MOLECULAR AND CONVENTIONAL TECHNIQUES TO DECIPHER GENETIC NETWORKS OF FIBER DEVELOPMENT TOWARD BREEDING IMPROVED COTTON GERMPLASM

by

JINESH DAHYABHAI PATEL

(Under the Direction of Andrew H. Paterson)

ABSTRACT

Study of complex biochemical pathway and genes involved in cotton fiber development is important to understanding the biology behind the longest plant cell known. Cotton seed epidermal cells go through initiation, elongation, secondary cell wall synthesis and maturation to become long lint fibers. Natural and artificial mutants are excellent tools to help decipher functions of genes involve in fiber development and/or a source to make improved germplasm. Here we used a natural mutant, *Ligon lintless-2* (*Li_2*), to identify gene/s in the fiber elongation process. In another experiment, we used mutant lines developed through EMS-mutagenesis to improve fiber quality of elite germplasm and study the effect of pyramiding novel alleles conferring improved fiber quality.

A population of 1,545 F2 plants derived from a cross between Pima S-7 (*G. barbadense*) and an *Li_2* mutant line (*G. hirsutum*) and 144 DNA markers were used to fine map the *Li_2* region on chromosome 18. We identified terminal deletion of the long arm of chromosome 18 to be the probable cause of the *Li_2* phenotype, identifying seven candidate genes. By Virus-Induced Gene Silencing (VIGS), knockdown of two genes, *GhUGT87A1-D1a* and *GhUGT87A*, showed *Li_2* like

phenotypes. This research provides new insight into the causal mutation of the *Li*$_2$ phenotype and identifies genes involved in fiber elongation.

A total of 12 mutant lines showing striking improvement in fiber attributes were studied for the stability of the mutant phenotype, heritability of the improved trait and interaction between novel alleles. A total of ten populations were developed, four involving crosses between TAM94L25 mutant lines and GA230, four from ACALA1517-99 mutant lines X GA230 and two from ACALA1517-99 mutant lines x TAM94L25 mutant lines. Each population had either a combination of two mutant lines in GA230 background or a combination of four mutant lines (ACALA1517-99 mutant lines x TAM94L25 mutant lines). Based on replicated trials in three environments we concluded that most mutant lines can be used to improve elite germplasm and thus mitigate the genetic bottleneck in cotton. In most instances, combining different mutant lines made it possible to improve two fiber attributes in a single cross.

INDEX WORDS: crop improvement, allele stacking, breeding, genomics

MOLECULAR AND CONVENTIONAL TECHNIQUES TO DECIPHER GENETIC NETWORKS OF FIBER DEVELOPMENT TOWARD BREEDING IMPROVED COTTON GERMPLASM

by

JINESH DAHYABHAI PATEL

B.S., Anand Agricultural University, India, 2008.

MS, The University of Georgia, 2011.

A Dissertation Submitted to the Graduate Faculty of The University of Georgia in Partial

Fulfillment of the Requirements for the Degree

DOCTOR OF PHILOSOPHY

ATHENS, GEORGIA

2017

MOLECULAR AND CONVENTIONAL TECHNIQUES TO DECIPHER GENETIC NETWORKS OF FIBER DEVELOPMENT TOWARD BREEDING IMPROVED COTTON GERMPLASM

by

JINESH DAHYABHAI PATEL

Major Professor: Andrew H Paterson

Committee: Peng W. Chee
John M Burke
Dayton Wilde
Russell L Malmberg

Electronic Version Approved:

Suzanne Barbour
Dean of the Graduate School
The University of Georgia
December 2017

DEDICATION

I would like to dedicate this work to my wife (Sejal J Patel), son (Krishav J Patel), my parents (Dr. Dahyabhai M Patel and Ranjenben D Patel), sister (Dr. Nicky N Seth), and brother-in-law (Dr. Nirav G Seth).

ACKNOWLEDGEMENTS

I would like to thank my mentor Dr. Andrew H. Paterson for his patience, invaluable guidance, encouragement, and financial support. I would also like to thank my advisory committee members Dr. Peng W Chee, John M Burke, Dayton Wilde and Russell L Malmberg for their advice on my dissertation research throughout my degree. Further, I am also thankful to my friends Rahul Chandnani, Sameer Khanal and Jeevan Adhikari for helping me in field work, data collection and for moral support. I am also grateful to all Plant Genome Mapping Laboratory members for their help in the field work and lab. I would like to extend my thanks to all the member of Cotton Molecular Breeding Lab for helping in field work at Tifton campus. Lastly, I am fortunate to have a wonderful family that provided me unconditional support and love during my PhD.

TABLE OF CONTENTS

Page

LIST OF TABLES

Page

LIST OF FIGURES

Page

CHAPTER 1

INTRODUCTION AND LITERATURE REVIEW

History and importance of cotton

Cotton (*Gossypium sp.*) is cultivated primarily for its fiber (BRUBAKER AND WENDEL 1994), though its edible oil makes it also a leading oilseed crop. Cotton was domesticated independently in both the Old and New World (WENDEL AND CRONN 2003; PICKERSGILL 2007; RENNY-BYFIELD *et al.* 2016). *Gossypium arboreum* and *Gossypium herbaceum* are Old World cotton species that have been cultivated for thousands of years in Africa and Asia (Viot 2017). Two New World cotton species have also been domesticated, with *Gossypium hirsutum* dated between 3400 BC to 2300 BC in Mexico, and *Gossypium barbadense* dated around 2500-1750 B.C in Peru (ABDEL-SALAM *et al.* 2009). Since the early 1900's, *G. hirsutum* has dominated world cotton production (WENDEL *et al.* 1992; CHEN *et al.* 2007).

Cotton is grown in around 100 countries with China, India and the United States of America being the top three cotton producing countries (Chapagain *et al.* 2006). Cotton is the fourth major crop in US in terms of acreage, grown in 17 states for fiber and cotton seed that generates about $8.3 billion (USDA, 2010; National Cotton Council of America. 2013). Around 80% of US cotton is exported and helps to reduce trade deficits. Most cotton fiber is used as raw material in textile industries to develop cotton clothes, towels and napkins, while some is used for home furnishing, paper, plastic, mattress padding, automobile cushions and explosives. Cotton seed are used for making cotton seed hull and cotton seed cake for livestock consumption while cotton

seed oil is consumed by human. Cotton seeds contain gossypol that is harmful to human consumption but recently cotton plants with ultra-low Gossypol Cottonseed were developed using RNAi technique, opening the door to cotton varieties with seed edible by humans (SUNILKUMAR *et al.* 2006; RATHORE *et al.* 2012; PALLE *et al.* 2013).

Genetic diversity in cotton

The cotton genus, *Gossypium*, consists of 52 species, with more being discovered. (WENDEL AND ALBERT 1992; SEELANAN *et al.* 1997; STEWART *et al.* 2015; WENDEL AND GROVER 2015; GALLAGHER *et al.* 2017). These *Gossypium* species are either diploid or tetraploid. The diploid Gossypium species have a haploid chromosome number of 13 and consist of eight monophyletic genome groups (A, B, C, D, E, F, G and K). The tetraploids are allopolyploid consisting of two diploid genomes, "A" and "D" (ENDRIZZI *et al.* 1985; PERCIVAL *et al.* 1999; WENDEL AND CRONN 2003 ; GROVER *et al.* 2007; CHEN *et al.* 2016). A total of seven allopolyploid species have been named so far, all of New World origin, including *Gossypium hirsutum* L., *G. barbadense* L., *G. tomentosum*, *G. mustelinum*, *G. darwinii*, *G. ekmanianum* and *Gossypium stephensii* (WENDEL AND GROVER 2015; GALLAGHER *et al.* 2017). Out of these seven, only *Gossypium hirsutum* L. and *G. barbadense* L. are cultivated.

Though *G. hirsutum* accounts for more than 90 % of the worldwide cotton crop, modern Upland cotton cultivars contain only half of the diversity available in *G. hirsutum* while *G. barbadense* cultivars have retained a much greater portion of available diversity, findings suggested to result from selection pressure and targeted breeding programs (WENDEL *et al.* 1992; WENDEL *et al.* 2009; TYAGI *et al.* 2014; ZHAO *et al.* 2015; AI *et al.* 2017). Research in which 250 DNA markers were screened in 320 cultivars/lines from the US National Plant Germplasm

collection found a very low level of genetic variation, lower than other major crops such as maize, sorghum, wheat and soybean (CHEE *et al.* 2004; LUBBERS *et al.* 2004).

There are natural and artificial sources available to increase the genetic diversity of Upland cotton. The natural sources can be divided into primary, secondary and tertiary gene pools based on degrees of compatibility with *G. hirsutum* (STEWART 1995; LUBBERS AND CHEE 2009). The primary gene pool consisting of seven races of *G. hirsutum*, namely, 'yucatanense', 'punctatum', 'palmeri', 'latifolium', 'marie-galante', 'morrilli', and 'richmondi' that are easily usable to increase the genetic diversity, save for some recalcitrance due to photoperiodism. (Hutchinson 1951; Lubbers and Chee 2009). Use of the secondary gene pool consisting of the remaining allopolyploid species is possible to increase genetic diversity, but progenies experience segregation distortion and carry many deleterious or unfavorable agronomical alleles. Use of the tertiary gene pool is very difficult as it contains diploid species which generally do not produce fertile hybrids in crosses with *G. hirsutum,* generally requiring chromosome doubling to be usable (LUBBERS AND CHEE 2009).

Artificial sources available to increase the genetic diversity of Upland cotton mainly consist of germplasm developed by mutation techniques or genetic modification (Genetically Modified Organisms, GMOs). Researchers have successfully release cotton mutant lines with improved fiber traits and yield components or herbicide tolerance (AULD *et al.* 2000; BECHERE *et al.* 2007a; BECHERE *et al.* 2009b; BECHERE *et al.* 2010; BECHERE *et al.* 2011; BROWN *et al.* 2012). Still mutant breeding has been rarely used in cotton breeding program due to multiple generation advancement to reduce the mutant load and stabilize the mutation.

Cotton life cycle and fiber development

Cotton is cultivated as annual crop, in contrast to the perennial nature of its wild ancestors. From planting, cotton requires around six months to be ready for harvest. During this six months some milestones of cotton development include emergence (around 3 to 5 days after planting or DAP), formation of the first true leaf (10 DAP), appearance of the first square (38 DPA), first flower (59 DPA), open bolls (120 DAP) and ready for harvesting (150 to 160 DAP) (OOSTERHUIS AND JERNSTEDT 1999). The major commercially important organ of cotton is bolls that takes about two months to open after the flower appears. These bolls contain cotton fibers that can grow up to 6 cm in some *G. barbadense* species (KIM AND TRIPLETT 2001; LEE *et al.* 2007b). Cotton fibers are unicellular unbranched cells.

The process of fiber development is complex and consist of four overlapping stages, namely, fiber cell initiation, elongation, secondary wall biosynthesis, and maturation (BASRA AND MALIK 1984). Fiber initiation starts around three to five days before anthesis and can last six to ten days (KIM AND TRIPLETT 2001; CHEN AND GUAN 2011). Many genes such as *FBP7* (auxin promoter), sucrose synthase (*Sus*), Fasciclin-Like Arabinogalactan protein, (*GhFLA1*), cotton JASMONATE ZIM-DOMAIN protein (*GhJAZ2*), and transcription factors such as *GhMYB25*, *GhMML3*, *GhMYB109*, and *GhMYB2* have been found to be involved in fiber initiation (RUAN *et al.* 2003; SUO *et al.* 2003; MACHADO *et al.* 2009; CHEN AND GUAN 2011; HUANG *et al.* 2013; GUAN *et al.* 2014; HU *et al.* 2016; WAN *et al.* 2016).

Fiber initiation is followed by fiber elongation that could last for 15 to 20 days. During this period a fiber can expand at a rate of 2mm/day (LEE *et al.* 2007b; XU *et al.* 2008). Fiber generally reaches its maximum length at the end of this process. Numerous complex and interconnected biochemical pathways participate in fiber elongation. For example, expression of

genes like *GhSCP2D* are positively related to expression of *GhPdBG3*, which controls sterol homeostasis. Maintenance of sterol homeostasis is essential for plasmodesmata permeability and fiber elongation (ZHANG *et al.* 2017). Another complex biochemical pathway is regulation of H_2O_2 and reactive oxygen species (ROS) which involve numerous genes such as *GhPK6*, Pyruvate kinase *(ZHANG AND LIU 2016)*; *GhFAnnxA*, an annexin (ZHANG *et al.* 2016a); *GhCaM7*, a calcium sensor (TANG *et al.* 2014); *GhAPX1*, ascorbate peroxidase (QIN *et al.* 2008b); *GhRac1* (KIM AND TRIPLETT 2004); *GhHD-1*, homeodomain leucine zipper (HD-ZIP) transcription factor (WALFORD *et al.* 2012) and many more. Other important processes taking place during fiber elongation regulating levels of ethylene, stress and Ca2+, water transportation, cell wall loosening, different secondary metabolic pathways, and pectin biosynthesis have been identified (HOVAV *et al.* 2008a; PANG *et al.* 2010; HAIGLER *et al.* 2012; LI *et al.* 2013; FANG *et al.* 2014; SHAN *et al.* 2014; TANG *et al.* 2014; YANG *et al.* 2014).

Cellulose (secondary wall) biosynthesis starts around 15 DPA in the middle of fiber elongation and lasts until 35 DPA. During this stage, cellulose is deposited causing cell wall thickening of up to ~3-6 μm which is highest percentage of cellulose deposition in a cell of plant kingdom (HAIGLER 2007; HAIGLER *et al.* 2009). Cellulose biosynthesis has a direct connection to fiber strength, an important component of fiber quality (PANG *et al.* 2010). Major genes expressed during cellulose biosynthesis include transcription factors (*GhMYB1* and *GhMYB7),* Endo 1,4-β-glucanase (*CEL*), Lipid transfer protein (*LTP3*) cellulose synthase (*GhCesA4*), species-specific expansin (*GbEXPATR*), and leucine-rich repeat protein kinase (*LRR RLK*) (MUNIS *et al.* 2010; KIM *et al.* 2011; MANSOOR AND PATERSON 2012; SUN *et al.* 2015; HUANG *et al.* 2016; ISLAM *et al.* 2016; LI *et al.* 2016c; FANG *et al.* 2017).

During fiber maturation, water potential of fiber decreases and minerals starts to accumulate (JOHN AND KELLER 1996). At the end of fiber development, one could expect a single fruit (boll) containing 30 to 35 seeds and around 500,000 elongated lint fibers (15 to 25% of cotton seed epidermal cells) (BASRA AND MALIK 1984; TIWARI AND WILKINS 1995; KIM AND TRIPLETT 2001; HOVAV *et al.* 2008c).

Fiber traits measure by High Volume Instrument (HVI) instrument

For many years, the major objectives of cotton breeding were biotic and abiotic stress and lint yield improvement. More recently, changes in the textile industries such as use of air jet spinning machines which are eight times faster than previous counterparts has placed greater emphasis on fiber quality (BRADOW AND DAVIDONIS 2000). There are two major instruments used to determine fiber properties, High-volume instrumentation (HVI) which generally measures bulk sample of fiber; and Advanced Fiber Information System (AFIS) which measures individual fibers (SASSER 1981; SHOFNER *et al.* 1990; SUH AND SASSER 1996; KRIFA AND ETHRIDGE 2006; KELLY *et al.* 2012). In 1960, the Plains Cotton Cooperative Association (PCCA) was a pioneer in developing HVI testing system for determining fiber properties (www.pcca.com). The major fiber quality attributes that are measured by the High Volume Instrument (HVI) are micronaire (fiber fineness), fiber length, fiber uniformity (uniformity index), fiber strength, fiber elongation, and short fiber content (SFC).

Fiber fineness or Micronaire (MIC) is determined based on the ability of compressed cotton fiber samples to allow air to pass through (www.cottoninc.com) The measurement is important because it identifies samples with coarse or immature fiber, both of which contribute to formation of ‘neps’ (small knots of entangled fibers in fabric) in yarn, thus reducing overall yarn

quality (Van der Sluijs and Hunter 1999; Ulloa 2006), and making dyeing problematic. A range of 3.7 to 4.2 of Micronaire is considered as the premium range (BRADOW AND DAVIDONIS 2000).

Fiber length or Upper Half Mean Length (UHML), measured in inches or mm, is one of the most important components of fiber quality (WAKEHAM 1955). Higher fiber length is better for high speed spinning machines as fibers can withstand more resistance. Also, higher fiber length is important for yarn quality (VAN DER SLUIJS AND HUNTER 1999; KRIFA AND ETHRIDGE 2006; ULLOA 2006). Fiber length is positively associated with multiple fiber attributes like fiber strength, uniformity and short fiber content (SFC). Thus, improvement of fiber length will affect overall fiber quality.

Fiber uniformity index is the ratio (%) of total mean length of fibers in a sample to that of the longest 50 percent of fibers in the sample (upper half mean length, or UHM). Uniformity of fiber is important for efficient yarn spinning and is positively related to yarn quality (MEREDITH 1994; ULLOA 2006).

Fiber strength is measured as the force (in grams) required to break a fiber bundle of one tex unit (COTTON 1998; HAIGLER 2010). Thus, its unit is grams/tex. Fiber strength has generally positive association with fiber length, micronaire and Fiber uniformity index. It has high impact on cotton fiber quality. It helps fiber to withstand the force generated by high speed spinning of yarn, reducing wastage and improving efficiency of fabric manufacturing. The USDA Agricultural Marketing Service (AMS), divides cotton fiber strength into various groups with 23.0 g/tex or below considered to be the lowest class and 31.0 g/tex and above the highest class.

Fiber elongation is the percentage by which a fiber could be stretched before it breaks. Fiber elongation provides the ability to withstand high throughput textile processing, adds toughness to yarn, improving yarn quality and work-to-break values (YANG AND GORDON ; WATERS *et al.*

1966). Research has suggested that there is broad scope to improve fiber elongation using conventional breeding, as it has a short breeding history (PATEL *et al.* 2014).

Short fiber content (SFC %) is measured as the percentage of fibers that are smaller than one-half inch (CUI *et al.* 2003). High SFC can cause problems in spinning, increasing irregularity in yarn, increasing wastage, number of neps and reducing strength (KRIFA AND ETHRIDGE 2006; ULLOA 2006; THIBODEAUX *et al.* 2008; CAI *et al.* 2011).

Genomic resources and QTL mapping

The basic steps to identify QTLs for a trait of interest are 1) develop a mapping population like F_2, backcross population, RILs (Recombinant Inbred Lines) or NILs (Near Isogenic Lines) 2) Phenotype the trait of interest 3) Genotype markers covering the whole genome, to permit a genome-wide scan for QTLs related to the trait.

Visible phenotypic markers, generally mutants, were used to develop the first map of the cotton genome (PERCY *et al.* 2015). Different types of markers like RFLP (Restriction Fragment Length Polymorphism), AFLP (Amplified Fragment Length Polymorphism), RAPD (Random Amplified Polymorphic DNA), SSR (simple sequence repeats), and SNP (Single Nucleotide Polymorphism) have been used to develop cotton linkage maps and conduct QTL mapping studies (RONG *et al.* 2004; LACAPE *et al.* 2005; RONG *et al.* 2005; RONG *et al.* 2007; YU *et al.* 2007; QIN *et al.* 2008a; XIAO *et al.* 2009; LACAPE *et al.* 2010; GORE *et al.* 2014; HULSE-KEMP *et al.* 2015; LI *et al.* 2016a; MEENA *et al.* 2017). In the 1990's RFLP markers were used to develop genetic maps (REINISCH *et al.* 1994; SHAPPLEY *et al.* 1996; SHAPPLEY *et al.* 1998). Later, a recombination map was developed using 3347 sequence-tagged site loci with average distance less than 2cM between markers (RONG *et al.* 2004). With the availability of SSRs, greater DNA polymorphism helped to construct several genetic maps (CHEN *et al.* 2008; YU *et al.* 2011). A

high density genetic map for cotton has been constructed using SSR and SNP in a RIL population derived from an interspecific cross between *G. hirsutum* and *G. barbadense* (JOHN et al. 2012). This map covers 3380 centiMorgans (cM) of the cotton genome (AD) by 2072 loci develop from 1825 SSRs and 247 SNPs.

The first QTL mapping study in cotton identified fourteen QTLs associated to different fiber traits (JIANG *et al.* 1998). QTL mapping for fiber traits in several other interspecific and intraspecific crosses has followed (ULLOA AND MEREDITH JR 2000; PATERSON *et al.* 2003; ZHANG *et al.* 2003a; MEI *et al.* 2004; RONG *et al.* 2004; LACAPE *et al.* 2005; SHEN *et al.* 2006; RONG *et al.* 2007; SHEN *et al.* 2007; QIN *et al.* 2008a; LACAPE *et al.* 2010; ZHANG *et al.* 2011; YUAN *et al.* 2014b; LI *et al.* 2016a; ZHANG *et al.* 2016b). Four meta-analyses have been performed to identify QTL hotspot and clusters for agronomically important traits (RONG *et al.* 2007; LACAPE *et al.* 2010; SAID *et al.* 2013; SAID *et al.* 2015). Recent meta-analysis used published information from 1,075 QTL and 1,059 QTL, respectively, on intraspecific crosses between *G. hirsutum* genotypes and interspecific cross between *G. hirsutum* × *G. barbadense* to identify QTL clusters and hotspots important for yield components, biotic and abiotic stress, and fiber attributes (SAID *et al.* 2015).

