catabolism genes (Sluis and

Hake 2015). BRs promote organ development; therefore they are directly antagonistic to the role of *KNOX1* genes to maintain stem cell indeterminacy (Sluis and Hake 2015).

1.11.2 Initiation of Leaves

Leaf primordia initiation in the PZ begins with the development of local auxin maxima established primarily using the auxin efflux carrier PIN-FORMED1 (PIN1) (Townsley and Sinha 2012, Bar and Ori 2014). Mutations to PIN1, other genes necessary for auxin transport, or genes involved in auxin biosynthesis (such as the *YUCCA* genes) will prevent organ formation, while exogenous auxin application can induce organ development (Bar and Ori 2014). The Hofmeister principle states that primordia will form as far away from the last primordia as possible (Bar and Ori 2014). This implies that the previous primordia can establish some sort of exclusion zone to prevent new primordia from forming near it and auxin may be a critical player in this process (Bar and Ori 2014).

Auxin response factors (ARFs), including MONOPTEROS (MP) in Arabidopsis, respond to the high auxin and mutations to ARFs can alter proper organ formation (Bar and Ori 2014). High auxin concentrations within the developing primordia activate the *MACCHI-BOU 4* (*MAB4*) gene (Bar and Ori 2014, Sluis and Hake 2015). MAB4, along with the related proteins MAB4/ENP/NPY1-LIKE1 (MEL1) and MEL2 then direct PIN1 patterning to move auxin towards the internal and basal part of the primordia where the mid-vein will form (Sluis and Hake 2015). This change in PIN1 localization is also mediated in part by the serine/threonine protein kinase PINOID (Bar and Ori 2014). The developing mid-vein eventually connects the developing leaf primordia to the existing vasculature (Bar and Ori 2014).

The down-regulation of *KNOX1* in the developing primordia is a critical step in the formation of leaves, allowing stem cells to begin to differentiate (Townsley and Sinha 2012). Auxin may play a role in the initial down-regulation of *KNOX1* in the developing leaf primordia (Bryne 2012). The continued repression of *KNOX1* involves *ASYMMETRIC LEAVES 1* (*AS1*) and *AS2* through chromatin remodeling (Ori et al. 2000, Byrne et al. 2000, Semiarti et al. 2001, Schofield et al. 2014, Sluis and Hake 2015). AS1 and AS2 work together to recruit the POLYCOMB-REPRESSIVE COMPLEX2 while AS1 recruits the histone deactylease HDA6 to *KNOX1* genes (Luo et al. 2012, Lodha et al. 2013, Sluis and Hake 2014). *AS1* is activated by the redundant Class I *TEOSINTE BRANCHED 1, CYCLOIDEA AND PCF TRANSCRIPTION FACTOR 3* (*TCP*) transcription factors (*TCP2, 4, 5, 10, 13, 17,* and *24*) (Koyama et al. 2010). The *TCP* genes are in turn activated by the redundant *YABBY* (*YAB*) family transcription factors *FILAMENTOUS FLOWER* (*FIL*), *YAB2, YAB3,* and *YAB5* that are themselves negatively regulated by miR319/JAGGED AND WAVY (JAW) (Sarojam et al. 2010, Palatnik et al. 2003). The auxin response genes *INDOLE-3-ACETIC ACID INDUCIBLE3* (*IAA3*) and *SMALL AUXIN UPREGULATED* (*SAUR*) as well as miR164 are also activated by the *TCPs* (Koyama et al. 2010). Loss-of-function of *AS1* leads to ectopic expression of *KNOX1* and abnormal leaf lobing (Byrne et al. 2000). BLADE-ON-PETIOLE1 (BOP1) and BOP2 directly activate AS2 but also work independently to down regulate *KNOX1* (Jun et al. 2012, Ha et al. 2010, Sluis and Hake

2015). *KNOX1* mediated repression of *BOP1* and *BOP2* in the SAM is required for its proper formation (Jun et al. 2010). *BOP1* and *BOP2* are also involved in proximal (petiole)-distal (blade) patterning (Townsley and Sinha 2012). As their name implies, mutants in BOP genes are compromised in the distinction between blade and petiole with leaf tissue developing where the petiole should be (Townsley and Sinha 2012).

Arabidopsis carries four *KNOX1* genes. The first, *SHOOT MERISTEMLESS* (*STM*), is a homolog of the maize *KNOTTED1* gene the family is named after (Barton and Poethig 1993, Long et al. 1996). Lincoln et al. (1994) identified two additional family members, *KNOTTED-LIKE FROM ARABIDOPSIS THALIANA1* (*KNAT1*) and *KNAT2* that behave in much the same way as *STM*. Later a fourth *KNOX1* gene in Arabidopsis, *KNAT6* was discovered (Semiarti et al. 2001).

1.11.3 Establishment of Adaxial-Abaxial Polarity

One of the first critical steps in the development of leaf is the establishment of adaxial (top, facing/closer to the SAM) vs. abaxial (bottom, facing away/farther from the SAM) polarity. The default fate of cells in developing primordia is abaxial (at least in Arabidopsis) and proper development depends on the hypothetical “anlagen factor” thought to originate in the SAM and travel to the developing primordia (Townsley and Sinha 2012, Tsukaya 2013). What the “anlagen factor” is, or whether it even exists at all, is not currently known (Tsukaya 2013).

The HD-ZIPIII transcription factors are required for adaxial identity (Townsley and Sinha 2012). There are three critical members of this group in Arabidopsis (*REVOLUTA* (*REV*), *PHABULOSA* (*PHB*), and *PHAVOLUTA* (*PHV*) that seem highly functionally redundant (Townsley and Sinha 2012). Triple mutants of *REV*, *PHB*, and *PHV* are seedling lethal (Townsley and Sinha 2012). The HD-ZIPIIIs activate the expression of little zipper proteins (ZPRs) (Townsley and Sinha 2012). ZPRs arose as truncated HD-ZIPIIIs that retained just the leucine zipper but none of the other functional domains (Townsley and Sinha 2012). ZPRs form heterodimers with the HD-ZIPIIIs to block their action, thus the HD-ZIPIIIs are able to self-regulate themselves in the adaxial domain (Townsley and Sinha 2012). *AS2* expression also appears critical for adaxial determination (Tsukaya 2013).

In the abaxial domain, miR165/166 target the HD-ZIPIIIs to repress them (Townsley and Sinha 2012, Sluis and Hake 2015). *KANADI* (*KAN*) transcription factors also promote abaxial cell fate (Townsley and Sinha 2012). KAN1 and KAN2 act redundantly to repress AS2 in the embryo while KAN1 suppresses AS2 in the abaxial domain of the leaf primordia (Townsley and Sinha 2012). KAN and the HD-ZIPIIIs have an antagonistic relationship in adaxial-abaxial polarity determination (Tsukaya 2013, Kalve et al. 2014). *ETTIN/AUXIN RESPONSE FACTOR 3* (*ETT/ARF3*) and *ARF4* act as transcriptional activators in the presence of auxin and play a key role in abaxial patterning (Townsley and Sinha 2012). *ETT* and *ARF4* arose as duplications of one another that have maintained many of the same roles and are highly functionally redundant (Townsley and Sinha 2012). *YABBY* transcription factors form positive feedback loops with *KAN* and *ARF4* in Arabidopsis (Kalve et al. 2014).

However, the role of YABBY genes in adaxial-abaxial polarity may be more variable across plant species (Townsley and Sinha 2012).

Back in the adaxial domain, *ETT* and *ARF4* are targeted by ta-siRNAs from the non-coding *TRANS-ACTING siRNA 3* (*TAS3*) gene (Townsley and Sinha 2012). *SUPRESSOR OF GENE SILENCING* (*SGS3*) regulates this interaction in an unknown way (Townsley and Sinha 2012). *ETT* is transcribed throughout the primordia but the ta-siRNAs prevent its translation in the adaxial domain (Townsley and Sinha 2012). *TAS3* itself is targeted by miR390 in an interaction mediated by ARGONAUTE 7/ZIPPY (AGO7/ZIP) (Townsley and Sinha 2012). AGO7 also stabilizes the ta-siRNAs to help in the degradation of *ETT* and *ARF4* (Kalve et al. 2014). *AGO1* is also required for proper functioning of the miRNA and siRNA systems (Townsley and Sinha 2012, Kalve et al. 2014). AGO10 also helps to clear miR165/166 signals from the adaxial domain, allowing HD-ZIPIII function (Townsley and Sinha 2012).

In addition to the adaxial and abaxial domains, current theory states that there exists a third, middle domain involving the expression of *WUSCHEL RELATED HOMEOBOX 1* (*WOX1*), *PRESSED FLOWER* (*PRS)/WOX3 and FIL* (Tsukaya 2013, Kalve et al. 2014). Expression of *WOX1* and *PRS* are specific to the middle domain and both are negatively regulated by KAN (Kalve et. 2014). Not too much else is known about the middle domain at this time (Kalve et al. 2014). This system is generally the same between rice, maize, and Arabidopsis but the different components may work together differently (Townsley and Sinha 2012).

1.11.4 Establishment of the SAM-Leaf Boundary

Another critical step in the development of the leaf is the proper definition of the organ boundary. The *CUP-SHAPED COTYLEDON* (*CUC*) genes of the NAC transcription factor family up-regulate *KNOX1* to create a boundary between the developing leaf primordia and the SAM (Townsley and Sinha 2012). Specifically, CUC1 and CUC2 act redundantly to up-regulate *STM* and *KNAT6,* which then proceed to activate *CUC3* (Aida et al. 1997, Aida et al. 1999, Takada et al. 2001, Vroemen et al. 2003,Townsley and Sinha 2012). CUC1 also activates *LIGHT SENSITIVE HYPOCOTYLS 3* (*LSH3*) and *LSH4* that theoretically inhibits differentiation in the boundary region (Sluis and Hake 2015). Further, other *KNOX1* genes (i.e. *KNAT1* and *KNAT2*) appear not to be involved in boundary formation indicating there is some degree of sub-functionalization within the *KNOX1* gene family (Townsley and Sinha 2012). The microRNA miR164 post-transcriptionally regulates *CUC1* and *CUC2* and this relationship is essential for proper organ patterning (Rhoades et al. 2002, Kasschau et al. 2003, Laufs et al. 2005). *CUC3* is likely the result of an ancient duplication event in *CUC* genes that has lost its ability to be regulated by miR164 like *CUC1* and *CUC2* (Townsley and Sinha 2012). The expression of *CUC* genes also appears to be regulated by *KNOX1* and auxin during boundary formation (Sluis and Hake 2015).

LATERAL ORGAN FUSION1 (*LOF1*) and *LOF2* are expressed at the bottom of developing leaf primordia and act through an unknown mechanism to promote proper leaf initiation from the SAM (Sluis and Hake 2015). *JAGGED LATERAL ORGANS* (*JLO*) is expressed at the boundary where it forms a heterodimer with AS2 to repress *KNOX1*

(Tsukaya 2013, Sluis and Hake 2015). JLO is also involved in auxin signaling (Sluis and Hake 2015). Organ differentiation also involves up-regulation of brassinosteroids (BRs) (Sluis and Hake 2015). BRs activate expression of *LATERAL ORGAN BOUNDARIES* (*LOB*), a gene that restricts growth in the boundary allowing the meristem and organ to grow apart properly (Sluis and Hake 2015). *LOB* functions by directly activating the expression of *PHYB ACTIVATION-TAGGED SUPPRESSOR 1* (*BAS1*), which in turn inactivates BRs (Sluis and Hake 2015).

1.11.5 Expansion of the Leaf: Cell Proliferation vs. Cell Enlargement

Leaves increase in size through cell division as well as the expansion of existing cells and the two processes are at least somewhat interrelated (Kalve et al. 2014). Leaf growth through cell division is primarily cytoplasmic while growth in expanding cells is largely through the vacuole (Kalve et al. 2014). Cytoplasmic growth consumes a substantial amount of resources in terms of energy and macromolecules (Kalve et al. 2014). These energy needs are met by the TORC1 complex in plants whose three main members are TARGET OF RAPAMYCIN (TOR), REGULATORY ASSOCIATED PROTEIN OF TOR (RAPTOR), and LETHAL WITH SEC13 8 (LST8) (Kalve et al. 2014). The TORC1 complex influences ribosome production, translation rates, autophagy, and the accumulation of storage products such as starch and lipids (Kalve et al. 2014).

Auxin in the primordia activates *ARGOS*, which in turn activates *AINTEGUMENTA* (*ANT*) and *CYCLIN D3* (*CYCD3*), both of which promote cell division (Kalve et al. 2014).

Cell division is also promoted by the transcription factors of the *TCP* family, which are negatively regulated by miR319 (Kalve et al. 2014, Bar and Ori 2014). The GROWTH-REGULATING FACTORS (GRFs) are a nine member family in Arabidopsis that influence cell division and leaf size while GRF-INTERACTING FACTOR 1/ANGUSTIFOLIA 3 (GIF1/AN3) interacts with GRF5 (Tsukaya 2013, Kalve et al. 2014). The TCPs also upregulate production of miR396, which negatively regulate the six of the GRFs and GIF, indicating the system has some negative feedback (Tsukaya 2013, Kalve et al. 2014). KLU and STRUWWELPETER (SWP) also independently promote cell division (Kalve et al. 2014).

Generally speaking, when cells stop dividing, they begin to expand (Kalve et al. 2014). A general theory in leaf development is that there exists a cell cycle “arrest front” that begins at the tip and gradually moves towards the leaf base (Kalve et al. 2014). Another theory dictates that the arrest front is held constant at some point during leaf development and then moves rapidly towards the base in more of a two-step process (Tsukaya 2013, Bar and Ori 2014). How exactly the arrest front moves is not known but it is hypothesized to be either a mobile signal that moves between cells or governed in a spatiotemporal context (Kalve et al. 2014). Expansion of cells is largely dependent on the arrest of cell proliferation (Kalve et al. 2014). BIG BROTHER (BB), PHYTOCHROME AND FLOWERING TIME 1/ MEDIATOR 25 (PFT1/MED25) and DA1 all restrict cell division in the leaf with DA1 being activated by ABA (Kalve et al. 2014). TOR and ABL both promote cell expansion while ORE1 SISTER 1 (ORS1) promotes both expansion and division (Kalve et al. 2014). Both sugar and brassinosteroids are thought to be involved in the division vs. elongation decision but their exact role is not known (Kalve et al. 2014).

Expansion in meristematic cells is driven largely by increases in cytoplasmic and nuclear size (Kalve et al. 2014). Nuclear size is enlarged by endoreduplication, or replication of DNA without cell division (Kalve et al. 2014). Differentiated cells largely increase in volume by enlarging their vacuole through the uptake of water and solutes (Kalve et al. 2014). For this to occur, cell wall expansion is necessary (Kalve et al. 2014). Relaxation of the cell wall allows for the uptake of water followed by wall extension by turgor pressure and cell wall stiffening (Kalve et al. 2014). Stiffening occurs through dehydration, cross-linking induced by the release of reactive oxygen species, and the insertion of cell wall components (Kalve et al. 2014).

1.11.6 Development of Marginal Leaf Structures

As the leaf develops and differentiates, it retains marginal blastozones (MBs) that can revert to stem cell potential at any point (Bar and Ori 2014). The MBs are thought to be controlled by the same factors that govern the SAM and leaf initiation (Bryne 2012, Bar and Ori 2014). The MBs are considered critical for the formation of “marginal structures” in the leaf including lobes, serrations, and leaflets (Bar and Ori 2014). These structures result in an increase in leaf shape complexity and result from longer periods of cell proliferation without differentiation (Bar and Ori 2014). Auxin plays a major role in development of leaflets in compound leaves and each leaflet can be thought of as a simple leaf on the rachis (Bar and

Ori 2014). Most leaf shape changes come through minor tweaks in the leaf development system that increase or decrease the length of indeterminacy in the developing leaf (Bar and Ori 2014). The re-establishment of *KNOX1* expression during later stages of leaf development is sufficient to create these longer periods of indeterminacy and is one of the major determinants of increases in leaf complexity although other mechanisms likely exist (Piazza et al. 2010, Townsley and Sinha 2012). Loss-of-function mutations to *KNOX1* genes to establish simple leaves are likely uncommon since these plants would fail to develop a SAM leading to a highly abnormal phenotype that aborts after producing no or a few above ground organs. Therefore, alterations in *KNOX1* expression that contribute to the diversity of leaf shapes seen throughout the plant kingdom likely arise through direct modifications to the promoter sequences of *KNOX1* genes (Piazza et al. 2010, Uchida et al. 2010) or to the other genes capable of influencing their expression.

Serrations in Arabidopsis form based on periodic spurts and slowdowns in cell proliferation along the developing leaf margin (Tsukaya 2013, Bar and Ori 2014). This patterning is influenced by auxin maxima and involves PIN1, miR164 and CUC2 (Nikovics et al. 2006, Tsukaya 2013, Bar and Ori 2014). As mentioned earlier, all three of these factors are involved in other aspects of leaf development demonstrating the multi-faceted nature of many of the genes in this process. Over-expression of miR164 or the inactivation of *CUC2* lead to Arabidopsis with smooth leaves (Nikovics et al. 2006). On the other hand, mutations to miR164 or mutations to *CUC2* that confer resistance to miR164, increase the depth of serrations (Nikovics et al. 2006). Ectopic expression of *CUC1* causes an increase in leaf complexity by forming a positive feedback loop with *KNOX1* genes (Hasson et al. 2011). *CUC3* has a lesser role in leaf serration, as it is involved in maintaining the growth of the serrations but not their initiation (Hasson et al. 2011). *DEVELOPMENT-RELATED PCG TARGET IN THE APEX 4* (*DPA4*) also negatively regulates *CUC2* independent of miR164 in a manner that also influences leaf margin serrations in Arabidopsis (Engelhorn et al. 2012). However, while the role of *CUC* genes in organ-boundary formation appear highly conserved, their role in influencing leaf shape may be more variable across species (Nikovics et al. 2006).

TCP factors also act redundantly to promote differentiation while being negatively regulated by miR319 (Bar and Ori 2014). Over-expression of miR319 produces more complex leaves while over-expression of *TCPs* results in simpler leaves (Bar and Ori 2014). *TCP*s promote simpler leaf shapes through repression of the *FRUITFUL* (*FUL*)-like genes and activation of gibberellins (Bar and Ori 2014). TCP4 also acts with BRM to bind to the promoter of *ARRs* and reduce the sensitivity to cytokinins (Tsukaya 2013, Bar and Ori 2014). Cytokinins act downstream of *KNOX1* to delay differentiation and thus promote leaf complexity (Bar and Ori 2014). Put succinctly, the interplay between *TCPs* and *KNOX1* controls leaf shape differentiation and complexity by regulating the cytokinin/gibberellin balance (Bar and Ori 2014).

Two *BEL1-LIKE HOMEODOMAIN* (*BLH*)/*SAWTOOTH* (*SAW*) genes, *BLH2/SAW1* and *BLH4/SAW2*, act redundantly to negatively regulate *KNOX1* in developing leaves by forming heterodimers with them (Kumar et al. 2007, Townsley and Sinha 2012). Loss-of-function of both *SAW1* and *SAW2* leads to an increase in leaf serrations in Arabidopsis much like ectopic expression of *KNOX1* (Kumar et al. 2007, Townsley and Sinha 2012). Truncated *KNOX1* genes exist that produce proteins with only the MEINOX domain for protein-protein interactions and not the DNA-binding homeodomain (Magnani and Hake 2008, Townsley and Sinha 2012). These shortened proteins can still form heterodimers with BLHs, sequester them in the cytoplasm, and prevent them from interacting with true *KNOX1* genes (Magnani and Hake 2008, Townsley and Sinha 2012). Examples of these special *KNOX1* genes include *KNATM* in Arabidopsis and *PTS* in tomato (Magnani and Hake 2008, Townsley and Sinha 2012). Over-expression of *KNATM* in Arabidopsis leads to increases in leaf serrations that are characteristic of *KNOX1* mis-expression and *BLH* loss-of-function (Magnani and Hake 2008). Thus, this represents a rare case of a truncated gene increasing the function of the gene from which it was derived rather than decreasing it (Townsley and Sinha 2012).

BOP1 and *BOP2* act redundantly to repress some *KNOX1* as well as *JAGGED* (*JAG*) and members of the *YAB* family (Norberg et al. 2005, Ha et al. 2010). The double mutant *bop1 bop2* displays phenotypes similar to those in *KNOX1* over-expressing lines including increased leaf lobing and dissection (Ha et al. 2010). *JAG* is a zinc-finger transcription factor involved in lateral organ development whose loss-of-function increases leaf serration (Dinneny et al. 2004, Ohno et al. 2004).

Members of the IRLC clade of the legumes have evolved a unique way to develop compound leaves that does not involve the re-expression of *KNOX1* genes during leaf development (Townsley and Sinha 2012). Instead, the *LEAFY* (*LFY*) orthologs fulfill the role of *KNOX1* in maintaining stem cell indeterminacy during cell proliferation and allowing the development of compound leaves (Townsley and Sinha 2012). *LFY* orthologs also may play a small role in leaf compounding in other legumes and tomato (Townsley and Sinha 2012).

The role of the HD-ZIP transcription factor *LATE MERISTEM IDENTITY1 (LMI1*) was originally proposed by Saddic et al. (2006). Knockout of *LMI1* in Arabidopsis lead to an increase in leaf complexity with one to two leaflets forming at the base of the leaf that was linked to an increase in expression of *KNAT1*. Within the Brassicaceae, *LMI1* underwent a tandem duplication resulting in a *LMI1*-like gene called *REDUCED COMPLEXITY* (*RCO*) (Vlad et al. 2014). *RCO* evolved a unique expression pattern through regulatory diversification and was shown to be responsible for the compound leaves of close Arabidopsis relative *Cardamine hirsuta*. Arabidopsis subsequently lost *RCO*, leaving it with simple leaves. Changes to the promoter region of the *RCO* gene are also responsible for the differences in leaf complexity seen between lobed *Capsella rubella* and simple leafed *Capsella grandiflora* in the Brassicaceae (Sicard et al. 2014). Thus, it has only recently been shown that *LMI1*-like genes are major players in the leaf shape pathway.

1.11.7 Evolution of Leaf Shape

The gene regulatory networks (GRNs) governing leaf shape does not appear to be highly conserved across species (Townsley and Sinha 2012). The system is at least somewhat similar to that seen in flower development since flowers arose as modified leaves and many of the same genes (or close orthologs) are involved in the development of both organs

(Townsley and Sinha 2012, Tsukaya 2013, Bar and Ori 2014) While most of the components and some general themes are shared, individual factors can have completely different roles in other species (Townsley and Sinha 2012). These GRNs need to variable in order to generate new phenotypes that natural selection can act on to push the species forward evolutionarily (Townsley and Sinha 2012). But changes cannot be drastic enough to cause large-scale disturbances to the plant (Townsley and Sinha 2012). Therefore, most of the changes come slowly through mutations that modify leaf shape slightly (Townsley and Sinha 2012). Most of these effects will therefore be selectively neutral or close to it and therefore stochastic processes likely play a larger role in the diverse shapes of leaf seen throughout nature rather than natural selection (Townsley and Sinha 2012). In addition to these GRNs, there are many biophysical and mechanical factors that influence leaf formation. This topic is beyond the scope of this section, but has been reviewed elsewhere (Ali et al. 2014).

Gene duplication may be a major way for species to experiment with new leaf shapes in a conserved manner. When a gene duplicates, it retains most of its original interactions within the GRN leading to redundancy (Townsley and Sinha 2012). This leaves one of the copies “free” to explore new roles that allow the plant to evolutionarily “tinker” with its leaf shape (i.e. neo-functionalization) without the usually catastrophic results of loss-of-function of a non-redundant gene (Townsley and Sinha 2012).

1.12 Statement of Objectives

Our hypothesis is that an ideal cotton cultivar will produce a lower canopy of normal leaves before transitioning to okra around the eighth node. The lower normal leaves should allow the variety to rapidly reach canopy closure and maximum light interception, establishing a basis for high yield. The transition to okra should then maintain the traditional benefits of an okra cultivar including earlier and higher flowering, increased spray penetration, and reduced incidence of boll rot to further maximize yield and profit. The first step in the development of such an ideotype is the dissection of the molecular and genetic control of leaf shape in cotton. In order to carry out this research, various molecular mapping techniques will be used to identify a small leaf shape gene candidate region. Expression analysis will resolve among the genes in the candidate region and strong candidates will be sequenced to identify causal polymorphisms. Ideally, a transgenic approach will be used to validate the candidate gene and develop a proof-of-concept of the ideotype.

REFERENCES

Adams KL, Wendel JF. (2005) Novel patterns of gene expression in polyploid plants. *Trends in Genetics* 21 (10): 539-543.

Adams KL, Wendel JF. (2005) Polyploidy and genome evolution in plants. *Current Opinion in Plant Biology* 8 (2): 135-141.

Aida M, Ishida T, Fukaki H, Fujisawa H, Tasaka M. (1997) Genes involved in organ separation in Arabidopsis: An analysis of the *cup-shaped cotyledon* mutant. *The Plant Cell* 9: 841-857.

Aida M, Ishida T, Tasaka M. (1999) Shoot apical meristem and cotyledon formation during Arabidopsis embryogenesis: Interaction among the *CUP-SHAPED COTYLEDON* and *SHOOT MERISTEMLESS* genes. *Development* 126: 1563-1570.

Albrecht T, Wimmer V, Auinger HJ, Erbe M, et al. (2011) Genome-based prediction of testcross values in maize. *Theoretical and Applied Genetics* 123 (2): 339-350.

Ali O, Mirabet V, Godin C, Traas J. (2014) Physical models of plant development. *Annual Reviews of Cell and Developmental Biology* 30: 59-78.

Allen RD, Aleman L. (2011) Abiotic stress and cotton fiber development pp. 156-167. In: Stress Physiology in Cotton (Ed: DM Oosterhuis).

Andries JA, Jones JE, Sloane LW, Marshall JG. (1969) Effects of okra leaf shape on boll rot, yield, and other important characters of Upland cotton, *Gossypium hirsutum* L. *Crop Science* 9: 705-710.

Andries JA, Jones JE, Sloane LW, Marshall JG. (1970) Effects of super okra leaf shape on boll rot, yield, and other characters of Upland cotton, *Gossypium hirsutum* L. *Crop Science* 10: 403-407.

Baker DN, Myhre DL. (1969) Effects of leaf shape and boundary layer thickness on photosynthesis in cotton (*Gossypium hirsutum*). *Physiologia Plantarum* 22: 1043-1049.

Bailey JC, Furr RE, Hanny BW, Meredith WR. (1978) Field populations of twospotted spider mites on sixteen cotton genotypes at Stoneville, MS 1977. *Journal of Economic Entomology* 71: 911-912.

Bailey JC, Meredith WR. (1983) Resistance of cotton, *Gossypium hirsutum* L., to natural field populations of twospotted spider mites (Acari: Tetranychidae). *Environmental Entomology* 12: 763-764.

Bar M, Ori N. (2014) Leaf development and morphogenesis. *Development* 141: 4219-4230.

Barton MK, Poethig RS. (1993) Formation of the shoot apical meristem in *Arabidopsis thaliana*: an analysis of development in the wild type and in the *shoot meristemless* mutant. *Development* 119: 823-831.

Blenda A, Fang DD, Rami JF, Garsmeur O, et al. (2012). A high density consensus genetic map of tetraploid cotton that integrates multiple component maps through molecular marker redundancy check. *PLOS One* 7 (9): 1-16.

Buckler ES, Holland JB, Bradbury PJ, Acharya CB et al. (2009) The genetic architecture of maize flowering time. *Science* 325: 714-718.

Butler GD, Wilson FD. (1984) Activity of adult whiteflies (Homoptera: Aleyrodidae) within plantings of different cotton strains and cultivars as determined by sticky-trap catches. *Journal of Economic Entomology* 77: 1137-1140.

Butler GD, Henneberry TJ, Wilson FD. (1986) *Bemisia tabaci* (Homoptera: Aleyrodidae) on cotton: adult activity and cultivar oviposition preference. *Journal of Economic Entomology* 79: 350-354.

Butler GD, Wilson FD. (1986) Whitefly adults in okra-leaf and normal-leaf cotton. *Cotton: A College of Agriculture Report* pp.223-226.

Butler GD, Rimon D, Henneberry TJ. (1988) *Bemisia tabaci* (Homoptera: Aleyrodidae): populations on different cotton varieties and cotton stickiness in Israel. *Crop Protection* 7: 43-47.

Butler GD, Wilson FD, Fishler G. (1991) Cotton leaf trichomes and populations of *Empoasca lybica* and *Bemisia tabaci. Crop Protection* 10: 461-464.

Buxton DR, Stapleton HN. (1970) Predicted rates of net photosynthesis and transpiration as affected by the microenvironment and size of cotton leaf. *Proceedings of the Beltwide Cotton Production Research Conference* pp. 31-33.

Byers RL, Harker DB, Yourstone SM, Maughan PJ, Udall JA. (2012) Development and mapping of SNP assays in tetraploid cotton. *Theoretical and Applied Genetics* 124: 1201-1214.
Byrne ME. (2012) Making leaves. *Current Opinion in Plant Biology* 15: 24-30.

Byrne ME, Barley R, Curtis M, Arroyo JM, et al. (2000) *Asymmetric leaves1* mediates leaf patterning and stem cell function in Arabidopsis. *Nature* 408: 967-971.

Byrne ME, Simorowski J, Martienssen RA. (2002) *ASYMMETRIC LEAVES1* reveals *knox* gene redundancy in Arabidopsis. *Development* 129: 1957-1965.

Cantrell RG, Davis DD. (2000) Registration of NM24016, a interspecific-derived cotton genetic stock. *Crop Science* 40: 1208.

Centintas R, McAuslane H. (2009) Effectiveness of parasitoids of *Bemisia tabaci* (Hemiptera: Aleyrodidae) on cotton cultivars differing in leaf morphology. *Florida Entomologist* 92: 538-547.

Chu C, Cohen AC, Natwick ET, Simmons GS, Henneberry TJ. (1999) *Bemisia tabaci* (Hemiptera: Aleyrodidae) biotype B colonization and leaf morphology relationships in Upland cotton cultivars. *Australian Journal of Entomology* 38: 127-131.

Chu C, Freeman T, Natwick ET, Buckner JS, et al. (2000) *Bemisia argentifolii* adult, nymph, and egg densities and egg distribution on selected Upland cottons. *Journal of Entomological Science* 35: 39-47.

Chu C, Natwick ET, Henneberry TJ. (2002) *Bemisia tabaci* (Homoptera: Aleyrodidae) biotype B colonization on okra- and normal-leaf Upland cotton strains and cultivars. *Journal of Economic Entomology* 95: 733-738.

Chuck G, Lincoln C, Hake S. (1996) *KNAT1* induces lobed leaves with ectopic meristems when overexpressed in Arabidopsis. *The Plant Cell* 8: 1277-1289.

Clement JD, Constable GA, Stiller WN, Liu SM. (2012) Negative associations still exist between yield and fibre quality in cotton breeding programs in Australia and USA. *Field Crops Research* 128: 1-7.

Cook JP, McMullen MD, Holland JB, Tian F. (2012) Genetic architecture of maize kernel composition in the nested association mapping and inbred association panels. *Plant Physiology* 158 (2): 824-834.

Cotton Incorporated (2015) Agricultural disciplines: Entomology. http://www.cottoninc.com/fiber/AgriculturalDisciplines/Entomology/

Dinneny JR, Yadegari R, Fischer RL, Yanofsky MF, Weigel D. (2004) The role of *JAGGED* in shaping lateral organs. *Development* 131: 1101-1110.

Dolan L, Poethig RS. (1991) Genetic analysis of leaf development in cotton. *Development Supplement* 1: 39-46.

Endrizzi JE, Brown MS. (1964) Identification of monosomes for six chromosomes in *Gossypium hirsutum*. *American Journal of Botany* 51 (2): 117-120.

Endrizzi JE, Kohel RJ. (1966) Use of telosomes in mapping three chromosomes in cotton. *Genetics* 54: 535-550.

Endrizzi JE, Stein R. (1975) Association of two marker loci with chromosome 1 in cotton. *The Journal of Heredity* 66: 75-78.

Engelhorn J, Reimer JJ, Leuz I, Gobel U, et al. (2012) *DEVELOPMENT-RELATED PcG TARGET IN THE APEX 4* controls leaf margin architecture in *Arabidopsis thaliana*. *Development* 139: 2566-2575.