With availability of more genetic markers and cheaper sequencing techniques such as Genotyping by Sequencing (GBS) (DAVEY *et al.* 2011; KIM *et al.* 2016) and whole genome sequencing of reference sequences for *G. raimondii* (PATERSON et al. 2012), *G. arboreum* (LI et al. 2014a), *G. hirsutum* (LI *et al.* 2015; ZHANG *et al.* 2015), fine mapping of fiber QTLs has become easier. A QTL, *qMi-C14*, related to Root-Knot Nematode (RKN) resistance was fine mapped and 20 candidate genes were identified in the 2.3 Mb flanking region (KUMAR *et al.* 2016a). A fiber length QTL, *qFL-chr1*, was fine mapped using 1672 BC_4F_2 genotypes and 24

PCR based polymorphic markers (XU *et al.* 2017). The region was narrowed down to 0.9 cM or 2.38Mb and gene expression of two candidate genes in the region showed positive correlation with fiber length (XU *et al.* 2017). An *LRR RLK* gene was identified residing in a stable QTL, qFS07.1, associated with fiber strength (FANG *et al.* 2017). A major QTL in the vicinity of the T_1 locus on chromosome 6 was identified controlling multiple yield components, fiber quality traits and resistance to spiny bollworm (*Earias spp.*) (WAN *et al.* 2007). A large population of 6975 F_2 individuals was used to fine map the QTL to 5.3 Mb of chromosome A06 of *G. hirsutum*. By using RNA-Seq and RT-PCR, three putative genes were identified that might explain its diverse effects on yield and different fiber quality traits (LIU *et al.* 2016). The next obvious step to these fine mapping efforts will be to functionally validate the candidate genes and transfer them to elite germplasm.

Natural and Artificial Mutants

Cotton mutants containing fiber anomalies are excellent tool to decipher the complex process of fiber development and identify genes involved. Several natural and artificial cotton mutants showing striking differences from wild type have been discovered, genetically mapped and/or used to identify and map other genes related to fiber development (RONG *et al.* 2005; HINCHLIFFE *et al.* 2011; PATEL *et al.* 2014; JIANG *et al.* 2015; PERCY *et al.* 2015; HINCHLIFFE *et al.* 2016; MA *et al.* 2016; PATEL *et al.* 2016; THYSSEN *et al.* 2016; WAN *et al.* 2016; NAOUMKINA *et al.* 2017; THYSSEN *et al.* 2017).

One of the earliest discovered types of natural cotton mutants affected color, of lint, fuzz, petal spots, and leaves (SHOEMAKER 1909; MCLENDON 1912). A total of 157 morphological characteristic loci have been reported in a recent survey of cotton species (PERCY *et al.* 2015). Rong *et al.* (2005) developed six populations and mapped seven fiber mutants. They were *Li*$_1$

(KOHEL *et al.* 1992), *Li*$_2$ (NARBUTH AND KOHEL 1990), *N*$_1$ (GRIFFEE AND LIGON 1929b) and *Fbl* (KEARNEY AND HARRISON 1927) which were genetically dominant and *n*$_2$ (HARLAND 1929), *sma-4(h*$_a$*)*, and *sma-4(f*$_z$*)* (BEASLEY AND EGLI 1977) which were recessive. Today several natural mutants have been fine mapped and candidate genes have been identified, some of which have also been validated through functional analysis. The 22-bp deletion in a pentatricopeptide repeat (PPR) gene was linked to the *im* (immature fiber) mutant (THYSSEN *et al.* 2016). Natural Antisense Transcripts (NATs) produced from GhMML3_A12, a MYBMIXTA-like transcription factor 3 /GhMYB25- like gene, were identified as the cause of a naked seed mutant (N1) (WAN *et al.* 2016). Gly65Val substitution in *GhACT_LI1* (actin gene), was identified as the causal mutation of the *Li*$_1$ phenotype (THYSSEN *et al.* 2017). A 33-bp tandem duplication in the promoter region of *LATE MERISTEM IDENTITY1-D1b* (*GhLMI1-D1b*) causes higher expression of the gene and okra leaf shape in cotton (ANDRES *et al.* 2017). To identify genetic sources of brown lint fiber, 595 F_2 progenies were screened using SNPs identified between parental lines. An inversion of 1.4 Mb located just upstream of transcription factor *GhTT2_A07* causes elevated expression of the gene and was tightly linked with colored fiber (HINCHLIFFE *et al.* 2016).

Artificial or man-made mutants are generally obtained through physical, chemical or insertional mutagenesis (BALCELLS *et al.* 1991; AULD *et al.* 1992; MALUSZYNSKI *et al.* 1995; AULD *et al.* 1998; AN *et al.* 2005; CHOPRA 2005; GAO *et al.* 2006; AULD *et al.* 2009; PATHIRANA 2011). Physical mutagens like X-ray, alpha, beta, gamma, and ultraviolet radiation generally cause large and small chromosomal rearrangements (KODYM AND AFZA 2003; AULD *et al.* 2009; KRISHNAN *et al.* 2009; MBA *et al.* 2010). Examples of chemical mutagens are ethylmethane sulphonate (EMS), methylmethane sulphonate (MMS), nitrosoguanidine (NTG, NG, MNNG),

hydrogen fluoride (HF), sodium azide, N-methyl-N-nitrosourea (MNU), diethylsulfate (DES), and hydroxylamine which generally cause point mutations with rare chromosomal aberrations (AWAN *et al.* 1980; AULD *et al.* 2009; PATHIRANA 2011; DE SERRES AND HOLLAENDER 2012; TALEBI *et al.* 2012). Transposons and T-DNA are used to perform insertional mutation, with the advantage that the mutation site can be easily identified using sequence information, in a similar fashion to gene tagging by endogenous transposable elements (JEON *et al.* 2000; AHLOOWALIA AND MALUSZYNSKI 2001; JEONG *et al.* 2002; ALONSO *et al.* 2003; PATHIRANA 2011). Several phenotypic abnormalities in cotton have been discovered through different mutation techniques such as cytoplasmic sterility (NGEMATOV *et al.* 1975), photoperiod insensitivity (RAUT *et al.* 1971), trichome variation (PATEL *et al.* 2016), naked seed (BECHERE *et al.* 2009a; PATEL *et al.* 2014; KONG DEPEI 2017), imazamox herbicide tolerance (BECHERE *et al.* 2009b), glandless plants (HUSSEIN *et al.* 1982; MEHETRE AND THOMBRE 1983), short fiber (KONG DEPEI 2017; NAOUMKINA *et al.* 2017), albino cotyledon and leaves, red or violet leaves and stems, and multilayered bracts (Kong Depei 2017). Striking improvement in yield components and fiber quality was also found through physical and chemical mutagenesis (MEHETRE AND THOMBRE 1983; AULD *et al.* 2000; AULD 2000; SHAMSUZZAMAN *et al.* 2003; BECHERE *et al.* 2007a; BROWN *et al.* 2012; PATEL *et al.* 2014). Although a vast source of phenotypic anomalies generated by artificial mutation are available, scarce effort has been taken to understand genetics of the cause. Only one gene expression study has been performed to study changes in gene expression of the li_y short fiber mutant. Recently, a TILLING (Targeting Induced Local Lesions IN Genomes) database in cotton was developed by targeted sequencing of M_2 mutant lines (ASLAM *et al.* 2016). Such a database will ease the effort to identify causal mutations for striking phenotypes in mutant lines.

Review of candidate genes found in the *Li₂* region

GhIRX7_D belongs to the gene family, glycosyltransferase family 47 (ZHONG AND YE 2003; GESHI *et al.* 2010). *IRX7* or *FRA8 (FRAGILE FIBER 8)* is involved in biosynthesis of the hemicellulose glucuronoxylan that is an essential component of secondary cell walls and is particularly activated during secondary wall thickening in fibers and vessels (BROWN *et al.* 2005; ZHONG *et al.* 2005; LI *et al.* 2014b). Disrupting the gene causes reduction in secondary wall thickness, decline in amounts of cellulose and xylan, collapse of xylem vessels, decrease in stem strength, and dwarf phenotype (BROWN *et al.* 2005; ZHONG *et al.* 2005; BROWN *et al.* 2007; LEE *et al.* 2007a).

GhETO1_D encodes an ethylene-overproduction protein. In Arabidopsis, *ETO1* negatively regulates 1-aminocyclopropane-1-carboxylic acid synthase (ACS), a protein that acts as catalyst in the rate-determining step of ethylene biosynthesis (WANG *et al.* 2004; YOSHIDA *et al.* 2005; CHRISTIANS *et al.* 2009)**.** Ethylene plays an important role in fiber elongation but excessive production of ethylene might curb fiber development (SHI *et al.* 2006). Ethylene was also found to enhance cotton fiber and *Arabidopsis* root hair growth, but a tenfold increase in ethylene concentration is seen in loss-of-function *eto1* mutations, further causing shorter seedlings, smaller leaves, reduced root and petiole lengths and inflorescence sizes (CHAE *et al.* 2003; WANG *et al.* 2004; CHRISTIANS *et al.* 2009; PANG *et al.* 2010; LUO *et al.* 2014)**.** Application of abscisic acid (ABA) enormously inhibited root growth in *eto1* mutants by promoting ethylene biosynthesis (LUO *et al.* 2014). Elevated levels of ABA have been observed during fiber development of *Li₂* mutant lines, thus an increased level of ABA with reduced expression level of *GhETO1_D* during fiber elongation might have contributed to the *Li₂* phenotype (CHEN *et al.* 1997; GILBERT *et al.* 2013).

GhUBE11_Db encodes Ubiquitin-activating enzyme E1 that catalyzes the first step of three consecutive enzymatic cascades in a ubiquitination reaction which is important for biological processes such as embryogenesis, plant growth and development, hormone signaling, response to environmental stress, and senescence (HATFIELD *et al.* 1997; MOON *et al.* 2004; PICKETT 2007). Genes participating in ubiquitin-mediated protein degradation are highly up-regulated during fiber development (ZHANG *et al.* 2003b; AL-GHAZI *et al.* 2009; HO *et al.* 2010). Interestingly, regulation of ethylene biosynthesis through ubiquitin-mediated protein degradation is assisted by ethylene-overproduction protein (LYZENGA AND STONE 2012).

GhEXPA8 belongs to the expansin superfamily, wall-loosening proteins which help with cell expansion (SAMPEDRO AND COSGROVE 2005). Multiple studies have discovered the role of expansin genes during fiber development, especially during the time of fiber elongation (ORFORD AND TIMMIS 1998; JI *et al.* 2003; SHI *et al.* 2006; SHAN *et al.* 2014; BAJWA *et al.* 2015). *GhEXPA1* is one of the two genes that are directly down stream of GhHOX3, a homeodomain transcription factor involved in fiber elongation (SHAN *et al.* 2014). *GhEXPA8* was introduced into a cotton variety (NIAB 846) through Agrobacterium-mediated gene transformation. Compared to NIAB 846, stable and significant improvement in fiber length and micronaire was seen in transgenic cotton plants with higher expression of *GhEXPA8* (BAJWA *et al.* 2015).

GhUGT87A1_Da, *GhUGT87A2_D* and *GhUGT87A1_Db* are three genes belonging to Glycosyltransferase Family 1, the largest GT family (YONEKURA-SAKAKIBARA AND HANADA 2011; HUANG *et al.* 2015). UGTs (UDP-glucuronosyltransferases) participate in germination, vegetative growth, flowering, fiber development, response and signaling to biotic and abiotic stress, regulation of hormonal homeostasis, biosynthesis of secondary metabolites and their stability, and detoxification and fragmentation of endogenous compounds and xenobiotics (TAI

et al. 2008; WANG *et al.* 2012; JUAN *et al.* 2015; LI *et al.* 2016b; MAMOON REHMAN *et al.* 2016). In soybean, genome-wide analysis revealed abundant expression of UGTs during seed, shoot, inflorescence, cotyledon, meristem and other development stages (MAMOON REHMAN *et al.* 2016). *GhUGT1* might be involved in responding to osmotic stress in plants and was highly expressed in fibers and roots (TAI *et al.* 2008). *UGT87A1* and *UGT87A2* are nearly identical, and might have similar function and redundant effects. Metabolic analysis in overexpressed *UGT87A2* lines suggest the possible function of *UGT87A2* is in ascorbic acid homeostasis or cell wall biosynthesis (SAINT PAUL 2010). Compare to ugt87a2 knockout mutant line and wild type, transgenic lines with overexpression of *UGT87A2* (*87A2OE*) showed notable improvement in seed germination, survival rate and root length during abiotic stress (LI *et al.* 2016b). *UGT87A2* has ROS (reactive oxygen species) scavenging activity, thus reduced levels of superoxide, H_2O_2 and cell damage was observed in *87A2OE* compare to *ugt87a2* knock out mutant line and wild type (LI *et al.* 2016b). This indicates that one of the functions of *UGT87A2* is to regulation of ROS like H_2O_2 activities for normal growth and development of plants. Hydrogen peroxide (H_2O_2) is important for cell loosening and elongation (LISZKAY *et al.* 2004). In cotton, it has been reported to promote fiber elongation and differentiation of secondary cell walls in cotton fiber (POTIKHA *et al.* 1999; LI *et al.* 2007; QIN *et al.* 2008b; MEI *et al.* 2009). Regulation of H_2O_2 or ROS is required to avoid the arrest the fiber elongation process, as higher levels of H_2O_2 could trigger secondary cell wall biosynthesis (LI *et al.* 2007; CHAUDHARY *et al.* 2008; HOVAV *et al.* 2008b; CHAUDHARY *et al.* 2009b; GUO *et al.* 2016). Hovav *et al.* (2008) found elevated concentration of H_2O_2 in *G. herbaceum* (A genome) and *G. longicalyx* (F genome) during early fiber elongation, but the latter was not able to reduce the H_2O_2 level and thus reduced fiber elongation by enhancing stress conditions and triggering secondary cell wall biosynthesis. A

similar result was also seen in wild (short fiber) and domesticated (long fiber) cotton species (CHAUDHARY *et al.* 2008; CHAUDHARY *et al.* 2009b). Thus, it seems that *UGT87A* mediated scavenging activity of H_2O_2 during fiber elongation is important for fiber expansion.

References

Abdel-Salam, M., M. Negm and C. S. Ardabb, 2009 The Egyptian cotton, current constraints and future opportunities. Textile Industries Holding Co., Modern Press-Alexandria-Egypt.

Ahloowalia, B., and M. Maluszynski, 2001 Induced mutations–A new paradigm in plant breeding. Euphytica 118**:** 167-173.

Ai, X., Y. Liang, J. Wang, J. Zheng, Z. Gong *et al.*, 2017 Genetic diversity and structure of elite cotton germplasm (Gossypium hirsutum L.) using genome-wide SNP data. Genetica**:** 1-8.

Al-Ghazi, Y., S. Bourot, T. Arioli, E. S. Dennis and D. J. Llewellyn, 2009 Transcript profiling during fiber development identifies pathways in secondary metabolism and cell wall structure that may contribute to cotton fiber quality. Plant Cell Physiol 2009.

Alonso, J. M., A. N. Stepanova, T. J. Leisse, C. J. Kim, H. Chen *et al.*, 2003 Genome-wide insertional mutagenesis of Arabidopsis thaliana. Science 301**:** 653-657.

An, G., D.-H. Jeong, K.-H. Jung and S. Lee, 2005 Reverse Genetic Approaches for Functional Genomics of Rice. Plant Molecular Biology 59**:** 111-123.

Aslam, U., H. M. Cheema, S. Ahmad, I. A. Khan, W. Malik *et al.*, 2016 COTIP: Cotton TILLING Platform, a resource for plant improvement and reverse genetic studies. Frontiers in plant science 7.

Auld, D., E. Bechere, M. Ethridge, W. Becker, E. Hequet *et al.*, 2000 Registration of TTU 202-1107-B and TTU 271-2155-C mutant germplasm lines of upland cotton with improved fiber quality. Crop science 40**:** 1835-1835.

Auld, D., M. Ethridge, J. Dever and P. Dotray, 1998 Chemical mutagenesis as a tool in cotton improvement. Beltwide Cotton Conferences (USA).

Auld, D., M. Heikkinen, D. Erickson, J. Sernyk and J. Romero, 1992 Rapeseed mutants with reduced levels of polyunsaturated fatty acids and increased levels of oleic acid. Crop science 32**:** 657-662.

Auld, D., G. G. Light, M. Fokar, E. Bechere and R. D. Allen, 2009 Mutagenesis Systems for Genetic Analysis of Gossypium. Genetics and Genomics of Cotton**:** 1-18.

Auld, D. L., 2000 Registration of TTU 202-1107-B and TTU 271-2155-C mutant germplasm lines of upland cotton with improved fiber quality. Crop science 40**:** 1835-1836.

Awan, M. A., C. Konzak, J. Rutger and R. Nilan, 1980 Mutagenic effects of sodium azide in rice. Crop Science 20**:** 663-668.

Bajwa, K. S., A. A. Shahid, A. Q. Rao, A. Bashir, A. Aftab *et al.*, 2015 Stable transformation and expression of GhEXPA8 fiber expansin gene to improve fiber length and micronaire value in cotton. Frontiers in Plant Science 6**:** 838.

Balcells, L., J. Swinburne and G. Coupland, 1991 Transposons as tools for the isolation of plant genes. Trends in Biotechnology 9**:** 31-37.

Barb, J. G., J. E. Bowers, S. Renaut, J. I. Rey, S. J. Knapp *et al.*, 2014 Chromosomal evolution and patterns of introgression in Helianthus. Genetics 197**:** 969-979.

Basra, A. S., and C. P. Malik, 1984 Development of the cotton fiber. Int. Rev. Cytol 89**:** 65-113.

Beasley, C., and E. Egli, 1977 Fiber production in vitro from a conditional fiberless mutant of cotton. Developmental biology 57**:** 234-237.

Bechere, E., D. Auld, R. Cantrell, E. Hequet, M. Krifa *et al.*, 2007a Registration of TTU 0774-3-3 and TTU 0808-1-6-1 upland cotton germplasm lines with improved fiber length and strength. Journal of plant registrations 1**:** 58-59.

Bechere, E., D. Auld and E. Hequet, 2009a Development of 'naked-tufted'seed coat mutants for potential use in cotton production. Euphytica 167**:** 333-339.

Bechere, E., D. Auld, M. Krifa, C. W. Smith and R. Cantrell, 2011 Registration of TTU 0782, an Upland Cotton Germplasm Line with Superior Fiber Quality. Journal of plant registrations 5**:** 207-210.

Bechere, E., D. L. Auld, R. G. Cantrell, E. Hequet, M. Krifa *et al.*, 2007b Registration of TTU 0774-3-3 and TTU 0808-1-6-1 Upland Cotton Germplasm Lines with Improved Fiber Length and Strength. J. Plant Reg. 1**:** 58-59.

Bechere, E., D. L. Auld, P. Dotray and H. Kebede, 2010 Registration of Four Upland Cotton (Gossypium hirsutum L.) Genetic Stock Mutants with Tolerance to Imazamox.

Bechere, E., D. L. Auld, P. A. Dotray, L. V. Gilbert and H. Kebede, 2009b Imazamox Tolerance in Mutation-Derived Lines of Upland Cotton.

Boopathi, N. M., and L. V. Hoffmann, 2016 Genetic Diversity, Erosion, and Population Structure in Cotton Genetic Resources, pp. 409-438 in *Genetic Diversity and Erosion in Plants*. Springer.

Bradow, J. M., and G. H. Davidonis, 2000 Quantitation of fiber quality and the cotton production-processing interface: a physiologist's perspective. J. Cotton Sci 4**:** 34-64.

Brown, D. M., F. Goubet, V. W. Wong, R. Goodacre, E. Stephens *et al.*, 2007 Comparison of five xylan synthesis mutants reveals new insight into the mechanisms of xylan synthesis. The Plant Journal 52**:** 1154-1168.

Brown, D. M., L. A. Zeef, J. Ellis, R. Goodacre and S. R. Turner, 2005 Identification of novel genes in Arabidopsis involved in secondary cell wall formation using expression profiling and reverse genetics. The Plant Cell 17**:** 2281-2295.

Brown, I., C. W. Smith, D. Auld, S. Hague, E. F. Hequet *et al.*, 2012 Registration of TAM 94L-25-M24, TAM 94L-25-M25, and TAM 94L-25-M30 Mutant Upland Cotton Germplasm with Improved Fiber Length and Strength. Journal of Plant Registrations 6**:** 195-199.

Brubaker, C., F. Bourland and J. Wendel, 1999a The origin and domestication of cotton. Cotton: Origin, History, Technology, and Production. John Wiley & Sons, New York**:** 3–32.

Brubaker, C. L., F. Bourland and J. F. Wendel, 1999b The origin and domestication of cotton. Cotton: Origin, history, technology, and production. John Wiley & Sons, New York**:** 3-31.

Brubaker, C. L., and J. F. Wendel, 1994 Reevaluating the origin of domesticated cotton (Gossypium hirsutum; Malvaceae) using nuclear restriction fragment length polymorphisms (RFLPs). American journal of botany**:** 1309-1326.

Cai, Y., X. Cui, J. Rodgers, D. Thibodeaux, V. Martin *et al.*, 2011 An investigation on different parameters used for characterizing short cotton fibers. Textile Research Journal 81**:** 239-246.

Cantrell, R., C. Roberts and C. Waddell, 2000 Registration of Acala 1517-99 Cotton. Crop Science 40**:** 1200-1200.

Chae, H. S., F. Faure and J. J. Kieber, 2003 The eto1, eto2, and eto3 Mutations and Cytokinin Treatment Increase Ethylene Biosynthesis in Arabidopsis by Increasing the Stability of ACS Protein. The Plant Cell 15**:** 545-559.

Chapagain, A. K., A. Y. Hoekstra, H. H. G. Savenije and R. Gautam, 2006 The water footprint of cotton consumption: An assessment of the impact of worldwide consumption of cotton products on the water resources in the cotton producing countries. Ecological Economics 60**:** 186-203.

Chaudhary, B., R. Hovav, L. Flagel, R. Mittler and J. F. Wendel, 2009 Parallel expression evolution of oxidative stress-related genes in fiber from wild and domesticated diploid and polyploid cotton (Gossypium). BMC Genomics 10.

Chaudhary, B., R. Hovav, R. Rapp, N. Verma, J. A. Udall *et al.*, 2008 Global analysis of gene expression in cotton fibers from wild and domesticated Gossypium barbadense. Evolution & development 10**:** 567-582.

Chee, P., E. Lubbers, O. May, J. Gannaway and A. H. Paterson, 2004 Changes in genetic diversity of the U.S. Upland cotton, pp. in *Beltwide Cotton Conference*. National Cotton Council, San Antonio, TX.

Chen, J.-G., X.-M. Du, X. Zhou and H.-Y. Zhao, 1997 Levels of cytokinins in the ovules of cotton mutants with altered fiber development. J Plant Growth Regul 16.

Chen, L., Z.-S. Zhang, M.-C. Hu, W. Wang, J. Zhang *et al.*, 2008 Genetic linkage map construction and QTL mapping for yield and fiber quality in upland cotton (Gossypium hirsutum L.). Acta Agron Sin 34**:** 1199-1205.

Chen, Z., K. Feng, C. E. Grover, P. Li, F. Liu *et al.*, 2016 Chloroplast DNA structural variation, phylogeny, and age of divergence among diploid cotton species. PloS one 11**:** e0157183.

Chen, Z. J., and X. Guan, 2011 Auxin boost for cotton. Nat Biotech 29**:** 407-409.

Chen, Z. J., B. E. Scheffler, E. Dennis, B. A. Triplett, T. Zhang *et al.*, 2007 Toward sequencing cotton (Gossypium) genomes. Plant physiology 145**:** 1303-1310.

Chopra, V., 2005 Mutagenesis: Investigating the process and processing the outcome for crop improvement. CURRENT SCIENCE-BANGALORE- 89**:** 353.

Christians, M. J., D. J. Gingerich, M. Hansen, B. M. Binder, J. J. Kieber *et al.*, 2009 The BTB ubiquitin ligases ETO1, EOL1 and EOL2 act collectively to regulate ethylene biosynthesis in Arabidopsis by controlling type-2 ACC synthase levels. The Plant Journal 57**:** 332-345.

Clement, J., G. Constable, W. Stiller and S. Liu, 2012 Negative associations still exist between yield and fibre quality in cotton breeding programs in Australia and USA. Field crops research 128**:** 1-7.

Clement, J. D., G. A. Constable, W. N. Stiller and S. M. Liu, 2015 Early generation selection strategies for breeding better combinations of cotton yield and fibre quality. Field Crops Research 172**:** 145-152.

Constable, G., D. Llewellyn, S. A. Walford and J. D. Clement, 2015 Cotton Breeding for Fiber Quality Improvement, pp. 191-232 in *Industrial Crops: Breeding for BioEnergy and Bioproducts*, edited by V. M. V. Cruz and D. A. Dierig. Springer New York, New York, NY.

Cotton, F. T. U., 1998 COTTON IMPROVEMENT.

Cui, X., T. A. Calamari, K. Q. Robert, J. B. Price and M. D. Watson, 2003 Measuring the short fiber content of cotton. Textile research journal 73**:** 891-895.

Culp, T., D. Harrell and T. Kerr, 1979 Some genetic implications in the transfer of high fiber strength genes to upland cotton. Crop Science 19**:** 481-484.

Davey, J. W., P. A. Hohenlohe, P. D. Etter, J. Q. Boone, J. M. Catchen *et al.*, 2011 Genome-wide genetic marker discovery and genotyping using next-generation sequencing. Nature reviews. Genetics 12**:** 499.

De Serres, F. J., and A. Hollaender, 2012 *Chemical mutagens: Principles and methods for their detection*. Springer Science & Business Media.

Dong, Z., Y. Shi, J. Zhang, S. Wang, J. Li *et al.*, 2009 Molecular marker-assisted selection and pyramiding breeding of major QTLs for cotton fiber length. Cotton Sci 21**:** 279-283.

Doudna, J. A., and E. Charpentier, 2014 The new frontier of genome engineering with CRISPR-Cas9. Science 346**:** 1258096.

Endo, T., B. Chiba, K. Wagatsuma, K. Saeki, T. Ando *et al.*, 2016 Detection of QTLs for cold tolerance of rice cultivar 'Kuchum'and effect of QTL pyramiding. Theoretical and applied genetics 129**:** 631-640.

Endrizzi, J., E. Turcotte and R. Kohel, 1985 Genetics, cytogenetics, and evolution of Gossypium. Adv. Genet 23**:** 271-375.

Fang, L., R. Tian, X. Li, J. Chen, S. Wang *et al.*, 2014 Cotton fiber elongation network revealed by expression profiling of longer fiber lines introgressed with different Gossypium barbadense chromosome segments. BMC Genomics 15**:** 838.

Fang, L., Q. Wang, Y. Hu, Y. Jia, J. Chen *et al.*, 2017a Genomic analyses in cotton identify signatures of selection and loci associated with fiber quality and yield traits. Nature genetics 49**:** 1089.

Fang, X., X. Liu, X. Wang, W. Wang, D. Liu *et al.*, 2017b Fine-mapping qFS07.1 controlling fiber strength in upland cotton (Gossypium hirsutum L.). Theoretical and Applied Genetics**:** 1-12.

Fukuoka, S., N. Saka, Y. Mizukami, H. Koga, U. Yamanouchi *et al.*, 2015 Gene pyramiding enhances durable blast disease resistance in rice. Scientific reports 5**:** 7773.

Gallagher, J. P., C. E. Grover, K. Rex, M. Moran and J. F. Wendel, 2017 A New Species of Cotton from Wake Atoll, Gossypium stephensii (Malvaceae). Systematic Botany 42**:** 115-123.

Gao, W., Z. J. Chen, Z. Y. John, R. J. Kohel, J. E. Womack *et al.*, 2006 Wide-cross whole-genome radiation hybrid mapping of the cotton (Gossypium barbadense L.) genome. Molecular genetics and genomics 275**:** 105-113.

Geshi, N., J. Harholt, Y. Sakuragi, J. Krüger Jensen and H. V. Scheller, 2010 Glycosyltransferases of the GT47 Family, pp. 265-283 in *Annual Plant Reviews*. Wiley-Blackwell.

Gilbert, M. K., J. M. Bland, J. M. Shockey, H. Cao, D. J. Hinchliffe *et al.*, 2013 A transcript profiling approach reveals an abscisic acid-specific glycosyltransferase (UGT73C14) induced in developing fiber of Ligon lintless-2 mutant of cotton (Gossypium hirsutum L.). PloS one 8**:** e75268.

Gore, M. A., D. D. Fang, J. A. Poland, J. Zhang, R. G. Percy *et al.*, 2014 Linkage map construction and quantitative trait locus analysis of agronomic and fiber quality traits in cotton. The Plant Genome 7.

Greene, E. A., C. A. Codomo, N. E. Taylor, J. G. Henikoff, B. J. Till *et al.*, 2003 Spectrum of chemically induced mutations from a large-scale reverse-genetic screen in Arabidopsis. Genetics 164**:** 731-740.

Griffee, F., and L. Ligon, 1929 Occurrence of "lintless" cotton plants and the inheritance of the character "lintless. J Am Soc Agron 21**:** 711-717.

Grover, C. E., H. Kim, R. A. Wing, A. H. Paterson and J. F. Wendel, 2007 Microcolinearity and genome evolution in the AdhA region of diploid and polyploid cotton (Gossypium). The Plant Journal 50**:** 995-1006.

Guan, X., M. Pang, G. Nah, X. Shi, W. Ye *et al.*, 2014 miR828 and miR858 regulate homoeologous MYB2 gene functions in Arabidopsis trichome and cotton fibre development. Nature communications 5**:** 3050.

Guo, K., X. Du, L. Tu, W. Tang, P. Wang *et al.*, 2016a Fibre elongation requires normal redox homeostasis modulated by cytosolic ascorbate peroxidase in cotton (Gossypium hirsutum). Journal of Experimental Botany 67**:** 3289-3301.

GUO, X.-h., C.-p. CAI, D.-d. YUAN, R.-s. ZHANG, J.-l. XI *et al.*, 2016b Development and identification of Verticillium wilt-resistant upland cotton accessions by pyramiding QTL related to resistance. Journal of Integrative Agriculture 15**:** 512-520.

Haigler, C., 2007 Substrate supply for cellulose synthesis and its stress sensitivity in the cotton fiber. Cellulose: Molecular and Structural Biology**:** 147-168.

Haigler, C., 2010 Physiological and anatomical factors determining fiber structure and utility. Physiology of cotton**:** 33-47.

Haigler, C., L. Betancur, M. Stiff and J. Tuttle, 2012 Cotton fiber: a powerful single-cell model for cell wall and cellulose research. Frontiers in Plant Science 3.

Haigler, C., B. Singh, G. Wang and D. Zhang, 2009 Genomics of cotton fiber secondary wall deposition and cellulose biogenesis. Genetics and Genomics of Cotton**:** 1-33.

Harland, S., 1929 The work of the genetics department of the Cotton Research Station, Trinidad. Emp Cotton Grow Rev 6**:** 304-314.

Hatfield, P. M., M. M. Gosink, T. B. Carpenter and R. D. Vierstra, 1997 The ubiquitin-activating enzyme (E1) gene family in Arabidopsis thaliana. The Plant Journal 11**:** 213-226.

Hattori, Y., K. Nagai, S. Furukawa, X.-J. Song, R. Kawano *et al.*, 2009 The ethylene response factors SNORKEL1 and SNORKEL2 allow rice to adapt to deep water. Nature 460**:** 1026.

Herring, A. D., D. L. Auld, M. D. Ethridge, E. F. Hequet, E. Bechere *et al.*, 2004 Inheritance of fiber quality and lint yield in a chemically mutated population of cotton. Euphytica 136**:** 333-339.

Hinchliffe, D. J., B. D. Condon, G. Thyssen, M. Naoumkina, C. A. Madison *et al.*, 2016 The GhTT2_A07 gene is linked to the brown colour and natural flame retardancy phenotypes of Lc1 cotton (Gossypium hirsutum L.) fibres. Journal of experimental botany 67**:** 5461-5471.

Hinchliffe, D. J., R. B. Turley, M. Naoumkina, H. J. Kim, Y. Tang *et al.*, 2011 A combined functional and structural genomics approach identified an EST-SSR marker with complete linkage to the Ligon lintless-2 genetic locus in cotton (Gossypium hirsutum L.). BMC genomics 12**:** 445.

Ho, M.-H., S. Saha, J. N. Jenkins and D.-P. Ma, 2010 Characterization and Promoter Analysis of a Cotton RING-Type Ubiquitin Ligase (E3) Gene. Molecular Biotechnology 46**:** 140-148.

Hovav, R., J. A. Udall, B. Chaudhary, E. Hovav, L. Flagel *et al.*, 2008a The evolution of spinnable cotton fiber entailed prolonged development and a novel metabolism. PLoS Genet 4.

Hovav, R., J. A. Udall, B. Chaudhary, E. Hovav, L. Flagel *et al.*, 2008b The Evolution of Spinnable Cotton Fiber Entailed Prolonged Development and a Novel Metabolism. PLoS Genet 4**:** e25.

Hovav, R., J. A. Udall, E. Hovav, R. Rapp, L. Flagel *et al.*, 2008c A majority of cotton genes are expressed in single-celled fiber. Planta 227**:** 319-329.

Hu, H., X. He, L. Tu, L. Zhu, S. Zhu *et al.*, 2016 GhJAZ2 negatively regulates cotton fiber initiation by interacting with the R2R3-MYB transcription factor GhMYB25-like. The Plant Journal 88**:** 921-935.

Huang, G.-Q., S.-Y. Gong, W.-L. Xu, W. Li, P. Li *et al.*, 2013 A fasciclin-like arabinogalactan protein, GhFLA1, is involved in fiber initiation and elongation of cotton. Plant physiology 161**:** 1278-1290.

Huang, J., F. Chen, S. Wu, J. Li and W. Xu, 2016 Cotton GhMYB7 is predominantly expressed in developing fibers and regulates secondary cell wall biosynthesis in transgenic Arabidopsis. Science China Life Sciences 59**:** 194-205.

Huang, J., C. Pang, S. Fan, M. Song, J. Yu *et al.*, 2015 Genome-wide analysis of the family 1 glycosyltransferases in cotton. Molecular Genetics and Genomics 290**:** 1805-1818.

Hulse-Kemp, A. M., J. Lemm, J. Plieske, H. Ashrafi, R. Buyyarapu *et al.*, 2015 Development of a 63K SNP Array for Cotton and High-Density Mapping of Intra-and Inter-Specific Populations of Gossypium spp. G3: Genes, Genomes, Genetics**:** g3. 115.018416.

Hussein, H., F. Al Enani and M. El Moghazi, 1982 Histological and morphological characteristics of a glandless cotton mutant induced with sodium azide. Egypt. J. Genet. Cytol 11**:** 167-174.

Hutchinson, J., 1951 Intra-specific differentiation in Gossypium hirsutum. Heredity 5**:** 161-193.

Iqbal, M., O. Reddy, K. El-Zik and A. Pepper, 2001 A genetic bottleneck in the'evolution under domestication'of upland cotton Gossypium hirsutum L. examined using DNA fingerprinting. TAG Theoretical and Applied Genetics 103**:** 547-554.

Islam, M. S., L. Zeng, G. N. Thyssen, C. D. Delhom, H. J. Kim *et al.*, 2016 Mapping by sequencing in cotton (Gossypium hirsutum) line MD52ne identified candidate genes for fiber strength and its related quality attributes. Theoretical and Applied Genetics 129**:** 1071-1086.

Jeon, J. S., S. Lee, K. H. Jung, S. H. Jun, D. H. Jeong *et al.*, 2000 T-DNA insertional mutagenesis for functional genomics in rice. The Plant Journal 22**:** 561-570.

Jeong, D.-H., S. An, H.-G. Kang, S. Moon, J.-J. Han *et al.*, 2002 T-DNA insertional mutagenesis for activation tagging in rice. Plant physiology 130**:** 1636-1644.

Ji, S. J., Y. C. Lu, J. X. Feng, G. Wei, J. Li *et al.*, 2003 Isolation and analyses of genes preferentially expressed during early cotton fiber development by subtractive PCR and cDNA array. Nucleic Acids Res 31.

Jiang, Y., M. Ding, Y. Cao, F. Yang, H. Zhang *et al.*, 2015 Genetic fine mapping and candidate gene analysis of the Gossypium hirsutum Ligon lintless-1 (Li1) mutant on chromosome 22(D). Molecular Genetics and Genomics 290**:** 2199-2211.

John, M. E., and G. Keller, 1996 Metabolic pathway engineering in cotton: Biosynthesis of polyhydroxybutyrate in fiber cells. Proceedings of the National Academy of Sciences of the United States of America 93**:** 12768-12773.

John, Z. Y., R. J. Kohel, D. D. Fang, J. Cho, A. Van Deynze *et al.*, 2012 A high-density simple sequence repeat and single nucleotide polymorphism genetic map of the tetraploid cotton genome. G3: Genes, Genomes, Genetics 2**:** 43-58.

Juan, H., F. ShuLi, S. MeiZhen, P. ChaoYou, W. HengLing *et al.*, 2015 Cloning and function analysis of uridine diphosphate glycosyltransferase gene GhUGT85O1 in cotton (Gossypium hirsutum). Journal of Agricultural Biotechnology 23**:** 701-710.

Kearney, T. H., and G. J. Harrison, 1927 Inheritance of smooth seeds in cotton. J Agric Res 35**:** 193-217.

Kelly, B., N. Abidi, D. Ethridge and E. F. Hequet, 2015 Fiber to fabric. Cotton**:** 665-744.

Kelly, C. M., E. F. Hequet and J. K. Dever, 2012 Interpretation of AFIS and HVI fiber property measurements in breeding for cotton fiber quality improvement. J Cotton Sci 16**:** 1-16.

Kim, C., H. Guo, W. Kong, R. Chandnani, L.-S. Shuang *et al.*, 2016 Application of genotyping by sequencing technology to a variety of crop breeding programs. Plant Science 242**:** 14-22.

Kim, H. J., N. Murai, D. D. Fang and B. A. Triplett, 2011 Functional analysis of Gossypium hirsutum cellulose synthase catalytic subunit 4 promoter in transgenic Arabidopsis and cotton tissues. Plant science 180**:** 323-332.

Kim, H. J., and B. A. Triplett, 2001 Cotton fiber growth in planta and in vitro. Models for plant cell elongation and cell wall biogenesis. Plant Physiology 127**:** 1361-1366.

Kim, H. J., and B. A. Triplett, 2004 Characterization of GhRac1 GTPase expressed in developing cotton (Gossypium hirsutum L.) fibers. Biochimica et Biophysica Acta (BBA)-Gene Structure and Expression 1679**:** 214-221.

Kodym, A., and R. Afza, 2003 Physical and chemical mutagenesis. Plant functional genomics**:** 189-203.

Kohel, R. J., E. V. Narbuth and C. R. Benedict, 1992 Fiber Development of Ligon Lintless-2 Mutant of Cotton. Crop Science 32**:** 733-735.

Kong Depei, Q. L., Zhang Xueyan, Liu Ji, Wang Peng, Li Fuguang, 2017 Optimization of EMS Mutagenesis Condition and Screening of Mutants in Gossypium arboretum. L. Cotton Science 29**:** 336-344.

Krifa, M., and M. D. Ethridge, 2006 Compact spinning effect on cotton yarn quality: Interactions with fiber characteristics. Textile Research Journal 76**:** 388-399.

Krishnan, A., E. Guiderdoni, G. An, C. H. Yue-ie, C.-d. Han *et al.*, 2009 Mutant resources in rice for functional genomics of the grasses. Plant physiology 149**:** 165-170.

Kumar, P., Y. He, R. Singh, R. F. Davis, H. Guo *et al.*, 2016 Fine mapping and identification of candidate genes for a QTL affecting Meloidogyne incognita reproduction in Upland cotton. BMC Genomics 17**:** 567.

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

Lacape, J.-M., T.-B. Nguyen, B. Courtois, J.-L. Belot, M. Giband *et al.*, 2005 QTL Analysis of Cotton Fiber Quality Using Multiple × Backcross Generations. Crop Sci. 45**:** 123-140.

Lee, C., R. Zhong, E. A. Richardson, D. S. Himmelsbach, B. T. McPhail *et al.*, 2007a The PARVUS gene is expressed in cells undergoing secondary wall thickening and is essential for glucuronoxylan biosynthesis. Plant and cell physiology 48**:** 1659-1672.

Lee, J. J., A. W. Woodward and Z. J. Chen, 2007b Gene expression changes and early events in cotton fibre development. Annals of Botany 100**:** 1391-1401.

Li, C., Y. Dong, T. Zhao, L. Li, C. Li *et al.*, 2016a Genome-Wide SNP Linkage Mapping and QTL Analysis for Fiber Quality and Yield Traits in the Upland Cotton Recombinant Inbred Lines Population. Frontiers in Plant Science 7**:** 1356.

Li, D.-D., X.-M. Ruan, J. Zhang, Y.-J. Wu, X.-L. Wang *et al.*, 2013 Cotton plasma membrane intrinsic protein 2s (PIP2s) selectively interact to regulate their water channel activities and are required for fibre development. New Phytologist 199**:** 695-707.

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

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

Li, F., G. Fan, K. Wang, F. Sun, Y. Yuan *et al.*, 2014a Genome sequence of the cultivated cotton Gossypium arboreum. Nature genetics 46**:** 567-572.