Flagel LE, Wendel JF. (2010) Evolutionary rate variation, genomic dominance, and duplicate gene expression evolution during allotetraploid cotton speciation. *New Phytologist* 186 (1): 184-193.

Flint HM, Parks NJ. (1990) Infestation of germplasm lines and cultivars of cotton in Arizona by whitefly nymphs (Homoptera: Aleyrodidae). *Journal of Entomological Science* 25:223-229.

Flint-Garcia SA, Thornsberry JM, Buckler ES. (2003) Structure of linkage disequilibrium in plants. *Annual Reviews of Plant Biology* 54: 357-374.

Gonias ED, Oosterhuis DM, Bibi AC. (2011) Light interception and radiation use efficiency of okra and normal leaf cotton isolines. *Environmental and Experimental Botany* 72: 217-222.

Gore MA, Percy RG, Zhang J, Fang DD, Cantrell RG. (2012) Registration of the TM-1/NM24016 cotton recombinant inbred line mapping population. *Journal of Plant Registrations* 6 (1): 124-127.

Gore MA, Fang DD, Poland JA, Zhang J, et al. (2014) Linkage map construction and quantitative trait locus analysis of agronomic and fiber quality traits in cotton. *The Plant Genome* 7 (1): doi:10.3835/plantgenome2013.07.0023

Green JM. (1953) Sub-okra, a new leaf shape in Upland cotton. *The Journal of Heredity* 44 (6): 229-232.

Grover CE, Zhu X, Grupp KK, Jareczk JJ, et al. (2015) Molecular confirmation of species status for the allopolyploid cotton species, *Gosspypium ekmanianum* Wittmack. *Genetic Resources and Crop Evolution* 62 (1): 103-114.

Guo W, Cai C, Wang C, Han Z et al. (2007) A microsatellite-based, gene-rich linkage map reveals genome structure, function and evolution in *Gossypium. Genetics* 176 (1): 527-541

Ha CM, Jun JH, Fletcher JC. (2010) Control of Arabidopsis leaf morphogenesis through regulation of the *YABBY* and *KNOX* families of transcription factors. *Genetics* 186: 197-206.

Hammond D. (1941a) The expression of genes for leaf shape in *Gossypium hirsutum* L. and *Gossypium arboreum* L. I. The expression of gene for leaf shape in *Gossypium hirsutum* L. *American Journal of Botany* 28: 124-137.

Hammond D. (1941b) The expression of genes for leaf shape in *Gossypium hirsutum* L. and *Gossypium arboreum* L. II. The expression of gene for leaf shape in *Gossypium arboreum* L. *American Journal of Botany* 28: 138-150.

Harland SC. (1932) The genetics of *Gossypium*. *Bibliographica Genetica* 9: 107-182.

Hasson A, Plessis A, Blein T, Adroher B, et al. (2011) Evolution and diverse roles of the *CUP-SHAPED COTYLEDON* genes in Arabidopsis leaf development. *The Plant Cell* 23: 54-68.

Heitholt JJ, Pettigrew WT, Meredith WR. (1992) Light interception and lint yied of narrow-row cotton. *Crop Science* 32: 728-733.

Heitholt JJ. (1993) Cotton boll retention and its relationship to lint yield. *Crop Science* 33: 486-490.

Heitholt JJ. (1994) Canopy characteristics associated with deficient and excessive cotton plant population densities. *Crop Science* 34: 1291-1297.

Heitholt JJ. (1995) Cotton flowering and boll retention in different planting configurations and leaf shapes. *Agronomy Journal* 87: 994-998.

Heitholt JJ, Meredith WR, Williford JR. (1996) Comparison of cotton genotypes varying in canopy characteristics in 76-cm vs. 102-cm rows. *Crop Science* 36: 955-960.

Heitholt JJ, Meredith WR. (1998) Yield, flowering, and leaf area index of okra-leaf and normal-leaf cotton isolines. *Crop Science* 38: 643-648.

Huang X, Zhao Y, Wei X, Li C et al. (2012) Genome-wide association study of flowering time and grain yield traits in a worldwide collection of germplasm. *Nature Genetics* 44: 32-39.

Hutchinson JB. (1934) The genetics of cotton. Part X. The inheritance of leaf shape in Asiatic *Gossypiums*. *Journal of Genetics* 28 (3): 437-513.

Hutchinson JB, Silow RA. (1939) Gene symbols for use in cotton genetics. *The Journal of Heredity* 30 (10): 461-464.

Jannink JF, Lorenz AJ, Iwata H. 2010. Genomic selection in plant breeding: from theory to practice *Briefings in Functional Genomcis* 9(2): 166-177.

Jasinski S, Piazza P, Craft J, Hay A, et al. (2005) KNOX action in Arabidopsis is mediated by cytokinin and gibberellin activities. *Current Biology* 15: 1560-1565.

Jiang C, Wright RJ, Woo SS, DelMonte TA, Paterson AH. (2000) QTL analysis of leaf morphology in tetraploid *Gossypium*. *Theoretical and Applied Genetics* 100: 409-418.

Jones JE. (1982) The present state of the art and science of cotton breeding for leaf-morphological traits. *Proceedings of the Beltwide Cotton Production Research Conferences* pp.93-99.

Jones JE, Andries JA. (1967) Okra leaf cotton for boll rot control. *Louisiana Agriculture* 10 (4): 4-5, 11.

Jones JE, Clower DF, Milam MR, Caldwell WD, Melville DR. (1975) Resistance in Upland cotton to the banded-winged whitefly *Trialeurodes abutilonea* (Haldeman). *Proceedings of the Beltwide Cotton Production Research Conferences* pp.98-99.

Jones JE, James D, Sistler FE, Stringer SJ. (1987) Spray penetration of cotton canopies as affected by leaf and bract isolines. *Louisiana Agriculture* 29: 14-15, 17.

Jun JH, Ha CM, Fletcher JC. (2010) BLADE-ON-PETIOLE1 coordinates organ determinacy and axial polarity in Arabidopsis by directly activating ASYMMETRIC LEAVES2. *The Plant Cell* 22: 62-76.

Kalve S, De Vos D, Beemster GTS. (2014) Leaf development: a cellular perspective. *Frontiers in Plant Science* Volume5, Article 362.

Karami E, Krieg DR, Quisenberry JE. (1980) Water relations and carbon-14 assimilation of cotton with different leaf morphology. *Crop Science* 20: 421-426.

Karami E, Weaver JB. (1972) Growth analysis of American Upland cotton, *Gossypium hirsutum* L., with different leaf shapes and colors. *Crop Science* 12: 317-320.

Kasschau KD, Xie Z, Allen E, Llave C, et al. (2003) P1/HC-Pro, a viral suppressor of RNA silencing, interferes with Arabidopsis development and miRNA function. *Developmental Cell* 4: 205-217.

Kerby TA, Buxton DR. (1976) Fruiting as affected by leaf type and population density. *Proceedings of the Beltwide Cotton Production Research Conferences* pp.67-70.

Kerby TA, Buxton DR, Matsuda K. (1980) Carbon source-sink relationships with narrow-row cotton canopies. *Crop Science* 20:208-213.

Koenning S. (2004) Cotton Disease Information Notes: Plant Pathology Extension, North Carolina State University. https://www.ces.ncsu.edu/depts/pp/notes/Cotton/cotton_contents.html

Koyama T, Mitsuda N, Seki M, Shinozaki K, Ohme-Takagi M. (2010) TCP transcription factors regulate the activities of ASYMMETRIC LEAVES1 and miR164, as well as the auxin response during differentiation of leaves in Arabidopsis. *The Plant Cell* 22: 3574-3588.

Kumar R, Kushalappa K, Godt D, Pidkowich MS, et al. (2007) The Arabidopsis BEL1-LIKE HOMEODOMAIN proteins SAW1 and SAW2 act redundantly to regulate *KNOX* expression spatially in leaf margins. *The Plant Cell* 19: 2719-2735.

Kump KL, Bradbury PJ, Wisser RJ, Buckler ES et al. (2011) Genome-wide association study of quantitative resistance to southern leaf blight in the maize nested association mapping population. *Nature Genetics* 43: 163-168.

Lacape JM, Nguyen TB, Thibivilliers S, Bojinov B, et al. (2003) A combined RFLP-SSR-AFLP map of tetraploid cotton based on a *Gossypium hirsutum* x *Gossypium barbadense* backcross population. *Genome* 46 (4): 612-626.

Lacape JM, Jacobs J, Arioli T, Derijcker R et al. (2009) A new interspecific, *Gossypium hirsutum* x *G. barbadense*, RIL population: towards a unified consensus linkage map of tetraploid cotton. *Theoretical and Applied Genetics* 119 (2): 281-292.

Lacape JM, Llewellyn D, Jacobs J, Arioli T et al. (2010) Meta-analysis of cotton fiber quality QTLs across diverse environments in *Gossypium hirsutum x G. barbadense* RIL population. *BMC Plant Biology* 10: 132.

Lacape JM, Gawrysiak G, Cao TV, Viot C, et al. (2013) Mapping QTLs for traits related to phenology, morphology, and yield components in an inter-specific *Gossypium hirsutum* x *G. barbadense* cotton RIL population. *Field Crops Research* 144: 256-267.

Landivar JA, Baker DN, Jenkins JN. (1983) Application of GOSSYM to genetic feasibility studies. Analyses of fruit abscission and yield in okra-leaf cotton. *Crop Science* 23: 497-504.

Laufs P, Peaucelle A, Morin H, Traas J. (2004) MicroRNA regulation of the *CUC* genes is required for boundary size control in Arabidopsis meristems. *Development* 131: 4311-4322.

Laurie CC, Chasalow SD, LeDeaux JR, McCarroll R et al. (2004) The genetic architecture of response to long-term artificial selection for oil concentration in maize kernel. *Genetics* 168: 2141-2155.

Li F, Fan G, Wang K, Sun F et al. (2014) Genome sequence of the cultivated cotton *Gossypium arboreum*. *Nature Genetics* 46 (6): 567-572.

Li F, Fan G, Lu C, Xiao G et al. (2015) Genome sequence of cultivated Upland cotton (*Gossypium hirsutum* TM-1) provides insight into genome evolution. *Nature Biotechnology* 33: 524-530.

Lin Z, Zhang Y, Zhang X, Gio X. (2009) A high-density integrative linkage map for *Gossypium hirsutum*. *Euphytica* 166: 35-45.

Lincoln C, Long J, Yamaguchi J, Serikawa K, Hake S. (1994) A *knotted1*-like homeobox gene in Arabidopsis is expressed in the vegetative meristem and dramatically alters leaf morphology when overexpressed in transgenic plants. *The Plant Cell* 6: 1859-1876.

Long JA, Moan EI, Medford JI, Barton MK. (1996) A member of the KNOTTED class of homeodomain proteins encoded by the *STM* gene of Arabidopsis. *Nature* 379: 66-69.

Magnani E, Hake S. (2008) *KNOX* lost the *OX*: The Arabidopsis *KNATM* gene defines a novel class of KNOX transcriptional regulators missing the homeodomain. *The Plant Cell* 20: 875-887.

Massman JM, Gordillo A, Lorenzana RE, Bernardo R. (2013) Genomewide predictions from maize single-cross data. *Theoretical and Applied Genetics* 126 (1): 13-22.
McClendon CA. (1912) Mendelian inheritance in cotton hybrids. *Georgia Agricultural Experiment Station Bulletins* 99: 139-228.

Mell PH. 1890. A microscopic study of the cotton plant. Bulletin No. 13 of the Agricultural Experiment Station of the Agricultural and Mechanical College, Auburn, AL.

Meredith WR. (1984) Influence of leaf morphology on lint yield of cotton – enhancement by the sub-okra trait. *Crop Science* 24 (5): 855-857.

Meredith WR. (1985) Lint yield genotype x environment interaction in Upland cotton as influenced by leaf canopy isolines. *Crop Science* 25: 509-512.

Meredith WR, Wells R. (1986) Normal vs okra leaf yield interactions in cotton. I. Performance of near-isogenic lines from bulk populations. *Crop Science* 26: 219-222.

Meredith WR, Wells R. (1987) Sub okra leaf influence on cotton yield. *Crop Science* 27:47-48.

Meredith WR, Pettigrew WT, Heitholt JJ. (1996) Sub-okra, semi-smoothness, and nectariless effect on cotton lint yield. *Crop Science* 36: 22-25.

National Cotton Council of America (NCCA). (2013) World of Cotton (http://www.cotton.org/econ/world/) and Frequently Asked Questions (http://www.cotton.org/edu/faq/index.cfm)

Nguyen TB, Giband M, Brottier P, Risterucci AM, Lacape JM. (2004) Wide coverage of the tetraploid cotton genome using newly developed microsatellite markers. *Theoretical and Applied Genetics* 109 (1): 167-175.

Nikovics K, Blein T, Peaucelle A, et al. (2006) The balance between the *MIR164A* and *CUC2* genes controls leaf margin serration in Arabidopsis. *The Plant Cell* 18: 2929-2945.

Norberg M, Holmlund M, Nilsson O. (2005) The *BLADE ON PETIOLE* genes act redundantly to control the growth and development of lateral organs. *Develpoment* 132: 2203-2213.

Novick RG, Jones JE, Anthony WS, Aguillard W, Dickson JI. (1991) Genetic trait effects of nonlint trash of cotton. *Crop Science* 31: 1029-1034.

Ohno CK, Reddy GV, Heisler MGB, Meyerowitz EM. (2004) The Arabidopsis *JAGGED* gene encodes a zinc finger protein that promotes leaf tissue development. *Development* 131: 1111-1122.

Ori N, Eshed Y, Chuck G, Bowman JL, Hake S. (2000) Mechanisms that control *knox* gene expression in Arabidopsis. *Development*: 127: 5523-5532.

Ozgur AF, Sekeroglu E. (1986) Population development of *Bemisia tabaci* (Homoptera: Aleurodidae) on various cotton cultivars in Cukurova, Turkery. *Agriculture, Ecosystems, and Environment* 17: 83-86.

Park Y, Alabady MS, Ulloa M, Sickler B et al. (2005) Genetic mapping of new cotton fiber loci using EST-derived microsatellites in an interspecific recombinant inbred line cotton population. *Molecular Genetics and Genomics* 274 (4): 428-441.

Paterson AH, Boman RK, Brown SM, Chee PW et al. (2004) Reducing the genetic variability of cotton. *Crop Science* 44: 1900-1901.

Paterson AH, Wendel JF, Gundlach H, Gui H, et al. (2012) Repeated polyploidization of *Gossypium* genomes and the evolution of spinnable fibers *Nature.* 492: 423-427.

Pegelow EJ, Buxton DR, Briggs RE, Muramoto H, Gensler WG. (1977) Canopy photosynthesis and transpiration of cotton as affected by leaf type. *Crop Science* 17: 1-4.

Peng S, Krieg DR. (1991) Single leaf and canopy photosynthesis response to plant age in cotton. *Agronomy Journal* 83: 704-708.

Perry SW, Kreig DR, Hutmacher RB. (1983) Photosynthetic rate control in cotton. *Plant Physiology* 73: 662-665.

Pettigrew WT. (2004) Physiological consequences of moisture deficit stress in cotton. *Crop Science* 44: 1265-1272.

Pettigrew WT, Heitholt JJ, Vaughn KC. (1993) Gas exchange differences and comparative anatomy among cotton leaf-type isolines. *Crop Science* 33: 1295-1299.

Piazza P, Bailey CD, Cartolano M, Krieger J, et al. (2010) *Arabidopsis thaliana* leaf form evolved via loss of *KNOX1* expression in leaves in association with a selective sweep. *Current Biology* 20: 2223-2228.

Pieters EP, Bird LS. (1977) Field studies of boll weevil resistant cotton lines possessing the okra leaf-frego bract characters. *Crop Science* 17: 431-433.

Meredith WR. (1983) Effect of environments and genetic backgrounds on evaluation of cotton isolines. *Crop Science* 23: 51-54.

Poland JA, Bradbury PJ, Buckler ES, Nelson RJ. (2011) Genome-wide nested association mapping of quantitative resistance to northern leaf blight in maize. *Proceedings of the National Academy of Sciences* 108 (17): 6893-6898.

Rao MJ, Weaver JB. (1976) Effect of leaf shape on response of cotton to plant population, N rate, and irrigation. *Crop Science* 68: 599-601.

Rapp RA, Udall JA, Wendel JF. (2009) Genomic expression dominance in allopolyploids. *BMC Biology* 7 (18) doi:10.1186/1741-7007-7-18

Reddy UK, Rong J, Nimmakayala P, Vajja G et al. (2011) Use of diversity array technology for integration into a cotton reference map and anchoring to a recombinant inbred line map. *Genome* 54: 349-359.

Reinisch AJ, Dong JM, Brubaker CL, Stelly DM, et al. (1994) A detailed RFLP map of cotton, *Gossypium hirsutum* x *Gossypium barbadense*: chromosome organization and evolution in a disomic polyploid genome. *Genetics* 138 (3): 829-847.

Rhoades MW, Reinhart BJ, Lim LP, Burge CB, et al. (2002) Prediction of plant microRNA targets. *Cell* 110: 513-520.

Riar R, Wells R, Edmisten K, Jordan D, Bacheler J. (2013) Cotton yield and canopy closure in North Carolina as influenced by row width, plant population, and leaf morphology. *Crop Science* 53: 1704-1711.

Rong J, Abbey C, Bowers JE, Brubaker CL, et al. (2004) A 3347-locus genetic recombination map of sequence-tagged sites reveals features of genome organization, transmission and evolution of cotton (*Gossypium*). *Genetics* 166:389-417.

Rong J, Feltus R, Waghmare VN, Pierce GJ et al. (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.

Saddic LA, Huvermann B, Bezhani S, Su Y, et al. (2006) The LEAFY target LMI1 is a meristem identity regulator and acts together with LEAFY to regulate expression of CAULIFLOWER. *Development* 133:1673-1682.

Said JI, Lin Z, Zhang X, Song M, Zhang J. (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.

Sarojam R, Sappl PG, Goldschmidt A, Efroni I, et al. (2010) Differentiating Arabidopsis shoots from leaves by combined YABBY activities. *The Plant Cell* 22: 2113-2130.

Scofield S, Dewitte W, Murray JAH. (2014) STM sustains stem cell function in the Arabidopsis shoot apical meristem and controls *KNOX* gene expression independently of the transcriptional repressor AS1. *Plant Signaling & Behavior* 9: e289634

Semiarti E, Ueno Y, Tuskaya H, Iwakawa H, et al. (2001) The *ASYMMETRIC LEAVES2* gene of *Arabidopsis thaliana* regulates formation of a symmetric lamina, establishment of venation, and repression of meristem-related homeobox genes in leaves. *Development* 128: 1771-1783.

Shappley ZW, Jenkins JN, Meredith WR, McCarty JC. (1998) An RFLP linkage map of Upland cotton, *Gossypium hirsutum* L. *Theoretical and Applied Genetics* 97: 756-761.

Shoemaker DN. (1909) A study of leaf characters in cotton hybrids. *The Journal of Heredity* 5 (1): 116-119.

Sicard A, Thamm A, Marona C, Lee YW, et al. (2014) Repeated evolutionary changes of leaf morphology caused by a mutation to a homeobox gene. *Current Biology* 24:1880-1886.

Sippell DW, Bindra OS, Khalifa H. (1987) Resistance to whitefly (*Bemisia tabaci*) in cotton (*Gossypium hirsutum*) in the Sudan. *Crop Protection* 6: 171-178.

Sluis A, Hake S. (2015) Organogenesis in plants: Initiation and elaboration of leaves. *Trends in Genetics* 31 (6): 300-306.

Song XL, Guo WZ, Han ZG, Zhang TZ. (2005) Quantitative trait loci mapping of leaf morphological traits and chlorophyll content in cultivated tetraploid cotton. *Journal of Integrative Plant Biology* 47 (11): 1382-1390.

Stephens SG. (1945) A genetic survey of leaf shape in new world cottons – a problem in critical identification of alleles. *Journal of Genetics* 46 (2-3): 313-330.

Stephens SG. (1949) The cytogenetics of speciation in *Gossypium*. I. Selective elimination of the donor parent genotype in interspecific backcrosses. *Genetics* 34 (5): 627-637.

Stephens SG. (1950) The internal mechanism of speciation in *Gossypium*. *Genetics* 16 (3): 115-149.

Stephens SG. (1955) Linkage in Upland cotton. *Genetics* 40 (6): 903-917

Stiller WN, Reid PE, Constable GA. (2004) Maturity and leaf shape as traits influencing cotton cultivar adaptation to dryland conditions. *Agronomy Journal* 96: 656-664.
Thomson NJ. (1971) Effects of the superokra leaf gene on cotton growth, yield, and quality. *Australian Journal of Agricultural Research* 23: 285-293.

Sunilkumar G, Campbell LM, Puckhaber L, Stipanovic RD, Rathore KS. (2006) Engineering cottonseed for use in human nutrition by tissue-specific reduction of toxic gossypol. *Proceedings of the National Academy of Sciences* 103 (48): 18054-18059.

Takada S, Hibara K, Ishida T, Tasaka M. (2001) The *CUP-SHAPED COTYLEDON1* gene of Arabidopsis regulates shoot apical meristem formation. *Development* 128: 1127-1135.

Thomson NJ, Reid PE, Williams ER. (1987) Effects of the okra leaf, nectariless, frego bract, and glabrous conditions on yield and quality of cotton lines. *Euphytica* 36: 545-553.

Tian F, Bradbury PJ, Brown PJ, Hung H. (2011) Genome-wide association study of leaf architecture in the maize nested association mapping population. *Nature Genetics* 43:159-62.

Townsley BT, Sinha NR. (2012) A new development: Evolving concepts in leaf ontogeny. *Annual Reviews in Plant Biology* 63: 535-562.

Tsukaya H. (2013) Leaf development. *The Arabidopsis Book* doi: 10.1199/tab.0163

Uchida N, Kimura S, Koening D, Sinha N. (2010) Coordination of leaf development via regulation of *KNOX1* genes. *Journal of Plant Research* 123: 7-14.

Ulloa M, Meredith WR, Shappley ZW, Kahler AL. (2002) RFLP genetic linkage maps from four $F_{2:3}$ populations and a joinmap of *Gossypium hirsutum* L. *Theoretical and Applied Genetics* 104: 200-208.

Vlad D, Kierzkowski D, Rast MI, Vuolo F, et al. (2014) Leaf shape evolution through duplication, regulatory diversification, and loss of a homeobox gene. *Science* 343: 780-783

Vroeman CW, Mordhorst AP, Albrecht C, Kwaaitaal MACJ, de Vries SC. (2003) The *CUP-SHAPED COTYLEDON3* gene is required for boundary and shoot meristem formation in Arabidopsis. *The Plant Cell* 15:1563-1577.

Walker GP, Natwick ET. (2006) Resistance to silverleaf whitefly, *Bemisia argentifolii* (Hem., Aleyrodidae), in *Gossypium thurberi*, a wild cotton species. *Journal of Applied Entomology* 130: 429-436.

Wallace JG, Larsson SJ, Buckler ES. (2014) Entering the second century of maize quantitative genetics. *Heredity* 112: 30-38.

Wallace TP, Bowman D, Campbell BT, Chee P et al. (2009) Status of the USA cotton germplasm collection and crop vulnerability. *Genetic Resource and Crop Evolution* 56 (4): 507-532.

Wang K, Wang Z, Li F, Ye W et al. (2012) The draft genome of a diploid cotton *Gossypium raimondii*. *Nature Genetics* 44 (10): 1098-1103.

Wells R, Meredith WR, Williford JR. (1986) Canopy photosynthesis and its relationship to plant productivity in near-isogenic cotton lines differing in leaf morphology. *Plant Physiology* 82: 635-640.

Wendel JF, Brubaker C, Alvarez I, Cronn R, Stewart JM. (2009) Evolution and natural history of the cotton genus. In: Genetics and Genomics of Cotton AH Paterson (ed.) 3-19.

White TG, Endrizzi JE. (1965) Tests for the association of marker loci with chromosomes in *Gossypium hirsutum* by the use of aneuploids. *Genetics* 51: 605-612.

Wilson FD. (1986) Pink bollworm resistance, lint yield, and lint yield components of okra-leaf cotton in different genetic backgrounds. *Crop Science* 26: 1164-1167.

Wilson FD, George BW. (1982) Effects of okra-leaf, frego-bract, and smooth-leaf mutants on pink bollworm damage and agronomic properties of cotton. *Crop Science* 22: 798-801.

Wilson FD, George BW, Fry KE, Szaro JL, et al. (1986) Pink bollworm (Lepidoptera: Gelechiidae) egg hatch, larval success, and pupal and adult survival of okra- and normal-leaf cotton. *Journal of Economic Entomology* 79 (6): 1671-1675.

Wilson FD, Szaro JL, Hefner BA. (1992) Behavior of pink bollworm larvae (Lepidoptera: Gelechiidae) on bolls of normal-leaf and okra-leaf cotton isolines under laboratory conditions. *Journal of Agricultural Entomology* 9 (1): 55-63.

Wilson LJ. (1993) Spider mites (Acari: Tetranychidae) affect yield and fiber quality of cotton. *Journal of Economic Entomology* 86: 566-585.

Wilson LJ. (1994) Resistance of okra-leaf cotton genotypes to twospotted spider mites (Acari: Tetranychidae). *Journal of Economic Entomology* 87: 1726-1735.

Wilson RL, Wilson FD, George BW. (1979) Mutants of *Gossypium hirsutum*: Effect on pink bollworm in Arizona. *Journal of Economic Entomology* 72 (2): 216-219.

Yanai O, Shani E, Dolezal K, Tarkowski P, Sablowski R, Sandberg G, Samach A, Ori N. (2005) Arabidopsis KNOX1 proteins activate cytokinin biosynthesis. *Current Biology* 15: 1566-1571.

Yu J, Yu S, Lu C, Wang W et al. (2007) High-density linkage map of cultivated allotetraploid cotton based on SSR, TRAP, SRAP, and AFLP markers. *Journal of Integrative Plant Biology* 49 (5): 716-724.

Yu JZ, Kohel RJ, Fang DD, Cho J et al. (2012) A high-density simple sequence repeat and single nucleotide polymorphism genetic map of the tetraploid cotton genome. *Genes, Genomes, Genetics* 2: 43-58.

Yu Y, Yuan D, Liang S, Li X et al. (2011) Genome structure of cotton revealed by a genome-wide SSR genetic map constructed from a BC_1 population between *Gossypium hirsutum* and *G. barbadense*. *BMC Genomics* 12 (15): doi:10.1186/1471-2164-12-15.

Zhang K, Zhang J, Ma J, Tang S. (2012) Genetic mapping and quantitative trait locus analysis of fiber quality traits using a three-parent composite population in Upland cotton (*Gossypium hirsutum* L.) *Molecular Breeding* 29 (2): 335-348.

Zhang T, Hu Y, Jiang W, Fang L, et al. (2015) Sequencing of allotetraploid cotton (*Gossypium hirsutum* L. acc. TM-1) provides a resource for fiber improvement. *Nature Biotechnology* 33: 531-537.

Zhu C, Gore M, Buckler ES, Yu J. (2008) Status and prospects of association mapping in plants. *The Plant Genome* 1: 5-20.

Zhu QH, Spriggs A, Taylor JM, Llewellyn D, Wilson I. (2014) Transcriptome and complexity-reduced, DNA-based identification of intraspecies single-nucleotide polymorphisms in the polyploidy *Gossypium hirsutum* L. *Genes ,Genomes, Genetics* 4 (10): 1893-1905.

Zhu W, Liu K Wang XD. (2008) Heterosis in yield, fiber quality, and photosynthesis of okra leaf oriented hybrid cotton (*Gossypium hirsutum* L.). *Euphytica* 164: 283-291.

CHAPTER 2: Mapping and genomic targeting of the major leaf shape gene (*L)* in Upland cotton (*Gossypium hirsutum* L.)

As adapted from: Andres RJ, Bowman DT, Kaur B, Kuraparthy V (2014) Mapping and genomic targeting of the major leaf shape gene (*L)* in Upland cotton (*Gossypium hirsutum* L.) *Theoretical and Applied Genetics* 127: 167-177.

Abstract

Leaf shape in cotton is an important trait that influences yield, earliness, flowering rate, disease resistance, and the efficacy of foliar chemical application. The leaves of okra leaf cotton display an enhanced lobing pattern as well as ectopic outgrowths along the lobe margins when compared to normal leaf cotton. To better understand the molecular and genetic processes underlying leaf shape in cotton, a normal leaf landrace accession (PI 607650) was crossed to an okra leaf breeding line (NC05AZ21). An F_2 population of 236 individuals confirmed the incompletely dominant single gene nature of the okra leaf shape trait in *Gossypium hirsutum* L. Molecular mapping with simple sequence repeat markers localized the leaf shape gene to a 5.4 cM interval in the distal region of the short arm of chromosome 15. Orthologous mapping of the closely linked markers to the sequenced diploid D genome donor (*Gossypium raimondii*) tentatively resolved the leaf shape locus to a small genomic region. By using the *G. raimondii* genome annotation within this region, two genes most similar to Arabidopsis *LATE MERISTEM IDENTITY1* (*LMI1*) are identified as possible candidates for the leaf shape trait in cotton. The linked molecular markers and delineated genomic region in the sequenced diploid D-genome will assist in the future high-resolution mapping and map-based cloning of the leaf shape gene in cotton.

Introduction

Cotton (*Gossypium* spp.) is the world's most important source of natural fiber as well as one of its leading oilseed crops. The majority of cultivated cottons (*G. hirsutum* and *G. barbadense*) are allotetraploid species ($2n = 4x = 52$, AADD) formed by the chance hybridization of two diploid progenitor species, *G. arboreum* ($2n = 2x =26$, AA) and *G. raimondii* ($2n = 2x = 26$, DD) (Wendel et al. 2009).

Nearly all Upland cotton (*G. hirsutum*) cultivars possess the normal or broad leaf shape. Along with normal (*l*), the mutant leaf types okra (L^O), sub-okra (L^U), and super-okra (L^S) make up an allelic series at the classical leaf shape locus (Chapter 1). This locus has been placed on the short arm of Chr15, a D sub-genome chromosome orthologous to *G. raimondii* Chr02, in a combination of cytogenetic and genetic mapping work (Chapter 1). Sea-islaned (L^E) is a possible fourth leaf shape mutant found commonly in the cultivars of extra-long staple cotton (*G. barbadense*). However, it has not been definitively proven that sea-island is a distinct leaf shape from sub-okra (Meredith 1984). Instead, the relatively minor differences between these two leaf types might be the result of different alleles at loci of minor effect throughout the genetic background of the two species.

The normal leaf shape in Upland cotton is broad and palmate with five readily observable, yet insipid lobes (Figure 1). Sub-okra is the least severe of the leaf shape mutants with a moderate increase in the degree of leaf lobing that generally does not become clear until later in the growing season. The okra leaf shape is characterized by a pronounced lobing pattern with deeper sinuses leading to a reduction in the photosynthetic area per leaf.

Numerous ectopic outgrowths along the lobe margins that appear random in shape, size, and position indicate that okra leaf is defective in cell patterning or differentiation. Super-okra is the most severe leaf shape mutation, approaching classification as a compound leaf with sinuses penetrating almost all the way to the petiole. Ectopic outgrowths occasionally appear in super-okra but much less frequently than in okra. At maturity, super-okra has been reduced to a single, narrow strip of leaf emanating from the petiole.

The advantages of okra leaf cultivars in production include reduced incidence of boll rot, accelerated flowering rates, early maturity, improved penetration of chemical sprays, resistance to the two-spotted spider mite *Tetranychus urticae*, and higher yield potential under certain environmental conditions (Chapter 1). Individual studies have also associated okra leaf shape with increased resistance to the pink bollworm (*Pectinophora gossypiella)*, the silverleaf/sweet potato whitefly (*Bemisia tabaci/argentifolli)*, and the banded-winged whitefly (*Trialeurodes abutilonea*) as well as enhanced photosynthetic rates and water use efficiency (Chapter 1). However, review of the entire literature on these topics reveals that the effect of leaf shape on these latter characters is small and/or inconsistent (Chapter 1). Sub-okra varieties generally do not perform differently than normal leaf varieties, limiting the agricultural relevance of this leaf shape. While super-okra can expand on most of the advantages of okra leaf cultivars, its leaf area is too small to yield competitively.

The reduced leaf area of the leaf shape mutants can delay the attainment of canopy closure when compared to normal, especially under adverse early-season growing conditions. This can lead to reduced yields in mutant leaf shape varieties and exacerbate their issues in regards to weed pressure and increased boll shed. It is likely that an ideal cotton cultivar would produce a lower canopy of normal leaves before switching over to an upper canopy of okra or super-okra leaves roughly about the time of the onset of flowering. While such a phenotype does not appear to be present in the natural variation of cotton, it is plausible that such plant architecture would combine the advantages of both leaf shapes while minimizing their disadvantages. In order to develop a plant with this transforming leaf phenotype, a more complete understanding of the genetic mechanism controlling leaf shape in cotton is needed.

Currently, there is no publicly available genome sequence available for allotetraploid cotton. However, draft sequences of the presumed D-genome donor diploid, *G. raimondii*, have recently been released (Paterson et al. 2012, Wang et al. 2012). Additionally, the A genome diploid *G. arboreum* has also been recently sequenced (Li et al. 2014). The availability of these draft genome sequences should help facilitate the fine mapping and cloning of genes in cotton by providing a physical map to integrate with constructed genetic maps. The gene annotation of these draft genomes, as well as the available protein homology and gene ancestry tools of the Paterson et al. (2012) genome, may allow the combination of physical and genetic maps to identify highly plausible candidate genes underlying phenotypes of interest.

The first objective of the present study was to use an intraspecific *G. hirsutum* mapping population to identify SSR or RFLP-based sequence-tagged site (STS) markers linked to the leaf shape locus. These linked markers were then to be used to construct the orthologous region in *G. raimondii* and identify potential candidate genes within this region.

These markers and candidate genes will then serve as the basis for future efforts to fine map and clone the major leaf shape gene in cotton. A detailed understanding of the genetic mechanism controlling leaf shape is needed in order to allow its manipulation through breeding and/or biotechnology for cotton cultivar improvement.

Materials and Methods

Plant Material: The okra leaf breeding line NC05AZ21 was crossed to the normal leaf accession NC11-2100 in the summer of 2011 at the Central Crops Research Station in Clayton, NC. NC05AZ21 is a recently released okra leaf line with *Fusarium* wilt resistance and superior yield and lint percent (Kuraparthy et al. 2013). NC11-2100 (TX-2324, PI 607650) was obtained from the USDA Cotton Germplasm Collection in College Station, TX. NC11-2100 was selected as a normal leaf parent because this line was a photoperiod insensitive landrace that was expected to show better molecular marker polymorphism than an elite cotton line. A single F_1 plant was self-fertilized to obtain F_2 seeds during the winter of 2011-2012 utilizing an off-season nursery program in Mexico. The F_2 population along with parental lines was grown in Clayton, NC over the summer of 2012. Large-scale leaf samples were collected individually from all 236 F_2 plants for DNA extraction. Leaf shape phenotype was scored on all F_2 individuals at multiple points throughout the growing season. Plants were characterized as either normal, okra, or heterozygous based on their similarity to the images in Figure 2.

SSR and STS Marker Analyses: Genomic DNA was isolated using a CTAB approach originally developed by Tel-Zur et al. (1999) and modified by Wendel (http://www.eeob.iastate.edu/faculty/WendelJ/dnaextraction.htm). Forty SSR markers mapped to Chr15 in a high-density consensus (HDC) genetic map between *G. hirsutum* and *G. barbadense* (Blenda et al. 2012) were selected and primer sequences were obtained from the Cotton Marker Database (http://www.cottonmarker.org/). Additionally, 23 RFLP markers placed on Chr15 by the HDC genetic map were converted into 40 STS markers. To develop STS markers from the RFLP markers, sequences of the restriction site region were obtained from the Cotton Genome Database (http://www.cottondb.org/) or from the Cotton Diversity Database (http://cotton.pgml.uga.edu/Cotton/index.aspx). Primers were then designed off of the RFLP marker sequence using the Primer 3 software (http://frodo.wi.mit.edu/). Multiple primers were designed from the same RFLP when possible. Two candidate leaf shape genes later determined to be present in the orthologous region of the *G. raimondii* genome delineated by the flanking SSR markers were also used to develop STS markers. Genomic sequences of these two genes were obtained from the draft sequence of the *G. raimondii* genome (Paterson et al. 2012, DOE Joint Genome Institute: Cotton D v2.0 (annotation v2.1); www.phytozome.net) and inserted into Primer 3 for primer design. Primers were designed in such a way that the entire length of each gene was covered by at least one primer pair. The M13 tail sequence 5'-CACGACGTTGTAAAACGAC-3' was added to the 5' end of all forward primers in order to facilitate capillary-based gel electrophoresis (Schuelke 2000). All primers were supplied by Integrated DNA Technologies (Coralville, IA, USA). All SSR and

STS markers were evaluated on the leaf shape mapping population parents NC05AZ21 and NC11-2100. Only markers polymorphic on the parents were run on the F_2 mapping population.

Each PCR reaction consisted of 40ng of genomic DNA, 1X reaction buffer with 15mM $MgCl_2$, 0.48 mM dNTPs, 1 unit of Taq DNA polymerase, 0.96µM forward primer, 7.2µM reverse primer, and 7.2µM M13 primer in a final volume of 12µL. M13 primer was labeled with either HEX (hexachlorofluorescein) or 6-FAM (6-carboxyfluorescein) fluorescent tags. All primers were amplified using a Touchdown PCR protocol consisting of an initial denaturation of 5 min at 95^{o}C followed by 1 cycle each of 94^{o}C for 45 sec, 65$\rightarrow$50^{o}C for 45 sec, and 72^{o}C for 1 min, then 25 cycles of 94^{o}C for 45 sec, 50^{o}C for 45 sec, and 72^{o}C for 1 min, with a final extension at 72^{o}C for 10 min.

PCR products were run on a high-resolution 3% agarose gel (GenePure HiRes, ISC Bioexpress, Kaysville, UT, USA) as well as an ABI 3730XL capillary-based gel electrophoresis sequencer (Applied Biosystems, Carlsbad, CA, USA) to compare PCR product size. Size standard for all capillary-based gel electrophoresis was GeneScan-500 LIZ (Applied Biosystems, Carlsbad, CA, USA). Data obtained from the ABI 3730XL were analyzed using GeneMarker V1.91 (SoftGenetics LLC, State College, PA, USA).

Linkage Analysis: The mapping software JoinMap 4.1 (Kyazma BV., Wageningen, Netherlands) and a LOD score $\geq$ 6.0 were used to develop the linkage map for the leaf shape gene. Chi-square ($\chi 2$) tests were performed to check for segregation distortion of markers and phenotypic data from expected Mendelian ratios.

Orthologous Mapping of the Leaf Shape Gene in the Diploid D-genome: Linked SSR and STS markers from the present genetic map and markers in the genomic region of the *L* locus from the HDC map were used to establish the orthologous genomic region in the sequenced *G. raimondii* genome. SSR sequences were obtained from the Cotton Marker Database while RFLP sequences were obtained as previously described above. Both sequence types were BLAST searched against the draft sequence of the *G. raimondii* genome. Base pair positions of the highest scoring matches on *G. raimondii* Chr02, the ortholog of *G. hirsutum* Chr15, were used to order relevant SSR and RFLP markers on the genome sequence. A comparative map correlating Chr02 of the *G. raimondii* draft genome, Chr15 of the *G. hirsutum* HDC map, and Chr15 of the *G. hirsutum* leaf shape mapping population was assembled using the Strudel software (JHI Plant Bioinformatics, Dundee, Scotland). The output map was further redrawn using MS PowerPoint to improve the resolution. The region encompassed by the closest proximal and distal markers, NAU2343 and Gh565 respectively, was annotated using the transcript, protein homolog, and gene ancestry features of the Phytozome *G. raimondii* draft genome.

Candidate Gene Mapping: Single genes known to have a major impact on leaf shape and complexity in Arabidopsis were identified in the literature. Full-length cDNA sequences of these genes were obtained from The Arabidopsis Information Resource (TAIR;

www.arabidopsis.org). The cDNA sequences were then BLAST searched against the *G. raimondii* genome (www. phtyozome.net) to analyze the gene ancestry of the corresponding Arabidopsis transcript for orthologous putative genes in *G. raimondii.* Transcript sequences of putative *G. raimondii* genes whose predicted protein sequence was most similar to a known a leaf shape modification protein were also BLAST searched against the DFCI Cotton Gene Index (http://compbio.dfci.harvard.edu/cgi-bin/tgi/Blast/index.cgi) to identify high-scoring expressed sequence tags (ESTs) and tentative contigs (TCs) in tetraploid cotton.

Results

Genetics of the Leaf Shape Trait: The phenotypic ratio of individuals within the F_2 population fit the expected 1:2:1 segregation ratio (χ^2 = 0.31, p = 0.86). This confirms the single gene nature of okra leaf shape reported in previous studies. Furthermore, the appearance of the intermediate leaf shape phenotype in the F_1 hybrid relative to the parental types indicates that alleles of the okra and normal leaf shape trait show incompletely dominant phenotypic expression in the heterozygote.

Molecular Mapping of the Leaf Shape Locus (L): Sixty-three SSR and RFLP markers genetically mapped on Chr15 in *G. hirsutum* were used to survey the polymorphism between the two parents, NC05AZ21 and NC11-2100. From the 31 SSR markers that amplified specifically in both parents, the five listed in Table 1 were polymorphic. Of the polymorphic SSRs, four (BNL2440, Gh565, DPL0402, and TMB1664) were co-dominant and one (NAU2343) was dominant for the NC11-2100 parent. From the 40 STS markers developed from the 23 RFLP sequences, only 24 amplified specifically and none were polymorphic. Of the 27 STS markers developed from the leaf shape candidate genes *Gorai.002G244000* and *Gorai.002G244200*, only one was polymorphic. This STS marker, 13-LS-195, produced distinct bands at both 172bp and 180bp in the normal leaf NC11-2100 parent but only one band at 180bp in the NC05AZ21 okra parent. Heterozygotes could be effectively scored for the 13-LS-195 marker by comparing the intensity of the 172bp peak relative to the 180bp peak (Figure 3). The five polymorphic SSR markers and one polymorphic STS marker were mapped in the F_2 population of 236 individuals (Figures 4 and 5).

Two of the polymorphic SSR markers, Gh565 and NAU2343, showed tight linkage with the leaf shape locus with Gh565 mapping 2.6cM distally and NAU2343 mapping 2.8cM proximally to the leaf shape locus. In the consensus genetic map, these flanking markers are localized toward the telomeric region of Chr15 of Upland cotton (Blenda et al. 2012). Thus, the leaf shape gene (*L*) was mapped to a 5.4cM region near the telomere on the short arm of Chr15 in *G. hirsutum*. The polymorphic STS marker 13-LS-195 designed off of the candidate gene *Gorai.002G244000* co-segregated with the leaf shape phenotype over all 236 F_2 individuals. This indicates that the 13-LS-195 marker is, at the very least, tightly linked to the leaf shape locus in Upland cotton. The 13-LS-195 marker covers the end of the second intron and the beginning of the third exon of *Gorai.002G244000*.

There were only two publicly available markers between Gh565 and NAU2343 in the HDC map, the RFLP marker A1485 and the SSR cluster CLU1520 (Blenda et al. 2012). No polymorphism between the two parents was detected in the single STS marker designed from the A1485 sequence. Similarly, no polymorphisms were detected in any of the five SSR markers that comprise CLU1520 (MGHES32, HAU2398, NAU3938, MON_CGR5326, and MON_CGR5902).

Genomic Targeting of the Leaf Shape Locus in the Diploid D-genome: SSR and RFLP marker sequences of interest with regard to the leaf shape locus were BLAST searched against the *G. raimondii* draft genome in order to establish the orthologous region in *G. raimondii* (Table 2). All markers that were used in the BLAST search had a high-scoring match within a ~2Mb region near the telomere of *G. raimondii* Chr02. The order of the markers and relative distances between them in the *G. raimondii* physical map were fairly consistent that of both the HDC map and the map developed here, although minor discrepancies did exist (Figure 5).

In the *G. raimondii* genome, the orthologous region between the most closely linked SSR markers, Gh565 and NAU2343, spanned a physical distance of 337kb. Annotation of this region identified 34 putative genes listed in Table 3. The highest scoring protein homologs in Arabidopsis were obtained for all 34 genes along with their putative functions (Table 3). Only two of these genes, *Gorai.002G244000* and *Gorai.002G244200*, showed substantial homology to genes implicated in leaf morphogenesis in Arabidopsis. These two genes are only 31.5kb apart in *G. raimondii* and separated by a single gene, a serine/threonine protein kinase. The predicted proteins from these two genes are most similar to one another (61.8 / 70.4% amino acid identity/similarity) and both showed high sequence similarity to the Arabidopsis meristem identity regulator *LATE MERISTEM IDENTITY1* (*LMI1*), also known as *ARABIDOPSIS THALIANA HOMEOBOX 51* (*ATHB51*). If not for a 1bp deletion in *Gorai.002G244000* that confers a frameshift mutation and premature stop codon, the amino acid identity/similarity of the two proteins would be 70.8 / 81.8%. *LMI1* genes carry two functional domains; a homeodomain (HD) for DNA binding and a leucine zipper (ZIP) for protein-protein interactions and are thus classified as HD-ZIP transcription factors. *LMI1* and its paralogs in related model species have a well-established role in leaf shape determination in model species (Saddic et al. 2006, Vlad et al. 2014, Sicard et al. 2014).

Candidate Gene Mapping: Genes known to influence leaf shape in Arabidopsis are summarized in Table 4. Analysis of homologous genes in *G. raimondii* identified 47 genes, seven of which were located on Chr02 (Table 4). Only *DPA4* and *KNATM* from Arabidopsis did not have a high scoring homolog in *G. raimondii*. Additionally, the miRNAs *miR164* and *miR319* could not be analyzed since their sequences are too short for BLAST and the *G. raimondii* genome has not yet been annotated for miRNAs.

Of the seven leaf shape gene homologs located on Chr02, four (*Gorai.002G054200*, *Gorai.002G064500*, *Gorai.002G067300*, and *Gorai.002G113300*) mapped to the long arm of Chr02 at physical positions 4.80, 7.55, 7.85, and 16.00Mb respectively, which is opposite the short of *G. raimondii* Chr02 proposed to carry the leaf shape locus in *G. hirsutum* (Table 4).

These genes were most similar, based on amino acid similarity, to the Arabidopsis leaf morphogenesis genes *YAB5*, *TCP4*, *CUC2*, and *CUC3* respectively (Table 4).

Of the remaining three, only *Gorai.002G244000* and *Gorai.002G244200* were within the 337kb physical region delineated by the flanking markers Gh565 and NAU2343 on Chr02 of *G. raimondii* (physical position 60.69 – 61.03Mb). As stated above, *Gorai.002G244000* and *Gorai.002G244200* are in close physical proximity and most similar to each other as well as the known leaf shape modification gene *LMI1*. The remaining candidate gene on *G. raimondii* Chr02 lays relatively close to, but outside of, the leaf shape locus region established here. This gene, *Gorai.002G226700* at physical position 58.35Mb, is most similar to the *BOP2* gene in Arabidopsis.

Based on both the annotation of genes in the delineated orthologous genomic region and in silico mapping of the Arabidopsis leaf shape candidate genes in the diploid D-genome only two genes, *Gorai.002G244000* and *Gorai.002G244200*, were found to be likely candidates for the leaf shape locus. The STS marker 13-LS-195 developed from the candidate gene *Gorai.002G244000* showed complete co-segregation with the leaf shape phenotype in the F_2 population suggesting this marker was tightly linked to the leaf shape locus. While this within-gene polymorphism established *Gorai.002G244000* as a stronger candidate gene, it does not definitively eliminate *Gorai.002G244200* from consideration.

Discussion

Okra leaf shape and its allelic variants in cotton have been widely assumed to be under single gene control that is considered incompletely dominant in nature. A F_2 population derived from the cross of an okra leaf breeding line to a normal leaf landrace accession confirms previous reports that leaf shape is controlled by a single nuclear gene and that alleles of okra and normal leaf shape show an incompletely dominant phenotypic expression in the heterozygote.

Five polymorphic markers showed varying degrees of linkage to the leaf shape locus on Chr15. The two closest markers, Gh565 and NAU2343, mapped 2.6cM distally and 2.8cM proximally, respectively. Only two other publicly available SSR markers mapped between Gh565 and NAU2343, neither of which were polymorphic. Comparison of the *G. hirsutum* marker sequences to the *G. raimondii* draft genome indicates that the physical distance of the 5.4cM region between Gh565 and NAU2343 is only ~337kb and contains 34 putative genes. Thus, the leaf shape gene is localized to a 337kb region on Chr15 of tetraploid cotton. Co-segregation of the candidate gene *Gorai.002G244000*-based STS marker 13-LS-195 with the leaf shape phenotype confirmed the orthologous region of the leaf shape gene between *G. hirsutum* and *G. raimondii.* That 13-LS-195 detects an 8bp polymorphism and in part covers the beginning of the third exon of *Gorai.002G244000* indicates that a frameshift mutation may be present in one of the two leaf shapes. Further, the genetic order of the co-segregating STS maker and flanking SSR markers of the leaf shape gene in *G. hirsutum* was consistent with the physical order of their orthologous sequences in the *G. raimondii* genome. This suggests that major rearrangements likely do not exist in the region delineated by the flanking SSR markers.

The leaf shape locus maps between 60.69 and 61.03Mb on *G. raimondii*'s Chr02, the ortholog of *G. hirsutum* Chr15 (Blenda et al. 2012). The end of *G. raimondii* Chr02 is projected at 62.8Mb, indicating that the leaf shape locus lies near the telomere and that its encompassing region is likely subject to high rates of recombination. Recombination was found to be exceptionally high in cotton (Rong et al. 2004) and most of the recombination was also reported to be localized to the gene-rich regions toward the distal or telomeric ends in plants (Werner et al. 1992, Gill et al. 1996). Therefore, the large genetic distance, but small physical distance, may be a result of the region falling in a recombination-rich telomeric region on Chr15. As recombination is a prerequisite for the map-based cloning of traits in plants (Martin et al. 1993, Faris et al. 2003, Yan et al. 2004, Lin et al. 2012), the mapping of the leaf shape locus to a gene and recombination-rich region of the cotton genome should make the genetic dissection of the leaf shape trait in tetraploid cotton amenable to a map-based cloning approach.

The publication of a high-density consensus (HDC) genetic map in cotton of AFLP, RFLP, and SSR markers allows for the comparison of the mapping studies of the major leaf shape gene (Blenda et al. 2012). Jiang et al. (2000) placed the major leaf shape gene in a confidence interval spanning the RFLP markers pAR1001 and pAR019 that corresponds to a genetic distance of 54cM in the HDC map. Song et al. (2005) found only a QTL of minor effect on lobe width in this region between the SSR markers BNL2440 and JESPR152. These two markers are separated by 14cM in the HDC map and lie entirely within the confidence interval from Jiang et al. (2000). A third interspecific mapping population placed the major leaf shape QTL between the SSR BNL2440 and the AFLP E43M52.M326.0 (Lacape et al. 2013). This corresponds to a 48cM distance in the HDC map. Here, the major leaf shape gene has been mapped between the SSR markers Gh565 and NAU2343. This corresponds to a genetic distance of only 4cM in the HDC and lies completely within the closest SSR markers previously associated with the leaf shape character in this region. Furthermore, additional markers placed within the QTL confidence intervals in previous reports (A1485 from Jiang et al. 2000 and BNL1693 (CLU16), MGHES32 (CLU1520), and M151 from Lacape et al. 2013) are in very close proximity in the HDC map to the closest flanking markers identified here.

Subsequently, an independent investigation has placed the major leaf shape gene in an 8.8cM region between two single nucleotide polymorphisms (SNPs) on Chr15 of *G. hirsutum* in an intraspecific okra x normal RIL population (Zhu et al. 2014). While these SNP markers cannot be correlated to the HDC map, they span a physical distance of 881kb on Chr02 in the *G. raimondii* genome that fully encompasses the 337kb candidate region identified here. Thus, all five studies have placed the leaf shape gene in relatively the same position leaving little doubt that the major leaf shape gene of cotton resides in this relatively narrow area of the genome.

Two genes, *Gorai.002G244000* and *Gorai.002G244200*, whose coding sequences predict proteins with high amino acid similarity to the known Arabidopsis leaf morphogenesis gene *LMI1* are present in a 5.4cM region shown to contain the Chr15 leaf shape locus. A STS marker developed from *Gorai.002G244000* showed complete co-segregation with the leaf shape phenotype in the F_2 population used in the current study. Furthermore, no other known leaf shape modification genes are likely present in the 5.4cM region delineated here. Taken together, the above findings and inferences indicate that modification of a *LMI1*-like gene is a strong candidate for the molecular mechanism underlying the Chr15 leaf shape locus in cotton.

Identification and molecular characterization of the genes involved in leaf shape are essential prerequisites to elucidate the molecular mechanism controlling leaf shape and permit its manipulation for cotton cultivar improvement. The following chapter will describe efforts to fine map the leaf shape locus, definitively determine the underlying gene, and identify the specific nucleotide differences between the alleles at this locus.

Figures

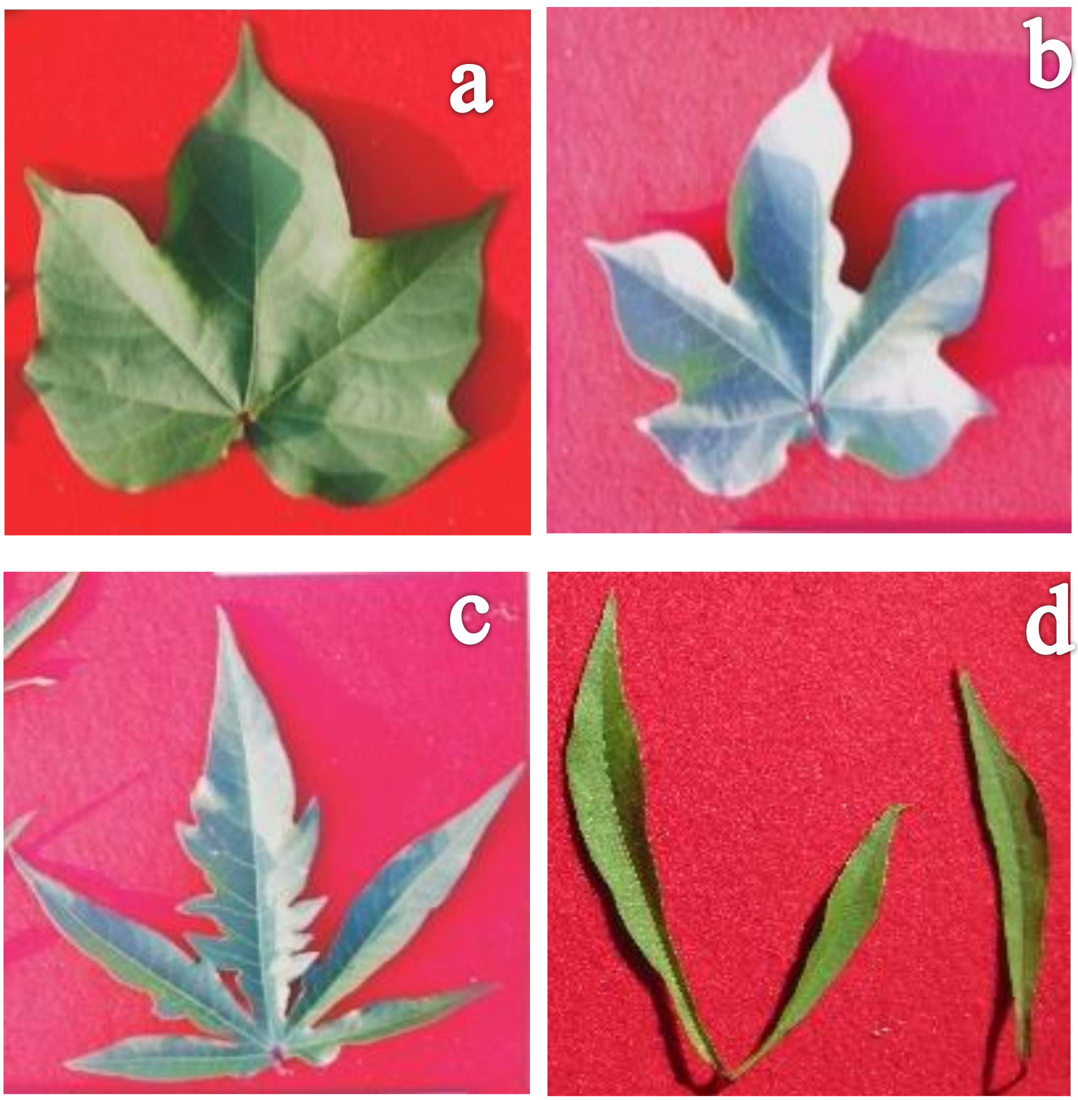


Figure 1: The major leaf shapes of Upland cotton (*Gossypium hirsutum*). **a)** Normal or broad leaf of the inbred line Texas Marker-1 (TM-1). **b)** Sub-okra leaf shape of the breeding line NC05AZ06. **c)** Okra leaf shape of the breeding line NC05AZ21 **d)** Super-okra leaf shape of the inbred line Super Okra UA 2-5.

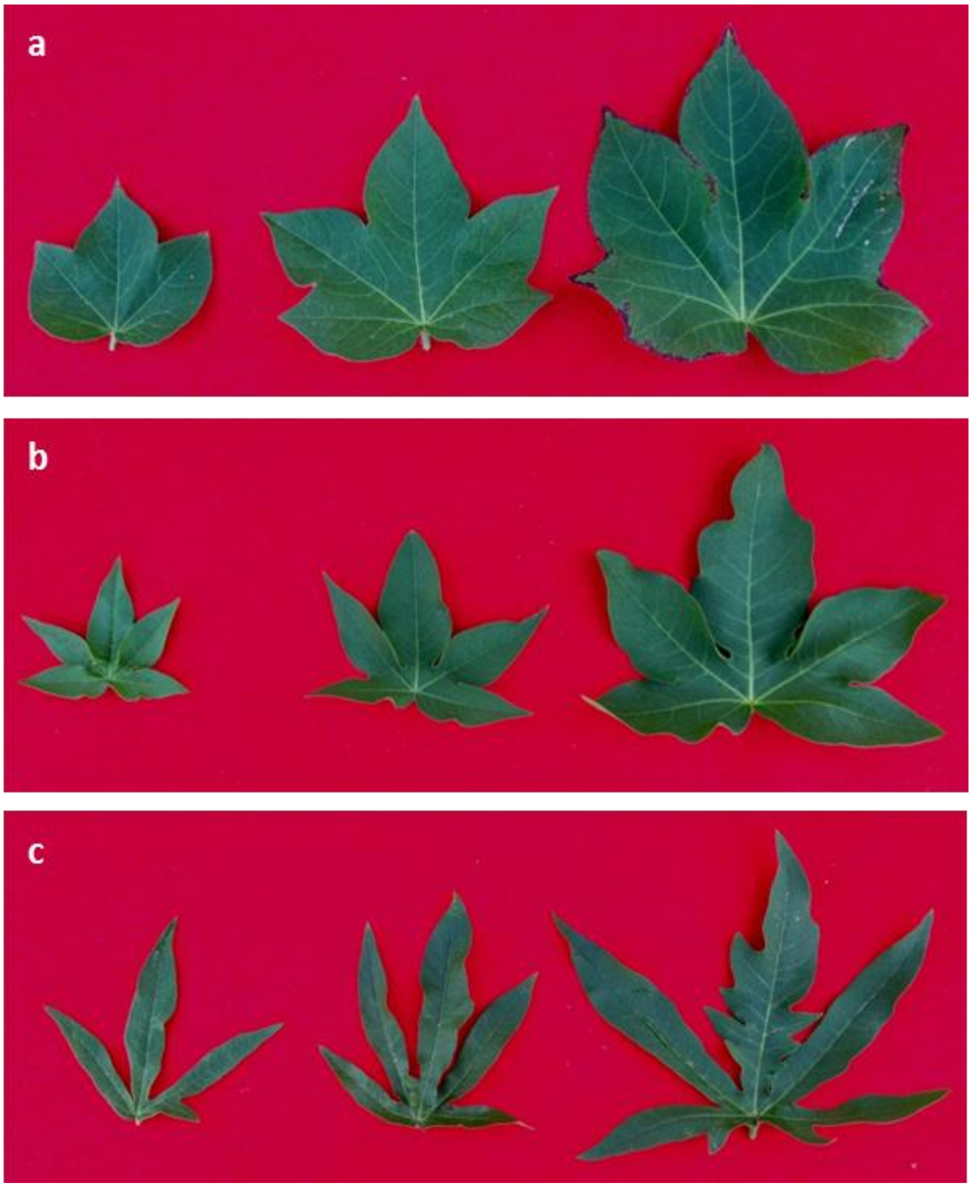


Figure 2: Leaf shape phenotypes from the upper, middle, and lower nodes of the leaf shape mapping population parents and their F_1 heterozygote. **a)** Normal shaped leaves of the landrace accession NC11-2011. **b)** Leaf shape of the F_1 heterozygote from the cross of NC11-2100 and NC05AZ21 that is intermediate between that of its parents. **c)** Okra leaves of the breeding line NC05AZ21. Note the increase in leaf lobing and the abundance of ectopic outgrowths in the rightmost leaf.

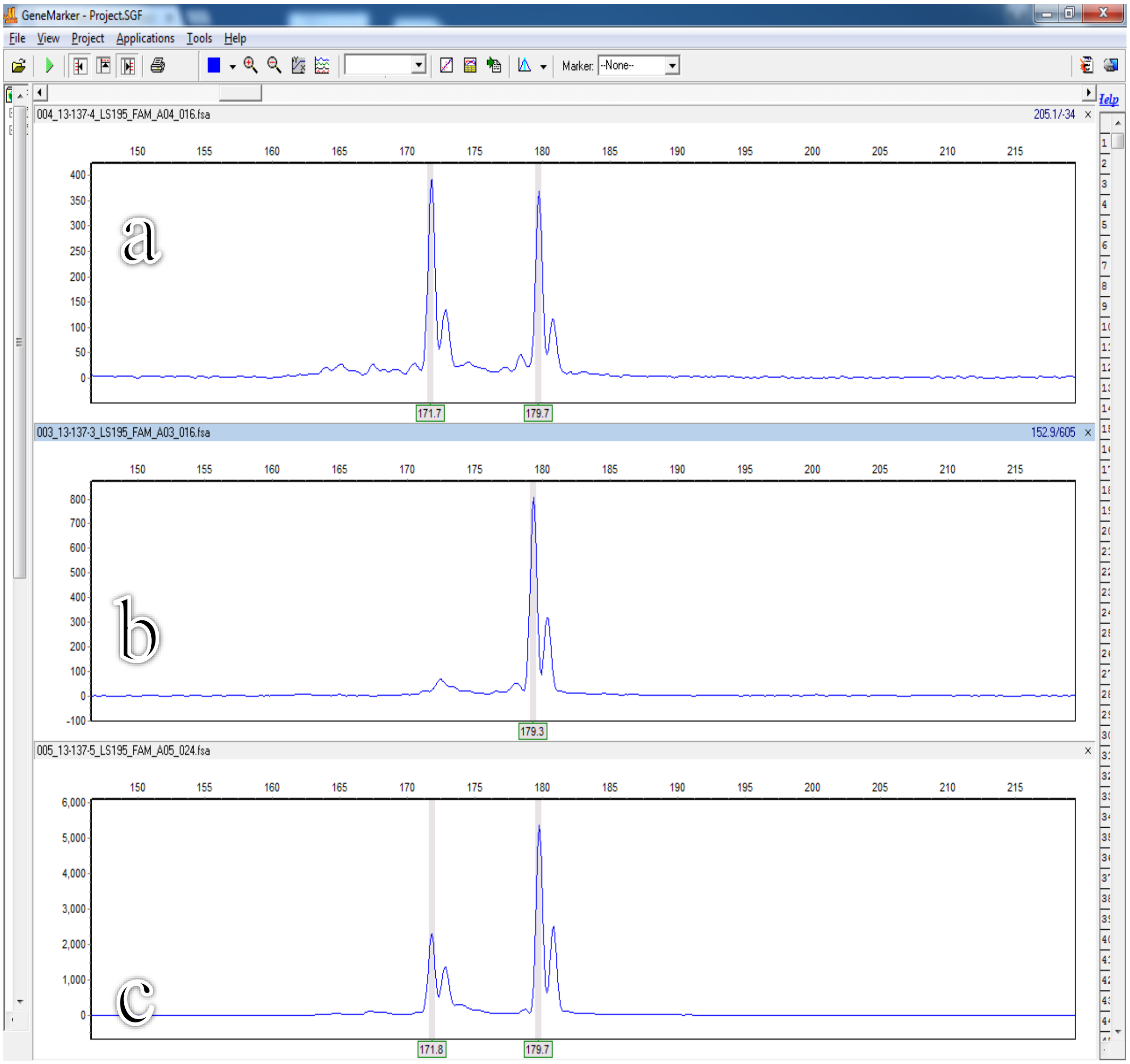


Figure 3: Scoring of STS marker 13-LS-195 in the F_2 leaf shape mapping population. X-axis is size of the PCR product in base pairs. Y-axis is the intensity of the signal. **a)** Characteristic signal of the normal leaf NC11-2100 parent. The peak at 180bp potentially represents the amplification of the A genome homeolog of 13-LS195 while the peak at 172bp is specific to the D genome. The two peaks always reach approximately the same intensity. **b)** Characteristic signal of the okra leaf parent NC05AZ21. Both the A and D genome homeologs of 13-LS-195 could be amplifying at 180bp and the 172bp band is absent. **c)** Characteristic signal of the F_1 heterozygote between NC11-2100 and NC05AZ21. Both the 180bp and the 172bp are present but the 172bp intensity reaches only ½ to 1/3 the height of the 180bp band.