Li, H. B., Y. M. Qin, Y. Pang, W. Q. Song, W. Q. Mei *et al.*, 2007 A cotton ascorbate peroxidase is involved in hydrogen peroxide homeostasis during fibre cell development. New Phytol 175.

Li, L., J. Huang, L. Qin, Y. Huang, W. Zeng *et al.*, 2014b Two cotton fiber-associated glycosyltransferases, GhGT43A1 and GhGT43C1, function in hemicellulose glucuronoxylan biosynthesis during plant development. Physiologia Plantarum 152**:** 367-379.

Li, P., Y. j. Li, B. Wang, H. m. Yu, Q. Li *et al.*, 2016b The Arabidopsis UGT87A2, a stress-inducible family 1 glycosyltransferase, is involved in the plant adaptation to abiotic stresses. Physiologia Plantarum.

Li, Y., L. Tu, F. A. Pettolino, S. Ji, J. Hao *et al.*, 2016c GbEXPATR, a species-specific expansin, enhances cotton fibre elongation through cell wall restructuring. Plant biotechnology journal 14**:** 951-963.

Liszkay, A., E. van der Zalm and P. Schopfer, 2004 Production of reactive oxygen intermediates (O2−, H2O2, and OH) by maize roots and their role in wall loosening and elongation growth. Plant Physiology 136**:** 3114-3123.

Liu, D., J. Zhang, X. Liu, W. Wang, D. Liu *et al.*, 2016 Fine mapping and RNA-Seq unravels candidate genes for a major QTL controlling multiple fiber quality traits at the T1 region in upland cotton. BMC Genomics 17**:** 295.

Lubbers, E., P. Chee, J. Gannaway, R. Wright, K. El-Zik *et al.*, 2004 Levels and patterns of genetic diversity in upland cotton, pp. in *Plant and Animal Genome XII Conference*, San Diego, CA.

Lubbers, E., S. Walker, L. May and P. Chee, 2006 Breeding cultivars and germplasm with enhanced yield and quality. In P. Roberts et al. (ed.) 2005 Georgia Cotton Research and Extension Reports. UGA/CPES Research – Extension Publication No. 6, University of Georgia, Athens, GA**:** 136-152.

Lubbers, E. L., and P. W. Chee, 2009 The Worldwide Gene Pool of G. hirsutum and its Improvement, pp. 23-52 in *Genetics and Genomics of Cotton*, edited by A. H. Paterson. Springer US, New York, NY.

Luo, X., Z. Chen, J. Gao and Z. Gong, 2014 Abscisic acid inhibits root growth in Arabidopsis through ethylene biosynthesis. The Plant Journal 79**:** 44-55.

Lyzenga, W. J., and S. L. Stone, 2012 Regulation of ethylene biosynthesis through protein degradation. Plant signaling & behavior 7**:** 1438-1442.

Ma, Q.-F., C.-H. Wu, M. Wu, W.-F. Pei, X.-L. Li *et al.*, 2016 Integrative transcriptome, proteome, phosphoproteome and genetic mapping reveals new aspects in a fiberless mutant of cotton. Scientific reports 6.

Machado, A., Y. Wu, Y. Yang, D. J. Llewellyn and E. S. Dennis, 2009 The MYB transcription factor GhMYB25 regulates early fibre and trichome development. The Plant Journal 59**:** 52-62.

Maluszynski, M., B. S. Ahloowalia and B. Sigurbjörnsson, 1995 Application of in vivo and in vitro mutation techniques for crop improvement. Euphytica 85**:** 303-315.

Mamoon Rehman, H., M. Amjad Nawaz, L. Bao, Z. Hussain Shah, J.-M. Lee *et al.*, 2016 Genome-wide analysis of Family-1 UDP-glycosyltransferases in soybean confirms their abundance and varied expression during seed development. Journal of Plant Physiology 206**:** 87-97.

Mansoor, S., and A. H. Paterson, 2012 Genomes for jeans: cotton genomics for engineering superior fiber. Trends in Biotechnology 30**:** 521-527.

May, O. L., 2002 Quality improvement of upland cotton (Gossypium hirsutum L.). Journal of crop production 5**:** 371-394.

Mba, C., R. Afza, S. Bado and S. M. Jain, 2010 Induced mutagenesis in plants using physical and chemical agents. Plant cell culture: essential methods 20**:** 111-130.

McCallum, C. M., L. Comai, E. A. Greene and S. Henikoff, 2000 Targeting induced locallesions in genomes (TILLING) for plant functional genomics. Plant physiology 123**:** 439-442.

McLendon, C. A., 1912 *Mendelian inheritance in cotton hybrids*. Georgia Experiment Station.

Meena, A. K., M. Ramesh, C. Nagaraju and B. L. Kumhar, 2017 A Review of QTL Mapping in Cotton: Molecular Markers, Mapping Populations and Statistical Methods. Int. J. Curr. Microbiol. App. Sci 6**:** 3057-3080.

Mehetre, S., and M. Thombre, 1983 Fibre properties of x-ray induced glandless mutants in American cotton. Journal of Maharashtra Agricultural Universities 8**:** 189-190.

Mei, M., N. Syed, W. Gao, P. Thaxton, C. Smith *et al.*, 2004 Genetic mapping and QTL analysis of fiber-related traits in cotton (Gossypium). TAG Theoretical and Applied Genetics 108**:** 280-291.

Mei, W., Y. Qin, W. Song, J. Li and Y. Zhu, 2009 Cotton GhPOX1 encoding plant class III peroxidase may be responsible for the high level of reactive oxygen species production that is related to cotton fiber elongation. J Genet Genomics 36.

Meredith, W., 1994 Genetics and management factors influencing textile fiber quality, pp. 256–261.

Meredith, W. R., 1984 Quantitative genetics. Cotton**:** 131-150.

Moon, J., G. Parry and M. Estelle, 2004 The Ubiquitin-Proteasome Pathway and Plant Development. The Plant Cell 16**:** 3181-3195.

Munis, M. F., L. Tu, F. Deng, J. Tan, L. Xu *et al.*, 2010 A thaumatin-like protein gene involved in cotton fiber secondary cell wall development enhances resistance against Verticillium dahliae and other stresses in transgenic tobacco. Biochem Biophys Res Commun 393**:** 38-44.

Naoumkina, M., E. Bechere, D. D. Fang, G. N. Thyssen and C. B. Florane, 2017 Genome-wide analysis of gene expression of EMS-induced short fiber mutant Ligon lintless-y (liy) in cotton (Gossypium hirsutum L.). Genomics 109**:** 320-329.

Naoumkina, M., D. J. Hinchliffe, R. B. Turley, J. M. Bland and D. D. Fang, 2013 Integrated metabolomics and genomics analysis provides new insights into the fiber elongation process in Ligon lintless-2 mutant cotton (Gossypium hirsutum L.). BMC Genomics 14.

Narbuth, E. V., and R. J. Kohel, 1990 Inheritance and linkage analysis of a new fiber mutant in cotton. Journal of Heredity 81**:** 131-133.

Ngematov, M., V. Kovalenko, V. Shumnyi and K. Asrorov, 1975 Induction of cytoplasmicmale sterility in cotton by the method of radiation mutagenesis. Soviet Genetics 11**:** 1593–1595.

Orford, S. J., and J. N. Timmis, 1998 Specific expression of an expansin gene during elongation of cotton fibres. Biochimica et Biophysica Acta (BBA)-Gene Structure and Expression 1398**:** 342-346.

Palle, S. R., L. M. Campbell, D. Pandeya, L. Puckhaber, L. K. Tollack *et al.*, 2013 RNAi-mediated Ultra-low gossypol cottonseed trait: performance of transgenic lines under field conditions. Plant biotechnology journal 11**:** 296-304.

Pang, C. Y., H. Wang, Y. Pang, C. Xu, Y. Jiao *et al.*, 2010 Comparative proteomics indicates that biosynthesis of pectic precursors is important for cotton fiber and Arabidopsis root hair elongation. Mol Cell Proteomics 9.

Pang, Y., K. Chen, X. Wang, W. Wang, J. Xu *et al.*, 2017 Simultaneous Improvement and Genetic Dissection of Salt Tolerance of Rice (Oryza sativa L.) by Designed QTL Pyramiding. Frontiers in Plant Science 8.

Patel, J. D., R. J. Wright, D. Auld, R. Chandnani, V. H. Goff *et al.*, 2014 Alleles conferring improved fiber quality from EMS mutagenesis of elite cotton genotypes. Theoretical and Applied Genetics 127**:** 821-830.

Patel, J. D., R. J. Wright, R. Chandnani, V. H. Goff, J. Ingles *et al.*, 2016 EMS-mutated cotton populations suggest overlapping genetic control of trichome and lint fiber variation. Euphytica 208**:** 597-608.

Paterson, A., Y. Saranga, M. Menz, C. X. Jiang and R. Wright, 2003 QTL analysis of genotype× environment interactions affecting cotton fiber quality. TAG Theoretical and Applied Genetics 106**:** 384-396.

Paterson, A. H., R. K. Boman, S. M. Brown and P. W. Chee, 2004 Reducing the genetic vulnerability of cotton. Crop science 44**:** 1900.

Paterson, A. H., J. F. Wendel, H. Gundlach, H. Guo, J. Jenkins *et al.*, 2012 Repeated polyploidization of Gossypium genomes and the evolution of spinnable cotton fibres. Nature 492**:** 423-427.

Pathirana, R., 2011 Plant mutation breeding in agriculture. Plant sciences reviews**:** 107-126.

Percival, A., J. Wendel and J. Stewart, 1999 Taxonomy and germplasm resources. Cotton: origin, history, technology, and production**:** 33–63.

Percy, R., B. Hendon, E. Bechere and D. Auld, 2015 Qualitative genetics and utilization of mutants. Cotton**:** 155-186.

Pickersgill, B., 2007 Domestication of Plants in the Americas: Insights from Mendelian and Molecular Genetics. Annals of Botany 100**:** 925-940.

Pickett, J., 2007 UBE1, you're not alone. Nat Rev Mol Cell Biol 8**:** 599-599.

Potikha, T. S., C. C. Collins, D. I. Johnson, D. P. Delmer and A. Levine, 1999 The involvement of hydrogen peroxide in the differentiation of secondary walls in cotton fibers. Plant physiology 119**:** 849-858.

Qin, H., W. Guo, Y.-M. Zhang and T. Zhang, 2008a QTL mapping of yield and fiber traits based on a four-way cross population in *Gossypium hirsutum L.* TAG Theoretical and Applied Genetics 117**:** 883-894.

Qin, Y.-M., C.-Y. Hu and Y.-X. Zhu, 2008b The ascorbate peroxidase regulated by H2O2 and ethylene is involved in cotton fiber cell elongation by modulating ROS homeostasis. Plant signaling & behavior 3**:** 194-196.

Rathore, K. S., S. Sundaram, G. Sunilkumar, L. M. Campbell, L. Puckhaber *et al.*, 2012 Ultra-low gossypol cottonseed: generational stability of the seed-specific, RNAi-mediated phenotype and resumption of terpenoid profile following seed germination. Plant biotechnology journal 10**:** 174-183.

Raut, R., H. Jain and R. Panwar, 1971 RADIATION-INDUCED PHOTO-INSENSITIVE MUTANTS IN COTTON, pp. Indian Agricultural Research Inst., New Delhi.

Reinisch, A. J., J. M. Dong, C. L. Brubaker, D. M. Stelly, J. F. Wendel *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**:** 829.

Renny-Byfield, S., J. T. Page, J. A. Udall, W. S. Sanders, D. G. Peterson *et al.*, 2016 Independent domestication of two Old World cotton species. Genome biology and evolution 8**:** 1940-1947.

Rong, J., C. Abbey, J. E. Bowers, C. L. Brubaker, C. Chang *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., F. A. Feltus, V. N. Waghmare, G. J. Pierce, P. W. Chee *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.

Rong, J. K., G. J. Pierce, V. N. Waghmare, C. A. J. Rogers, A. Desai *et al.*, 2005 Genetic mapping and comparative analysis of seven mutants related to seed fiber development in cotton. Theoretical and Applied Genetics 111**:** 1137-1146.

Ruan, Y.-L., D. J. Llewellyn and R. T. Furbank, 2003 Suppression of sucrose synthase gene expression represses cotton fiber cell initiation, elongation, and seed development. The Plant Cell 15**:** 952-964.

Said, J. I., Z. Lin, X. Zhang, M. Song and J. Zhang, 2013 A comprehensive meta QTL analysis for fiber quality, yield, yield related and morphological traits, drought tolerance, and disease resistance in tetraploid cotton. BMC genomics 14**:** 776.

Said, J. I., M. Song, H. Wang, Z. Lin, X. Zhang *et al.*, 2015 A comparative meta-analysis of QTL between intraspecific Gossypium hirsutum and interspecific G. hirsutum × G. barbadense populations. Molecular Genetics and Genomics 290**:** 1003-1025.

Saint Paul, V. v., 2010 Stress inducible glycosyltransferases in Arabidopsis thaliana and their impact on plant metabolism and defense mechanisms, pp. lmu.

Sampedro, J., and D. J. Cosgrove, 2005 The expansin superfamily. Genome Biology 6**:** 242.

Sasser, P., 1981 Basics of high volume instruments for fiber testing, pp. in *Proceedings-Beltwide Cotton Production Research Conferences*.

Seelanan, T., A. Schnabel and J. F. Wendel, 1997 Congruence and consensus in the cotton tribe (Malvaceae). Systematic Botany**:** 259-290.

Shalem, O., N. E. Sanjana and F. Zhang, 2015 High-throughput functional genomics using CRISPR-Cas9. Nature reviews. Genetics 16**:** 299.

Shamsuzzaman, K., M. Hamid, M. Azad, M. Hussain and M. Majid, 2003 Varietal improvement of cotton (Gossypium hirsutum) through mutation breeding. Improvement of new and traditional industrial crops by induced mutations and related biotechnology**:** 78.

Shan, C.-M., X.-X. Shangguan, B. Zhao, X.-F. Zhang, L.-m. Chao *et al.*, 2014 Control of cotton fibre elongation by a homeodomain transcription factor GhHOX3. Nature communications 5.

Shappley, Z., J. Jenkins, C. Watson, A. Kahler and W. Meredith, 1996 Establishment of molecular markers and linkage groups in two F2 populations of upland cotton. Theoretical and applied genetics 92**:** 915-919.

Shappley, Z., J. N. Jenkins, J. Zhu and J. C. McCarty Jr, 1998 Quantitative trait loci associated with agronomic and fiber traits of upland cotton.

Shen, X., W. Guo, Q. Lu, X. Zhu, Y. Yuan *et al.*, 2007 Genetic mapping of quantitative trait loci for fiber quality and yield trait by RIL approach in Upland cotton. Euphytica 155**:** 371-380.

Shen, X., T. Zhang, W. Guo, X. Zhu and X. Zhang, 2006 Mapping Fiber and Yield QTLs with Main, Epistatic, and QTL × Environment Interaction Effects in Recombinant Inbred Lines of Upland Cotton. Crop Sci. 46**:** 61-66.

Shi, Y. H., S. W. Zhu, X. Z. Mao, J. X. Feng, Y. M. Qin *et al.*, 2006 Transcriptome profiling, molecular biological, and physiological studies reveal a major role for ethylene in cotton fiber cell elongation. Plant Cell 18.

Shoemaker, D., 1909 A study of leaf characters in cotton hybrids. Journal of Heredity**:** 116-118.

Shofner, F. M., Y. Chu and D. Thibodeaux, 1990 An overview of the advanced fiber information system, pp. 173-181 in *Proc. Int. Cotton Conf., Faserinstitut, Bremen, Germany*.

Smith, C. W., 2003 Registration of TAM 94L-25 and TAM 94J-3 germplasm lines of upland cotton with improved fiber length. Crop science 43**:** 742-744.

Stephens, S., 1967 Evolution under domestication of the New World cottons (Gossypium spp.). Cienc Cult 19**:** 118-134.

Stewart, J., 1995 Potential for crop improvement with exotic germplasm and genetic engineering, pp. 313-327 in *Proceeding of the world cotton research conference-I, Brishbane, Australia, February 14-17, Melbourne, 1995*.

Stewart, J. M., L. A. Craven, C. Brubaker and J. F. Wendel, 2015 Gossypium anapoides (Malvaceae), a new species from Western Australia. Novon 23**:** 447-451.

Suh, M., and P. Sasser, 1996 The technological and economic impact of high volume instrument (HVI) systems on the cotton and cotton textile industries. Journal of the Textile Institute 87**:** 43-59.

Sun, X., S. Y. Gong, X. Y. Nie, Y. Li, W. Li *et al.*, 2015 A R2R3-MYB transcription factor that is specifically expressed in cotton (Gossypium hirsutum) fibers affects secondary cell wall biosynthesis and deposition in transgenic Arabidopsis. Physiologia plantarum 154**:** 420-432.

Sunilkumar, G., L. M. Campbell, L. Puckhaber, R. D. Stipanovic and K. S. Rathore, 2006 Engineering cottonseed for use in human nutrition by tissue-specific reduction of toxic gossypol. Proceedings of the National Academy of Sciences 103**:** 18054-18059.

Suo, J. F., X. O. Liang, L. Pu, Y. S. Zhang and Y. B. Xue, 2003 Identification of GhMYB109 encoding a R2R3 MYB transcription factor that expressed specifically in fiber initials and elongating fibers of cotton (Gossypium Hirsutum L.). Biochimica Et Biophysica Acta-Gene Structure and Expression 1630**:** 25-34.

Tai, F. J., X. L. Wang, W. L. Xu and X. B. Li, 2008 Characterization and expression analysis of two cotton genes encoding putative UDP-Glycosyltransferases. Molecular Biology 42**:** 44-51.

Talebi, A. B., A. B. Talebi and B. Shahrokhifar, 2012 Ethyl methane sulphonate (EMS) induced mutagenesis in Malaysian rice (cv. MR219) for lethal dose determination. American Journal of Plant Sciences 3**:** 1661.

Tang, W., L. Tu, X. Yang, J. Tan, F. Deng *et al.*, 2014 The calcium sensor GhCaM7 promotes cotton fiber elongation by modulating reactive oxygen species (ROS) production. New Phytologist 202**:** 509-520.

Tanksley, S. D., and S. R. McCouch, 1997 Seed banks and molecular maps: unlocking genetic potential from the wild. Science 277**:** 1063-1066.

Thibodeaux, D., H. Senter, J. Knowlton, D. Mcalister and X. Cui, 2008 The impact of short fiber content on the quality of cotton ring spun yarn. J Cotton Sci 12**:** 368-377.

Thyssen, G. N., D. D. Fang, R. B. Turley, C. B. Florane, P. Li *et al.*, 2017 A Gly65Val substitution in an actin, GhACT_LI1, disrupts cell polarity and F-actin organization resulting in dwarf, lintless cotton plants. The Plant Journal.

Thyssen, G. N., D. D. Fang, L. Zeng, X. Song, C. D. Delhom *et al.*, 2016 The Immature fiber mutant phenotype of cotton (Gossypium hirsutum) is linked to a 22-bp frame-shift deletion in a mitochondria targeted pentatricopeptide repeat gene. G3: Genes| Genomes| Genetics 6**:** 1627-1633.

Tiwari, S. C., and T. A. Wilkins, 1995 Cotton (*Gossypium hirsutum*) seed trichomes expand via diffuse growing mechanism. Canadian Journal of Botany 73**:** 746-757.

Tyagi, P., M. A. Gore, D. T. Bowman, B. T. Campbell, J. A. Udall *et al.*, 2014 Genetic diversity and population structure in the US Upland cotton (Gossypium hirsutum L.). Theoretical and Applied Genetics 127**:** 283-295.

Ulloa, M., 2006 Heritability and correlations of agronomic and fiber traits in an okra-leaf upland cotton population. Crop science 46**:** 1508-1514.

Ulloa, M., and W. R. Meredith Jr, 2000 Genetic linkage map and QTL analysis of agronomic and fiber quality traits in an intraspecific population. Journal of Cotton Science 4**:** 161-170.

Van der Sluijs, M., and L. Hunter, 1999 Neps in cotton lint. Textile Progress 28**:** 1-47.

Van Esbroeck, G., and D. T. Bowman, 1998 Cotton germplasm diversity and its importance to cultivar development. Journal of Cotton Science.

Viot, C., 2017 Evolution and domestication of diploid cultivated cottons: molecular genetics and agronomic evidence.

Waghmare, V. N., J. Rong, C. J. Rogers, J. E. Bowers, P. W. Chee *et al.*, 2016 Comparative transmission genetics of introgressed chromatin in Gossypium (cotton) polyploids. American journal of botany 103**:** 719-729.

Wakeham, H., 1955 Cotton fiber length distribution—an important quality factor. Textile Research Journal 25**:** 422-429.

Walford, S.-A., Y. Wu, D. J. Llewellyn and E. S. Dennis, 2012 Epidermal cell differentiation in cotton mediated by the homeodomain leucine zipper gene, GhHD-1. The Plant Journal 71**:** 464-478.

Wan, Q., X. Guan, N. Yang, H. Wu, M. Pan *et al.*, 2016 Small interfering RNAs from bidirectional transcripts of GhMML3_A12 regulate cotton fiber development. New Phytologist.

Wan, Q., Z. Zhang, M. Hu, L. Chen, D. Liu *et al.*, 2007 T1 locus in cotton is the candidate gene affecting lint percentage, fiber quality and spiny bollworm (Earias spp.) resistance. Euphytica 158**:** 241-247.

Wang, B., S. H. Jin, H. Q. Hu, Y. G. Sun, Y. W. Wang *et al.*, 2012 UGT87A2, an Arabidopsis glycosyltransferase, regulates flowering time via FLOWERING LOCUS C. New Phytologist 194**:** 666-675.

Wang, K. L.-C., H. Yoshida, C. Lurin and J. R. Ecker, 2004 Regulation of ethylene gas biosynthesis by the Arabidopsis ETO1 protein. Nature 428**:** 945-950.

Wangzhen, G., Z. Tianzhen and D. Yezhang, 2005 Molecular marker assisted selection and pyramiding of two QTLs for fiber strength in upland cotton. Acta Genetica Sinica 32**:** 1275-1285.

Waters, W. T., J. Phillips and L. A. Fiori, 1966 The Effect of Fiber-Bundle Elongation of Medium Staple Cottons on Processing Performance and Y arn Properties. Textile Research Journal 36**:** 1004-1012.

Wendel , J., and R. Cronn, 2003 Polyploidy and the evolutionary history

of cotton. Adv Agron 78**:** 139–186.

Wendel, J. F., and V. A. Albert, 1992 Phylogenetics of the Cotton Genus (Gossypium): Character-State Weighted Parsimony Analysis of Chloroplast-DNA Restriction Site Data and Its Systematic and Biogeographic Implications. Systematic Botany 17**:** 115-143.

Wendel, J. F., C. Brubaker, I. Alvarez, R. Cronn and J. M. Stewart, 2009 Evolution and natural history of the cotton genus, pp. 3-22 in *Genetics and genomics of cotton*. Springer.

Wendel, J. F., C. L. Brubaker and A. E. Percival, 1992 Genetic diversity in Gossypium hirsutum and the origin of upland cotton. American Journal of Botany**:** 1291-1310.

Wendel, J. F., and R. C. Cronn, 2003 Polyploidy and the evolutionary history of cotton. Advances in agronomy 78**:** 139-186.

Wendel, J. F., and C. E. Grover, 2015 Taxonomy and evolution of the cotton genus, Gossypium. Cotton**:** 25-44.

Xiao, J., K. Wu, D. Fang, D. M. Stelly, J. Yu *et al.*, 2009 New SSR Markers for Use in Cotton (*Gossypium* spp.*)* Improvement. The Journal of Cotton Science 13**:** 75-157.

Xu, P., J. Gao, Z. Cao, P. W. Chee, Q. Guo *et al.*, 2017 Fine mapping and candidate gene analysis of qFL-chr1, a fiber length QTL in cotton. Theoretical and Applied Genetics 130**:** 1309-1319.

Xu, Z., R. J. Kohel, G. Song, J. Cho, M. Alabady *et al.*, 2008 Gene-rich islands for fiber development in the cotton genome. Genomics 92**:** 173-183.

Yang, S., and S. Gordon, A study on cotton fibre elongation measurement.

Yang, Z., C. Zhang, X. Yang, K. Liu, Z. Wu *et al.*, 2014 PAG1, a cotton brassinosteroid catabolism gene, modulates fiber elongation. New Phytologist 203**:** 437-448.

Yonekura-Sakakibara, K., and K. Hanada, 2011 An evolutionary view of functional diversity in family 1 glycosyltransferases. The Plant Journal 66**:** 182-193.

Yoshida, H., M. Nagata, K. Saito, K. L. Wang and J. R. Ecker, 2005 Arabidopsis ETO1 specifically interacts with and negatively regulates type 2 1-aminocyclopropane-1-carboxylate synthases. BMC Plant Biology 5**:** 14.

Yu, J. W., S. X. Yu, C. R. Lu, W. Wang, S. L. Fan *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**:** 716-724.

Yu, Y., D. Yuan, S. Liang, X. Li, X. Wang *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.

Yuan, Y., T. Wang, Y. Shi, H. Shang, A. Liu *et al.*, 2014 Molecular marker-assisted selection and pyramiding effect of major QTLs for cotton fiber strength. New Biotechnology 31**:** S14.

Zhang, B., and J.-Y. Liu, 2016 Cotton cytosolic pyruvate kinase GhPK6 participates in fast fiber elongation regulation in a ROS-mediated manner. Planta 244**:** 915-926.

Zhang, F., X. Jin, L. Wang, S. Li, S. Wu *et al.*, 2016a GhFAnnxA affects fiber elongation and secondary cell wall biosynthesis associated with Ca2+ influx, ROS homeostasis and actin filament reorganization. Plant Physiology.

Zhang, J., and R. Percy, 2007 Improving Upland cotton by introducing desirable genes from Pima cotton, pp. 10-14 in *World Cotton Research Conf.*

Zhang, S.-W., X.-F. Zhu, L.-C. Feng, X. Gao, B. Yang *et al.*, 2016b Mapping of fiber quality QTLs reveals useful variation and footprints of cotton domestication using introgression lines. Scientific Reports 6**:** 31954.

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

Zhang, T., Y. Yuan, J. Yu, W. Guo and R. J. Kohel, 2003a Molecular tagging of a major QTL for fiber strength in Upland cotton and its marker-assisted selection. Theoretical and Applied Genetics 106**:** 262-268.

Zhang, X.-D., J. N. Jenkins, F. E. Callahan, R. G. Creech, Y. Si *et al.*, 2003b Molecular cloning, differential expression, and functional characterization of a family of class I ubiquitin-conjugating enzyme (E2) genes in cotton (Gossypium). Biochimica et Biophysica Acta (BBA) - Gene Structure and Expression 1625**:** 269-279.

Zhang, Z., J. Rong, V. N. Waghmare, P. W. Chee, O. L. May *et al.*, 2011 QTL alleles for improved fiber quality from a wild Hawaiian cotton, Gossypium tomentosum. Theoretical and applied genetics 123**:** 1075.

Zhang, Z., Y.-L. Ruan, N. Zhou, F. Wang, X. Guan *et al.*, 2017 Suppressing a Putative Sterol Carrier Gene Reduces Plasmodesmal Permeability and Activates Sucrose Transporter Genes during Cotton Fiber Elongation. The Plant Cell.

Zhao, Y., H. Wang, W. Chen, Y. Li, H. Gong *et al.*, 2015 Genetic diversity and population structure of elite cotton (Gossypium hirsutum L.) germplasm revealed by SSR markers. Plant systematics and evolution 301**:** 327-336.

Zhong, R., M. J. Peña, G.-K. Zhou, C. J. Nairn, A. Wood-Jones *et al.*, 2005 Arabidopsis Fragile Fiber8, Which Encodes a Putative Glucuronyltransferase, Is Essential for Normal Secondary Wall Synthesis. The Plant Cell 17**:** 3390-3408.

Zhong, R., and Z.-H. Ye, 2003 Unraveling the functions of glycosyltransferase family 47 in plants. Trends in Plant Science 8**:** 565-568.

CHAPTER 2

IMPROVEMENT OF ELITE UPLAND COTTON GERMPLASM FOR MULTIPLE FIBER TRAITS BY TRANSFERRING NOVEL ALLELES FROM EMS-GENERATED MUTANT LINES

Patel, Jinesh[1]; Chandnani, Rahul[1]; Khanal, Sameer[1]; Adhikari, Jeevan[1]; Brown, Ismail [2]**;** Chee, Peng W[2]; Jones, Don[3] and Paterson, Andrew H[1 *] To be submitted to Theoretical and Applied Genetics

Abstract

EMS-mutagenesis offers important advantages for improvement of crops such as cotton that have limited diversity in elite gene pools, as EMS-induced point mutations are less frequently associated with deleterious traits than alleles from wild or exotic germplasm. Here, we further investigated nine mutant lines out of 157 found to have significantly improved fiber properties. A total of eight populations were developed by crossing mutant lines in different combinations into GA230 (GA2004230) background. Multiple lines in each population were significantly improved for the fiber trait that distinguished the donor parent(s), demonstrating that an elite breeding line (GA230) could be improved for fiber qualities using these mutant lines. Genotypes improved for multiple fiber traits suggest that allele pyramiding may confer further improvements. Compared to mid parent values, mutant lines conferred fiber quality improvements to progeny of as high as 31.7% (O013) for micronaire, 16.1% (P058) for length, 22.4% (K92) for strength, 4.1% (Q47) for uniformity, 45.8% (N068) for elongation and 13.9% (O014) for lint percent. While further testing for phenotypic stability and estimation of yield potential is necessary, mutation breeding shows promise as an approach to reduce the genetic bottleneck of Upland cotton. The populations developed here may also contribute to identifying candidate genes and causal mutations for fiber quality improvement.

Keywords: crop improvement, functional genomics, TAM94L25, ACALA1517-99, allele stacking

Introduction

The history of cotton cultivation and improvement has included several genetic bottlenecks that constrain variation available for breeders to utilize to develop improved cultivars. A total of four Gossypium species were domesticated (WENDEL AND CRONN 2003; PICKERSGILL 2007; RENNY-BYFIELD *et al.* 2016) as sources of textile fiber, two Old World diploids (*G. arboreum L and G. herbaceum L.*) and two New World tetraploids (*G. hirsutum* and *G. barbadense*) (BRUBAKER *et al.* 1999; WENDEL AND CRONN 2003). Due to longer and stronger fiber and higher yield, tetraploid cotton species *(especially G. hirsutum)* are now preferred and account for 95% of world cotton fiber production (WENDEL et al. 1992; CHEN et al. 2007). Cotton breeding focused largely on developing high yielding and stress tolerant varieties, until increasing use of air jet and vortex spinning technologies in the textile industry motivated improving fiber quality traits (BRADOW AND DAVIDONIS 2000; MAY 2002). Major fiber quality traits measured by HVI (High Volume Instrument) or AFIS (Advanced Fiber Information System) are fiber length, uniformity, strength, fineness, maturity ratio, fiber color characteristic (color as reflectance (Rd) and yellowness (+b)) and short fiber content (KELLY et al. 2012; KELLY et al. 2015).

Intensive breeding within a narrow sampling of germplasm, together with several decades of focus on transgenic lines, has rendered the elite cotton gene pool one of the narrowest among the major crops (LUBBERS AND CHEE 2009) and has been suggested to leave very little room for fiber quality improvement (VAN ESBROECK AND BOWMAN 1998; IQBAL *et al.* 2001; PATERSON *et al.* 2004; TYAGI *et al.* 2014). Wild species can provide desirable agronomical traits and be used in breeding program to broaden the genetic diversity, but often carry deleterious alleles that must be eliminated. Moreover, reproductive barriers, meiotic drive, hardship in adaptability to new habitats and differences in photoperiodic response and flowering time, all complicate selection

within crosses involving wild relatives to achieve desired characteristics without reducing yield (ZHANG AND PERCY 2007; BARB *et al.* 2014; WAGHMARE *et al.* 2016). Cloning of alleles affecting fiber traits and transferring them into elite cotton cultivars involves prohibitive time, cost, and regulatory concerns regarding GMO (Genetically Modified Organisms).

Chemical mutagens such as EMS (Ethyl methanesulfonate) that generate single nucleotide 'point' mutations offer a means to create new alleles that are expected to be free from problems associated with linkage drag of unfavorable alleles during the cross of wild species. EMS-mutagenesis generally creates point mutation by converting G/C-to-A/T through DNA mispairing, thus reducing the chances of lethal mutation (GREENE *et al.* 2003). In a single experiment, one can use EMS to create and screen for multiple mutant lines that have significantly improved fiber qualities as compared to parental lines. In addition, one can expect to find other agronomical traits of interest such as fuzzless seeds, biotic and abiotic stress tolerance, and glabrous or hairy varieties along with improved fiber quality without affecting the lint yield (BECHERE *et al.* 2009a; PATEL *et al.* 2016). While appreciable screening of large numbers of candidate lines is necessary, one might discover genotypes with novel alleles that could contribute to desirable agronomical traits. Mutant populations can be subjected to either forward genetics or reverse genetics approaches such as TILLING to identify causal mutations (MCCALLUM *et al.* 2000). Several successes in using EMS to enhance elite cotton germplasm have been published (AULD *et al.* 2000; BROWN *et al.* 2012; PATEL *et al.* 2014).

Here we build on prior work to identify desirable EMS-mutated cotton lines (PATEL *et al.* 2014), transferring novel alleles from their source backgrounds (TAM94L25 and Acala 1517-99) to an elite breeding line, GA2004230. We also investigated the consequences of 'pyramiding' such novel alleles, in eight populations from different combinations of mutant lines crossed to

GA2004230. Significant improvement in fiber qualities of progenies developed by crosses between GA2004230 and selected mutant lines provided motivation to transfer novel alleles from mutant lines to additional elite germplasm, and more generally provided support for increased use of EMS mutants for improvement of cotton fiber quality.

Materials and Methods

Plant sources and population development

Based on initial screening of large populations of mutant lines generated using EMS in two different genetic backgrounds of *G. hirsutum*, TAM94L25 (SMITH 2003) and Acala 1517-99 (CANTRELL *et al.* 2000) in Texas (2007) and Georgia (2008), a subset of 157 mutant lines having significant improvements in one or more components of fiber quality were tested along with control lines at two locations in a replicated trial (Texas and Georgia, 2009). The data from the four environments supports that mutant lines with significantly improved fiber length, strength, elongation, fineness, lint percent and other quality and yield components can be achieved (PATEL *et al.* 2014). Nine striking mutants were selected from the 157 mutant lines showing improved fiber length (1 mutant line from Acala 1517-99), fiber elongation (1 from TAM94L24), fiber fineness (1 each from TAM94L24 and Acala1517-99), fiber uniformity (1, Acala1517-99), fiber strength (1, Acala1517-99), lint percent (1 each from TAM94L24 and Acala1517-99) and Rd value (1 from TAM94L25) (Table 1). Each of these nine mutant lines were crossed with a non-transgenic cotton variety adapted to the southeastern cotton belt, GA2004230 (PVP 201500309; from now on, GA230) (LUBBERS *et al.* 2006) to develop F1 hybrids in a greenhouse in Athens, GA (Summer 2012). The F1 hybrids were further crossed with each other to pyramid novel alleles from mutant lines contributing to various fiber traits in GA230 background (Table 2), in an off-season nursery in Mexico.

Field trial and data collection

A total of 100 BC1F1 (estimated at 50% source and 50% GA230 genetic background) progenies from each cross were grown in Watkinsville, GA in 2013. In 2014, progenies with sufficient seed were planted in a randomized complete block design (RCB) with two replicates at two locations (Watkinsville and Tifton, GA). The soil type at Watkinsville, GA was Appling Coarse Sandy Loam (fine, kaolinitic, thermic typic kanhapludults) and at Gibbs farm, Tifton, GA was Tifton loamy sand (fine, loamy, siliceous, thermic Plinthic kandiudult). A total of 35 seeds were planted in plot sizes of 3 m, spaced 1 m apart. Agronomic practices like weeding, irrigation, fertilizer application and pest management followed local recommendations to commercial growers. To obtain fiber samples, bolls were hand-picked in Athens (November 25, 2013) and Tifton (October 26, 2014) while seed cotton samples were collected from machine harvested cotton in Athens (November 19, 2014) and ginned using a 20-saw gin (DENNIS MFG. CO., INC.). Lint and seed weight (seed plus fuzz) were measured and lint percent (lint weight X 100/seed cotton weight) was calculated. Samples of 10 grams of lint were sent to Cotton Inc. to measure six fiber properties using HVI, including upper half mean fiber fineness or micronaire (MIC), fiber length (LEN), fiber strength (STR), fiber elongation (ELON), Uniformity index (UI) and Short Fiber Content (SFC).

Data analysis

Data was analyzed using **SAS 9.4** (**SAS** Institute Inc., Cary, NC, USA). The program statement “Proc CORR” was used to determine correlations between fiber traits. Heritabilities of fiber traits were calculated using parent-offspring regression by the SAS “Proc REG” statement. The contribution and significance of genotype, environment and interaction between genotype and environment for fiber traits was calculated using the SAS statement “Proc GLM”. Among

the total of eight populations, four involved crossing TAM94L25 mutant lines to GA230 (pops J, M, N, and O) and four involved crossing Acala1517-99 mutant lines to GA230 (pops K, L, P, and Q) (Table 2). The means of fiber traits of these populations were compared with the 'parental' (wild type progenitor and GA230) means using Fisher's LSD test at alpha level of 0.05 to determine if there was significant improvement in a sub-population for a fiber trait. To determine if pyramiding of two alleles for different traits produced germplasm with improvement for both fiber traits (for example MIC and ELON for pop J), Z scores for each line were calculated for each trait and the ten lines with the highest summed Z score for the two traits were compared with parental lines. The analysis considered genotype, environment, replication and selection (based on Z score of two traits) as fixed variables.

Results

Heritability between generation and correlation between fiber traits

Heritability was calculated between F2 and F3 generations for each fiber trait based on parent-offspring regression. Moderate heritability for fiber traits was seen in both populations (Table 3). In TAM94L25 X GA230 (hereafter 'TAGA') populations, LEN showed the maximum heritability of 0.43 whereas UI showed the minimum heritability of 0.17. Similarly, for Acala1517-99 X GA230 (from here ACGA) population, STR showed the maximum heritability of 0.44 whereas UI showed the minimum heritability of 0.27.

Correlation between traits helps to determine if two traits can be simultaneously improved or if improving one trait might impair another trait. Positive correlation generally suggests that two traits can be improved simultaneously, except for correlations involving MIC or SFC, for which negative values are desirable. MIC showed moderate negative correlation with fiber length and positive correlation with lint% in both populations; and small but significant negative correlation

with STR and UI in ACGA only. LEN showed moderate positive correlation with STR and UI but small negative correlation with ELON and lint%. UI had positive correlation with STR and ELONG. STR had negative correlation with ELONG in TAGA and lint% in ACGA. Moderate positive correlation between ELON and lint% was seen in both populations. SFC had negative correlation with each fiber trait except MIC and lint% (in ACGA). It had strong negative correlations (-0.79 in TAGA; -0.80 in ACGA) with UNIF and weak negative correlation of -0.17 with lint% in TAGA (Table 4).

Genotype and environment effects

ANOVA in both populations showed significant differences between genotypes and between environments but no significant genotype X environment interaction (Table 5). In TAGA the variance explained for different fiber traits by genotype ranged from 19% (ELON) to 48% (LEN), similarly in ACGA variance explained by genotype range from across trait with 21% (ELON) to 56% (LEN).

Fiber trait improvement

Improvement for MIC and ELON (Pop J)

A total of 100 lines were evaluated from 'population J', placing TAM94L25 mutants for improved MIC and ELON in GA230 background, with overall means that showed significant ($p< 0.01$) improvement over the parental lines and exceeded the midparent values by 4% (MIC) and 13.6% (ELON, Table 6). Significant improvement over the parents was realized for MIC by 30 lines with line J29 showing 19% improvement over the midparent value; and for ELON by 9 lines with line J45 showing 32.2% improvement over the midparent value (Table 6, Figure 2). The top 10 lines based on the sum of Z scores reflecting enhanced UNIF and STR showed an

average improvement of 9.9% in MIC and 25.2% in ELON relative to midparent values, each significant (Figure 3).

Improvement for LEN and STR (Pop K)

A total of 98 lines were evaluated from 'population K", placing ACALA 1517-99-derived mutants for LEN and STR in GA230 background, with overall means that showed significant improvement ($p<0.01$) over the parental lines and exceeded the mid-parent values for LEN by 5.7% and STR by 8.4% (Table 6). Significant improvement over the parents was realized for LEN by 40 lines with line K68 showing a 15.8% improvement; and STR for 54 lines with line K92 showing 22.4% improvement (Table 6, Figure 2). The top 10 lines based on the sum of Z scores reflecting enhanced LEN and STR significantly exceeded both parents and showed improvements over the midparent value of 10.7% in LEN and 16.6% in STR (Figure 3).

Improvement for MIC and lint% (Pop L & O)

A total of 89 and 87 lines were evaluated from "pop L" and "pop O", placing ACALA1517-99 and TAM94L25-derived mutants, respectively, for MIC and lint% into GA230 background. The average of individuals in Pop O showed significant improvement for MIC compared to parental lines while there was no significant difference between pop L and parents. For lint%, mean of both populations (pop L & O) were significantly better than their respective mutant sources or parent (ACALA1517-99 or TAM94L25), but not exceeding GA230. Significant improvements over parents was realized for MIC by 25 and 7 lines in pop O and pop L respectively. With respect to mid-parent MIC value, the most positive changes were in line O013 (pop O, 31.7%) and L078 (pop L, 10.4%). Likewise, for lint% significant improvement over the parents was realized by three lines in pop O with line O014 showing a 13.9% improvement and in two lines in pop L with line L071 showing a 13.1% improvement. Due to

negative association of MIC and lint% (HERRING *et al.* 2004; PATEL *et al.* 2014), only five pop O lines showed promise for improving MIC and lint%. The MIC mean of these lines was 8.2% lower (better) than the mid-parent value and lint% mean was 6.8% higher than mid-parent value (but not significantly improved over GA230). For pop L, we could identify only three lines that showed overall improvement in lint% and MIC within the population, but unfortunately their means were not significantly different from parents.