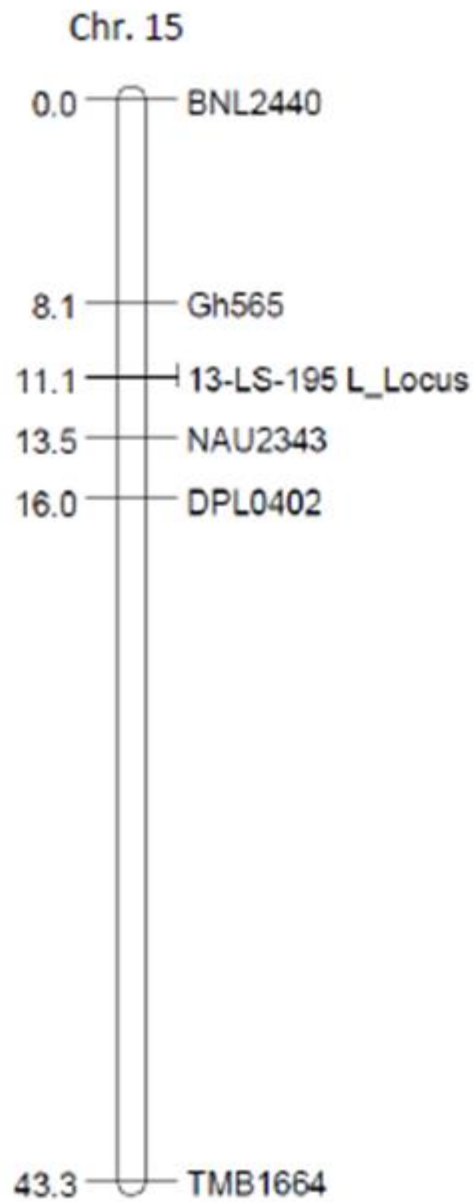


Figure 4: Molecular markers associated with the leaf shape (*L*) locus on *G. hirsutum* Chr15. Genetic distance is on the left with marker name on the right. Orientation of the map is in agreement with Blenda et al. (2012) with the top of the map toward the telomere.

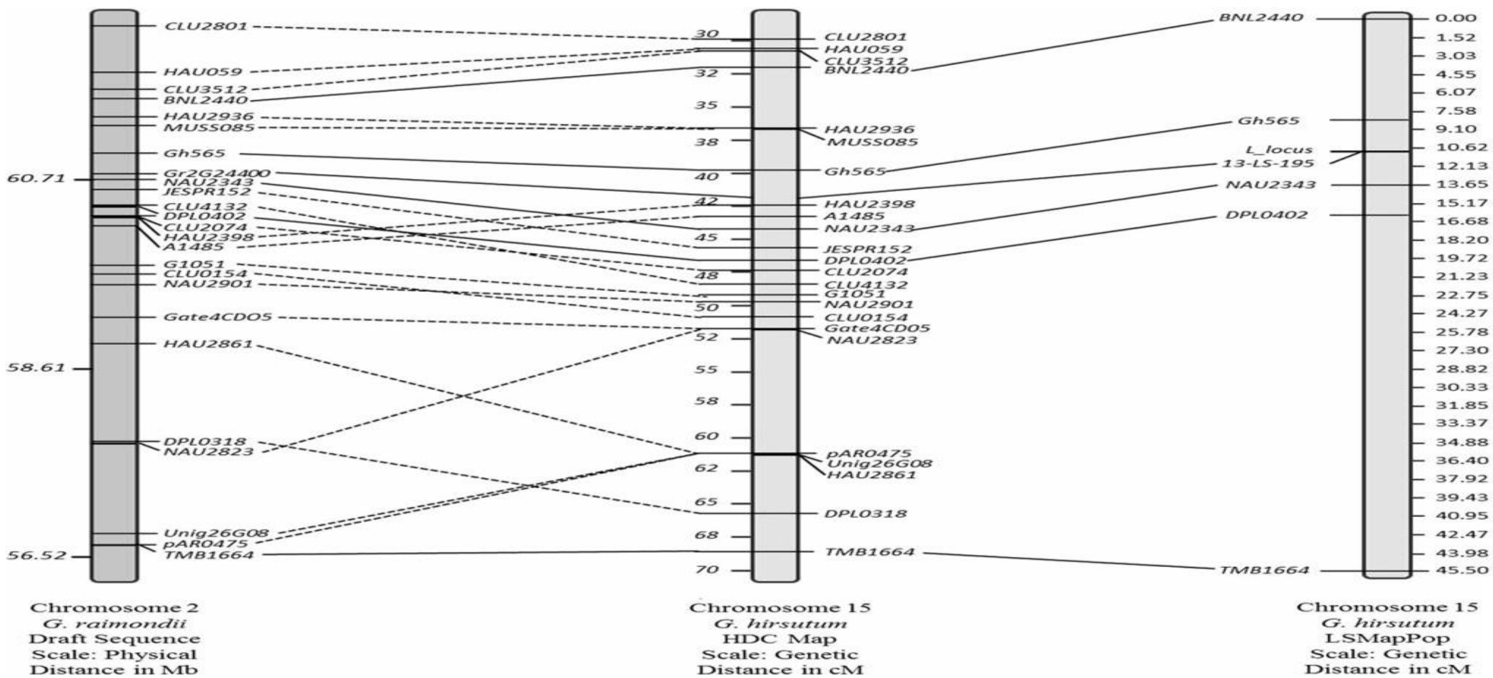


Figure 5: Orthologous relationship at the region encompassing the leaf locus of Upland cotton. Comparison between Chr02 of the *G. raimondii* draft sequence (Paterson et al. 2012), Chr15 of the *G. hirsutum* high-density consensus (HDC) map (Blenda et al. 2012) and Chr15 of the *G. hirsutum* map of the present study (LSMapPop) at the region encompassing the leaf shape locus of Upland cotton. The leaf shape gene (*L*) maps between the SSR markers NAU2343 and Gh565 in the LSMapPop. The *LMI1*-like candidate gene (*Gorai.002G244000*) based STS marker 13-LS-195 co-segregates with the leaf shape phenotype in the LSMapPop. Markers that show conserved order and relative distances across all three maps are indicated with solid lines. Dashed lines represent markers only present in the *G. raimondii* draft genome and the HDC map.

Tables

Table 1: Primer sequence and allele sizes for SSR and STS markers associated with the leaf shape gene in Upland cotton.

Marker	**Forward Primer 5'→3'**	**Reverse Primer 5'→3'**	**NC05AZ21 Allele Size (bp)**	**NC11-2100 Allele Size (bp)**
BNL2440	TGTTAAGCATACATTAGTTTCACTCG	CCGGCACCACAAAAGTAAAT	245	219
Gh565	AAAGACTCGGGTACCACCTAATC	GTCCTTCTCATTTATCTGAATTCACC	154	129
NAU2343	GCTTTGCTTTGGAATGAGAT	ATACTGCAACCCCTCACACT	266	280
DPL0402	TTACAAGCGAATTTAGGATGCC	ACTTGAGGTGCAATTGACGAG	260-263	272-275
TMB1664	AAATACCGGAACTTGATTGGG	AATTTGGTTGGGTTTCCACA	190-198	198-204
13-LS-195	ACCTTTTACGCAGGTGATGG	TCGGATATAGTCGTTTCCTGCT	180	172, 180

Table 2: Establishing the orthologous relationship between Chr15 of *G. hirsutum* and Chr02 of *G. raimondii* at the region encompassing the leaf shape locus. SSR and RFLP markers either linked to the leaf shape locus in the current study or near the locus in the HDC map were BLAST searched against the draft *G. raimondii* sequence.

Markers from Chr15	**Type**	**Orthologous relationship with *G. raimondii* genome (BLASTn)**		
		Score	**E-value**	**Physical Position on Chr02 (bp)**
BNL2440	SSR	257.4	9.4e-67	61,637,363 – 61,637,509
Gh565	SSR	342.1	3.9e-92	61,031,975 – 61,032,199
NAU2343	SSR	205.1	7.3e-51	60,694,487 – 60,699,332
DPL0402	SSR	291.6	2.8e-77	60,397,282 – 60,397,471
NAU2814 (CLU4132)	SSR	994.9	0	60,385,657 – 60,386,318
HAU2398 (CLU1520)	SSR	625.3	5.9e-177	60,297,166 – 60,301,417
A1485	RFLP	998.5	0	60,213,673 – 60,214,255
G1051	RFLP	742.5	0	59,761,625 – 59,761,830
TMB1664	SSR	224.9	1e-57	59,715,822 – 59,715,970

Table 3: Annotated gene sequences in the genomic region of *G. raimondii* Chr02 delineated by the markers closely linked to the leaf shape gene and their homology to Arabidopsis genes. Highest scoring expressed sequence tag (EST) from the Cotton Gene Index are also included. Candidate leaf shape genes are indicated with an asterisk.

G. raimondii		***G. hirsutum* EST**		**Homology with *Arabidopsis***		
Locus Name	**Physical Location**	**EST/TC**	**E-value**	**Protein**	**Similarity**	**Putative function**
Gh565 – SSR	61031975 - 61032199					
Gorai.002G246200	61027658 - 61032210	TC229763	4.4E-90	AT3G53190	66.4	Pectate lyase
Gorai.002G246100	61015799 - 61018395	TC239893	7.0E-86	ACA7	63.4	Carbonic anhydrase
Gorai.002G246000	61009256 - 61011927	TC239893	1.1E-200	ACA4	53.1	Carbonic anhydrase
Gorai.002G245900	61004759 - 61006420	TC233148	2.1E-104	TIP2	90.8	Aquaporin transporter
Gorai.002G245800	60992779 - 60993648	TC253115	1.8E-30	AT2G28200	53.6	Metal ion binding
Gorai.002G245700	60987773 - 60990298	TC262237	0.56	AT3G09890	59.3	Ankyrin repeat
Gorai.002G245600	60979680 - 60986732	TC253684	1.9E-220	PIP5K9	83.3	Phosphate kinase
Gorai.002G245500	60950960 - 60959459	TC274111	5.1E-179	AT2G36720	55.9	PH-D finger
Gorai.002G245400	60934232 - 60942403	TC274111	6.0E-162	AT2G36720	57.3	PH-D finger
Gorai.002G245300	60918534 - 60922736	ES851936	3.2E-184	AT2G36670	75.4	Aspartyl protease
Gorai.002G245200	60915361 - 60918548	CO494244	2.8E-54	AT3G10060	68.2	Isomerase
Gorai.002G245100	60909449 - 60912281	TC265610	5.8E-155	AT5G22810	76.1	Lipase/Acylhydrolase
Gorai.002G245000	60902520 - 60910968	DR462456	6.7E-99	UPL7	73.4	Ubiquitin transferase
Gorai.002G244900	60889583 - 60893220	TC232237	5.9E-174	UGE5	89.9	Epimerase
Gorai.002G244800	60882963 - 60886378	TC237927	0	ADSS	83.6	Adenylosuccinate synthetase

Gorai.002G244700	60878611 - 60881268	CO089290	5.7E-192	EMB166	68.0	Pentatricopeptide repeat
Gorai.002G244600	60867000 - 60871561	TC244038	1.4E-177	AT5G03795	62.8	Exostosin
Gorai.002G244500	60863949 - 60866523	TC269303	6.1E-221	AT5G03795	43.7	Exostosin
Gorai.002G244400	60856168 - 60863202	CO090499	1.1E-173	SRPK4	64.0	Ser/Thr protein kinase
Gorai.002G244300	60852082 - 60854077	TC239774	3.6E-167	STV1	58.5	Ribosomal protein L24e
Gorai.002G244200	60848207 - 60849955	TC252073	9.8E-166	LMI1	55.5	HD-Zip transcription factor*
Gorai.002G244100	60818675 - 60821101	TC277482	1.5E-39	RLK1	44.6	Ser/Thr protein kinase
Gorai.002G244000	60816695 - 60817565	TC252073	2.0E-66	LMI1	42.4	HD-Zip transcription factor*
Gorai.002G243900	60802515 - 60804472	TC252850	1.7E-120	AT3G10040	59.5	Trihelix transcription factor
Gorai.002G243800	60793342 - 60798916	TC239482	1.2E-275	AT3G52990	95.4	Pyruvate kinase
Gorai.002G243700	60784258 - 60789349	TC244120	0	AT3G52990	89.4	Pyruvate kinase
Gorai.002G243600	60775425 - 60777957	ES804942	3.2E-215	AT2G36570	75.2	Ser/Thr protein kinase
Gorai.002G243500	60763514 - 60767171	TC253760	5.1E-131	AT5G04550	63.9	Unknown
Gorai.002G243400	60739751 - 60745494	TC236797	1.2E-273	RAT4	89.6	Glycosyl transferase
Gorai.002G243300	60734924 - 60737571	ES803238	4.5E-220	AT5G12100	63.5	Pentatricopeptide repeat
Gorai.002G243200	60733892 - 60734923	TC238398	1.5E-188	AT3G52960	74.0	Redoxin
Gorai.002G243100	60725536 - 60729527	TC259411	0	LOS2	94.4	Enolase
Gorai.002G243000	60721265 - 60724366	TC274400	3.5E-243	HDT3	20.8	Histone deacetylase
Gorai.002G242900	60700190 - 60718341	DR452531	3.8E-111	FK	78.1	Ergosterol biosynthesis
NAU2343 - SSR	60694487 - 60699332					

Table 4: Physical location of *G. raimondii* genes most similar to the major Arabidopsis leaf shape genes.

***G. raimondii* Gene**	**Chr**	**Position (Mb)**	**Putative Class**	**Arabidopsis Homolog**	**% Similarity**
Gorai.002G244000	2	60.81	HD-Zip	LMI1	42.4
Gorai.002G244200	2	60.84	HD-Zip	LMI1	55.5
Gorai.009G409600	9	61.50	HD-Zip	LMI1	31.0
Gorai.004G236400	4	57.47	Class I KNOX	KNAT2	62.6
Gorai.005G098100	5	15.05	Class I KNOX	KNAT1	60.0
Gorai.005G180500	5	52.99	Class I KNOX	KNAT6	84.7
Gorai.007G306500	7	52.00	Class I KNOX	KNAT6	52.8
Gorai.008G296800	8	56.95	Class I KNOX	KNAT6	68.0
Gorai.009G181500	9	13.95	Class I KNOX	STM	61.5
Gorai.009G223200	9	17.38	Class I KNOX	KNAT1	67.5
Gorai.009G336900	9	36.51	Class I KNOX	KNAT6	73.8
Gorai.010G029000	10	2.37	Class I KNOX	KNAT1	72.0
Gorai.010G183800	10	53.31	Class I KNOX	STM	62.0
Gorai.011G011700	11	0.82	Class I KNOX	STM	64.4
Gorai.013G129400	13	33.74	Class I KNOX	KNAT6	52.6
Gorai.012G104100	12	23.31	MYB Domain	AS1	61.0
Gorai.007G350300	7	58.03	LOB Domain	AS2	67.6
Gorai.002G064500	2	7.55	Class I TCP	TCP4	61.7
Gorai.006G009800	6	2.18	Class I TCP	TCP13	46.0
Gorai.008G199700	8	48.46	Class I TCP	TCP17	38.5
Gorai.009G153900	9	11.75	Class I TCP	TCP2	43.9
Gorai.009G373000	9	50.50	Class I TCP	TCP4	65.8
Gorai.011G046000	11	3.52	Class I TCP	TCP2	44.7

Gorai.012G048500	12	6.24	Class I TCP	TCP13	31.1
Gorai.013G172800	13	45.97	Class I TCP	TCP4	63.6
Gorai.001G042700	1	4.03	YAB Family	YAB5	82.0
Gorai.001G174800	1	25.60	YAB Family	FIL	74.5
Gorai.001G127000	1	16.01	YAB Family	YAB2	78.7
Gorai.002G054200	2	4.80	YAB Family	YAB5	72.5
Gorai.006G117800	6	36.72	YAB Family	YAB2	79.5
Gorai.007G089700	7	6.52	YAB Family	FIL	71.3
Gorai.008G236500	8	52.22	YAB Family	FIL	78.7
Gorai.009G016700	9	1.31	YAB Family	YAB5	73.1
Gorai.009G367100	9	49.18	YAB Family	YAB5	66.7
Gorai.010G062900	10	7.84	YAB Family	YAB5	77.8
Gorai.002G067300	2	7.85	NAC Domain	CUC2	71.0
Gorai.007G323900	7	54.20	NAC Domain	CUC2	69.7
Gorai.013G171300	13	45.79	NAC Domain	CUC2	69.6
Gorai.002G113300	2	16.00	NAC Domain	CUC3	49.5
Gorai.009G186000	9	14.30	NAC Domain	CUC3	46.1
Gorai.002G226700	2	58.35	BTB/POZ Domain	BOP2	78.9
Gorai.006G133600	6	38.79	BTB/POZ Domain	BOP2	82.2
Gorai.005G191500	5	55.83	C2H2 Zinc Finger	JAG	57.5
Gorai.012G122600	12	28.07	C2H2 Zinc Finger	JAG	51.7
Gorai.003G018100	3	1.33	BEL-like HD	SAW1	48.1
Gorai.007G207200	7	21.77	BEL-like HD	SAW1	57.8
Gorai.008G070800	8	12.16	BEL-like HD	SAW1	60.0

References

Andres RJ, Bowman DT, Kaur B, Kuraparthy V. (2014) Mapping and genomic targeting of the major leaf shape gene (*L*) in Upland cotton (*Gossypium hirsutum* L.). *Theoretical and Applied Genetics* 127: 167-177.

Blenda A, Fang DD, Rami JF, Garsmeur O, et al. (2012). A high density consensus genetic map of tetraploid cotton that integrates multiple component maps through molecular marker redundancy check. *PLOS One* 7 (9): 1-16.

Faris JD, Fellers JP, Brooks SA, Gill BS. (2003) A bacterial artificial chromosome contig spanning the major domestication locus Q in wheat and identification of a candidate gene. *Genetics* 164:311-321.

Gill KS, Gill BS, Endo TR, Boyko EV. (1996) Identification and high-density mapping of gene rich regions in chromosome group 5 of wheat. *Genetics* 143:1001–1012.

Jiang C, Wright RJ, Woo SS, DelMonte TA, Paterson AH. (2000) QTL analysis of leaf morphology in tetraploid *Gossypium* (cotton). *Theoretical and Applied Genetics* 100:409-18.

Kuraparthy V, Bowman DT, Jones DC. (2013) Registration of NC05AZ21 okra-leaf and NC05-11 normal-leaf Upland cotton germplasm. *Journal of Plant Registrations* 7:334-338.

Lacape JM, Gawrysiak G, Cao TV, Viot C, et al. (2013) Mapping QTLs for traits related to phenology, morphology, and yield components in an inter-specific *Gossypium hirsutum* x *G. barbadense* cotton RIL population. *Field Crops Research* 144: 256-267.

Li F, Fan G, Wang K, Sun F et al. (2014) Genome sequence of the cultivated cotton *Gossypium arboreum*. *Nature Genetics* 46 (6): 567-572.

Lin Z, Li X, Shannon LM, Yeh CT, et al. (2012) Parallel domestication of the Shattering1 genes in cereals. *Nature Genetics* 44:720-724.

Martin GB, Brommonschenkel SH, Chunwongse J, Frary A, et al. (1993) Map-based cloning of a protein kinase gene conferring disease resistance in tomato. *Science* 262:1432-1436.

Meredith WR. (1984) Influence of leaf morphology on lint yield of cotton – enhancement by the sub-okra trait. *Crop Science* 24 (5): 855-857.

Paterson AH, Wendel JF, Gundlach H, Gui H, et al. (2012) Repeated polyploidization of *Gossypium* genomes and the evolution of spinnable fibers *Nature.* 492: 423-427.

Rong J, Abbey C, Bowers JE, Brubaker CL, et al . (2004) A 3347-locus genetic recombination map of sequence-tagged sites reveals features of genome organization, transmission and evolution of cotton (*Gossypium*). *Genetics* 166:389-417.

Saddic LA, Huvermann B, Bezhani S, Su Y, et al. (2006) The LEAFY target LMI1 is a meristem identity regulator and acts together with LEAFY to regulate expression of *CAUILIFLOWER*. *Development* 133: 1673-1682.

Schuelke M. 2000. An economic method of the fluorescent labeling of PCR fragments. *Nature Biotechnology* 18: 233-234.

Sicard A, Thamm A, Marona C, Lee YW, et al. 2014. Repeated evolutionary changes of leaf morphology caused by a mutation to a homeobox gene. *Current Biology* 24:1880-1886.

Song XL, Guo WZ, Han ZG, Zhang TZ. (2005) Quantitative trait loci mapping of leaf morphological traits and chlorophyll content in cultivated tetraploid cotton. *Journal of Integrative Plant Biology* 47 (11): 1382-1390.

Tel-Zur N, Abbo S, Myslabodski D, Mizrahi Y. (1999) Modified CTAB procedure for DNA isolation from epiphytic cacti of the genera Hylocereus and Selenicereus (Cactaceae) *Plant Molecular Biology Reporter* 17:249-254.

Vlad D, Kierzkowski D, Rast MI, Vuolo F, et al. (2014) Leaf shape evolution through duplication, regulatory diversification, and loss of a homeobox gene. *Science* 343: 780-783.

Wang K, Wang Z, Li F, Ye W et al. (2012) The draft genome of a diploid cotton *Gossypium raimondii*. *Nature Genetics* 44 (10): 1098-1103.

Wendel JF, Brubaker C, Alvarez I, Cronn R, Stewart JM. (2009) Evolution and natural history of the cotton genus. In: Genetics and Genomics of Cotton AH Paterson (ed.) 3-19.

Werner JE, Endo TR, Gill BS. (1992) Toward a cytogenetically based physical map of the wheat genome. *Proceedings of the National Academy of Sciences* 89:11307–11311.

Yan L, Loukoianov A, Blechl A, Tranquilli G, et al. (2004) The wheat VRN2 gene is a flowering repressor down-regulated by vernalization. *Science* 303:1640-1644.

Zhu QH, Spriggs A, Taylor JM, Llewellyn D, Wilson I. (2014) Transcriptome and complexity-reduced, DNA-based identification of intraspecies single-nucleotide polymorphisms in the polyploidy *Gossypium hirsutum* L. *Genes ,Genomes, Genetics* 4 (10): 1893-1905

CHAPTER 3: Fine Mapping and Cloning of the Major Leaf Shape Gene (*L-D*$_1$) in Upland Cotton (*Gossypium hirsutum* L.)

Abstract

Leaf shape in cotton (*Gossypium hirsutum*) is an important trait that influences yield, flowering, disease resistance, and foliar chemical application. A complete understanding of the genetic mechanism controlling leaf shape is essential for its proper manipulation to develop a cotton ideotype that maximizes yield while minimizing inputs. Previously, bi-parental F_2 and orthologous mapping were used to place the major leaf shape gene (*L-D*$_1$) in a 337kb, 34 gene candidate region near the telomere of chromosome 15. Shuttle mapping with the diploid A-genome species *G. arboreum* narrowed the leaf shape locus to a 112kb interval containing nine genes. Fine mapping using a large association mapping panel and two sets of isogenic lines resolved the *L-D*$_1$ locus to a genomic region of 52kb and four genes, *Gorai.002G243900* through *Gorai.002G244200*. Two *LMI1*-like genes, *GhLMI1-D1a* (*Gorai.002G244200*) and *GhLMI1-D1b* (*Gorai.002G244000*), were identified as possible candidates for the leaf shape trait in cotton based on work in model species. Expression analysis revealed that neither *Gorai.002G243900* nor *Gorai.002G244100* was expressed in young leaf tissue while *GhLMI1-D1a* was not differentially expressed among leaf shapes. Conversely, *GhLMI1-D1b* was upregulated ~800 fold in the young leaves of okra cotton compared to normal. Virus-Induced Gene Silencing (VIGS) of *GhLMI1-D1b* in an okra leaf variety of transformation-recalcitrant cotton reduced *GhLMI1-D1b* transcript levels by three fold and led to a reduction in the degree of leaf lobing. Sanger sequencing of *GhLMI1-D1b*

identified two notable polymorphisms: a promoter duplication in okra likely responsible for the altered expression and an exonic deletion in normal that likely conditions a frameshift mutation and premature stop codon. Interestingly, an intermediate leaf shape allele (sub-okra) lacks both the promoter duplication and the exonic deletion with an expression pattern identical to normal. A fourth extreme leaf shape (super-okra) is identical to okra in terms of sequence and expression of *GhLMI1-D1b*. Thus, fine mapping using a large F_2 population, shuttle mapping, association mapping, gene sequence and expression analysis, and complementation using VIGS unequivocally established that *GhLMI1-D1b* is responsible for the major leaf shapes of cotton. Establishing the molecular genetic basis of leaf shape variation in the current study could help utilize the leaf shape phenotypic diversity to develop superior cultivars with ideal leaf shape towards sustainable and profitable cotton production.

Introduction

Cotton (*Gossypium* spp.) is the world's most important source of natural fiber as well as a leading oilseed crop. The majority of cultivated cottons (*G. hirsutum* and *G. barbadense*) are allotetraploid species (AADD) formed from the hybridization of diploids *G. arboreum* (AA) and *G. raimondii* (DD)[1]. Genome sequences of *G. raimondii*, *G. arboreum*, and *G. hirsutum* have recently been completed, allowing the integration of physical and genetic maps to facilitate map-based gene cloning in this agronomically important species[2-6].

There exist four major leaf shapes of cotton categorized by an increasing degree of lobing (Figure 1): normal (*l*), sub-okra/sea-island (L^U/L^E), okra (L^O), and super-okra (L^S). All are considered monogenic, incompletely dominant, and allelomorphic at the *L* locus (re-named L-D_1 here) [7-12]. Okra leaf shapes have been involved in classical leaf development studies where the underlying factor has been shown to act early in leaf development and cell autonomously[13-15]. Agronomically, leaf shape has consistent effects on chemical spray penetration, boll rot resistance, lint trash, earliness, flowering rate, boll retention, and yield (Figure S1)[16-36]. Reported effects on various insect resistances, photosynthetic rate, water use efficiency, and fiber quality are not consistent across the literature[37-69]. It is hypothesized that an ideal cotton cultivar would produce a lower canopy of normal leaves before transitioning to an upper canopy of okra. Such an ideotype may combine the advantages of the two leaf shapes while minimizing their disadvantages.

The L-D_1 locus was placed on the short arm of chromosome 15-D1 (Chr15) in cytogenetic work[70-71]. This location was later confirmed by multiple QTL mapping studies[72-

[75]. The first study to map cotton leaf shape as a qualitative trait placed the locus in a 5.4cM interval near the telomere of Chr15[76]. Orthologous mapping to *G. raimondii* identified two strong candidate genes, *Gorai.002G244200* and *Gorai.002G244000.* Both showed high sequence similarity to the *Arabidopsis* gene *LATE MERISTEM IDENTITY1* (*LMI1*) and have been re-named *GhLMI1-D1a* and *GhLMI1-D1b* respectively for this study. An STS marker from *GhLMI1-D1b* co-segregated with the leaf shape phenotype in a small F_2 population further validating *GhLMI1-D1a* and *GhLMI1-D1b* as strong candidates for the leaf shape gene in cotton. Shuttle mapping utilizing the duplicate laciniate locus (*L-A₂*) from *G. arboreum* was able to reduce the candidate region to 112kb and nine genes[77].

LMI1 is a Class 1 homeodomain (HD) leucine zipper (ZIP) transcription factor (TF) with a *LEAFY* (*LFY*)-dependent but redundant role in the vegetative to reproductive switch in meristem identity [78]. *LMI1* also possesses a separate *LFY*-independent role in leaf morphogenesis as *lmi1* plants produce one to two abnormal leaflets at the base of the leaf[78]. A lost-in-*Arabidopsis* duplication of *LMI1* called *REDUCED COMPLEXITY* (*RCO*) is responsible for the compound leaves of close relative *Cardamine hirsuta* owing to regulatory diversification[79]. *RCO p*romoter modifications are also responsible for the differences between lobed *Capsella rubella* and simple-leafed *Capsella grandiflora*[80]. Thus, *LMI1*-like genes are well-established modifiers of leaf shape in model species with effects similar to those seen at the *L-D₁* locus in cotton.

The objectives of the current study were to: 1) fine map the major leaf shape locus (*L-D₁*) in cotton, 2) identify the causal nucleotide variations underlying these leaf shapes, 3) study the expression of leaf shape candidate genes, 4) genetically and functionally confirm the leaf shape candidate gene through transgenics, and 5) attempt to develop the transforming leaf shape cotton ideotype.

Materials and Methods

Fine Mapping in F_2 Population: A total of 1,027 F_2 plants derived from the original okra (NC05AZ21) x normal (NC11-2100) cross was used to identify recombinants for fine mapping the *L-D_1* locus. All individuals where grown, phenotyped, genotyped, and analyzed as described in Andres *et al.* (2014)[76].

Association Mapping Population: The association mapping population consisted of the 384 member cotton diversity panel[81] and 181 additional leaf shape mutants as classified by the USDA Cotton Germplasm Collection, College Station, TX. Ninety of these leaf shape mutants were photoperiod insensitive and grown in the field along with the cotton diversity panel. The remaining 91 leaf shape mutants were photoperiod-sensitive and grown in 10-inch single pots in the greenhouse under short-day conditions. Included in the association panel were two sets of isogenic lines: a BC_8 set that included all four leaf shapes (normal, sub-okra, okra and super okra) in the Stoneville 213 background and a BC_3 pair of normal and okra in the Stoneville 7A background[15, 82]. All plants were phenotyped and had DNA collected as described in Andres *et al.* (2014)[76]. Out of 47 STS markers from nine genes, three were polymorphic and run on the association mapping panel as described in Andres *et al.* (2014)[76]. Additionally, a single SNP was converted into a Kompetitive Allele Specific PCR (KASP) assay and analyzed on the population by the Eastern Regional Small Grains Genotyping Laboratory (Raleigh, NC, USA). Marker locations are summarized in Figure S3.

Semi-quantitative Expression Analysis: Total RNA was collected from three field-grown plants each of six varieties: NC05AZ21 (okra), NC11-2100 (normal), LA 213-63 Normal (Recurrent Parent), LA 213 Sea Island Leaf (sub-okra), LA 213 Okra, and LA 213 Super-Okra. Samples were taken approximately 90 days after planting. Leaves were taken at the earliest possible time they could reliably be distinguished from the shoot apical meristem without the help of any equipment. At this time-point, leaves were ~30-50mm in length from tip to base.