Improvement for lint% (Pop M) and Elon (Pop N)

Two populations were developed to improve lint% (pop M), Elon (pop N) and Rd value (both) in GA230 background. Mutant lines M276 (lint%), M1251 (Rd), and M2925 (Elon) used in the breeding scheme were from TAM94L25 background. Unfortunately, Cotton Inc. discontinued measurement of Rd value, so no pyramiding was further considered for these population. Based on performance of progenies in 2013 for lint% (in pop M) and Elon (in pop N), only selected genotypes from pop M (45 lines) and pop N (58 lines) were evaluated to see if there was improvement in progenies for lint% (pop M) and Elon (pop N).

The average lint% of pop M was significantly higher than TAM94L25 but not GA230. Only two lines were significantly improved over both parents, with M63 showed 11.6% improvement over the mid-parent value (Table 6). The top 10 lines for lint% in pop M showed an average 9.9% improvement over the mid-parent value and were significantly higher than both parents.

The average ELON of pop N was 19.9% higher than the mid-parent value and significantly better than both parents. A total of 11 lines showed improvement over parents, with line N068 showing the greatest improvement, 45.8% higher than the mid-parental value. The top 10 lines in the population for ELON showed an average 35.6% improvement over the mid-parent value.

Improvement for MIC and LEN (Pop P)

A total of 97 lines were evaluated from population P, placing ACALA 1517-99-derived mutants for LEN and MIC into GA230, with overall means that showed significant improvement ($p<0.01$) over the parental lines and exceeded the mid-parent values for LEN by 7.4% while MIC showed no significant improvement over parental lines (Table 6). Significant improvement over the parents was realized for LEN by 67 lines with line P058 showing a 16.1% improvement; and MIC by 18 lines with line P007 showing 12.8% improvement (Table 6, Figure 2). The top 10 lines based on the sum of Z scores reflecting enhanced LEN and MIC significantly exceeded both parents and showed improvements over the mid-parent value of 12% in LEN and 9.6% in MIC (Figure 3).

Improvement for STR and UNIF (Pop Q)

A total of 87 lines were evaluated from population Q, placing ACALA 1517-99-derived mutants for STR and UNIF in GA230 background, with overall means that showed significant improvement ($p<0.01$) over the parents and exceeded the mid-parent values for STR by 6.8% and UNIF by 1.6% (Table 6). Significant improvement over the parents was realized for STR by 36 lines with line Q85 showing a 17.9% improvement; and UNIF for 15 lines with line Q47 showing a 4.1% improvement (Table 6, Figure 2). The top 10 lines based on the sum of Z scores reflecting enhanced LEN and STR significantly exceeded both parents and showed improvements over the mid-parent value of 3.1% in UNIF and 12.6% in STR (Figure 3).

Discussion

Building on a few prior efforts that have produced improved cotton germplasm lines through EMS-generated mutations (AULD *et al.* 2000; BECHERE *et al.* 2009a; BROWN *et al.* 2012; PATEL *et al.* 2014; PATEL *et al.* 2016), from a total of 157 lines validated as possessing superior

quantitative fiber quality parameters (PATEL *et al.* 2014), here we further evaluated 9 striking mutants, confirming that 5 of the 8 (improvement of Rd value by TAM 94L25 M-1251 was not studied) mutant lines significantly improve fiber properties in a new elite genetic background (Table 6), and in most cases in two different combinations (Table 2). Further, comparison to two commercial checks (Delta Pine 393 and Fiber Max 832) showed mutants to be superior in several cases, including mean STR and LEN of pop K, STR and UNIF of pop Q, Lint_% of pop M, and LEN of Pop P (Table 6). These findings provide additional strong support for the value of these putatively novel alleles for improving fiber qualities of elite cottons.

Each of the eight test populations in the study was based on intercrossing of two F1's that involved different fiber quality mutants -- all populations contained multiple genotypes in the population improved for either one of two target traits, and from four populations (J, K, P, and Q) the average of ten lines selected based on summed z scores for the two traits were better for both traits than the parental lines (Figure 3), providing a natural foundation for pyramiding of fiber quality traits. For Pop O only five lines and for Pop L only three lines were obtained for such comparison, and the mean for MIC of pop O only was significantly improved compared to both parents and for lint% the selected lines of pop L and O were not significantly improved over the parents (Table 4). Thus, based on results from Pop O and L, it seems simultaneous improving of two traits like MIC and lint% that have negative association is arduous to achieve. Still the conclusion is based on four mutant lines, two for MIC (1 from ACALA1517-99, 1 from TAM94L25) and two for lint% (1,1) from many candidate mutant lines for MIC (18, 5) and lint% (7, 10) (PATEL *et al.* 2014) that can be further investigated. Testing more combinations with different mutant lines might identify alleles that can be used to improve both lint% and MIC in a single cross. Development of higher fiber strength and higher yield lines was possible

despite the traditional negative association between fiber strength and yield, suggesting that linkage between genes rather than pleiotropic effects cause this negative correlation (CULP *et al.* 1979; ZHANG *et al.* 2003a). On the other hand, crosses involving traits having positive correlation like LEN and STR (pop K), LEN and MIC (pop P), and STR and UNIF (pop Q) produced lines that were significantly improved for both traits. Crosses made between MIC and ELON (Pop J) showed no correlation between these traits in our population and we found genotypes that were improved for both traits (Figure 3), suggesting that simultaneous improvement of MIC and ELON is possible.

Within a population, the number of lines showing significant improvement for a trait relative to both parents ranges from two (lint% in pop L) to 64 (LEN in pop P) (Table 6). Pops L and P were developed using a common mutant line, Acala 1517-99-M3010, for MIC and Acala 1517-99-M1524 (pop L for lint%) or Acala 1517-99-M1903 (pop P for LEN). One reason for this variation may be that yield components are controlled by a more complex genetic network than fiber quality components. It can be Acala 1517-99-M1524 being relatively ineffective to improve lint% in GA230 or it might be also possible that GA230, which is elite germplasm, has less scope for improvement of lint% than fiber length. Another factor might be that some populations combine negatively associated traits like lint% and MIC, such as pop L. The two populations using mutant line Acala 1517-99-3010 also differ in the efficacy of MIC improvement, with 18 (18.6%) improved lines in pop P but only 7 (8%) in pop L.

Pops J and N were developed to improve ELON, based on mutant line TAM 94L25-2925 which previously showed 50% improvement over the parental lines control lines. Compared to mid-parental values, the top 10 lines showed average improvement of 27.6% and 35.6% for ELON in pop J and N, respectively, with the best lines showing 32.2% and 45.8%

improvement. This result continues to support the hypothesis that there is particularly large scope for improvement of ELON in elite germplasm such as GA230, perhaps because ELON has only recently become a priority in cotton breeding and has a short history of selection in elite germplasm (ZHANG *et al.* 2011).

Many lines showed remarkable improvement for additional fiber traits beyond those known to be conferred by the mutants used in population development. For example, line K68 of pop K showed improvement of 9.2% for MIC, 15.8% for LEN, 3.4% for UNIF, 19.6% for STR, and 22.7% for SFC with no effect on lint percent. Similarly, line P07 of pop P showed 12.8% for MIC, 11.75% for LEN, 3.3% for UNIF, 17.6% for STR and 19% for SFC. Such effects might be due to strong positive correlation between traits, thus mutants identified for single fiber traits might also contribute to improvement of other traits. Similar results were seen in the pilot study for mutant lines 3010 and 1903 from Acala 1517-99 background (Patel et al. 2014). Lines such as K68 and P07 that combine multiple desirable traits may warrant release as germplasm after verifying stability of the phenotypes and determining their yield potential, as done by previous cotton breeding programs in which mutant lines were crossed with cultivars to develop new germplasm (BECHERE *et al.* 2007a; BECHERE *et al.* 2011).

In summary, we validated that novel EMS-derived alleles confer improved fiber quality to elite cultivars beyond those in which they were identified, and can be combined to obtain genotypes that are improved for multiple fiber traits. Populations segregating for multiple mutations can be used to develop new germplasm and to map novel alleles conferring improved fiber quality, accelerating progress by using high-quality cotton genome sequences (PATERSON *et al.* 2012; LI *et al.* 2015) and contemporary molecular techniques like GBS (Genotype-By-Sequencing) which can provide plentiful numbers of SNPs (DAVEY *et al.*

2011). The availability of cotton genome sequence will further support the process (PATERSON *et al.* 2012; LI *et al.* 2015). Moreover, the 157 different mutant lines (103 Acala 1517-99, 54 TAM 94L-25) confirmed to be superior to parental or control lines for fiber traits including fiber length, uniformity, strength, fineness, elongation, Rd value and in some cases multiple attributes (PATEL *et al.* 2014) provide rich scope for further improvements beyond the alleles studied here, and may not represent saturation of the potential alleles that can be EMS-mutagenized.

Acknowledgements

We thank the Georgia Cotton Commission and Cotton Incorporated for financial support, and members of the Plant Genome Mapping Laboratory (PGML, Athens) and Molecular Cotton Breeding Laboratory (Tifton) for help in agronomic practices and harvesting.

Conflict of interest**:** The authors declare that they have no conflicts of interest.

References

Auld, D., E. Bechere, M. Ethridge, W. Becker, E. Hequet *et al.*, 2000 Registration of TTU 202-1107-B and TTU 271-2155-C mutant germplasm lines of upland cotton with improved fiber quality. Crop science 40**:** 1835-1835.

Barb, J. G., J. E. Bowers, S. Renaut, J. I. Rey, S. J. Knapp *et al.*, 2014 Chromosomal evolution and patterns of introgression in Helianthus. Genetics 197**:** 969-979.

Bechere, E., D. Auld, R. Cantrell, E. Hequet, M. Krifa *et al.*, 2007 Registration of TTU 0774-3-3 and TTU 0808-1-6-1 upland cotton germplasm lines with improved fiber length and strength. Journal of plant registrations 1**:** 58-59.

Bechere, E., D. Auld and E. Hequet, 2009 Development of 'naked-tufted'seed coat mutants for potential use in cotton production. Euphytica 167**:** 333-339.

Bechere, E., D. Auld, M. Krifa, C. W. Smith and R. Cantrell, 2011 Registration of TTU 0782, an Upland Cotton Germplasm Line with Superior Fiber Quality. Journal of plant registrations 5**:** 207-210.

Bradow, J. M., and G. H. Davidonis, 2000 Quantitation of fiber quality and the cotton production-processing interface: a physiologist's perspective. J. Cotton Sci 4**:** 34-64.

Brown, I., C. W. Smith, D. Auld, S. Hague, E. F. Hequet *et al.*, 2012 Registration of TAM 94L-25-M24, TAM 94L-25-M25, and TAM 94L-25-M30 Mutant Upland Cotton Germplasm with Improved Fiber Length and Strength. Journal of Plant Registrations 6**:** 195-199.

Brubaker, C. L., F. Bourland and J. F. Wendel, 1999 The origin and domestication of cotton. Cotton: Origin, history, technology, and production. John Wiley & Sons, New York**:** 3-31.

Cantrell, R., C. Roberts and C. Waddell, 2000 Registration of Acala 1517-99 Cotton. Crop Science 40**:** 1200-1200.

Chen, Z. J., B. E. Scheffler, E. Dennis, B. A. Triplett, T. Zhang *et al.*, 2007 Toward sequencing cotton (Gossypium) genomes. Plant physiology 145**:** 1303-1310.

Culp, T., D. Harrell and T. Kerr, 1979 Some genetic implications in the transfer of high fiber strength genes to upland cotton. Crop Science 19**:** 481-484.

Davey, J. W., P. A. Hohenlohe, P. D. Etter, J. Q. Boone, J. M. Catchen *et al.*, 2011 Genome-wide genetic marker discovery and genotyping using next-generation sequencing. Nature reviews. Genetics 12**:** 499.

Doudna, J. A., and E. Charpentier, 2014 The new frontier of genome engineering with CRISPR-Cas9. Science 346**:** 1258096.

Greene, E. A., C. A. Codomo, N. E. Taylor, J. G. Henikoff, B. J. Till *et al.*, 2003 Spectrum of chemically induced mutations from a large-scale reverse-genetic screen in Arabidopsis. Genetics 164**:** 731-740.

Herring, A. D., D. L. Auld, M. D. Ethridge, E. F. Hequet, E. Bechere *et al.*, 2004 Inheritance of fiber quality and lint yield in a chemically mutated population of cotton. Euphytica 136**:** 333-339.

Iqbal, M., O. Reddy, K. El-Zik and A. Pepper, 2001 A genetic bottleneck in the'evolution under domestication'of upland cotton Gossypium hirsutum L. examined using DNA fingerprinting. TAG Theoretical and Applied Genetics 103**:** 547-554.

Kelly, B., N. Abidi, D. Ethridge and E. F. Hequet, 2015 Fiber to fabric. Cotton**:** 665-744.

Kelly, C. M., E. F. Hequet and J. K. Dever, 2012 Interpretation of AFIS and HVI fiber property measurements in breeding for cotton fiber quality improvement. J Cotton Sci 16**:** 1-16.

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

Lubbers, E., S. Walker, L. May and P. Chee, 2006 Breeding cultivars and germplasm with enhanced yield and quality. In P. Roberts et al. (ed.) 2005 Georgia Cotton Research and Extension Reports. UGA/CPES Research – Extension Publication No. 6, University of Georgia, Athens, GA**:** 136-152.

Lubbers, E. L., and P. W. Chee, 2009 The Worldwide Gene Pool of G. hirsutum and its Improvement, pp. 23-52 in *Genetics and Genomics of Cotton*, edited by A. H. Paterson. Springer US, New York, NY.

May, O. L., 2002 Quality improvement of upland cotton (Gossypium hirsutum L.). Journal of crop production 5**:** 371-394.

McCallum, C. M., L. Comai, E. A. Greene and S. Henikoff, 2000 Targeting induced locallesions in genomes (TILLING) for plant functional genomics. Plant physiology 123**:** 439-442.

Patel, J. D., R. J. Wright, D. Auld, R. Chandnani, V. H. Goff *et al.*, 2014 Alleles conferring improved fiber quality from EMS mutagenesis of elite cotton genotypes. Theoretical and Applied Genetics 127**:** 821-830.

Patel, J. D., R. J. Wright, R. Chandnani, V. H. Goff, J. Ingles *et al.*, 2016 EMS-mutated cotton populations suggest overlapping genetic control of trichome and lint fiber variation. Euphytica 208**:** 597-608.

Paterson, A. H., R. K. Boman, S. M. Brown and P. W. Chee, 2004 Reducing the genetic vulnerability of cotton. Crop science 44**:** 1900.

Paterson, A. H., J. F. Wendel, H. Gundlach, H. Guo, J. Jenkins *et al.*, 2012 Repeated polyploidization of Gossypium genomes and the evolution of spinnable cotton fibres. Nature 492**:** 423-427.

Pickersgill, B., 2007 Domestication of Plants in the Americas: Insights from Mendelian and Molecular Genetics. Annals of Botany 100**:** 925-940.

Renny-Byfield, S., J. T. Page, J. A. Udall, W. S. Sanders, D. G. Peterson *et al.*, 2016 Independent domestication of two Old World cotton species. Genome biology and evolution 8**:** 1940-1947.

Shalem, O., N. E. Sanjana and F. Zhang, 2015 High-throughput functional genomics using CRISPR-Cas9. Nature reviews. Genetics 16**:** 299.

Smith, C. W., 2003 Registration of TAM 94L-25 and TAM 94J-3 germplasm lines of upland cotton with improved fiber length. Crop science 43**:** 742-744.

Tyagi, P., M. A. Gore, D. T. Bowman, B. T. Campbell, J. A. Udall *et al.*, 2014 Genetic diversity and population structure in the US Upland cotton (Gossypium hirsutum L.). Theoretical and Applied Genetics 127**:** 283-295.

Van Esbroeck, G., and D. T. Bowman, 1998 Cotton germplasm diversity and its importance to cultivar development. Journal of Cotton Science.

Waghmare, V. N., J. Rong, C. J. Rogers, J. E. Bowers, P. W. Chee *et al.*, 2016 Comparative transmission genetics of introgressed chromatin in Gossypium (cotton) polyploids. American journal of botany 103**:** 719-729.

Wendel, J. F., C. L. Brubaker and A. E. Percival, 1992 Genetic diversity in Gossypium hirsutum and the origin of upland cotton. American Journal of Botany**:** 1291-1310.

Wendel, J. F., and R. C. Cronn, 2003 Polyploidy and the evolutionary history of cotton. Advances in agronomy 78**:** 139-186.

Zhang, J., and R. Percy, 2007 Improving Upland cotton by introducing desirable genes from Pima cotton, pp. 10-14 in *World Cotton Research Conf*.

Zhang, T., Y. Yuan, J. Yu, W. Guo and R. J. Kohel, 2003 Molecular tagging of a major QTL for fiber strength in Upland cotton and its marker-assisted selection. Theoretical and Applied Genetics 106**:** 262-268.

Zhang, Z., J. Rong, V. N. Waghmare, P. W. Chee, O. L. May *et al.*, 2011 QTL alleles for improved fiber quality from a wild Hawaiian cotton, Gossypium tomentosum. Theoretical and applied genetics 123**:** 1075.

Tale 2.1- Mutant lines from PATEL *et al.* (2014) selected for the breeding scheme

Line #	Mean	Control mean	Parent	Selected for	Percent difference
1903	1.3	1.18	Acala 1517-99	LEN	9.60%
1793	37.1	33.8	Acala 1517-99	STR	9.70%
1524	41.4	38.7	Acala 1517-99	Lint%	7.20%
3010	4	4.71	Acala 1517-99	MIC	15.00%
2455	85.9	84.3	Acala 1517-99	UNIF	1.90%
2925	8.68	5.78	TAM 94L25	ELON	50.00%
276	43.1	40.7	TAM 94L25	Lint%	5.90%
2877	3.94	4.83	TAM 94L25	MIC	18.40%
1251	80.5	77.4	TAM 94L25	Rd value	4.10%

Table 2.2- Crossing scheme between F1 hybrids for mutant pyramiding

Pop ID	Crosses between F1 hybrid	Fiber trait targeted	Mutant parental lines	Population size
K	M1903 X GA230 M1793 x GA230	LEN X STR	Acala 1517-99	98
L	M3010 X GA230 M1524 X GA 230	MIC X Lint%	Acala 1517-99	89
P	M1930 X GA230 M3010 X GA230	LEN X MIC	Acala 1517-99	97
Q	M1793 XGA230 M2455 X GA 230	STR X UNIF	Acala 1517-99	86
J	M2925 X GA 230 M2877 X GA230	ELON X MIC	TAM 94L25	100
M	M276 X GA230 M1251 x GA230	Lint% X Rd value	TAM 94L25	45
N	M2925 X GA230 M1251 X GA230	ELONX Rd value	TAM 94L25	58
O	M2877 X GA230 M276 X GA 230	MIC X Lint%	TAM 94L25	87

Table 2.3- Heritability estimates based on parent-offspring regression for fiber quality traits in EMS-mutant-derived cotton populations

Population	MIC	LEN	UI	STR	ELON	SFC	LINT%
TAM94L25(mutants) x GA230	0.39**	0.43**	0.17*	0.30**	0.38**	0.26*	0.33**
ACALA1517-99(mutants) x GA230	0.32**	0.40**	0.27**	0.44**	0.34**	0.36**	0.41**

** and * in significant at $p < 0.01$ and at $p < 0.05$ respectively

Table 2.4-Correlations between fiber quality traits in crosses between mutant and elite cottons

	MIC	UHM	UI	STR	ELO	SFC%
TAM94L25 (mutants) x GA230						
UHM	-0.30*					
UI	0.04	0.39*				
STR	0.04	0.53*	0.44*			
ELO	0	-0.17*	0.28*	-0.07		
SFC%	-0.09	-0.42*	-0.79*	-0.51*	-0.35*	
Lint%	0.19*	-0.11*	0.19*	-0.07	0.59*	-0.17*
ACALA1517-99 (mutants) x GA230						
UHM	-0.41*					
UI	-0.10*	0.49*				
STR	-0.17*	0.58*	0.51*			
ELO	-0.01	-0.15*	0.21*	-0.12*		
SFC%	0.11*	-0.54*	-0.80'	-0.55*	-0.26*	
Lint%	0.19*	-0.19*	0.05	-0.23*	0.47*	0.01

* shows significance at $p < 0.0001$

Table 2.5- Variance components for fiber quality traits in crosses between mutant and elite cottons
(a) TAM94L25 (mutants) X GA230