RNA was isolated using the Spectrum™ Plant Total RNA Kit (Sigma-Aldrich, St. Louis, MO, USA) following the manufacturer's instructions. Total RNA was converted to cDNA using the ImProm-II™ Reverse Transcription System (Promega) per the manufacturer's instructions. cDNA was then used as template in 50 µL PCR reactions using 1 µL cDNA template, 25 µL 2X GoTaq® Colorless Master Mix (Promega), 5 µL each of forward and reverse primer (10µM each), and 14 µL nuclease-free water (Promega). *GhLMI1* primers span introns so that true RNA products are substantially shorter than potential DNA background amplification. Cycling conditions were an initial denaturation of 95°C for 2 min followed by 35 cycles of 95°C for 45 sec, 55°C for 45 sec, and 72°C for 1 min with a final

72^oC extension step for 5 min. PCR products were visualized on 3% HiRes agarose (GeneMate) gels. *GAPDH* was used as a positive control and reference gene.

Real-Time Quantitative PCR Analysis of GhLMI1-like Gene Expression: cDNA from the LA213 isolines[83] in the preceding section was used in Real-Time qPCR. The positive control/reference gene was changed to *UBI14*[82]. Real-Time qPCR utilized 25 µL reactions consisting of 1 µL of cDNA template, 12.5 µL Power SYBR® Green PCR Master Mix (Applied Biosystems), 1.9 µL each of forward and reverse primer (2 µM), and 7.7 µL autoclaved water. Reactions were carried out in a MicroAmp® Optical 96-Well Reaction Plate with Barcode (Applied Biosystems) and MicroAmp® Optical 8-Cap Strip caps (Applied Biosystems). The ABI 7300 Real-Time PCR System (Applied Biosystems) was used for the qPCR experiments. qPCR cycling conditions were 2 min at 50^oC then 10 min at 95^oC followed by 40 cycles of 95^oC for 15 sec and 60^oC for 1 min. Following cycling, a dissociation stage of 95^oC for 15 sec, 60^oC for 1 min, 95^oC for 15 sec, and 60^oC for 15 sec was performed. Ct values from real-time qPCR were analyzed using the $\Delta\Delta$Ct method. Fold changes and their standard deviations were plotted using Microsoft Excel.

Sequencing of GhLMI1-like Genes: Sanger sequencing was used to obtain the genomic DNA sequence of *GhLMI1-D1a* and *GhLMI1-D1b* in 20 different tetraploid *Gossypium* accessions listed in Table S1. Sanger sequencing was performed by the North Carolina State University Genomic Sciences Laboratory (Raleigh, NC, USA). These 20 different accessions included five cultivars each of the four major leaf shapes of cotton: normal, sub-okra, okra, and super-okra. Included are the LA 213 isolines for normal, okra, and super okra. Genome-specific primers were developed by aligning homeologous sequences from *G. raimondii* and *G. arboreum* and targeting differences between the two donor diploid genomes. Genome-specificity of the primers was confirmed by analyzing amplification in a panel of diploid species from both genomes (Table S2). For the sequencing of the 5' end of *GhLMI1-D1a*, nested PCR was needed in order to produce a single, genome specific band suitable for Sanger sequencing. The DNA Clean & Concentrator™ kit (Zymo Research, Irvine, CA, USA) was used to purify PCR products prior to the second round of amplification in the nested PCR. Sanger sequencing results were analyzed and assembled using Sequencher 5.2.3 (Gene Codes, Ann Arbor, MI, USA) with additional alignments performed using Clustal Omega Multiple Sequence Alignment (http://www.ebi.ac.uk/Tools/msa/). Exon/intron structure of *GhLMI1*-like genes was estimated based off of the publicly available predictions for *Gorai.002G244200* and *Gorai.002G244000* (Phytozome, www.phytozome.net) and expanded to the stop codon when necessary. Protein translations were performed using the ExPASy Translate Tool (http://web.expasy.org/translate/). Renderings of *GhLMI1*-like genes were drawn with fancyGENE (http://bio.ieo.eu/fancygene/) and modified in MS PowerPoint.

Virus-Induced Gene Silencing of GhLMI1-D1b in LA 213 Okra: A 461bp fragment of the *GhLMI1-D1b* gene was amplified from cDNA derived from LA 213 Okra. This fragment, along with the pCambia0390 TRV2 pYL156 vector (provided by Dr. Rich Tuttle), was then digested with the restriction enzyme Acc65I (New England Biolabs, Ipswich, MA, USA) overnight at 37^{o}C. Digestion reactions totaled 30 µL in volume consisting of 2 µg of DNA, 10 units of Acc65I, 3 µL of 10X buffer 3.1, 0.3 µL 100X BSA, and the remainder nuclease-free water. Following the digestion, the restriction enzyme was inactivated by placing the reactions at 65^{o}C for 20 minutes.

Digested vector DNA was then de-phosphorylated with Antarctic Phosphatase (New England Biolabs) by mixing 3.4 µL 10X Antarctic Phosphatase Reaction Buffer and 1 µL Antarctic Phosphatase with the 30 µL digestion reaction and incubating at 37^{o}C for 30 minutes.

Both digested vector and *GhLMI1-D1b* fragment were separated on a 0.8% agarose gel, excised, and purified using the Zymoclean™ Gel DNA Recovery Kit (Zymo Research, Irvine, CA, USA) per the manufacturer's instructions. In order to carry out the ligation reaction, 100 ng of digested vector was combined with 15 ng of digested *GhLMI1-D1b* fragment, 2 µL 10X DNA Ligase buffer, 1 µL T4 DNA Ligase, and the remainder nuclease-free water in a 20 µL reaction that was allowed to incubate overnight at 4^{o}C.

The following morning, 5 µL of cold ligation mixture were added to a 50 µL aliquot of high efficiency NEB 10-beta competent *E. coli* cells (New England Biolabs) and allowed to sit on ice for 30 minutes. They were then transferred to a 42^{o}C water bath for 30 seconds and then placed back on ice for five minutes. Then 950 µL of SOC media was added and the solution was incubated for one hour in a shaking incubator set to 37^{o}C and 200 rpm. Using Luria Bertani (LB) plates with kanamycin selection (50 µg/mL), 80 µL was plated and grown overnight at 37°C. Insert presence was then confirmed via colony PCR and Sanger sequencing of PCR product. A positive selection was renamed TRV2:*GhLMI1-D1b*.

Cotton plants were agro-inoculated as described in Tuttle et al. (2012)[84]. Plants were grown under greenhouse conditions in the NCSU Phytotron. Starting at three weeks post-inoculation plants were photographed weekly. At four weeks post-inoculation, leaves of tissue stage B (described above) were collected from each plant. Real-time qPCR was performed on these leaf samples as described above. VIGS experiments were successfully repeated twice more in a growth room.

Phylogenetic Analysis of LMI1-like Genes in Gossypium: In addition to the sequencing of *GhLMI1*-like genes in tetraploid *Gossypium hirsutum, GhLMI1-D1b* was sequenced in lobed D-genome diploid species *G. thurberi* (PIs 530766 and 530789) and *G. trilobum* (PI 530967). This was done in order to facilitate comparison to the publicly available *GhLMI1*-like sequences in the broad leaved diploid *G. raimondii* and those of the *G.hirsutum* leaf shapes developed here. Alignment and phylogeny were performed using Clustal Omega. Helix-turn-helix prediction was carried out using https://npsa-prabi.ibcp.fr/cgi-bin/primanal_hth.pl Leucine zipper was predicted using http://2zip.molgen.mpg.de/.

GhLMI1-D1b cDNA isolated from the okra leaf mapping population parent NC05AZ21 was sequenced via Sanger sequencing as described previously. This was done in order to confirm the predicted exon/intron structure of *GhLMI1-D1b* and to confirm the exonic placement of the 8bp deletion. Analysis, assembly, and alignment of the sequence were as described previously for the genomic sequencing of the *LMI1*-like genes. Extraction of RNA and conversion to cDNA are described under the semi-quantitative RT-PCR section.

Co-localization of GhLMI1-D1b^{Okra}-GFP in Nicotiana benthamiana: Fluorescent protein fusions of GhLMI1-D1b were generated by Gateway cloning (Life Technologies, Carlsbad, CA, USA). The GhLMI1-D1b_Okra coding sequence without the STOP codon was amplified from Okra cDNA using primers GhLMI1-D1b-Okra-TOPO-F (5' – CACCATGGATTGGGATGGCACCATTCGACCCTTT - 3') and GhLMI1-D1b-Okra-STOP-R (5' – GGGATAAGAAGGGAGTTGAA - 3'), and cloned into pENTR/D/TOPO (Life Technologies) to generate pENTR::GhLMI1-D1b-Okra-stop. LR recombination was carried out between pENTR::GhLMI1-D1b-Okra-stop and the following destination vectors: pGWB5[102], pGWB8[102], and pUBQ10-C-GFP. pUBQ10-C-GFP is a modified version of pUBC-GFP[103] but includes the full pUBQ10 promoter. The following constructs were obtained and confirmed by sequencing: 35S::GhLMI1-D1b-GFP, 35S::GhLMI1-D1b-His, and pUBQ10::GhLMI1-D1b-GFP. GhLMI1-D1b-GFP fusions driven by native promoters were generated as follows: A 1,095 bp kb fragment from "normal" genotype and 1,227 bp from Okra cotton genomic DNA was amplified by PCR. PCR products were cloned into pUBQ10::GhLMI1-D1b-GFP by digestion with PmeI and AseI and ligation to generate pNormal::GhLMI1-D1b-GFP and pOkra::GhLMI1-D1b-GFP. Expression constructs were introduced into Arabidopsis wild type Col-0 plants and the DR5::GUS line by floral dip, and transformants were selected in media containing Kanamycin or Basta, accordingly. The

35S::GhLMI1-D1b-GFP and pUBQ10::GhLMI1-D1b-GFP were introduced by transient Agrobacterium transformation into *Nicotiana benthamiana* plants carrying a RFP-Histone2B marker[104] for co-localization analysis.

A Zeiss LSM 710 confocal microscope with a 40x water objective (1.1 N.A.) was used to image fluorescence protein fusions. The excitation/emission wavelengths during acquisition were 488 nm/492–570 nm for GFP and 561 nm/588–696 nm for RFP. All of this work was kindly performed by Sang-Won Han, a Ph.D. student in Plant Biology at N.C. State University under the direction of Dr. Marcela Rojas-Pierce.

Results

Fine mapping of the $L\text{-}D_1$ locus in an Expanded F_2 Population: An additional 1,027 F_2 plants were phenotyped and genotyped using the closest, co-dominant flanking SSRs of *$L\text{-}D_1$*, Gh565 and DPL0402[76]. The phenotypic ratio showed the expected values for the single gene inheritance of okra leaf shape (Table S3). Flanking SSRs showed a genetic distance of 2.9cM (Gh565) and 3.7cM (DPL0402), similar to the preliminary mapping data (Figure 2). Of the 1,027 additional F_2 individuals analyzed, 122 recombinants were detected between Gh565 and DPL0402. The leaf shape co-segregating marker 13-LS-195 was run on all 122 recombinants. This marker continued to co-segregate with the leaf shape phenotype, confirming and strengthening the notion that 13-LS-195 is at least very tightly linked to the leaf shape locus (*$L\text{-}D_1$*) in cotton. The combined population (1,263 plants) of the original leaf shape mapping population (236 plants) with the large leaf shape mapping population (1,027 plants) produced a genetic map comparable to the populations treated separately (Figure 2). Thus, fine mapping in the large F_2 population failed to narrow the candidate region from the 112kb and nine genes determined previously[76,77].

Fine Mapping Using Association Mapping and Isogenic Lines: Of the original 565 lines planted, five failed to germinate and 16 greenhouse single pots were either heterozygous or ambiguous as to leaf shape resulting in a final panel of 544 accessions. This updated list includes 381 lines from the cotton diversity panel, 90 additional photoperiod insensitive lines, and 73 day-length sensitive accessions. Included in the panel were two sets of leaf shape isogenic lines: a BC_8 set that included all four major leaf shapes in the Stoneville 213 background and a BC_3 pair of normal and okra in the Stoneville 7A background[15,82].

Using *G. raimondii* sequence information, four novel markers within the nine gene candidate region were developed: 1) a 13bp STS marker close to the 3' end of *GhLMI1-D1a*, 2) a C→A SNP in the intron of *Gorai.002G244100* converted into a KASP assay, 3) an 8bp STS marker within *GhLMI1-D1b* (13-LS-195 from above) and 4) a large (>100bp) STS marker in the promoter region of *GhLMI1-D1b*. Novel marker development is summarized in Figure S3. The results of phenotyping and genotyping the LSAM with the three STS markers and one KASP marker are summarized in Table 1.

The 13bp STS marker fell just outside of *GhLMI1-D1a*, although it was within 300bp of the stop codon. As seen in Table 1, there was no association between this marker and the two most common leaf shapes, normal and okra. A third relatively rare size variant was detected that also showed no association with leaf shape. The complete association of super-okra leaf shape with the 458bp variant, as well as the strong association of the sub-okra leaf with the 471bp band, were likely artifacts of the small number of lines sampled with these leaf shapes and/or population structure resulting from their development.

Overall, the lack of association of this STS marker with leaf shape was sufficient to reduce the nine gene, 112kb candidate region from shuttle mapping to a four gene, 52kb region between *GhLMI1-D1a* and *Gorai.002G243900*. *Gorai.002G243900* is predicted to produce a protein with no annotated domains but is most similar in *Arabidopsis* to a trihelix transcription factor associated with hypoxia response called *HRA1*. *Gorai.002G244100* is a serine/threonine leucine-rich repeat receptor-like protein kinase most similar in Arabidopsis to the pathogen response gene *RLK1*. The two remaining genes were the leaf shape modification genes *GhLMI1-D1a* (*Gorai.002G244200*) and *GhLMI1-D1b* (*Gorai.002G244000*). The complete genetic resolution of the $L\text{-}D_1$ locus through the four molecular mapping techniques is summarized in Figure 3.

The SNP variant within *Gorai.002G244100* showed complete association with normal, okra, and super-okra leaf shape (Table 1). Whether or not the lack of association of this SNP marker with sub-okra leaf is sufficient to further reduce the candidate region awaits the result of a genome-wide association study currently in progress. Therefore, this SNP marker was unable to further narrow down the four gene, 52kb candidate region.

All 124 lines classified as either okra or super-okra possessed the larger *GhLMI1-D1b* promoter while the 419 normal and sub-okra lines carried the smaller promoter (Table 1).

Therefore, this promoter polymorphism may be responsible for the difference between okra/super-okra and normal/sub-okra by altering gene expression. Meanwhile, the 172bp variant of the STS marker within *GhLMI1-D1b* was found only in normal while only the 180bp variant was found in the other three leaf shapes (Table 1). Furthermore, this marker (13-LS-195) had co-segregated with leaf shape phenotype over 1,263 F_2 plants. Such a polymorphism within the gene likely influences gene structure and may be responsible for separating normal from the other three leaf shapes. Taken together, the complete association of two sizable polymorphisms in a known leaf shape modification gene with leaf shape phenotypes over a wide diversity of cotton germplasm provided strong evidence that modifications to *GhLMI1-D1b* are responsible for the various leaf shapes of cotton.

Expression Analysis of Leaf Shape Candidate Genes: Expression analysis was performed in order to further resolve among the four remaining candidate genes and determine if the promoter polymorphism had any effect on *GhLMI1-D1b* expression. Semi-quantitative expression analysis revealed that neither *Gorai002G243900* nor *Gorai002G244100* were expressed in young leaf tissue (Figure 4). The lack of expression of these genes in the critical tissues, combined with their lack of sequence polymorphisms from novel STS marker development, was deemed sufficient to eliminate these two genes from consideration

GhLMI1-D1a appeared equally expressed among leaf shapes while *GhLMI1-D1b* expression was detected only in leaf shapes with the larger promoter, okra and super-okra (Figure 4). qRT-PCR confirmed that expression of *GhLMI1-D1a* was not significantly different among leaf shapes while *GhLMI1-D1b* was upregulated ~800 fold in okra and super-okra compared to normal and sub-okra (Figure 5). This indicates that up-regulation and/or ectopic expression of *GhLMI1-D1b* in okra and super-okra may be responsible for these leaf shapes. In a separate test, there was no significant difference in expression of *GhLMI1-D1b* between okra and super-okra at the single time point tested. Therefore, there was no evidence to indicate what might be differentiating these two leaf shapes. However, scanning electron microscopy results from collaborators demonstrated that super-okra leaf shape is determined even early in development than okra, pointing to differences in expression at specific time points (Figure S4).

Sequencing of GhLMI1-like Genes: In order to determine the nature of the *GhLMI1-D1b* promoter polymorphism and pinpoint the location of the genic polymorphism, *GhLMI1-D1b* was sequenced in 20 varieties (five of each leaf shape) via Sanger. Comparison of these sequences among leaf shapes, set relative to sub-okra, is summarized in Figure 6.

The large *GhLMI1-D1b* promoter synonymous with okra and super-okra was caused by a 133bp tandem duplication ~800bp upstream of the transcription start codon. The implications of such a large duplication in the promoter region were not immediately clear but it is probable that this could result in the altered expression of *GhLMI1-D1b* seen earlier (Figures 4 & 5).

The genic polymorphism within *GhLMI1-D1b* was caused by an 8bp deletion near the beginning of the proposed 3rd exon that was found in all normal leaf cotton but never any of the other three leaf shapes (Figure 6). This polymorphism had co-segregated with leaf shape phenotype through 1,263 F_2 individuals and the association mapping study. The exonic location of this deletion was confirmed through Sanger sequencing of *GhLMI1-D1b* in okra leaf cDNA. Translation of the resulting coding sequence from normal leaf cotton produces a frameshift mutation 156 amino acids (aa) into the protein and a premature stop codon that truncates the protein from 228aa to 178aa (Figure 7). Neither of the conserved functional domains of an HD-ZIP transcription factor appear directly impacted by the deletion. However, the frameshift introduces multiple leucines that may disturb the characteristic spacing of the leucine zipper and alter protein-protein interactions (Figure 7).

Sanger sequencing revealed that normal *GhLMI1-D1b* was considerably different from the other three leaf shapes (Figure 6). Normal carries six unique promoter SNPs and two SNPs in the second exon, both of which cause amino acid changes. Normal also carries a 1bp deletion in the 3rd exon, an additional third exon SNP, and a SNP in the 3' region. The 1bp deletion would also cause a frameshift and premature stop codon if not for the preceding 8bp deletion. Outside of the promoter duplication sub-okra, okra, and super-okra were identical except for one promoter SNP unique to both okra and super-okra (Figure 6).

Sanger sequencing of *GhLMI1-D1a* in the same twenty lines identified two variants (Figure 8).Variant 1 was found in two of the normal leaf lines, all five okra lines, and all five sub-okra lines. Variant 2 was found in the remaining three normal leaf lines and all five of the super-okra leaf lines. There are a large number of small polymorphisms but only one, a SNP at 1,411bp, lies within the coding region of the gene and results in an amino acid change. The 13bp STS marker from association mapping that showed no association with leaf shape was shown to fall ~100bp from the end of the 3' UTR. Although this polymorphism lies just outside of *GhLMI1-D1a*, its proximity, as well as the lack of other suitable size-based polymorphisms within the gene provide strong evidence against *GhLMI1-D1a* as the leaf shape gene.

Functional Characterization of GhLMI1-D1b Using Virus-Induced Gene Silencing (VIGS) in Okra Leaf Cotton: Expression analysis indicated that elevated or mis-expression of *GhLMI1-D1b* may be responsible for okra leaf shape (Figures 4-5). Therefore, it was hypothesized that silencing of *GhLMI1-D1b* in okra would reduce transcript levels of *GhLMI1-D1b* and confer a simpler leaf shape through reduced lobing. In the VIGS experiment, a 461bp silencing fragment of *GhLMI1-D1b* included the last 29bp of the second exon, the entire 239bp of the third exon, and the first 193bp of the proposed 3' UTR from an okra variety was used. The 3' UTR was purposefully targeted for the silencing fragment in order to minimize off-target silencing of other *G. hirsutum LMI1*-like genes, especially *GhLMI1-D1a.*

Two week old okra seedlings were inoculated with various control and experimental treatments. Starting at 21 days post-infection (DPI) there was no discernible difference in leaf shape between the treatments. However, by 28 DPI the newly developing leaves of the *GhLMI1-D1b* treatment began to show a pronounced reduction in the degree of the leaf lobing compared to the other treatments. By 35 DPI these leaves had fully expanded (Figure 9). Occasionally leaves produced by plants in the *GhLMI-D1b* treatment approached classification as normal (Figures 9-10). However most were an intermediate between normal and okra, similar to their heterozygote or sub-okra. In the more normal looking leaves, the appearance and severity of ectopic outgrowths was reduced and oftentimes absent (Figures 9-10). By 42 DPI silencing had been abolished and newly forming leaves had reverted to the original okra leaf shape. At 49 DPI, plants were obtained that had briefly produced a lower canopy of normal leaves before reverting to an okra leaf shape (Figure 11). Thus, the leaf shape ideotype was created, albeit in a transient, non-heritable manner. VIGS using TRV2: *GhLMI1-D1b* confirmed that *GhLMI1-D1b* is the gene controlling leaf shape variation at the *L-D*$_1$ locus in Upland cotton. The TRV2: *GhLMI1-D1b* treatment did not appear to affect the onset or rate of flowering although specific measurements were not taken.

qRT-PCR Analysis of GhLMI1-like Expression in VIGS Experiment: At 28 DPI, a newly expanding leaf (~30-50mm in length) was collected from each plant for RNA isolation. Leaves were chosen when silencing was hypothesized to be strongest but before they were big enough to be readily apparent in weekly photographs. It is likely that some of the potentially simplest leaves were sacrificed for the expression analysis. Therefore, the reduction in leaf lobing seen in the TRV2: *GhLMI1-D1b* phenotype may have been even greater than that seen in Figure 6 had qPCR not been performed.

The transcript abundance of *GhLMI1-D1b* was reduced four fold in the TRV2: *LMI1-D1b* treatment compared to the two controls (Figure 12). This was significantly different while the difference in *GhLMI1-D1b* levels between the two negative control treatments was not. Levels of *GhLMI1-D1b* in the uninfected LA213-63 Recurrent Parent (normal leaves) were negligible in comparison as expected (Figure 12). This indicated that VIGS utilizing the TRV2: *LMI1-D1b* vector was effective in reducing the abundance of the *GhLMI1-D1b* transcript and that this knockdown effect was sufficient to reduce leaf shape complexity in okra leaf cotton. Furthermore, the levels of *GhLMI1-D1a* were unaffected by the TRV2: *LMI1-D1b* treatment (Figure 12). This demonstrates that the construct was specific to *GhLMI1-D1b* and that off-target silencing of *GhLMI1-D1a* was not responsible for the observed reduction in leaf complexity. In summary, phenotyping and transcript profiling of the VIGS silenced okra line unequivocally established that *GhLMI1-D1b* is controlling the leaf shape variation at the $L\text{-}D_1$ locus and over-expression of *GhLMI1-D1b* was responsible for the okra leaf shape in Upland cotton.

Phylogenetic Analysis of LMI1-like Genes in Diploid and Tetraploid Cotton: The high degree of similarity between sub-okra, okra, and super-okra sequences for *GhLMI1-D1b* indicate that these leaf shapes diverged recently from one another (Figure 6). The normal leaf sequence of *GhLMI1-D1b* however, was relatively distinct from the other leaf shapes with a greater number of polymorphisms. Within the 3rd exon, slightly downstream of the 8bp deletion, normal leaf *GhLMI1-D1b* carries a 1bp deletion that would also condition a frameshift and premature stop codon. This 1bp deletion is identical to the one carried by the simple leafed D genome donor *G. raimondii* in gene *Gorai.002G244000.* It is unclear what effect, if any, this 1bp deletion might have on leaf shape in the tetraploid in the absence of the preceding 8bp deletion. It also has not yet been proven that the 1bp deletion in *Gorai.002G244000* plays any role in the simple leaf shape of *G. raimondii* (Figure S2). In addition to the 1bp deletion, normal leaf *GhLMI1-D1b* shares many of its polymorphisms with *Gorai.002G244000* including four out of the six promoter SNPs, the G→A SNP in the 2nd exon, the T→A SNP in the third exon, and the 3' UTR SNP. The remaining two promoter SNPs, the C→A SNP in the second exon, and the 8bp deletion in the 3rd exon are unique to normal leaf *GhLMI1-D1b*. This indicates that the normal leaf *GhLMI1-D1b* allele is more closely related to the *G. raimondii LMI1-D1b* gene than it is to any of the leaf shapes and likely diverged from them some time ago.

GaLMI1-A1b is predicted to produce a full-length *LMI1*-like protein that is identical in length to those produced by the non-normal leaf shapes of tetraploid cotton. The leaf of *G. arboreum* is moderately lobed and most similar in shape to the sub-okra leaf of *G. hirsutum* although an okra-type leaf shape called laciniate exists in *G. arboreum* (Figure S2).

In order to further assess the variability of *LMI1-D1b* genes in *Gossypium*, *GhLMI1-D1b* was Sanger sequenced in two additional D genome diploid species, *G. thurberi* and *G. trilobum*. Both of these species have a highly lobed leaf similar to *G. hirsutum* okra leaf (Figure S2). Both *G. thurberi* and *G. trilobum* possess full-length *LMI1-D1b* genes similar to those found in the *G. hirsutum* leaf shape mutants and *G. arboreum*. Thus, while only a handful of species from the large *Gossypium* genus have been analyzed, there exists a trend that species with lobed leaves carry full-length *LMI1*-1b genes while those with simpler leaves produce truncated LMI1-1b proteins. Therefore, *LMI1-1b* genes may play a broad role in the diversity of leaf shapes seen throughout the genus *Gossypium.*

In phylogenetic analysis, the normal *GhLMI1-D1b* coding sequence (CDS) was most closely related to the *GrLMI1-D1b* coding sequence (Figure S3). Interestingly, the *LMI1-D1b* CDSs of the non-normal leaf shapes were more closely related to these two CDSs than they were to the *LMI1-D1b* CDSs of both *G. thurberi* and *G. trilobum* (Figure S3). This indicates that the source of the mutant may not be one of these lobed D genome diploids but rather a species more closely related to *G. raimondii.* As expected, *G. arboreum LMI1-A1b* is the most distantly related member of the group analyzed and formed its own subgroup.

Phylogenetic analysis of promoter sequences largely corroborated the findings with the CDSs (Figure S3). *G. arboreum* was most distantly related followed by *G. thurberi* and *G. trilobum* together in their own node. Consistent with the inferences made earlier from sequencing results, super-okra and okra where most closely related and arose within the sub-okra lineage. The only major difference between the two analyses is that normal promoter is more closely related to the other *G. hirsutum* promoters than it is to the *G. raimondii* promoter. This may be due to the accumulation of neutral mutations in non-critical parts of the promoter over evolutionary time.

Predicted protein sequence alignment of *G. arboreum LMI1-A1* (Cotton_A_00505), *G. raimondii LMI1-D1* (*Gorai.002G244200*), and both variants of *G. hirsutum LMI1-D1* indicate that all three species carry highly conserved, full-length *LMI1*-like genes at this locus. The first exon of *GaLMI1-A1* was extended 27bp (9 codons) to resolve issues in exon/intron predictions between the *G. arboreum* and *G. raimondii* genomes.

Co-localization of GhLMI1-D1b^{Okra}-GFP in Nicotiana benthamiana: Transient expression of 35S::GhLMI1-D1b-GFP and pUBQ10::GhLMI1-D1b-GFP were used to determine the subcellular localization of GhLMI1-D1b^{Okra}. GhLMI1-D1b^{Okra} -GFP was detected in the nucleus of transformed plants (Figure 13A) in a pattern that co-localized with the nuclear marker RFP-Histone2B (Figure 13B-C), indicating that GhLMI-D1b is efficiently targeted to the nucleus. Similar results were found for GhLMI1-D1b-GFP when expressed under the control of the UBQ10 promoter. A strong GFP signal is observed in the nucleus expressing pUBQ10::GhLMI1-D1b-GFP (Figure 13D) and this signal co-localizes with the RFP-Histone2B marker (Figure 13E). Co-localization of GhLMI1-D1b to the nucleus is consistent with the notion that *LMI1*-like genes are transcription factors[78]. This analysis and Figure 13 were prepared by Sang-Won Han, a Ph.D. student in Plant Biology at N.C. State University under Dr. Marcela Rojas-Pierce.

Discussion

Map-Based Cloning of the Major Leaf Shape Gene in Cotton: Previously, a combination of bi-parental F_2 and orthologous mapping placed the major leaf shape gene in a 337kb, 5.4cM candidate region near the telomere of Chr15[76]. This corresponds to a small physical-to-genetic distance ratio of 62kb/cM given the genome-wide estimates of 328kb/cM for *G.*

raimondii and 276-594kb/cM for *G. hirsutum*[3,6,85,86]. In *G. raimondii* euchromatic values were estimated at 170kb/cM while ratios for QTL have been reported as low as 17-90kb/cM in *G. hirsutum*[3,85]. Therefore, these ratios can be small in recombination hotspots. As mentioned above, the leaf shape gene has been localized near the telomere of Chr15. The orthologous region in *G. raimondii* has been identified as a recombination hotspot consistent with the notions that recombination is high both near the telomere and in cotton[6, 87-89]. Since high levels of recombination are a prerequisite for the map based-cloning of traits in plants, the location of the $L\text{-}D_1$ gene in a recombination hotspot made it particularly amenable to map-based cloning[90-93].

Shuttle mapping utilizing the duplicate laciniate ($L\text{-}A_2$) locus from *G. arboreum* and fine mapping combining association mapping and isogenic lines later resolved the leaf shape locus to a 52kb, four gene candidate region (Figure 3). Two of these four genes, *Gorai.002G244200/GhLMI1-D1a* and *Gorai.002G244000/GhLMI1-D1b*, had previously been implicated in leaf shape modification in model plants[78-80]. Semi-quantitative expression analysis revealed that the other two candidate genes (*Gorai.002G243900* and *Gorai.002G244100*) were not expressed in young leaf tissue (Figure 4). Meanwhile qRT-PCR showed no significant difference in the expression of *GhLMI1-D1a* among leaf shapes but a ~800 fold increase in the expression of *GhLMI1-D1b* in okra and super-okra (Figure 5). VIGS of *GhLMI1-D1b* in okra resulted in a reduction of both transcript level and degree of leaf lobing (Figures 9-10).

Duplications of LMI1-like Genes May Be a Major Evolutionary Driver of Leaf Shape: Recently, *LMI1*-like genes have been proposed as evolutionary hotspots with a major effect on leaf shape in the model plant species of the Brassicaceae[79-80]. In particular, the duplication and regulatory diversification of *LMI1*-like genes may play a major role in the evolution and manipulation of leaf shape in plants. The evidence presented here indicates that the major leaf shapes of cultivated cotton may be controlled by this system as well. In the model species, *LMI1*-like genes commonly exist in tandem duplications[79-80]. The duplication of *LMI1*-like genes in *Gossypium* (Figure 3) is likely unique from those experienced in the model plants as the two lineages diverged prior to the duplications in the Brassicaceae[79]. Furthermore, the duplications in the Brassicaceae exist in tandem while the one in cotton is separated by a single gene, a serine/threonine protein kinase. Therefore, it appears *Gossypium* experienced a separate *LMI1*-like duplication event, possibly indicating convergent evolution and strengthening the notion that *LMI1*-like genes may act as leaf shape hotspot genes throughout evolutionary time.

Truncations of LMI1-like Genes Result in Simpler Leaf Shapes: Sanger sequencing of *GhLMI1-D1b* in a set of twenty cultivars elucidated the two major polymorphisms that coincided with leaf shape in association mapping (Table 1). First, there was an 8bp deletion near the beginning of the third exon found only in normal leaf (Table 1, Figure 6). This 8bp deletion corresponded directly with the 13-LS-195 marker that co-segregated with the leaf shape phenotype in over 1,200 F_2 plants (Figure 1). This deletion results in a frameshift and premature stop codon in the putative protein sequence of normal leaf *GhLMI1-D1b* that may interfere with the function of the leucine zipper (Figure 7). Alternatively, the C-terminal truncation may impact protein stability by removing residues necessary for proper folding.