	Source	DF	SS	MS	F Value	Contribution
MIC	**G**	294	115.40	0.39	2.88*	0.38
	E	2	10.89	5.45	40.01*	0.04
	G*E	586	64.58	0.11	0.81	0.21
	Error	815	110.93	0.14		
Len	**G**	294	2.86	0.01	4.27*	0.48
	E	2	0.09	0.04	18.9*	0.01
	G*E	586	1.12	0.00	0.84	0.19
	Error	815	1.86	0.00		
UI	**G**	294	858.96	2.92	1.81*	0.26
	E	2	240.79	120.40	74.77*	0.07
	G*E	586	901.25	1.54	0.96	0.27
	Error	815	1312.27	1.61		
STR	**G**	294	2256.48	7.68	2.63*	0.35
	E	2	232.37	116.19	39.83*	0.04
	G*E	586	1669.44	2.85	0.98	0.26
	Error	815	2377.29	2.92		
ELON	**G**	294	346.38	1.18	5.85*	0.19
	E	2	1219.66	609.83	3026.25*	0.66
	G*E	586	116.73	0.20	0.99	0.06
	Error	815	164.23	0.20		
SFC	**G**	294	282.57	0.96	1.68*	0.26
	E	2	90.94	45.47	79.39*	0.08
	G*E	586	250.38	0.43	0.75	0.23
	Error	815	466.79	0.57		
Lint%	**G**	294	4851.50	16.50	2.86*	0.23
	E	2	8794.74	4397.37	763.38*	0.41
	G*E	586	2994.04	5.11	0.89	0.14
	Error	815	4694.74	5.76		

* shows significance at $p < 0.0001$, G= genotype and E=environment

(b) ACALA1517-99 (mutants) X GA230

	Source	DF	SS	MS	F Value	Contribution
MIC	**G**	373	108.81	0.29	2.53*	0.35
	E	2	10.35	5.18	44.82*	0.03
	G*E	744	78.75	0.11	0.92	0.25
	Error	973	112.39	0.12		
Len	**G**	373	4.44	0.01	5.8*	0.56
	E	2	0.05	0.02	11.8*	0.01
	G*E	744	1.49	0.00	0.97	0.19
	Error	973	2.00	0.00		
UI	**G**	373	1781.84	4.78	3.24*	0.38
	E	2	425.27	212.63	144.04*	0.09
	G*E	744	1082.83	1.46	0.99	0.23
	Error	973	1436.38	1.48		
STR	**G**	373	5356.63	14.36	5.14*	0.51
	E	2	435.49	217.75	78*	0.04
	G*E	744	2053.16	2.76	0.99	0.19
	Error	973	2716.28	2.79		
ELON	**G**	373	416.47	1.12	5.72*	0.21
	E	2	1206.64	603.32	3089.75*	0.61
	G*E	744	168.78	0.23	1.16	0.09
	Error	973	189.99	0.20		
SFC	**G**	373	464.47	1.25	2.92*	0.38
	E	2	111.07	55.53	130.2*	0.09
	G*E	744	246.34	0.33	0.78	0.20
	Error	973	415.03	0.43		
Lint%	**G**	373	6888.53	18.47	3.53*	0.28
	E	2	8890.88	4445.44	849.4*	0.36
	G*E	746	3730.77	5.00	0.96	0.15
	Error	973	5092.29	5.23		

* shows significance at $p < 0.0001$

Table 2.6 - Comparison between parentals or checks and populations from crosses between elite cotton line GA230 and mutants for fiber quality traits

Pop id	Trait	Pop mean	Bkgrd (Acala or TAM)	GA230	Midparent (Bkgrd + GA230 avg.)	% improvement over midparent	Fiber Max 832	Delta Pine 393	Total # of improved lines in pop	Best line	% improv to mid-parent
ACALA1517-99 (mutants) X GA230											
K	LEN	1.21	1.12*	1.17*	1.15	5.7	1.14*	1.13*	40	k68	15.8
K	STR	32.31	29.15*	30.48*	29.82	8.4	30.6*	30.3*	54	k92	22.4
L	MIC	4.57	4.54	4.55	4.55	0.6	4.26¥	4.57	7	L078	-10.4
L	Lint%	40.56	38.01*	41.04	39.53	2.6	40.2	39.5	2	L071	13.1
P	LEN	1.23	1.12*	1.17*	1.15	7.4	1.14*	1.13*	67	P058	16.1
P	MIC	4.46	4.54	4.55	4.55	-1.9	4.26¥	4.57	18	P007	-12.8
Q	STR	31.83	29.15*	30.48*	29.82	6.8	30.6*	30.3*	36	Q85	17.9
Q	UNIF	84.18	82.11*	83.6*	82.86	1.6	83*	82.7*	15	Q47	4.1
TAM94L25 (mutants) XGA230											
J	ELON	5.36	4.48*	4.96*	4.72	13.6	5.11	5.28	9	J45	32.2
J	MIC	4.36	4.53*	4.55*	4.54	-4	4.26	4.57*	30	J29	-19
M	Lint%	42.2	38.93*	41.04	39.99	5.5	40.2*	39.5*	2	M63	11.6
N	ELON	5.66	4.48*	4.96*	4.72	19.9	5.11*	5.28	11	N068	45.8
O	MIC	4.32	4.53*	4.55*	4.54	-4.8	4.26	4.57*	25	O013	-31.7
O	Lint%	41.13	38.93*	41.04	39.99	2.9	40.2	39.5*	3	O014	13.9

Parental or checks cells with "*" are significantly inferior to population average by $p<0.01$ and ¥ are significantly superior to population average by $p<0.01$

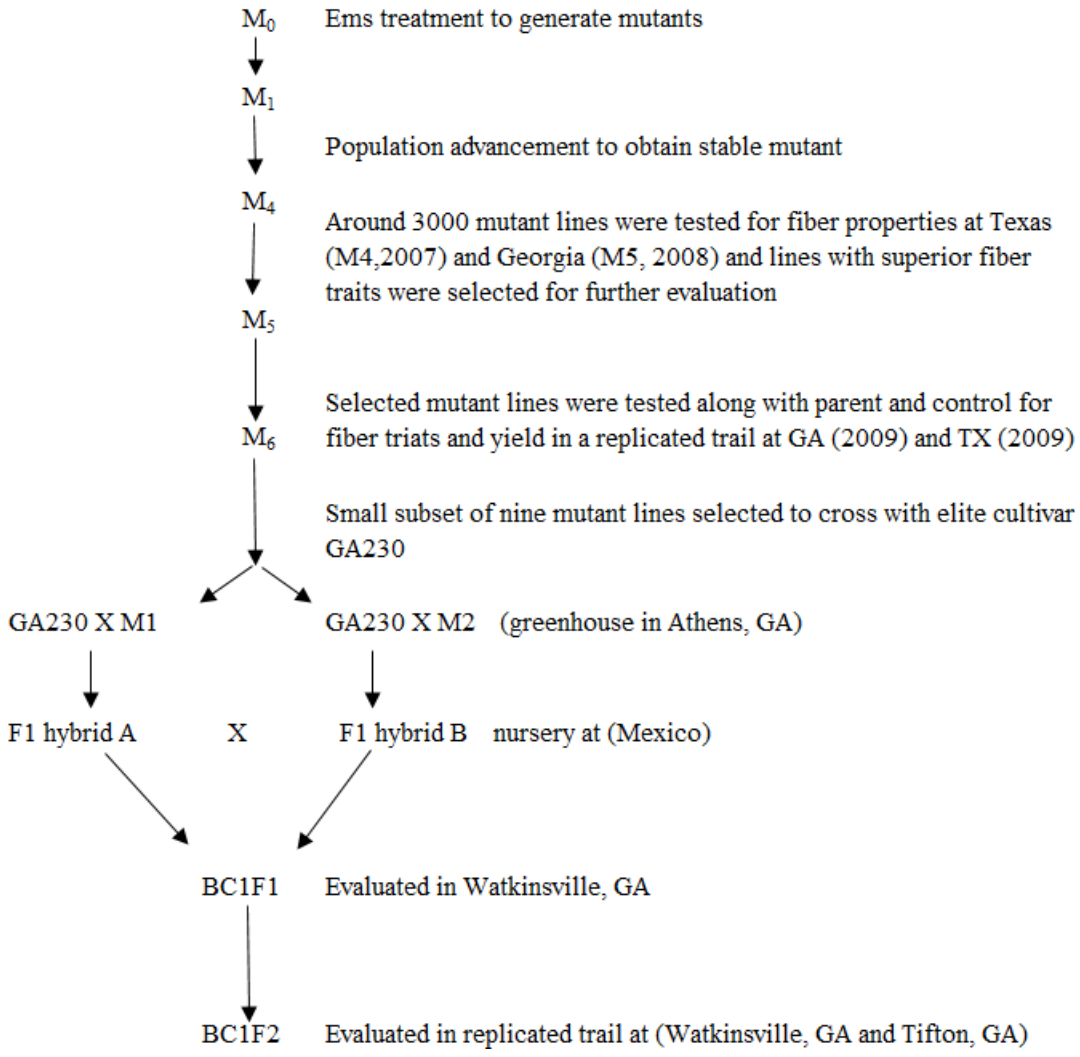


Figure 2.1- Development of populations for evaluation of multiple EMS-induced mutants

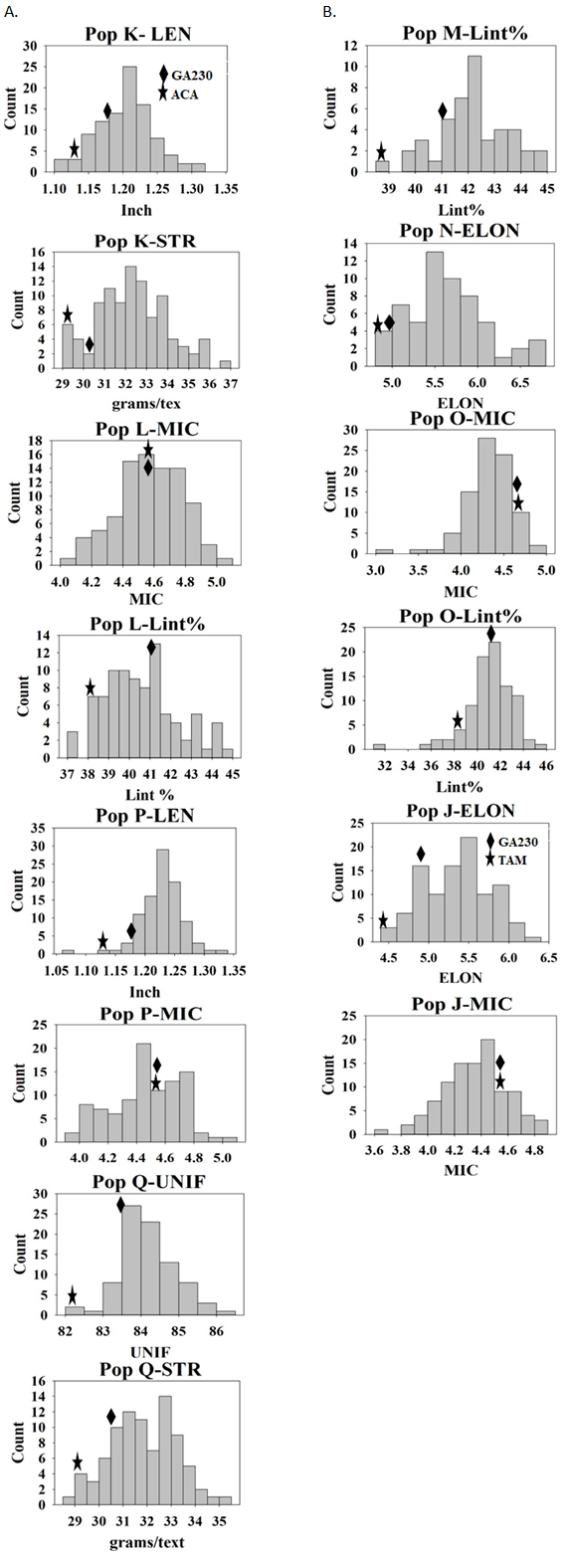
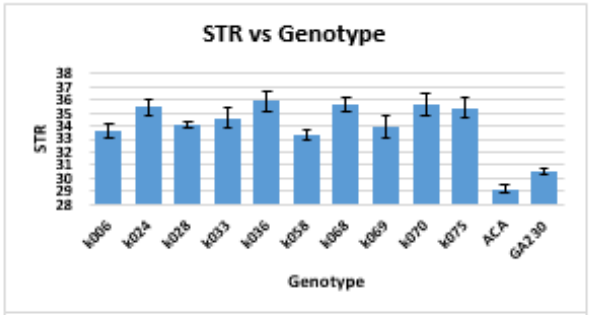
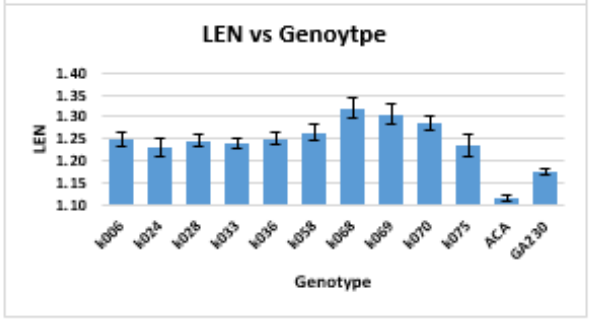


Figure 2.2 - Distribution of genotypes in populations for fiber traits they were developed. (A) populations from ACALA1517-99 (mutant) X GA230 and (B) populations from TAM94L25 (mutant) X GA230

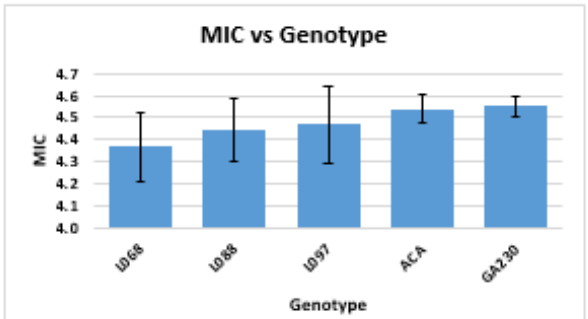
A.

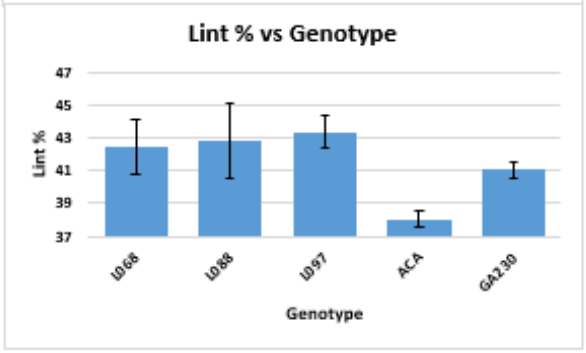
a.


STR vs Genotype
STR
38
37
36
35
34
33
32
31
30
29
28
k006
k024
k028
k033
k036
k058
k068
k069
k070
k075
ACA
GA230
Genotype

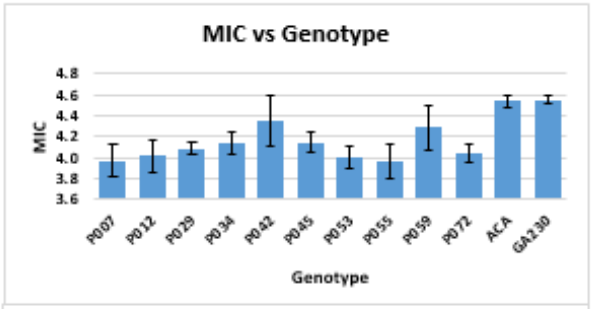

LEN vs Genoytpe
LEN
1.40
1.35
1.30
1.25
1.20
1.15
1.10
k006
k024
k028
k033
k036
k058
k068
k069
k070
k075
ACA
GA230
Genotype

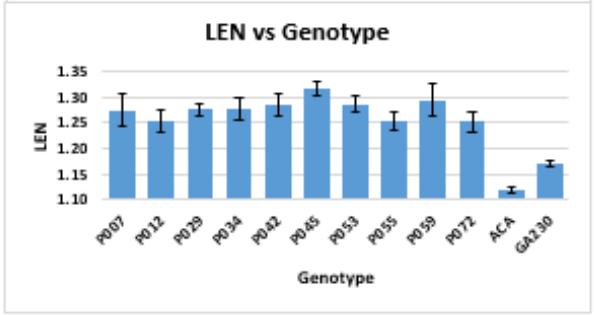
b.


MIC vs Genotype
MIC
4.7
4.6
4.5
4.4
4.3
4.2
4.1
4.0
L068
L088
L097
ACA
GA230
Genotype

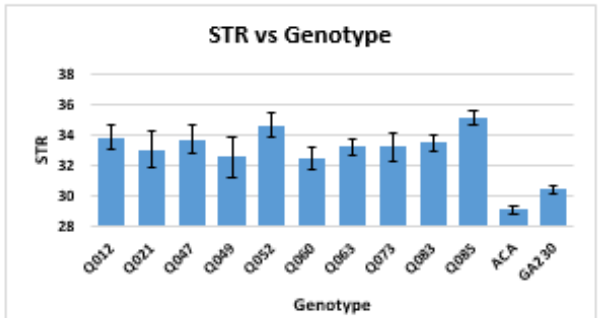

Lint % vs Genotype
Lint %
47
45
43
41
39
37
L068
L088
L097
ACA
GA230
Genotype

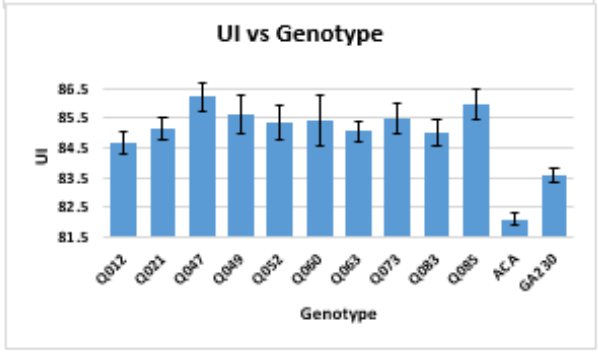
c.


MIC vs Genotype
MIC
4.8
4.6
4.4
4.2
4.0
3.8
3.6
P007
P012
P029
P034
P042
P045
P053
P055
P059
P072
ACA
GA230
Genotype


LEN vs Genotype
LEN
1.35
1.30
1.25
1.20
1.15
1.10
P007
P012
P029
P034
P042
P045
P053
P055
P059
P072
ACA
GA230
Genotype

d.


STR vs Genotype
STR
38
36
34
32
30
28
Q012
Q021
Q047
Q049
Q052
Q060
Q063
Q073
Q083
Q085
ACA
GA230
Genotype


UI vs Genotype
UI
86.5
85.5
84.5
83.5
82.5
81.5
Q012
Q021
Q047
Q049
Q052
Q060
Q063
Q073
Q083
Q085
ACA
GA230
Genotype

B.

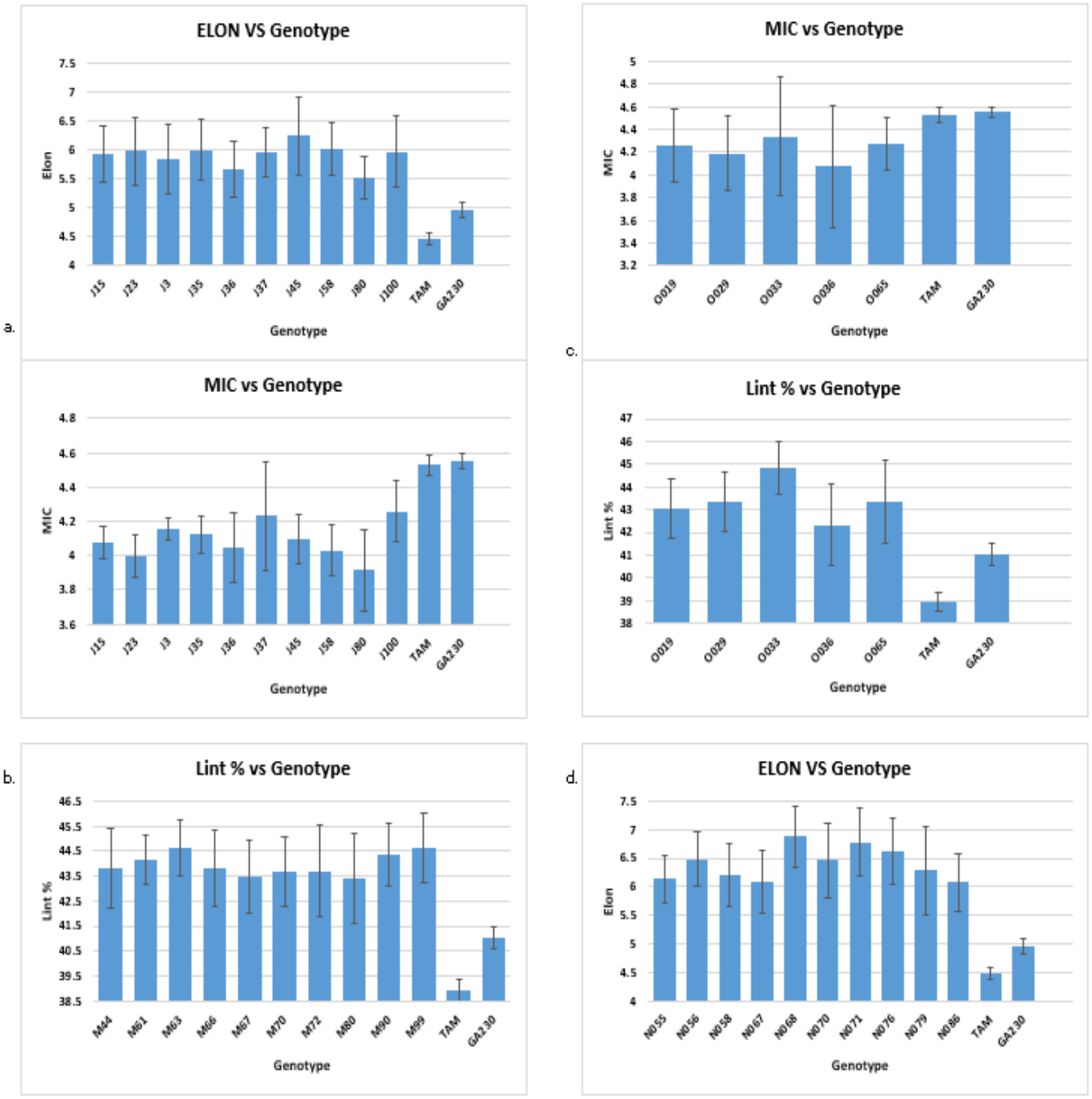


Figure 2.3 - Selected lines showing improvements over parental lines for multiple fiber traits (A) ACALA1517-99 (mutant) X GA230; and (B) TAM94L25 (mutant) X GA230. In graphs A-a is pop K, A-b is pop L, A-c is pop P, A-d is pop Q and B-a is pop J, B-b is pop M, B-c is pop O, and B-d is pop N. Error bar indicates standard error (S.E)

Chapter 3

PYRAMIDING NOVEL EMS-GENERATED MUTANT ALLELES TO IMPROVE FIBER QUALITY COMPONENTS OF ELITE UPLAND COTTON GERMPLASM.