EMS mutagenesis has been used to knockout a *LMI1*-like gene in *C.hirsuta* that reduces leaf shape complexity[79]. This truncated protein lacked the leucine zipper motif indicating that this domain was essential for proper function. A non-functional *LMI1*-like gene from *C. rubella* also carries a C-terminal deletion truncating the protein near the leucine zipper but an *A.lyrata LMI1*-like gene with the same mutation retains functionality[79]. Thus, it is currently unclear how the cotton deletion may affect protein function other than to hypothesize non- or partially functionality based on phenotype.

Altered Expression of LMI1-like Genes Can Drastically Increase Leaf Shape Complexity: The second co-segregating polymorphism was a 133bp tandem duplication ~800bp upstream of the *GhLMI1-D1b* transcription start site (Table 1, Figure 6). The origin of this promoter duplication is currently unclear but it may have arisen via unequal crossing over or gene conversion[105]. This duplication was present in all super-okra and okra cultivars but absent from normal and sub-okra varieties (Table 1). In *C. rubella*, three large promoter insertions in an *LMI1*-like gene result in differential expression leading to increased leaf complexity[80]. Likewise, it has been demonstrated that modifications to the promoter region of an *LMI1*-like gene increased leaf complexity in *C. hirsuta*[79]. Thus, modification to the promoters of *LMI1*-like genes can cause differential expression that is sufficient to alter leaf shape.

Here it is currently unclear how the promoter tandem duplication may be able to increase gene expression. The effect of a duplicated promoter element on expression appears situation dependent in the literature[94-99]. However, it is well-established that *LMI1* plays an important role in the vegetative-to-reproductive phase change of Arabidopsis[78]. In Arabidopsis LMI1 is directly activated by LFY in the flowering pathway but LMI1 appears to act independently of LFY in regards to leaf shape. Since LFY does not appear to activate LMI1 expression to influence leaf shape, what transcription factor(s) do so is currently unknown. Investigation of the okra/super-okra promoter repeat using the JASPAR CORE Plantae database revealed that this sequence was enriched for putative binding sites of the *Arabidopsis* MADS-box transcription factor *SUPPRESSOR OF OVEREXPRESSION OF CONSTANTS1 (SOC1)*. SOC1 is known to act with AGAMOUNS-LIKE24 (AGL24) to directly regulate LFY[100]. Therefore, while a direct link between SOC1 and LMI1 has not been demonstrated previously, both genes exist in the complex pathway regulating flowering time[101] and thus may also interact to influence leaf shape.

Altered Expression of LMI1-like Genes May Also Alter Flowering Patterns: Perhaps not coincidentally, okra flowers 2-12 days earlier and 25-50% more often than normal near-isogenic lines while super-okra flowers 9-12 days earlier and 85-100% more than normal[14-18, 22, 24, 25]. This pleiotropy has historically been attributed to physiological reasons; namely that okra plants required less photosynthate to both grow new leaves and maintain existing ones and were therefore able to contribute more photosynthate to initiate flowers and to do so earlier[21]. However, the finding that okra and super-okra cotton over-express an *Arabidopsis* flowering promoter, indicates that this pleiotropy may have more of a biochemical and molecular undertone than previously thought. In agreement with results obtained in *Arabidopsis*, loss-of-function of *GhLMI1-D1b* in normal cotton may not influence flowering due to compensation by redundantly acting genes[78]. This may explain why sub-okra and normal leaf cotton flower at relatively the same time and rate. However, over-expression of *LMI1* may be sufficient to lead to precocious and prolific flowering in Arabidopsis, although this has not been tested[78]. Therefore, the over-expression of *GhLMI1-D1b* demonstrated in okra and super-okra cotton may lead to the altered expression of other flowering time pathway genes causing precocious and elevated flowering. Thus, these two leaf shapes may be able to influence flowering independently of physiology.

Cause of Super-Okra Leaf Shape is Still Unknown: Super-okra does not express either *GhLMI1-D1a* or *GhLMI1-D1b* at a higher rate than okra cotton in the one time-point analyzed here (Figure 5). However, super-okra may over-express *GhLMI1-D1b* earlier in development as supported by evidence that super-okra is determined earlier than okra (Figure S3). This differential expression may then cause super-okra to flower earlier and more often than okra. Alternatively, super-okra may be caused by undetected changes in the expression of *GhLMI1-D1a* in conjunction with the okra allele at *GhLMI1-D1b*. The close physical proximity of the two genes may have precluded an observed recombination event between them and led to the determination that their inheritance is determined by a single gene.

Proposed Evolutionary Origin of Leaf Shape Alleles in Gossypium: Based off phylogenetic analysis (Figure S4), it appears the normal allele originated from *G. raimondii LMI1-D1b* (*Gorai.002G244000*). Owing to their shared 1bp deletion, frameshift, and premature stop codon, this protein may already have been non- or partially functional. What effect the subsequent 8bp deletion, which causes an earlier frameshift and premature stop codon, had on the protein function is unclear. This newer deletion may have had no effect or made leaf shape simpler. Under this hypothesis, wild tetraploids may exist that carry just the 1bp deletion and a unique leaf shape. Sequence analysis indicates the other three leaf shape alleles did not arise within *G. raimondii LMI1-D1b.* This ancestral allele may have come from a lobed D genome diploid although not *G. thurberi* or *G. trilobum* (Figure S4).

There are some seemingly unlikely scenarios under which a full-length functional *LMI1* allele became available at the *L-D₁* locus: 1) A separate polyploidization event involving a different D genome diploid. While additional polyploidization events are generally rejected by the cotton community[1], a *G. arboreum* x *G. thurberi* cross resulted in a synthetic tetraploid with sub-okra leaf shape[12]. 2) A D genome diploid somehow hybridized with the tetraploid in order to transfer in the functional *LMI1-D1b* or 3) the D genome plant in the original polyploid event was heterozygous at this locus. It may have contained one truncated allele from *G. raimondii* and one a functional allele from another source, both of which persisted in the ensuing tetraploid lineage.

In any case, once the sub-okra allele arrived in the D genome, it likely gave rise to okra via a single mutation that resulted in the 133bp tandem promoter duplication. Consistent with expression results and gene silencing results obtained here; this promoter duplication led to ectopic and/or over-expression of *GhLMI1-D1b* and an increase in the degree of leaf lobing and complexity. The finding of only a single additional promoter SNP in the entire *GhLMI1-D1b* genic region between sub-okra and okra indicates this event happened relatively recently. Only two promoter SNPs differentiate super-okra from okra (Figure 6) consistent with the knowledge that super-okra originated from okra within the last 90 years[9].

It is Unclear if the Laciniate Leaf Shape Has Been Transferred to Tetraploid Cotton: The laciniate leaf shape of *G. arboreum* is controlled by the $L\text{-}A_2$ locus orthologous to $L\text{-}D_1$ of Upland cotton[77]. The sequenced *G. arboreum* carries two full-length *LMI1*-like genes in its leaf shape candidate region and produces a moderately lobed, “recessive broad” leaf shape. Thus, it is likely that the over-expression of at least one of these genes is responsible for the laciniate leaf shape. In any case, it does not appear that the laciniate allele is common in the tetraploid cotton germplasm pool. Eight “laciniate” lines were included in association mapping. However, all carried the *GhLMI1-D1b* polymorphisms present in okra/super-okra cotton and were phenotypically indistinguishable from them. Evaluation of F_2 populations from crosses of purported laciniate lines to other leaf shapes failed to segregate as would be expected if these lines carried the A genome mutation. Thus, they are all likely misclassifications of okra or super-okra and it is not clear if any lines currently available still retain the laciniate allele.

An Ideotype for Cotton Production: An ideal cotton cultivar would produce a lower canopy of normal leaves in order to attain canopy closure as quickly as possible. Once canopy closure was obtained, this ideal cultivar would switch over to the production of okra leaves in order to minimize shading of existing leaves, maximize allocation of resources to reproductive organs, and capture the other benefits of an open cultivar crop. Virus-Induced Gene Silencing (VIGS) of *GhLMI1-D1b* in okra leaf cotton temporarily produced a lower canopy of normal leaves before reverting to okra leaf shape as silencing was abolished (Figure 11). Thus, the ideotype was successfully created here, albeit in a transient, non-heritable manner. Since over-expression of *GhLMI1-D1b* is responsible for the formation of okra leaf shape, placement of this gene under the control of an appropriate promoter may be sufficient to stably produce the desired plant architecture. This promoter may be an inducible system such as the ethanol-based AlcR/*AlcA* system where application of the induction signal activates the expression of the target gene. An inducible system would give the grower complete control over the phase change but off-target effects of the signal and application frequency are potential drawbacks. An alternative promoter strategy would be growth stage specific, thereby precluding an induction signal. However, this system would give the grower less control and finding a suitable promoter may be difficult.

Conclusions and Future Directions: Fine mapping using a large F_2 population, an association mapping panel, sequence and gene expression analysis, and complementation using VIGS unequivocally established that *GhLMI1-D1b* is responsible for the major leaf shapes of cotton. Based off phylogenetic analysis and the nature of the polymorphisms, sub-okra is the "wild-type" or "normal" leaf shape of tetraploid cotton, rather than a mutant. Classical "normal" leaf shape is actually a mutant caused by an 8bp exonic deletion in *GhLMI1-D1b* that results in a frameshift and truncated protein that may be partially functional or non-functional. Okra leaf shape arose as a separate mutation to sub-okra *GhLMI1-D1b* resulting in a 133bp tandem repeat of a promoter element that causes ectopic and/or over-expression of the gene. Super-okra also carries the promoter duplication and differentially expresses *GhLMI1-D1b,* although what separates okra from super-okra is not currently known. This research represents the first reported map-based cloning of a gene in cotton.

Ongoing and future research includes two techniques to further validate the role of *GhLMI1-D1b* in cotton leaf shape, gamma ray mutagenesis and Arabidopsis transformations. Seeds of okra cotton were gamma irradiated with the hopes of inducing knockout mutations in *GhLMI1-D1b* that could be functionally characterized. This approach has yielded several promising segmental mutants that were self-fertilized in the summer of 2015. Attempts to over-express okra *GhLMI1-D1b in* Arabidopsis and observe changes in leaf shape are also currently underway. RNA Seq analysis is in progress with collaborators at the Danforth Plant Sciences Center to identify genes acting downstream of *GhLMI1-D1b* over-expression in okra. Yeast-One Hybrid analysis is also planned to identify binding elements of the duplicated promoter element. Establishing the molecular genetic processes underlying leaf shape and development will lead to fundamental discoveries towards developing cotton cultivars with ideal leaf and canopy architecture and improved productivity while ensuring ecological sustainability.

Figures

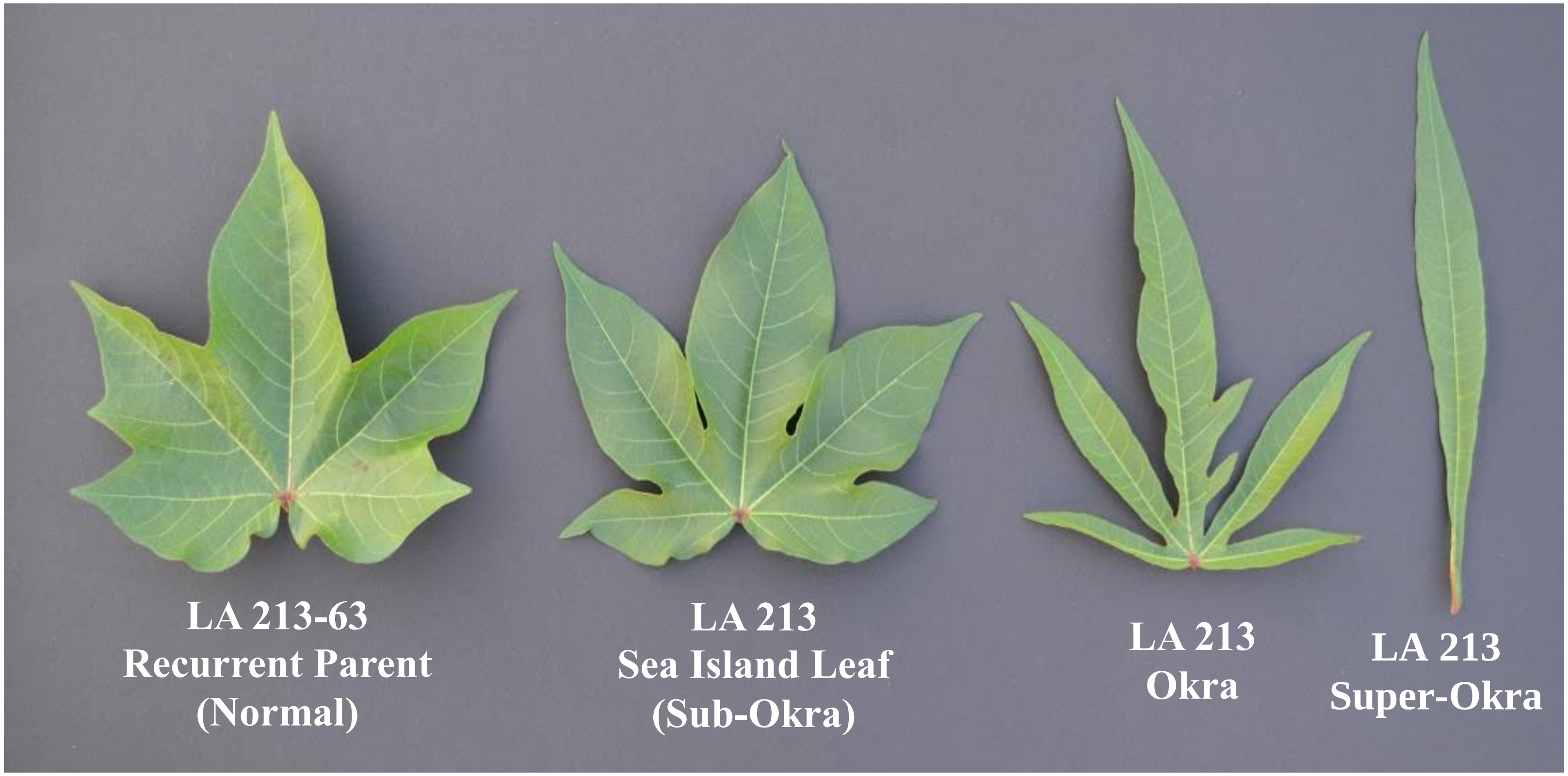
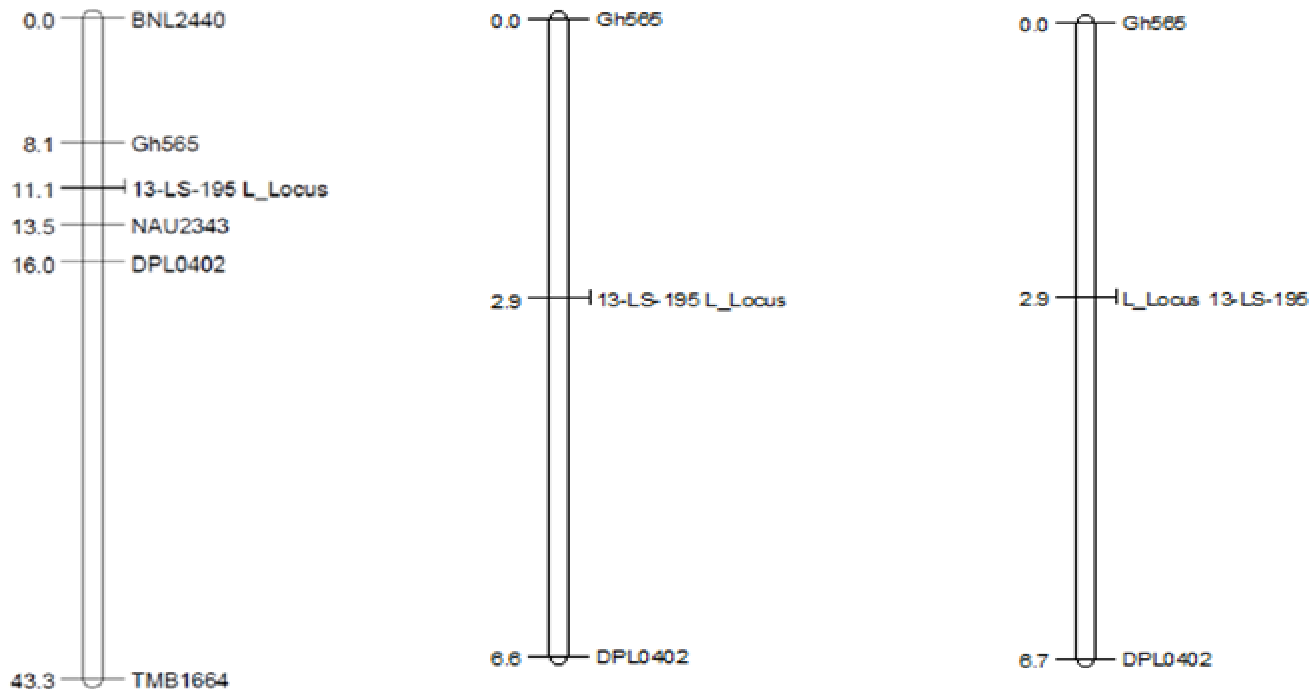


Figure 1 - Major Leaf Shapes of Cotton: Characteristic leaves taken from the four members of the BC_8 LA 213 isoline series[82].



Small Leaf Shape Mapping Population from Chapter1 Andres et al. 2014 (236 F_2 plants)

Large Leaf Shape Mapping Population (1,027 F_2 plants)

Combined Leaf Shape Mapping Population (1,263 F_2 plants)

Figure 2 - Comparison of Genetic Maps between Mapping Populations: All three maps depict *G. hirsutum* chromosome 15. Genetic distance in centi Morgans is on the left and marker names are on the right. Leaf shape mapping population parents are NC05AZ21 (okra) and NC11-2100 (normal).

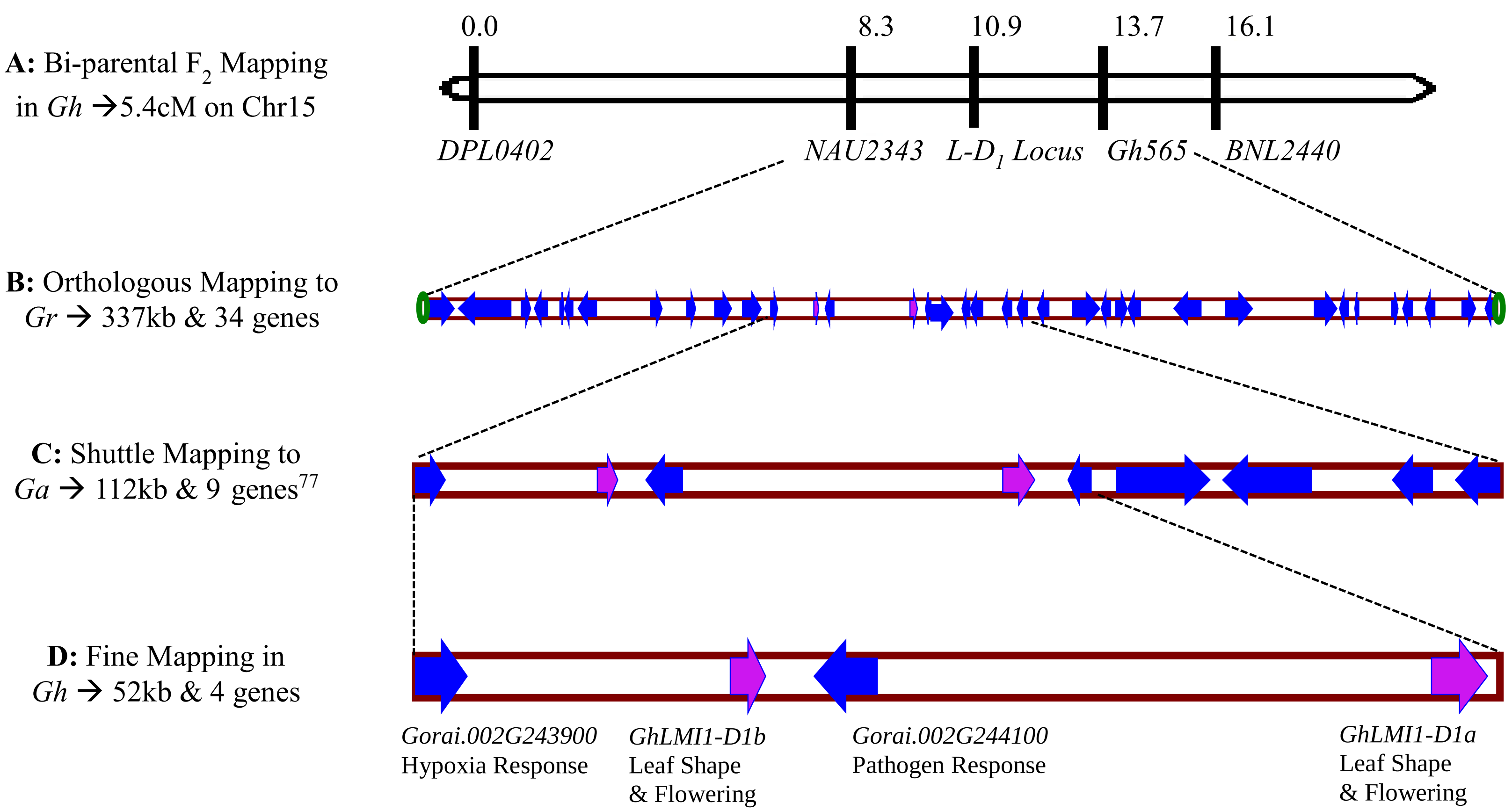


Figure 3 – Genetically Resolving the L-D_1 Locus in Upland Cotton: Four molecular mapping techniques were used to identify the leaf shape candidate gene region. **A:** Bi-parental F2 mapping in *G. hirsutum* **B:** Orthologous mapping to the sequenced D genome donor diploid *G. raimondii*. **C:** Shuttle mapping utilizing the duplicate laciniate (*L-A_2*) locus from donor diploid *G. arboreum* **D:** Fine mapping combining an association mapping panel and BC_8 isogenic lines.

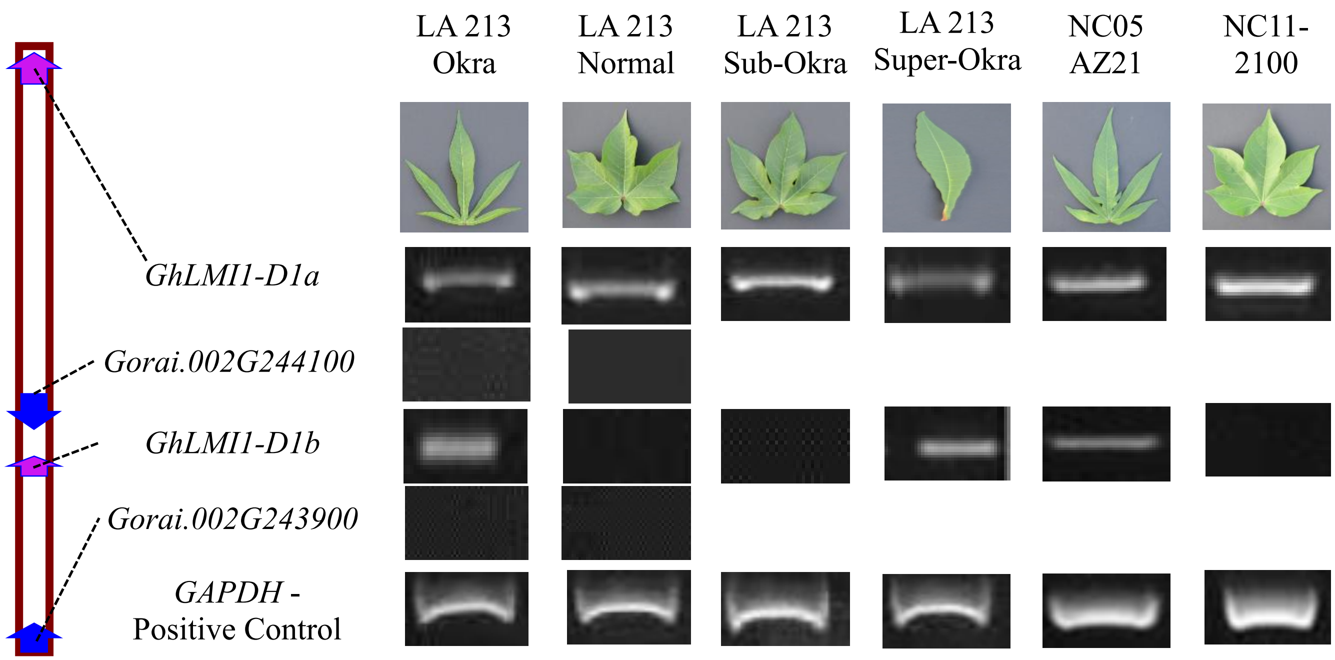
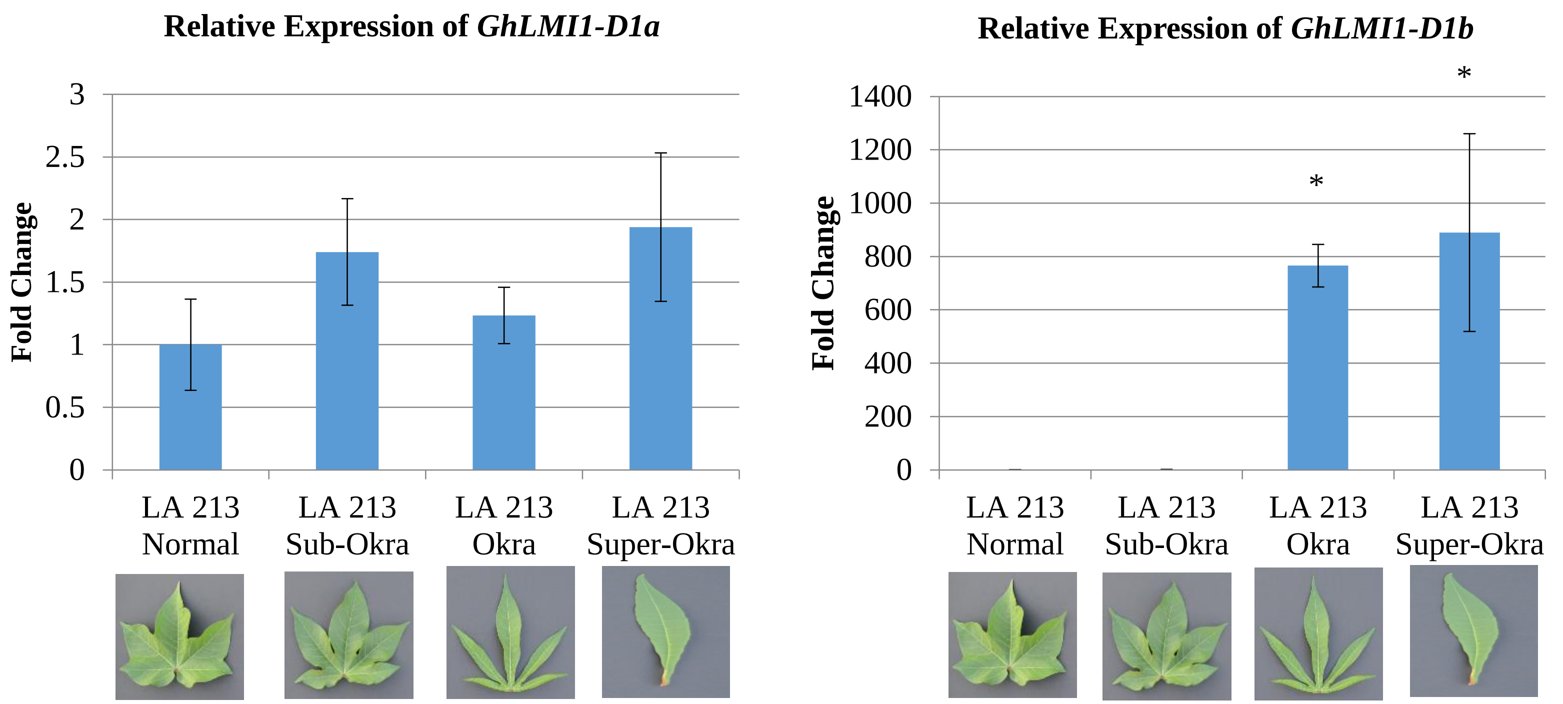


Figure 4 - Semi-quantitative Expression Analysis of Leaf Shape Candidate Genes: Neither *Gorai.002G243900* nor *Gorai.002G244100* was expressed in the critical young leaf tissue, eliminating these two genes from consideration. *GhLMI1-D1a* appeared equally expressed across leaf shapes. *GhLMI1-D1b* was expressed only in leaf shapes (okra and super-okra) with the larger promoter (Table 1) indicating differential expression of this gene plays a major role in leaf shape. *GAPDH* was the positive control and reference gene.



Data collected from three plants per genotype with three technical replicates per plant. Fold change was calculated via the ΔΔCt method with *UBI*14 as the reference gene. LA 213 Normal was used as the control treatment. Error bars represent the standard deviation of the fold change. * Represent statistical differences as determined by unpaired t-tests a $p < .05$

Figure 5 – qRT-PCR of GhLMI1-like Genes in Leaf Shape Isolines: qRT-PCR confirms that *GhLMI1-D1a* was not differentially expressed among leaf shapes. Expression levels of *GhLMI1-D1b* were shown to be ~800 fold high in okra and super-okra cotton compared to normal and sub-okra. This strengthens the notion that differential expression of *GhLMI1-D1b* is a major factor in cotton leaf shape. No significant difference in *GhLMI1-D1b* expression was shown between okra and super-okra at this time-point indicating the difference between those two leaf shapes may be spatial and/or temporal.

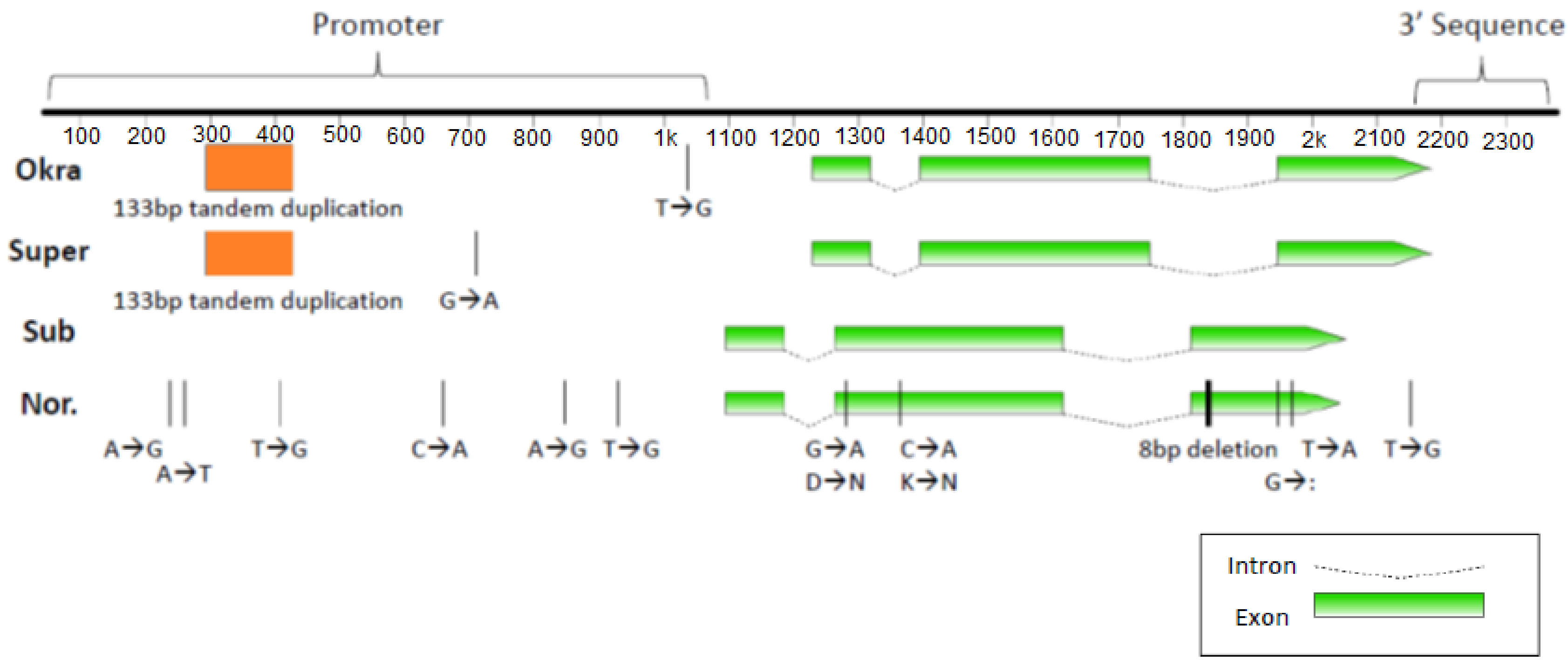


Figure 6 - Polymorphisms among GhLMI1-D1b sequences from the four major leaf shapes of G. hirsutum: Each sequence was a consensus formed from five accessions of that leaf shape. The 133bp tandem promoter duplication was found only in okra and super okra and likely responsible for the elevated expression of *GhLMI1-D1b* seen in these leaf shapes. The 8bp deletion was found only in normal (Nor.) and causes a frameshift mutation and premature stop codon. All other polymorphisms are SNPs. Sub-okra is set as the standard that the other three leaf shapes are compared.

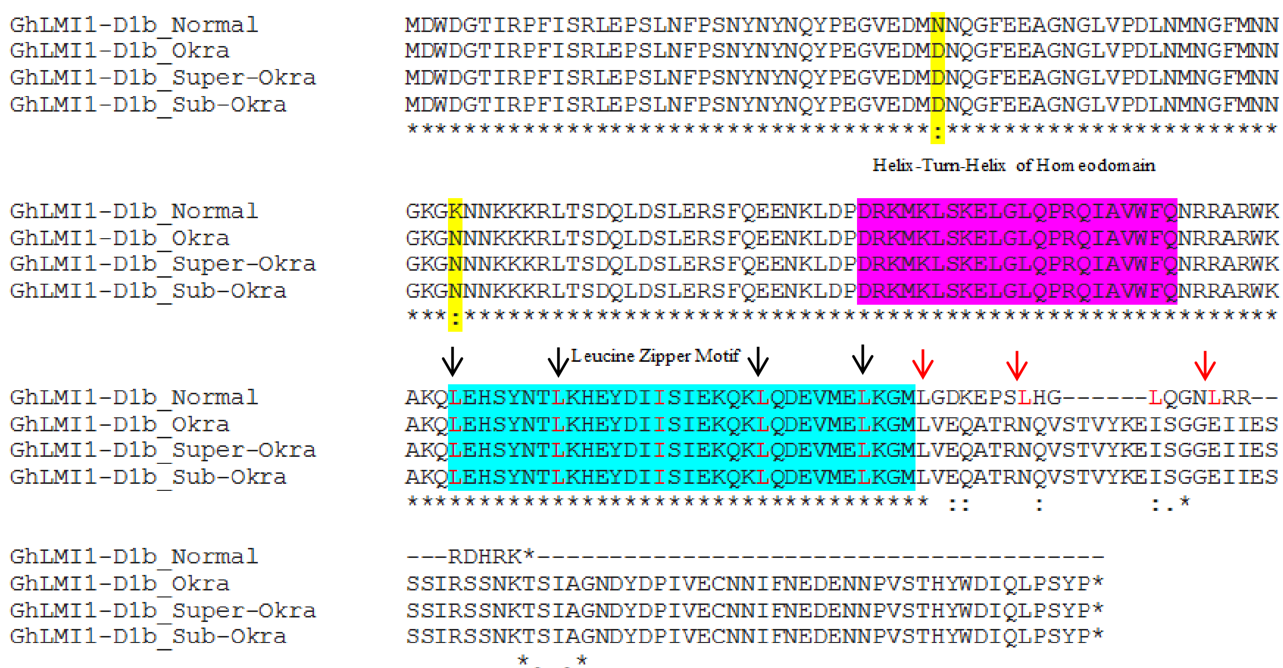


Figure 7 – Impact of Deletion on Protein Sequence of GhLMI1-D1b: Sub-okra, okra, and super-okra produce identical proteins. Normal has two amino acid substitutions highlighted in yellow. The two conserved domains of *LMI1*-like proteins, the helix-turn-helix of the homeodomain (purple) and leucine zipper motif (blue) are also highlighted. The 8bp deletion results in a frameshift at amino acid 156 causinga premature stop codon 178aa into the 228aa protein. Leucine zippers are characterized by leucine residues in every seventh position (highlighted in red with black arrows). The frameshift inserts multiple leucines at seven residue intervals (highlighted in red with red arrows) not present in the full-length protein. This alteration may interfere in protein-protein interactions.

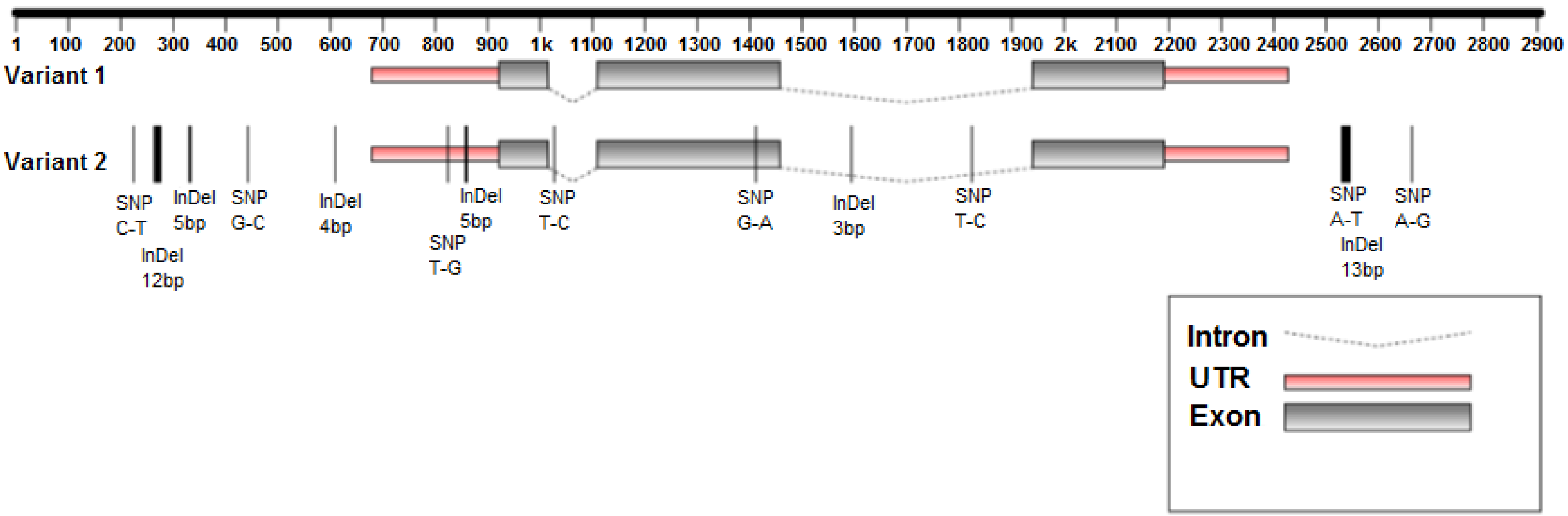


Figure 8 - Polymorphisms among GhLMI1-D1a sequences from the four major leaf shapes of G. hirsutum: Variant 1 was found in all 5 okra lines, all 5 sub-okra lines, and 2 of the normal lines. Variant 2 was found in the three remaining normal leaf lines and all 5 super-okra lines. The 13bp InDel in the 3' region corresponds to the STS marker used in association mapping that had no association with leaf shape. The only SNP within the coding sequence, the G→A at position 1,411bp in the second exon, does result in an E→K amino acid substitution. Axis along the top is in base pairs (bp).

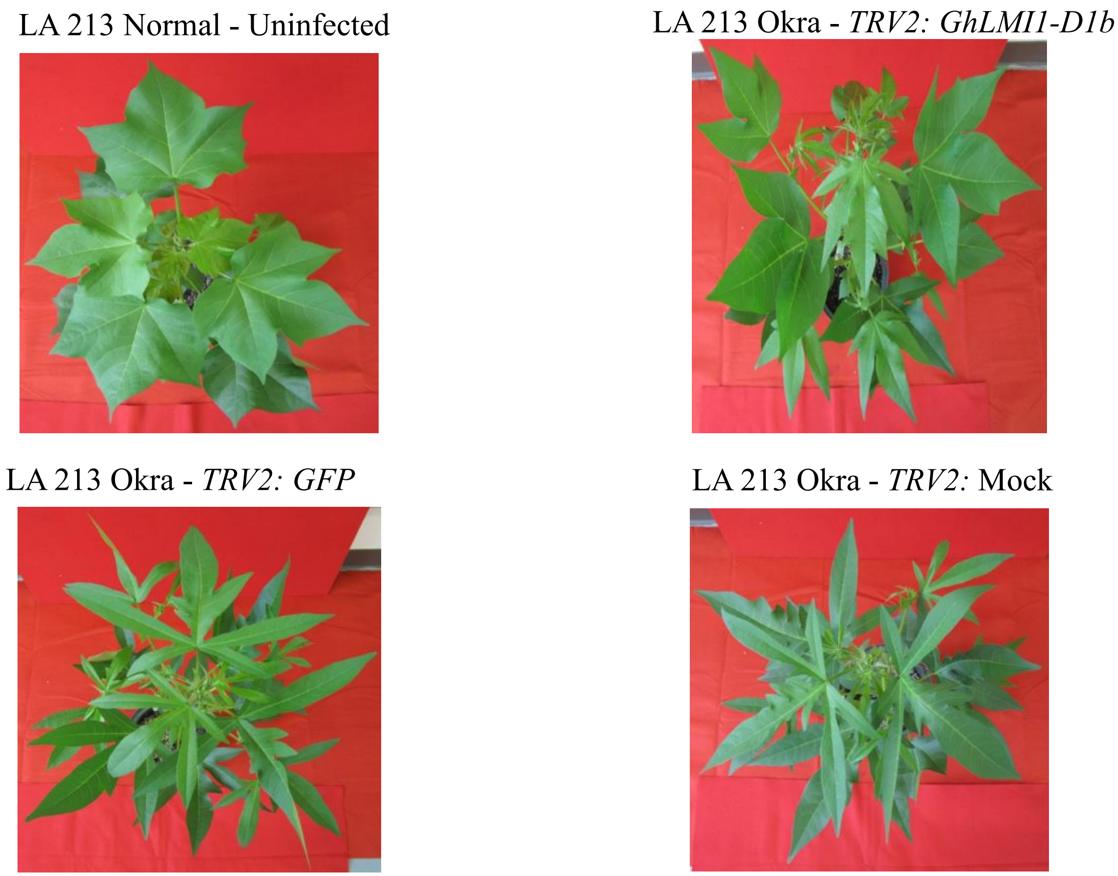


Figure 9 – VIGS Plants 35 DPI: LA 213 Normal was uninfected and included for reference. *TRV2: GFP* and *TRV2*: Mock are negative controls of LA 213 Okra that produce typical okra leaf shape. *TRV2: GhLMI1-D1b* is LA 213 Okra infected with a TRV engineered to silence *GhLMI1-D1b*. Leaf lobing and ectopic outgrowths are drastically reduced confirming that *GhLMI1-D1b* is responsible for okra leaf shape in cotton.

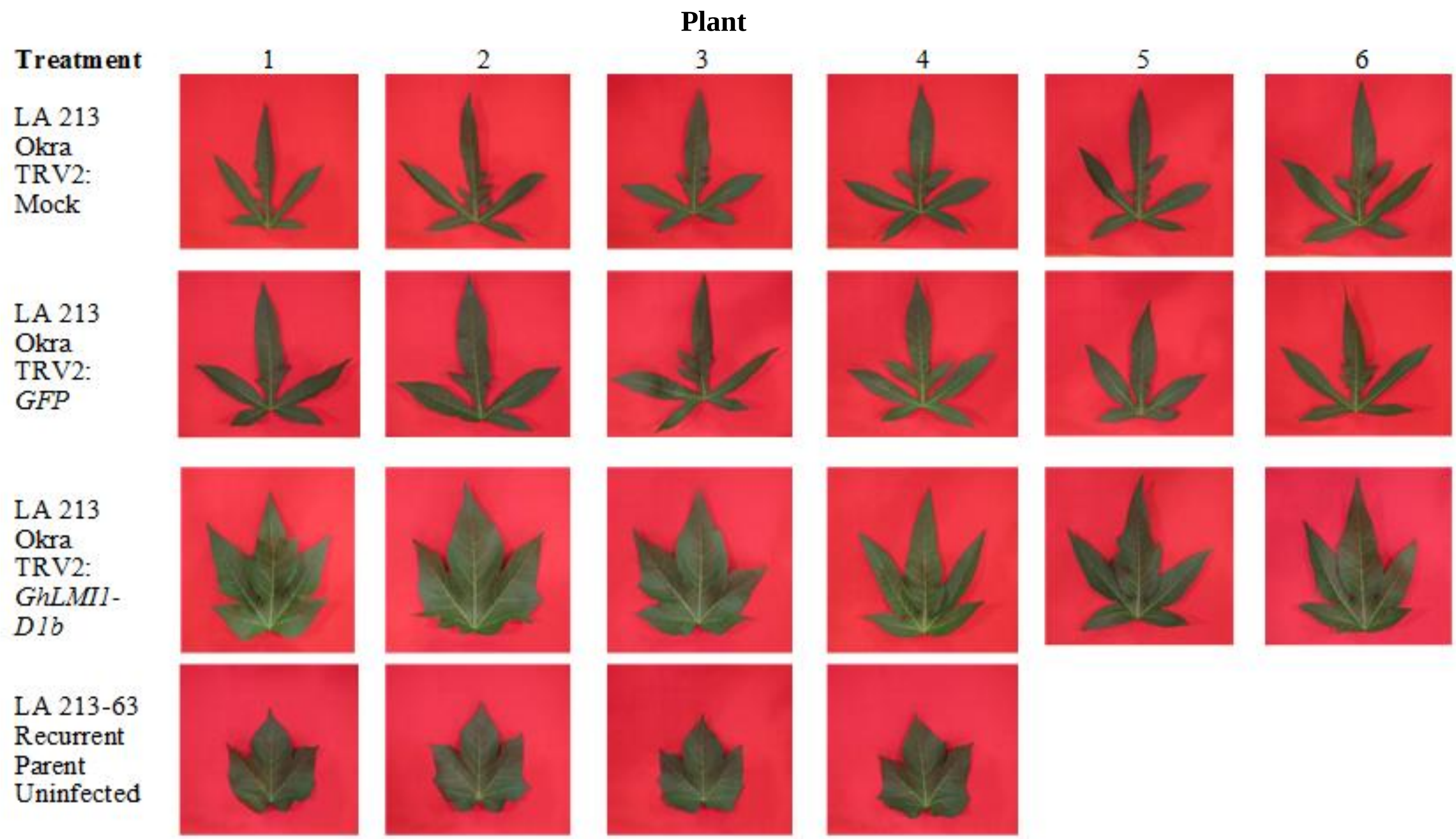


Figure 10 – Sixth True Leaf from VIGS Plants: Silencing of *GhLMI1-D1b* results in reduced leaf shape lobing of LA 213 Okra that approaches leaf shape of LA 213 Normal confirming that *GhLMI1-D1b* is responsible for the various leaf shapes of cotton.

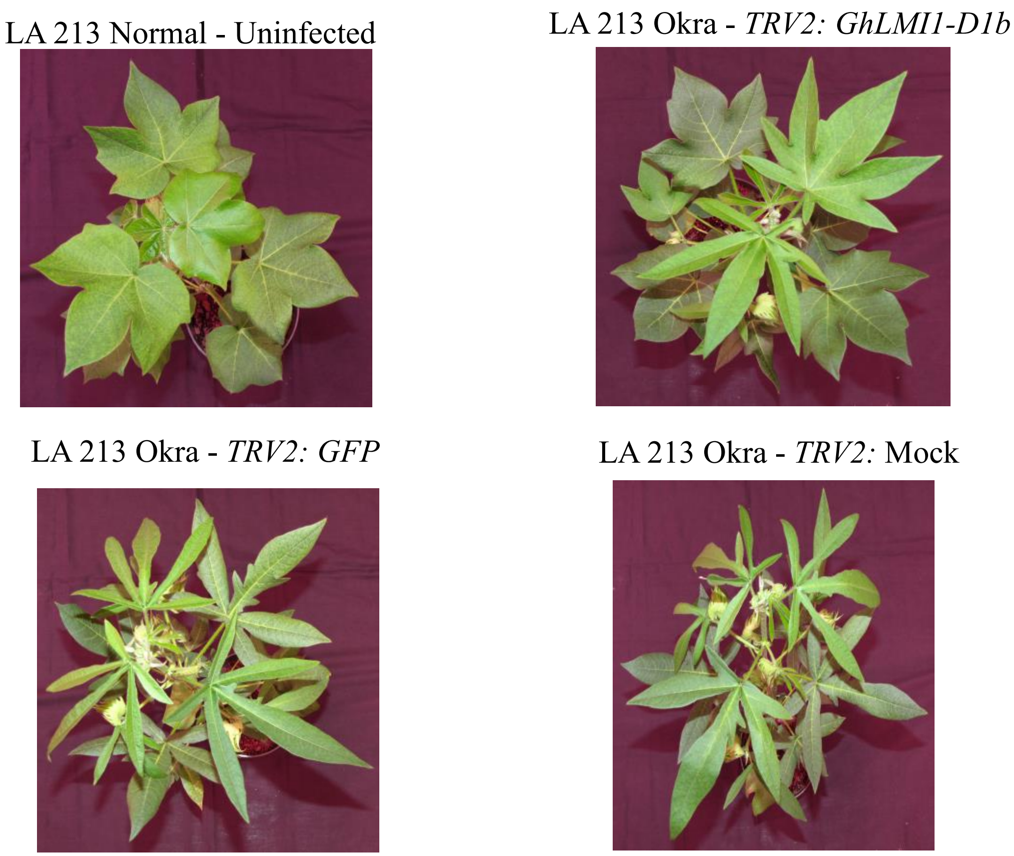
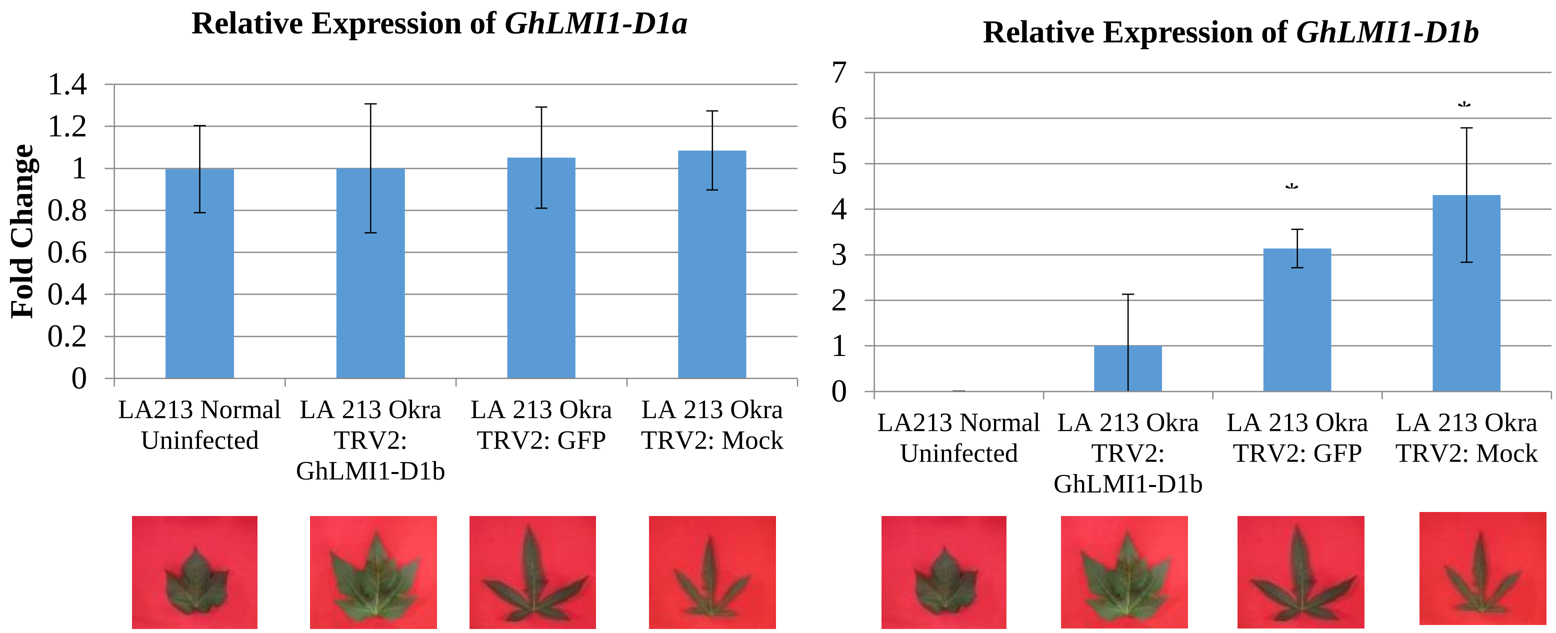


Figure 11 – Cotton Ideotype Developed through VIGS: The LA 213 Okra *TRV2: GhLMI1-D1b* was able to create the cotton leaf shape ideotype by producing a lower canopy of normal leaves before reverting to okra leaf shape. Such a variety has an intermediate amount of leaf area between LA 213 Normal and the negative control treatments that may be optimal for cotton production.



Expression data were collected from three plants per genotype with three technical replicates per plant. Fold change was calculated via the ΔΔCt method with *UBI*14 as the reference gene. LA 213 Okra *TRV2: GhLMI1-D1b* was used as the control treatment. Error bars represent the standard deviation of the fold change. * Represent statistical differences as determined by unpaired t-tests a $p < .05$

Figure 12 - Expression of GhLMI1-like genes from the VIGS experiment: *GhLMI1-D1a* was equally expressed among treatments indicating the VIGS construct was specifc to *GhLMI1-D1b* and off-target silencing of *GhLMI1-D1a* was not responsible for change in leaf shape. Expression of *GhLMI1-D1b* was reduced roughly four fold in the TRV2: *GhLMI1-D1b* treatment compared to TRV2: *GFP* and TRV2: Mock negative controls. Therefore change in leaf shape phenotype can be directly attributed to the knock down in *GhLMI1-D1b* further confirming that *GhLMI1-D1b* is responsible for okra leaf shape in cotton.

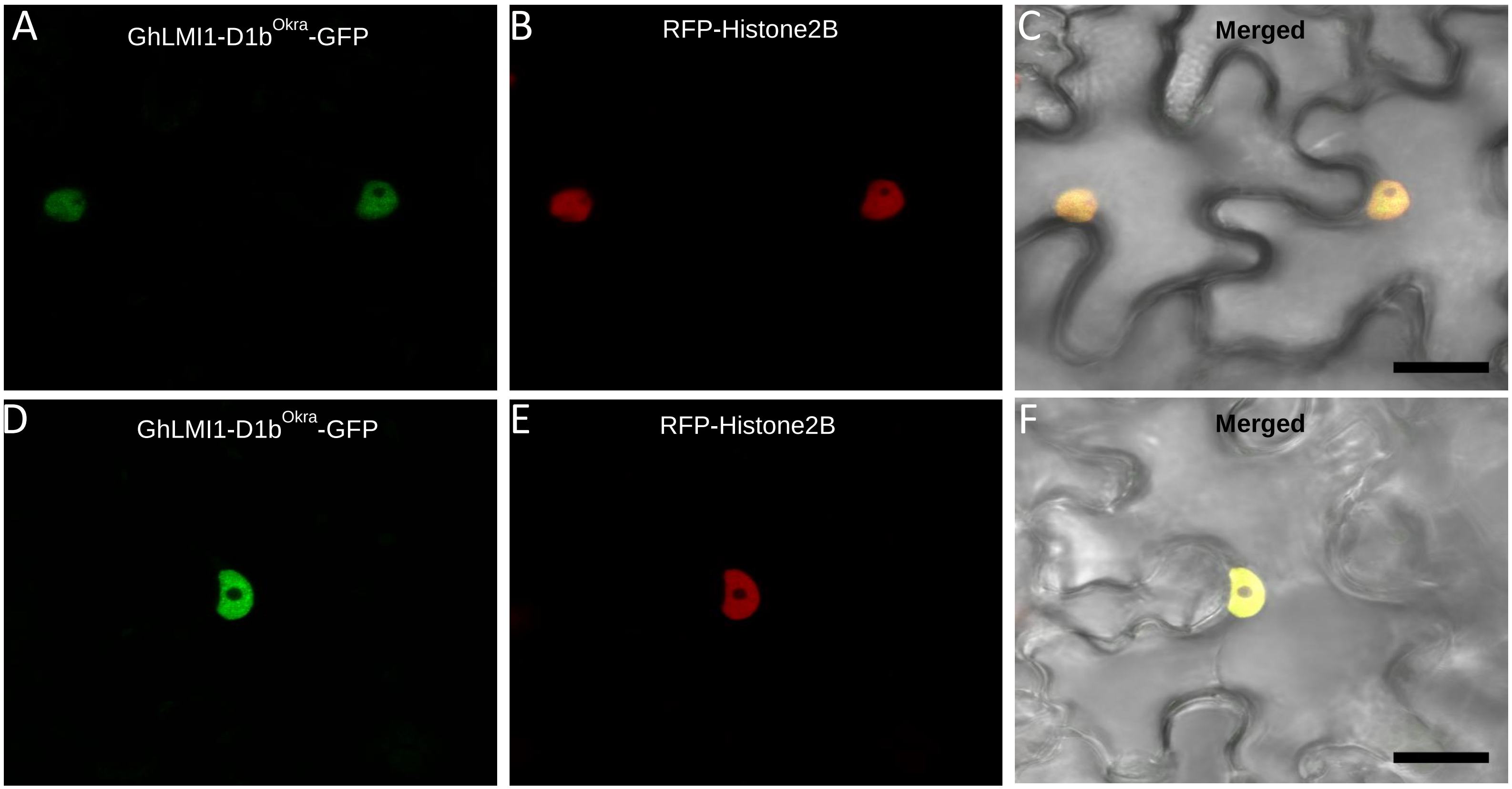


Figure 13 - GhLMI1-D1b^{Okra}-GFP localizes to the nucleus: A GhLMI1-D1b^{Okra}-GFP fusion construct driven by the 35S (A-C) or pUBQ10 (D-F) promoter was transiently expressed in *Nicotiana benthamiana* leaves. Signal from GhLMI1-D1b^{Okra}-GFP (A, D), the nuclear marker RFP-Histone2B (B, E) or the merged images including brightfield (C, F) are shown. Scale bar: 20 μm.

	Spray Pen.	Boll Rot	Lint Trash	Earliness	Flowering	Boll Reten.	Yield
Sub-Okra	?	None	None	None	None	+3%	0→ +4.8%
Okra	+26%	-44%	+15%	-6.5 days	+38%	-13%	-8%→ +20%
Super	?	-55%	-20%	-10.5 days	+93%	-21%	0→ -9%

Figure S1 – Agronomic Effect of Leaf Shape in Cotton: Values are set relative to normal or broad leaf. Data are averages of values reported across references 13-33.

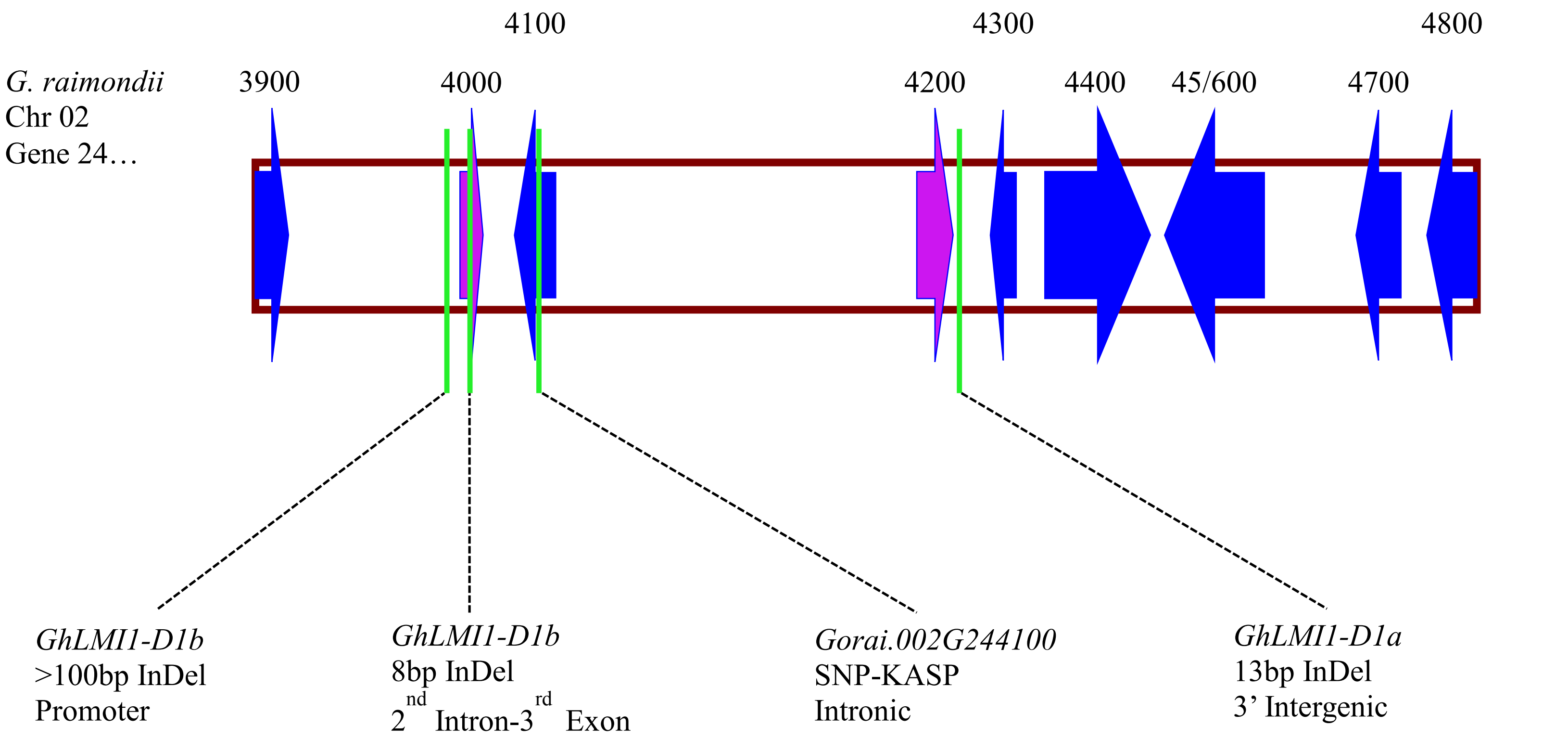


Figure S2 – Novel Marker Development for Leaf Shape Association Mapping: Four markers from the 112kb, nine gene candidate region identified by Kaur et al. (2015)[77] were run on the leaf shape association mapping panel. The 13bp InDel from the 3' intergenic region showed no association with leaf, allowing the candidate region to be narrowed to 52kb and four genes.

A)

B)

C)

D)

Figure S3: Leaf shapes of Gossypium diploids: **A)** “Normal” leaf of *G. arboreum* (AA) NC501 SMA4 (PI 529740). **B)** Laciniate leaf of *G. arboreum* (AA) NC505 Chinese Narrow Leaf (PI 615700). **C)** Spade-shaped leaf of *G. raimondii* (DD) NC174 $D_{5\text{-}1}$ (PI 530920) **D)** Lobed leaf of *G. thurberi* (DD) NC221 $D_{1\text{-}1}$ (PI 530765)

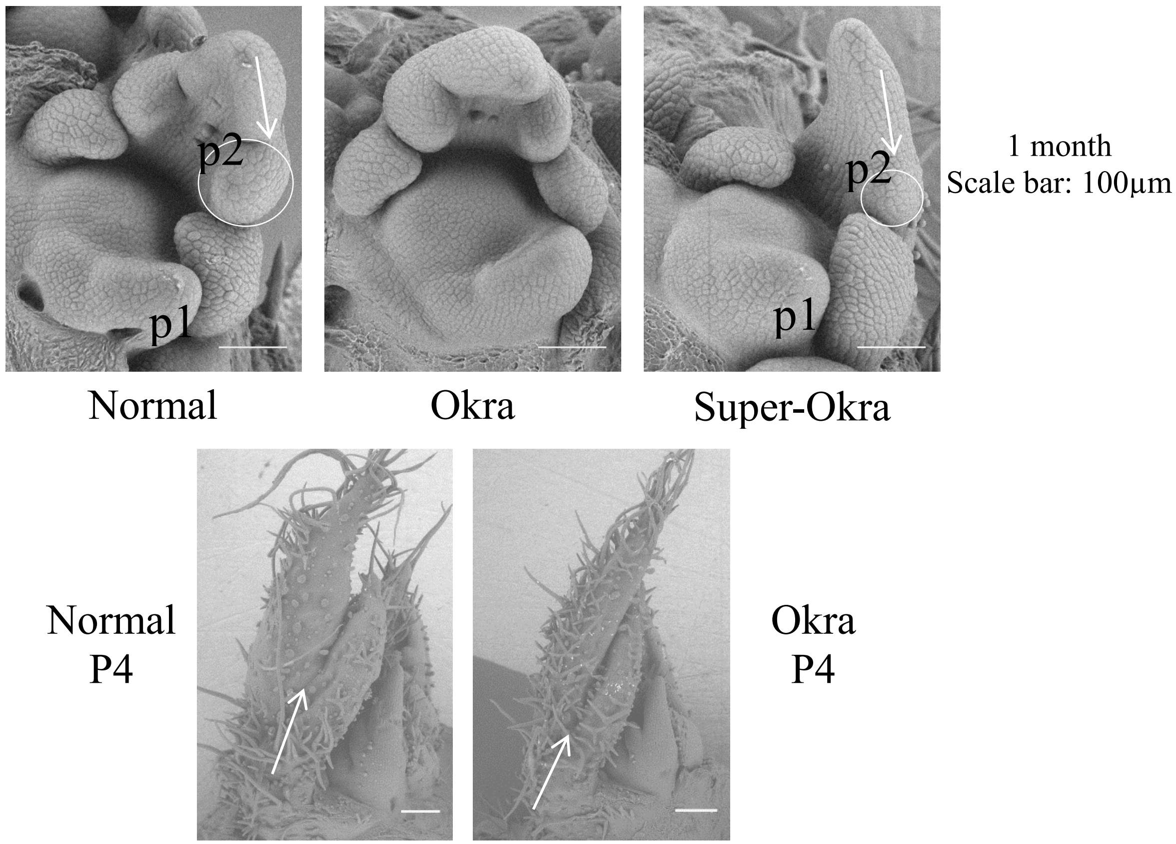


Figure S4 – Scanning Electron Microscopy (SEM) of Leaf Shapes in Cotton: **A)** Deep lobing and highly reduced distal lobe are evident in super-okra by the plastochron 2 (p2) stage of leaf development (white arrow). **B)** Deep lobing in okra is not evident in okra until the p4 stage (white arrow). Photos courtesy of Dan Chitwood, Viktoriya Coneva, and Margaret Frank, Danforth Plant Sciences Center, St. Louis, MO.

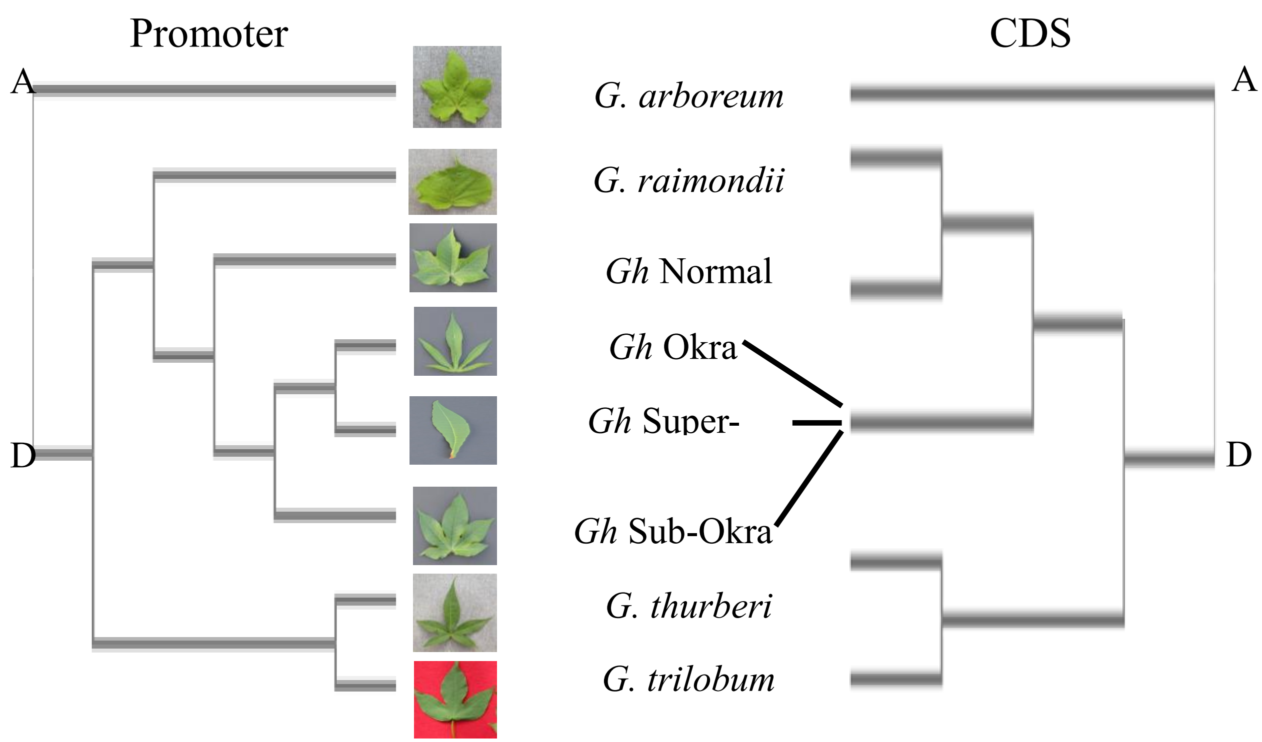
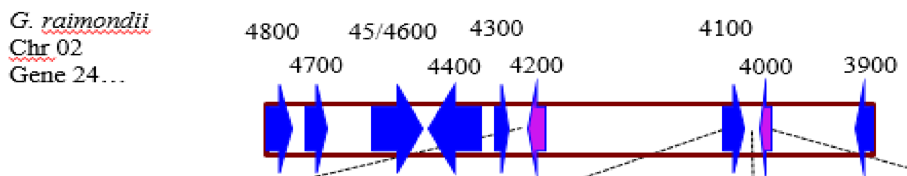


Figure S5 – Phylogenetic Analysis of Gossypium LMI1-1b Genes: The *G. hirsutum* normal CDS appears more closely related to *G . raimondii.* The other three leaf shapes are closely related and are likely not descended from either *G. trilobum* or *G. thurberi*. As expected *G. arboreum* forms its own outgroup.

Tables

Table 1 – Leaf Shape Association Mapping: There was no association between leaf shape and the *GhLMI1-D1a* polymorphism allowing the candidate region to be reduced to the 52 kb, four gene interval between *Gorai.002G243900* and *GhLMI1-D1a*.The near complete association between leaf shape and SNP within *Gorai.002G244100* was unable to further reduce the candidate region. The large *GhLMI1-D1b* promoter was completely associated with okra and super-okra leaf shape while the smaller genic polymorphism of *GhLMI1-D1b* was completely associated with normal leaf shape. Individual marker data may be missing so numbers may not sum completely.



	***GhLMI1-D1a* 3' Region**			***Gr.02G244100* SNP**		***GhLMI1-D1b* Promoter**		***GhLMI1-D1b* Genic**		
	471bp	**458bp**	**450bp**	**C**	**A**	**Large**	**Small**	**172bp**	**180bp**	**Total**
Normal	172	215	4	365	0	0	395	393	0	**395**
Okra	36	66	8	0	102	111	0	0	111	**111**
Sub-Okra	22	1	1	6	17	0	24	0	25	**25**
Super-Okra	0	13	0	0	7	13	0	0	13	**13**
Total	230	295	13	371	126	124	419	393	149	**544**

Table S1 **-** *The 20 Gossypium accessions used to sequence GhLMI1-like genes*:

Entry No.	**VARIETY**	**LEAF SHAPE**	**Species**
1	NC11-2100	Normal	*G. hirsutum*
2	NC11-2091	Normal	*G. hirsutum*
3	Coker 312	Normal	*G. hirsutum*
4	TM-1	Normal	*G. hirsutum*
5	LA 213 Frego Bract	Normal	*G. hirsutum*
6	NC05AZ21	Okra	*G. hirsutum*
7	Stoneville 7A Okra	Okra	*G. hirsutum*
8	Acala Red Okra Leaf	Okra	*G. hirsutum*
9	LA213 Okra	Okra	*G. hirsutum*
10	Aub Okra-16	Okra	*G. hirsutum*
11	Super Okra UA2-5	Super Okra	*G. hirsutum*
12	Acala 6010-15-U Okra	Super Okra	*G. hirsutum*
13	LA213 Super Okra	Super Okra	*G. hirsutum*
14	P62Lo	Super Okra	*G. barbadense*
15	Acala 6010-27-10 Okra	Super Okra	*G. hirsutum*
16	TAMCOT CAMD-ES	Sub-Okra	*G. hirsutum*
17	DPL 5540-85	Sub-Okra	*G. hirsutum*
18	MD 65-11	Sub-Okra	*G. hirsutum*
19	DES 422	Sub-Okra	*G. hirsutum*
20	NC05AZ06	Sub-Okra	*G. hirsutum*

Table S2 - Gossypium Diploid Accessions Used in the Genome Specificity Checks

NC Accession No.	Species	Genome	PI No.	Name
174	*G. raimondii*	DD	530920	D_{5-1}
175	*G. raimondii*	DD	530921	D_{5-2}
176	*G. raimondii*	DD	530928	D_{5-4}
235	*G. trilobum*	DD	530967	D_{8-5}
222	*G. thurberi*	DD	530767	D_{1-3}
223	*G. thurberi*	DD	530768	D_{1-4}
260	*G. thurberi*	DD	530766	D_{1-2}
262	*G. thurberi*	DD	530782	D_{1-18}
265	*G. thurberi*	DD	530789	D_{1-25}
501	*G. arboreum*	AA	167906	N/A
503	*G. arboreum*	AA	529740	SMA4
505	*G. arboreum*	AA	615700	Chinese Narrow Leaf
506	*G. arboreum*	AA	615701	Soudanense
507	*G. arboreum*	AA	615750	Punjabi 39
508	*G. arboreum*	AA	615759	231R

Table S3 - Chi-square Analysis of Large F_2 Population: Chi-square analysis of individual plant phenotype confirms the assumption that leaf shape in cotton is controlled by a single, incompletely dominant gene. *Three plants were classified as ambiguous owing to stunted growth.

Phenotype	**Observed**	**Expected (1:2:1)**	**Chi-Square**
Okra	262	256	0.141
Heterozygous	509	512	0.018
Normal	253	256	0.035
Total	**1,024***	**1,024**	**0.194 (p =0.91)**

References

1. Wendel JF, Brubaker C, Alvarez I, Cronn R, Stewart JM. (2009) Evolution and natural history of the cotton genus. In: Genetics and Genomics of Cotton AH Paterson (ed.) 3-19.

2. Wang K, Wang Z, Li F, Ye W et al. (2012) The draft genome of a diploid cotton *Gossypium raimondii*. *Nature Genetics* 44 (10): 1098-1103.

3. Paterson AH, Wendel JF, Gundlach H, Gui H, et al. (2012) Repeated polyploidization of *Gossypium* genomes and the evolution of spinnable fibers *Nature.* 492: 423-427.

4. Li F, Fan G, Wang K, Sun F et al. (2014) Genome sequence of the cultivated cotton *Gossypium arboreum*. *Nature Genetics* 46 (6): 567-572.

5. Li F, Fan G, Lu C, Xiao G et al. (2015) Genome sequence of cultivated Upland cotton (*Gossypium hirsutum* TM-1) provides insight into genome evolution. *Nature Biotechnology* 33: 524-530.

6. Zhang T, Hu Y, Jiang W, Fang L, et al. (2015) Sequencing of allotetraploid cotton (*Gossypium hirsutum* L. acc. TM-1) provides a resource for fiber improvement. *Nature Biotechnology* 33: 531-537.

7. Shoemaker DN. (1909) A study of leaf characters in cotton hybrids. *The Journal of Heredity* 5 (1): 116-119.

8. McClendon CA. (1912) Mendelian inheritance in cotton hybrids. *Georgia Agricultural Experiment Station Bulletins* 99: 139-228.

9. Harland SC. (1932) The genetics of *Gossypium*. *Bibliographica Genetica* 9: 107-182.

10. Hutchinson JB, Silow RA. (1939) Gene symbols for use in cotton genetics. *The Journal of Heredity* 30 (10): 461-464.

11. Stephens SG. (1945) A genetic survey of leaf shape in new world cottons – a problem in critical identification of alleles. *Journal of Genetics* 46 (2-3): 313-330.

12. Green JM. (1953) Sub-okra, a new leaf shape in Upland cotton. *The Journal of Heredity* 44 (6): 229-232.

13. Jones JE, James D, Sistler FE, Stringer SJ. (1987) Spray penetration of cotton canopies as affected by leaf and bract isolines. *Louisiana Agriculture* 29: 14-15, 17.

14. Jones JE, Andries JA. (1967) Okra leaf cotton for boll rot control. *Louisiana Agriculture* 10 (4): 4-5, 11.

15. Andries JA, Jones JE, Sloane LW, Marshall JG. (1969) Effects of okra leaf shape on boll rot, yield, and other important characters of Upland cotton, *Gossypium hirsutum* L. *Crop Science* 9: 705-710.

16. Andries JA, Jones JE, Sloane LW, Marshall JG. (1970) Effects of super okra leaf shape on boll rot, yield, and other characters of Upland cotton, *Gossypium hirsutum* L. *Crop Science* 10: 403-407.

17. Karami E, Weaver JB. (1972) Growth analysis of American Upland cotton, *Gossypium hirsutum* L., with different leaf shapes and colors. *Crop Science* 12: 317-320.

18. Rao MJ, Weaver JB. (1976) Effect of leaf shape on response of cotton to plant population, N rate, and irrigation. *Crop Science* 68: 599-601.

19. Novick RG, Jones JE, Anthony WS, Aguillard W, Dickson JI. (1991) Genetic trait effects of nonlint trash of cotton. *Crop Science* 31: 1029-1034.

20. Kerby TA, Buxton DR. (1976) Fruiting as affected by leaf type and population density. *Proceedings of the Beltwide Cotton Production Research Conferences* pp.67-70.

21. Landivar JA, Baker DN, Jenkins JN. (1983) Application of GOSSYM to genetic feasibility studies. I. Analyses of fruit abscission and yield in okra-leaf cotton. *Crop Science* 23: 497-504.

22. Heitholt JJ. (1993) Cotton boll retention and its relationship to lint yield. *Crop Science* 33: 486-490.

23. Heitholt JJ. (1995) Cotton flowering and boll retention in different planting configurations and leaf shapes. *Agronomy Journal* 87: 994-998.

24. Heitholt JJ, Meredith WR. (1998) Yield, flowering, and leaf area index of okra-leaf and normal-leaf cotton isolines. *Crop Science* 38: 643-648.

25. Gonias ED, Oosterhuis DM, Bibi AC. (2011) Light interception and radiation use efficiency of okra and normal leaf cotton isolines. *Environmental and Experimental Botany* 72: 217-222.

26. Meredith WR. (1983) Effect of environments and genetic backgrounds on the evaluation of cotton isolines. *Crop Science* 23 (1): 51-54.

27. Meredith WR. (1984) Influence of leaf morphology on lint yield of cotton – enhancement by the sub-okra trait. *Crop Science* 24 (5): 855-857.

28. Meredith WR. (1985) Lint yield genotype x environment interaction in Upland cotton as influenced by leaf canopy isolines. *Crop Science* 25: 509-512.

29. Meredith WR, Wells R. (1986) Normal vs okra leaf yield interactions in cotton. I. Performance of near-isogenic lines from bulk populations. *Crop Science* 26: 219-222.

30. Wells R, Meredith WR, Williford JR. (1986) Canopy photosynthesis and its relationship to plant productivity in near-isogenic cotton lines differing in leaf morphology. *Plant Physiology* 82: 635-640.

31. Thomson NJ, Reid PE, Williams ER. (1987) Effects of the okra leaf, nectariless, frego bract, and glabrous conditions on yield and quality of cotton lines. *Euphytica* 36: 545-553.

32. Heitholt JJ, Meredith WR, Williford JR. (1996) Comparison of cotton genotypes varying in canopy characteristics in 76-cm vs. 102-cm rows. *Crop Science* 36: 955-960.

33. Riar R, Wells R, Edmisten K, Jordan D, Bacheler J. (2013) Cotton yield and canopy closure in North Carolina as influenced by row width, plant population, and leaf morphology. *Crop Science* 53: 1704-1711.

34. Wilson RL, Wilson FD, George BW. (1979) Mutants of *Gossypium hirsutum*: Effect on pink bollworm in Arizona. *Journal of Economic Entomology* 72 (2): 216-219.

35. Wilson FD, George BW. (1982) Effects of okra-leaf, frego-bract, and smooth-leaf mutants on pink bollworm damage and agronomic properties of cotton. *Crop Science* 22: 798-801.

36. Wilson FD. (1986) Pink bollworm resistance, lint yield, and lint yield components of okra-leaf cotton in different genetic backgrounds. *Crop Science* 26: 1164-1167.

37. Wilson FD, George BW, Fry KE, Szaro JL, et al. (1986) Pink bollworm (Lepidoptera: Gelechiidae) egg hatch, larval success, and pupal and adult survival of okra- and normal-leaf cotton. *Journal of Economic Entomology* 79 (6): 1671-1675.

38. Wilson FD, Szaro JL, Hefner BA. (1992) Behavior of pink bollworm larvae (Lepidoptera: Gelechiidae) on bolls of normal-leaf and okra-leaf cotton isolines under laboratory conditions. *Journal of Agricultural Entomology* 9 (1): 55-63.

39. Jones JE, Clower DF, Milam MR, Caldwell WD, Melville DR. (1975) Resistance in Upland cotton to the banded-winged whitefly *Trialeurodes abutilonea* (Haldeman). *Proceedings of the Beltwide Cotton Production Research Conferences* pp.98-99.

40. Butler GD, Wilson FD. (1984) Activity of adult whiteflies (Homoptera: Aleyrodidae) within plantings of different cotton strains and cultivars as determined by sticky-trap catches. *Journal of Economic Entomology* 77: 1137-1140.

41. Butler GD, Henneberry TJ, Wilson FD. (1986) *Bemisia tabaci* (Homoptera: Aleyrodidae) on cotton: adult activity and cultivar oviposition preference. *Journal of Economic Entomology* 79: 350-354.

42. Butler GD, Wilson FD. (1986) Whitefly adults in okra-leaf and normal-leaf cotton. *Cotton: A College of Agriculture Report* pp.223-226.

43. Ozgur AF, Sekeroglu E. (1986) Population development of *Bemisia tabaci* (Homoptera: Aleurodidae) on various cotton cultivars in Cukurova, Turkery. *Agriculture, Ecosystems, and Environment* 17: 83-86.

44. Sippell DW, Bindra OS, Khalifa H. (1987) Resistance to whitefly (*Bemisia tabaci*) in cotton (*Gossypium hirsutum*) in the Sudan. *Crop Protection* 6: 171-178.

45. Butler GD, Rimon D, Henneberry TJ. (1988) *Bemisia tabaci* (Homoptera: Aleyrodidae): populations on different cotton varieties and cotton stickiness in Israel. *Crop Protection* 7: 43-47.

46. Flint HM, Parks NJ. (1990) Infestation of germplasm lines and cultivars of cotton in Arizona by whitefly nymphs (Homoptera: Aleyrodidae). *Journal of Entomological Science* 25:223-229.

47. Butler GD, Wilson FD, Fishler G. (1991) Cotton leaf trichomes and populations of *Empoasca lybica* and *Bemisia tabaci*. *Crop Protection* 10: 461-464.

48. Chu C, Cohen AC, Natwick ET, Simmons GS, Henneberry TJ. (1999) *Bemisia tabaci* (Hemiptera: Aleyrodidae) biotype B colonization and leaf morphology relationships in Upland cotton cultivars. *Australian Journal of Entomology* 38: 127-131.

49. Chu C, Freeman T, Natwick ET, Buckner JS, et al. (2000) *Bemisia argentifolii* adult, nymph, and egg densities and egg distribution on selected Upland cottons. *Journal of Entomological Science* 35: 39-47.

50. Chu C, Natwick ET, Henneberry TJ. (2002) *Bemisia tabaci* (Homoptera: Aleyrodidae) biotype B colonization on okra- and normal-leaf Upland cotton strains and cultivars. *Journal of Economic Entomology* 95: 733-738.

51. Walker GP, Natwick ET. (2006) Resistance to silverleaf whitefly, *Bemisia argentifolii* (Hem., Aleyrodidae), in *Gossypium thurberi*, a wild cotton species. *Journal of Applied Entomology* 130: 429-436.

52. Centintas R, McAuslane H. (2009) Effectiveness of parasitoids of *Bemisia tabaci* (Hemiptera: Aleyrodidae) on cotton cultivars differing in leaf morphology. *Florida Entomologist* 92: 538-547.

53. Bailey JC, Furr RE, Hanny BW, Meredith WR. (1978) Field populations of twospotted spider mites on sixteen cotton genotypes at Stoneville, MS 1977. *Journal of Economic Entomology* 71: 911-912.

54. Bailey JC, Meredith WR. (1983) Resistance of cotton, *Gossypium hirsutum* L., to natural field populations of twospotted spider mites (Acari: Tetranychidae). *Environmental Entomology* 12: 763-764.

55. Wilson LJ. (1993) Spider mites (Acari: Tetranychidae) affect yield and fiber quality of cotton. *Journal of Economic Entomology* 86: 566-585.

56. Wilson LJ. (1994) Resistance of okra-leaf cotton genotypes to twospotted spider mites (Acari: Tetranychidae). *Journal of Economic Entomology* 87: 1726-1735.

57. Pieters EP, Bird LS. (1977) Field studies of boll weevil resistant cotton lines possessing the okra leaf-frego bract characters. *Crop Science* 17: 431-433.

58. Jones JE. (1982) The present state of the art and science of cotton breeding for leaf-morphological traits. *Proceedings of the Beltwide Cotton Production Research Conferences* pp.93-99.

59. Baker DN, Myhre DL. (1969) Effects of leaf shape and boundary layer thickness on photosynthesis in cotton (*Gossypium hirsutum*). *Physiologia Plantarum* 22: 1043-1049.

60. Buxton DR, Stapleton HN. (1970) Predicted rates of net photosynthesis and transpiration as affected by the microenvironment and size of cotton leaf. *Proceedings of the Beltwide Cotton Production Research Conference* pp. 31-33.

61. Pegelow EJ, Buxton DR, Briggs RE, Muramoto H, Gensler WG. (1977) Canopy photosynthesis and transpiration of cotton as affected by leaf type. *Crop Science* 17: 1-4.

62. Karami E, Krieg DR, Quisenberry JE. (1980) Water relations and carbon-14 assimilation of cotton with different leaf morphology. *Crop Science* 20: 421-426.

63. Peng S, Krieg DR. (1991) Single leaf and canopy photosynthesis response to plant age in cotton. *Agronomy Journal* 83: 704-708.

64. Perry SW, Kreig DR, Hutmacher RB. (1983) Photosynthetic rate control in cotton. *Plant Physiology* 73: 662-665.

65. Pettigrew WT, Heitholt JJ, Vaughn KC. (1993) Gas exchange differences and comparative anatomy among cotton leaf-type isolines. *Crop Science* 33: 1295-1299.

66. Pettigrew WT. (2004) Physiological consequences of moisture deficit stress in cotton. *Crop Science* 44: 1265-1272.

67. Hammond D. (1941a) The expression of genes for leaf shape in *Gossypium hirsutum* L. and *Gossypium arboreum* L. I. The expression of gene for leaf shape in *Gossypium hirsutum* L. *American Journal of Botany* 28: 124-137.

68. Hammond D. (1941b) The expression of genes for leaf shape in *Gossypium hirsutum* L. and *Gossypium arboreum* L. II. The expression of gene for leaf shape in *Gossypium arboreum* L. *American Journal of Botany* 28: 138-150.

69. Dolan L, Poethig RS. (1991) Genetic analysis of leaf development in cotton. *Development Supplement* 1: 39-46.

70. Endrizzi JE, Brown MS. (1964) Identification of monosomes for six chromosomes in *Gossypium hirsutum*. *American Journal of Botany* 51 (2): 117-120.

71. Endrizzi JE, Kohel RJ. (1966) Use of telosomes in mapping three chromosomes in cotton. *Genetics* 54: 535-550.

72. Jiang C, Wright RJ, Woo SS, DelMonte TA, Paterson AH. (2000) QTL analysis of leaf morphology in tetraploid *Gossypium* (cotton). *Theoretical and Applied Genetics* 100: 409-418.

73. Song XL, Guo WZ, Han ZG, Zhang TZ. (2005) Quantitative trait loci mapping of leaf morphological traits and chlorophyll content in cultivated tetraploid cotton. *Journal of Integrative Plant Biology* 47 (11): 1382-1390.

74. Lacape JM, Gawrysiak G, Cao TV, Viot C, et al. (2013) Mapping QTLs for traits related to phenology, morphology, and yield components in an inter-specific *Gossypium hirsutum* x *G. barbadense* cotton RIL population. *Field Crops Research* 144: 256-267.

75. Zhu QH, Spriggs A, Taylor JM, Llewellyn D, Wilson I. (2014) Transcriptome and complexity-reduced, DNA-based identification of intraspecies single-nucleotide polymorphisms in the polyploidy *Gossypium hirsutum* L. *Genes ,Genomes, Genetics* 4 (10): 1893-1905.

76. Andres RJ, Bowman DT, Kaur B, Kuraparthy V. (2014) Mapping and genomic targeting of the major leaf shape gene (*L*) in Upland cotton (*Gossypium hirsutum* L.). *Theoretical and Applied Genetics* 127: 167-177.

77. Kaur B, Andres RJ, Kuraparthy V. (2015) Major leaf shape genes, laciniate ($L\text{-}A_2^L$) in diploid (*Gossypium arboreum*) and okra (L^O) in polyploid Upland (*G. hirsutum*) cotton, map to an orthologous genomic region. *Crop Science* Submitted.

78. Saddic LA, Huvermann B, Bezhani S, Su Y, et al. (2006) The LEAFY target LMI1 is a meristem identity regulator and acts together with LEAFY to regulate expression of CAULIFLOWER. *Development* 133:1673-1682.

79. Vlad D, Kierzkowski D, Rast MI, Vuolo F, et al. (2014) Leaf shape evolution through duplication, regulatory diversification, and loss of a homeobox gene. *Science* 343: 780-783.

80. Sicard A, Thamm A, Marona C, Lee YW, et al. (2014) Repeated evolutionary changes of leaf morphology caused by a mutation to a homeobox gene. *Current Biology* 24:1880-1886.

81. Tyagi P, Gore MA, Bowman DT, Campbell BT, et al. (2014) Genetic diversity and population structure in the US Upland cotton (*Gossypium hirsutum* L.) *Theoretical and Applied Genetics* 127: 283-295.

82. Kennedy CW, Smith WC, Jones JE. (1986) Effect of early season square removal on three leaf types of cotton. *Crop Science* 26: 139-145.

83. Artico S, Nardeli SM, Brilhante O, Grossi-de-Sa MF, Alves-Ferreira M. (2010) Identification and evaluation of new reference genes in *Gossypium hirsutum* for accurate normalization of real-time quantitative RT-PCR data. *BMC Plant Biology* 10:49.

84. Tuttle JR, Haigler CH, Roberston D. (2012) Method: low-cost delivery of the cotton leaf crumple virus-induced gene silencing system. *Plant Methods* 8: 27

85. Xu Z, Kohel R, Song G, Cho J et al. (2008) An integrated genetic and physical map of homoeologous chromosomes 12 and 26 in Upland cotton (*G. hirsutum* L). *BMC Genomics* 9: 108.

86. Wang S, Chen J, Zhang W, Hu Y, et al. (2015) Sequence-based ultra-dense genetic and physical maps reveal structural variations of allopolyploid cotton genomes. *Genome Biology* 16: 108.

87. Werner JE, Endo TR, Gill BS. (1992) Toward a cytogenetically based physical map of the wheat genome. *Proceedings of the National Academy of Sciences* 89:11307–11311.

88. Gill KS, Gill BS, Endo TR, Boyko EV. (1996) Identification and high-density mapping of gene rich regions in chromosome group 5 of wheat. *Genetics* 143:1001–1012.

89. Rong J, Abbey C, Bowers JE, Brubaker CL, et al. (2004) A 3347-locus genetic recombination map of sequence-tagged sites reveals features of genome organization, transmission and evolution of cotton (*Gossypium*). *Genetics* 166:389-417.

90. Martin GB, Brommonschenkel SH, Chunwongse J, Frary A, et al. (1993) Map-based cloning of a protein kinase gene conferring disease resistance in tomato. *Science* 262:1432-1436.

91. Faris JD, Fellers JP, Brooks SA, Gill BS. (2003) A bacterial artificial chromosome contig spanning the major domestication locus Q in wheat and identification of a candidate gene. *Genetics* 164:311-321.

92. Yan L, Loukoianov A, Blechl A, Tranquilli et al. (2004) The wheat VRN2 gene is a flowering repressor down-regulated by vernalization. *Science* 303:1640-1644.

93. Lin Z, Li X, Shannon LM, Yeh CT, Wang ML, et al. (2012) Parallel domestication of the Shattering1 genes in cereals. *Nature Genetics* 44:720-724.

94. Kay R, Chan A, Daly M, McPherson J (1987) Duplication of CaMV 35S promoter sequences creates a strong enhancer for plant genes. *Science* 236 (4806): 1299-1302.

95. Ebert PR, Ha SB, An G. (1987) Identification of an essential upstream element in nopaline synthase promoter by stable and transient assays. *Proceedings of the National Academy of Sciences* 84 (16): 5745-5749.

96. Li ZT, Jayasankar S, Gray DJ. (2004) Bi-directional duplex promoters with duplicated enhancers significantly increase transgene expression in grape and tobacco. *Transgenic Research* 13 (2): 143-154

97. Tantikanjana T, Nasrallah ME, Nasrallah JB. (1996) The *Brassica S* gene family: molecular characterization of the *SLR2* gene. *Sexual Plant Reduction* 9 (2): 107-116.

98. Li ZT, Jayasankar S, Gray DJ. (2004) Bi-directional duplex promoters with duplicated enhancers significantly increase transgene expression in grape and tobacco. *Transgenic Research* 13 (2): 143-154.

99. Henderson IA, Jacobsen SE. (2008) Tandem repeats upstream of the *Arabidopsis* endogene *SDC* recruit non-CG DNA methylation and initiate siRNA spreading. *Genes & Development* 22: 1597-1606.

100. Lee J, Oh M, Park H, Lee I. (2008) SOC1 translated to the nucleus by interaction with AGL24 directly regulates *LEAFY*. *The Plant Journal* 55: 832-843.

101. Chandler JW. (2012) Floral meristem initiation and emergence in plants. *Cellular and Molecular Life Sciences* 69 (22): 3807-3818.

102. Nakagawa T, Kurose T, Hino T, Tanaka K, et al. (2007) Development of series of gateway binary vectors, pGWBs, for realizing efficient construction of fusion genes for plant transformation. *Journal of Bioscience and Bioengineering* 104: 34-41.

103. Grefen C, Donald N, Schumacher K, Blatt MR. (2010) A Ubiquitin-10 promoter-based vector set for fluorescent protein tagging facilitates temporal stability and native protein distribution in transient and stable expression studies. *Plant Journal* 64: 355-365

104. Martin K, Kopperud K, Chakrabarty R, Banerjee R, et al. (2009) Transient expression in *Nicotiana benthamiana* fluorescent marker lines provides enhanced definition of protein localization, movement and interactions in planta. *Plant Journal* 59: 150-162.

105. Ohno S. (2013) Evolution by gene duplication. *Springer Science+Business Media*, New York. NY.