Patel, Jinesh[1]; Chandnani, Rahul[1]; Khanal, Sameer[1]; Adhikari, Jeevan[1]; Brown, Ismail [2]**;** Chee, Peng W[2]; Jones, Don[3] and Paterson, Andrew H[1 *] To be submitted to Field Crops Research

Abstract

Improvement of cotton fiber quality, hampered by historical genetic bottlenecks, may benefit from the use of EMS-induced mutants that are largely free of linked unfavorable alleles often associated with the use of secondary and tertiary *Gossypium* gene pools. Here we intercrossed seven of 157 EMS-generated improved fiber quality mutant lines to produce two populations, one (pop R) focused on improving four fiber attributes (micronaire, length, strength and elongation) and the other (pop S) to pyramid novel alleles for fiber length alone. The overall average of both populations was significantly improved for micronaire, fiber length, fiber strength, uniformity and short fiber content compared to parental lines, with 39 lines in pop. R and 71 in pop. S showing significant improvement for four or more traits. Lines R048, S003, S042 and S050 showed improvement for all six fiber traits tested. Fiber length of populations S and R was significantly higher than EMS mutated parent (ACALA1517-99, TAM94L25), local elite germplasm (GA230) and other commercial checks (DeltaPine 393 and Fiber Max 832). As expected average fiber length of pop. S was significantly higher by 4.2% than pop. R. Surprisingly, pop S was also significantly better than pop R in micronaire, fiber strength, uniformity and short fiber content, adding further support to hypotheses about the complex nature of cotton fiber QTLs and the corollary that selection for one fiber quality trait may also increase values of other traits. New alleles from these mutant lines show promise for improving fiber qualities beyond the levels of current elite varieties.

Keywords: crop improvement, functional genomics, EMS mutagenesis, allele stacking, genetic diversity

Introduction

Despite its global importance in contributing about one-third of the raw material used by textile industries and its cultivation in about 100 countries (CONSTABLE *et al.* 2015), allotetraploid Upland cotton (*Gossypium hirsutum L.*) improvement has been constrained by several historical genetic bottlenecks (ULLOA AND MEREDITH JR 2000; PATERSON *et al.* 2004; TYAGI *et al.* 2014; BOOPATHI AND HOFFMANN 2016). These bottlenecks are results of polyploid formation, domestication, human movement of small germplasm samples and intensive breeding for yield components (LUBBERS AND CHEE 2009).

Further increasing the need for new genetic variation, the era of high throughput rotor and air jet spinning machines in the textile industry (BRADOW AND DAVIDONIS 2000) motivated increased focus on cotton fiber quality parameters including length, strength, uniformity, micronaire and elongation, in addition to long-standing efforts to improve yield and yield components. Fiber fineness (micronaire) and elongation have particularly short histories of selection, making it rare to find an elite line with a high concentration of favorable alleles suitable to improve such attributes (ZHANG *et al.* 2011; PATEL *et al.* 2014).

Chemical mutagens such as EMS (Ethyl methanesulfonate) that generate single nucleotide 'point' mutations offer a means to create new alleles that are expected to experience fewer problems associated with linkage drag of unfavorable alleles than alleles introgressed from wild species (REINISCH *et al.* 1994; TANKSLEY AND MCCOUCH 1997; BARB *et al.* 2014; WAGHMARE *et al.* 2016). Researchers have identified EMS mutations conferring discrete morphological traits such as trichome variations (PATEL *et al.* 2016), naked seed (BECHERE *et al.* 2009a; PATEL *et al.* 2014; KONG DEPEI 2017), short fiber mutants (KONG DEPEI 2017; NAOUMKINA *et al.* 2017), albino cotyledons and leaves, red-violet leaves and stems, and multilayered bracts (KONG DEPEI

2017). Multiple years of field trials showed that mutant lines with superior fiber properties can be developed through mutation breeding (AULD *et al.* 2000; BECHERE *et al.* 2007b; BROWN *et al.* 2012; PATEL *et al.* 2014) but only a handful of attempts have been made to transfer EMS alleles into elite backgrounds and none to our knowledge have investigated combining such alleles in elite or mutant backgrounds (BECHERE *et al.* 2007b; BECHERE *et al.* 2011).

Experiments on gene or QTL pyramiding have been conducted in different plant species with a major focus on developing lines resistant to biotic or abiotic stresses (GREGORIO *et al.* 2002; ATKINSON AND URWIN 2012). In cotton, QTL pyramiding have been reported to improve fiber qualities such as fiber length and strength (WANGZHEN *et al.* 2005; DONG *et al.* 2009; YUAN *et al.* 2014b). In our prior work, we discovered multiple mutant lines that were improved for fiber length, strength, elongation, fineness, uniformity, Rd value and lint percent (PATEL *et al.* 2014). Here, we developed two populations, one combining four mutant lines that have shown improvement for multiple fiber traits (fiber length, strength, fineness and elongation) and another combining four mutant lines that all had strikingly improved fiber length. Such populations have allowed us to investigate interactions between different fiber traits, effects of allele pyramiding for the same or different fiber traits, and the possibility to break negative associations between yield and fiber quality components by crossing novel alleles generated by EMS mutagenesis.

Materials and Methods

Plant sources and population development

A total of seven mutant lines (Table 1) were used to develop two populations. Four lines, namely Acala 1517-99-M1903 (fiber length, herein abbreviated LEN), Acala 1517-99-M1793 (fiber strength, STR), TAM94L25-M2925 (fiber elongation, ELON), and TAM94L25-M2877(micronaire, MIC) were used to develop ‘pop R’; and four lines, namely, Acala 1517-99 -

M1903, Acala 1517-99 -3028, TAM94L25-M926, and TAM94L25-M2888 showing improved LEN were used to develop pop S. These lines were selected from a set of 157 mutant lines in two different genetic backgrounds [*G. hirsutum* viz. TAM94L25 (SMITH 2003) and Acala 1517-99 (CANTRELL *et al.* 2000)] that showed striking improvement over wild type progenitor ('parent') or control lines. The pilot results were supported by replicated trials in multiple environments (PATEL *et al.* 2014). In a greenhouse at Athens, GA (Summer, 2012), two mutant lines were crossed with one another to develop F1's and in an off-season nursery in Tecoman, Mexico (Winter, 2012), the F1 hybrids were further crossed with each other, thus developing populations that behave like F2 and combine four mutant lines (Table 2).

Field trial and data collection

A total of 100 F2 progenies from each population were grown in Watkinsville, GA (soil type-fine, kaolinitic, thermic typic kanhapludults) in May 2013. In 2014, a total of 95 individuals from pop R and 94 from pop S were evaluated at two locations with two replications (i.e., Watkinsville and Tifton), in a randomized complete block design (RCB). The soil type at Watkinsville, GA was Appling Coarse Sandy Loam (fine, kaolinitic, thermic typic kanhapludults) and at Gibbs farm, Tifton, GA was Tifton loamy sand (fine, loamy, siliceous, thermic Plinthic kandiudult). For all three environments both parents TAM94L25 and ACALA1517-99, TXA (TAM94L25 x ACALA1517-99), plus three checks GA230 (PVP 201500309), Fiber Max 832 (PVP 9800258), and Delta Pine 393 (PVP 200400266) were replicated 10 times for each replication in the field. A total of 35 seeds were planted in a single-row plot of 3m, with plots spaced 1 m apart. Agronomic practices like weeding, irrigation, fertilizer application and pest management were conducted as per standard cotton growing practices. To obtain fiber samples, bolls were hand-picked in Athens (November 25, 2013) and

Tifton (October 26, 2014) while seed cotton samples were collected from machine harvested cotton in Athens (November 19, 2014) and ginned using a 20-saw gin (DENNIS MFG. CO., INC.). Lint weight and seed weight (seed plus fuzz) were measured and lint percent (lint weight X 100/seed cotton weight) was calculated. Samples of 10 grams of cotton fiber were sent to Cotton Inc. to measure HVI fiber properties, namely upper half mean fiber fineness or micronaire (MIC), fiber length (LEN), fiber strength (STR), fiber elongation (ELON), Uniformity index (UI) and Short Fiber Content (SFC).

Data analysis

Data was analyzed using SAS 9.4 (**SAS** Institute Inc., Cary, NC, USA). The program statement, "Proc CORR" was used to determine correlations between fiber traits. Heritabilities of fiber traits were calculated using parent-offspring regression by the SAS "Proc REG" statement. The contribution and significance of genotype, environment and interaction between genotype and environment for fiber traits was calculated using the SAS statement "Proc GLM". The means of each fiber trait of pop R and pop S were compared with means of wild type progenitors, checks and each other by Fisher's LSD test at alpha level of 0.01 to identify significant differences. Only, feasible number of lines from a population which eliminates error of by chance improvement are carried forward for further testing and releasing as germplasm. The average of the top 10 lines for the focal fiber trait(s) of the respective populations was compared with TXA (TAM94L25 X ACALA1517-99) to see percentage gains in these lines and assess their potential merit in breeding programs to improve fiber traits. For the analysis, genotype, environment, replication and selection (top 10 lines) were considered fixed variables.

Results

Heritability of fiber traits and association between the fiber traits

Parent-offspring regression was done to calculate heritability between F2 and F3 for each fiber trait. Overall Lint_% showed the lowest heritability (0.25) while LEN showed highest heritability (0.44) (Table 3). Trends of heritability for different fiber traits were consistent with previous reports (HERRING *et al.* 2004).

Correlation coefficients were used to study association between fiber traits. In both populations, positive correlation was seen between LEN and UNIF, LEN and STR, and STR and UNIF, indicating that simultaneous improvement of LEN, STR and UNIF is possible. For MIC and SFC, a negative correlation of these traits with other fiber traits is favorable as low values of each of these traits are preferred. In both populations, SFC showed negative correlation with LEN, UNIF, STR and ELON, which means improving one or more of these fiber traits may also improve SFC. There was a negative correlation between MIC and LEN in both populations. In pop R, no correlation was observed between MIC and STR, MIC and SFC, and MIC and lint%. In pop S, there was favorable correlation between MIC and STR, and MIC and SFC; and unfavorable correlation between MIC and lint%. In pop R, ELON and UNIF had no correlation but a positive correlation in pop S. lint% was positively correlated with ELON in both populations. STR had minor negative correlation to lint% in pop S (Table 4).

Analysis of Variance

Analysis of variance in both populations showed significant difference between genotypes and between environments but no significant genotype X environment interaction. In pop R, the contribution of genotype to overall variance was lowest for ELON (19.3%) and highest for LEN

(50.5%) and in pop S, contribution of genotype to overall variance was lowest for ELON (21.5%) and highest for MIC (48.8%) (Table 5).

Fiber traits

MIC (fiber fineness)

Both populations showed significant improvement for MIC when compared to TAM (TAM94L25), ACA (ACALA1517-99), TXA and DP (Delta Pine393), but no significant difference was found between these populations and FM (Fiber Max 832). The overall mean of Pop R showed 4.5 % improvement compared to TXA, and pop S showed 7.3% improvement with significantly better MIC than pop R. Totals of 23 and 49 lines of pop R and pop S respectively showed significant improvement over both parents and TXA. The greatest improvements were shown by R093 in pop R (17.2%) and S079 in pop S (19.4%: Table 6).

LEN, UNIF, STR AND SFC

Both populations showed significant improvement for LEN, UNIF, STR and SFC with respect to all parents and checks, with the average of pop S significantly better than pop R for LEN, UNIF, STR and SFC. Compared to TXA, pop R showed an average 3.4%, 1.3%, 6.4%, and 8.8% improvement for LEN, UNIF, STR and SFC, respectively; while pop S showed 7.8%, 2.1%, 12.8% and 13.7%. The number of lines exceeding parental values for LEN, UNIF, STR and SFC, respectively was 55, 43, 38, and 59 for pop R; and 78, 74, 87, and 87 for pop S. The best lines for LEN, UNIF, STR and SFC, respectively, were R001 (10.5% improvement), R073 (4%), R090 (16.6%), and R073 (19.5%) in pop R; and S044 (14.9% improvement), S027 (4.3%), S018 (21.8%), and S044 (24.8%) in pop S (Table 6).

ELON

For ELON, pop R showed significant improvement compared to TAM and TXA, but no significant difference from ACA and the two elite checks. Pop S showed significant improvement over TAM but no significant improvement over TXA and FM, and was significantly inferior to ACA, DP and pop R. Compared to TXA, the mean of pop R showed 15.3% improvement whereas pop S showed 5.2% improvement. No line in either population had significantly higher ELON than ACA, but 39 and 9 genotypes in pop R and pop S, respectively, showed significant improvement over TXA. Line R038 of pop R showed 32% improvement compared to TXA while S025 showed 21.3% improvement.

Lint %

Both populations had significantly lower lint% compared to all the parental and checks. Average of pop R was 6.5% lower and pop S was 7.7% lower than TXA. No line in either population was significantly better than the parental lines. However, 60 and 41 genotypes in pop R and pop S, respectively, had improved fiber quality and were not significantly different for lint% than the parental lines.

Discussion

Building on recent evidence that EMS-induced mutants may contribute substantially to mitigating a lack of genetic diversity owing to genetic bottlenecks during cotton evolution, domestication, selection and crop breeding practices (BECHERE *et al.* 2007b; BROWN *et al.* 2012; PATEL *et al.* 2014), the present research validates additional EMS-mutants for roles in cotton fiber quality (beyond what were validated in a companion study, Patel et al. unpublished) and explores the effects of pyramiding multiple mutants.

Surprisingly, pop S, successfully combining multiple mutations to improve LEN alone, also had better MIC, STR, UNIF and SFC than pop R, combining mutants to improve LEN, MIC, STR and ELON, which suggests that it is possible to improve multiple fiber attributes by targeting single fiber quality traits such as LEN. This might be due to the presence of fiber QTL hotspots comprised of dozens of genes with coordinated expression during different stages of fiber development (PATERSON *et al.* 2012). Thus, by editing a single gene through EMS-mutagenesis or other mechanisms, we might affect the function of other genes that might produce pleiotropic effects. There might be a negative effect of one of the mutant lines TAM94L25-M2925 (selected for ELON) which was used to develop pop R as ELON generally have negative association with other fiber traits.

SFC and UNIF were not directly targeted in this breeding program (although SFC is clearly related to LEN) but we still found striking improvement in both mutant populations compared to all parents and checks. SFC is a major factor contributing to irregularity in yarn and reducing its strength (THIBODEAUX *et al.* 2008; CAI *et al.* 2011). The number of neps (small knot of entangled fibers in fabric) that reduce the overall quality of yarn is also positively associated with SFC (VAN DER SLUIJS AND HUNTER 1999; ULLOA 2006). Here we found lines such as R073 and S44 with reduced SFC content by 19.5% and 24.8% respectively with respect to TXA, that can be used to improve yarn quality.

Lint % is an important component of cotton yield. Pop R had better lint% than pop S, but both populations had lower lint% than parents and checks. This was expected as yield components are strongly negatively associated with fiber quality (MEREDITH 1984; CLEMENT *et al.* 2012; CONSTABLE *et al.* 2015). Still, multiple lines in both populations had improved fiber qualities with no adverse effect on lint%, suggesting that negative association between fiber traits

and yield components could be overcome, as also suggested in previous research (CLEMENT *et al.* 2015). Intermating among such lines coupled with recurrent selection may weaken negative associations and produce lines with superior fiber qualities and adequate Lint%. Similar strategies have been suggested by CLEMENT *et al.* (2012) to break negative correlations between yield components and fiber quality.

Multiples lines in each population showed improvement for more than one fiber trait. In pop S, 79% of total lines in the population showed significant improvement in at least four fiber traits over TXA, with 29 lines were improved for four fiber traits, 39 for five fiber traits, and three (S3, S42 and S50) for six fiber traits (LEN, STR, UNIF, MIC, ELON and SFC), with multiple lines showing no significant difference for ELON and lint% when compared to parental and TXA. This further supports our hypothesis that breeding programs using such mutant lines for one fiber trait can simultaneously improve other fiber traits. For pop R, 41% of total lines in the population showed significant improvement over TXA in at least four fiber traits, with 22 lines improved for four fiber traits 16 for five and one (R048) for six (LEN, STR, UNIF, MIC, ELON and SFC).

A few lines stood out as being particularly well suited for direct use to improve fiber quality in mainstream breeding programs. For example, genotype S050 of pop S showed striking improvement of 14.1% in MIC, 13.5% in LEN, 3.8% in UNIF, 20.1% in STR, 16.1% in ELON and 23% in SFC content compared to TXA but had significantly lower lint%. Genotype S032 showed improvement of 12.6% for MIC, 12.7% in LEN, 2.8% in UNIF, 17.4% in STR, 11.8% in ELON and 21.8% in SFC and no difference in lint% with respect to TXA. Similarly, S046 showed improvement of 15.9% for MIC, 12.1% in LEN, 2.3% in UNIF, 13.2% in STR and 18% in SFC and no difference in ELON and lint% with respect to TXA. In pop R, R021 showed improvement of 6.8% in LEN, 2.1% in UNIF, 12.5% in STR, 21.3% in ELON, and 14.7% in

SFC content compared to TXA and line R090 showed 12.9% in MIC 6.2% in LEN, 2% in UNIF, 16.6% in STR, and 12.9% in SFC content with respect to TXA. Such lines were not significantly different for lint% as compare to TXA. Thus, lines like R021, R090, S032, S046 and other genotypes showing improvement in multiple fiber traits with no adverse effect on lint% can be further tested for yield trials and for stability of the improvement before releasing them as germplasm lines.

In summary, the present research shows the opportunity for simultaneously improving multiple traits and the merit of pyramiding independent EMS-induced mutants for a trait. The overall mean of pop S was better for LEN, MIC and STR than pop R and more lines with improvement in multiple fiber traits were found in pop S than R, despite that pop R was developed to improve four fiber traits (MIC, LEN, STR and ELON) simultaneously while pop S was developed to improve LEN alone. Further work is needed to determine if this is a general trend or peculiar to these particular sets of mutants, and to investigate consequences for yield components and other traits. It would of course be interesting to identify such mutants that presumably have pleiotropic effects on multiple fiber traits. The ability to manipulate germplasm containing discrete mutations affecting fiber traits provides new insight into cotton breeding strategies, that may inform fiber improvement programs using natural or induced alleles, as well as CRISPR/CAS9 genome editing (DOUDNA AND CHARPENTIER 2014; SHALEM *et al.* 2015).

Acknowledgements

We thank the Georgia cotton commodity commission and Cotton Incorporated for financial support, and members of the Plant Genome Mapping Laboratory (PGML, Athens) and Molecular Cotton Breeding Laboratory (Tifton) for helping in agronomic practices and harvesting.

Conflict of interest: The authors declare that they have no conflicts of interest.

References

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

Barb JG, Bowers JE, Renaut S, Rey JI, Knapp SJ, Rieseberg LH, Burke JM (2014) Chromosomal evolution and patterns of introgression in Helianthus. Genetics 197:969-979

Bechere E, Auld D, Hequet E (2009) Development of 'naked-tufted'seed coat mutants for potential use in cotton production. Euphytica 167:333-339

Bechere E, Auld D, Krifa M, Smith CW, Cantrell R (2011) Registration of TTU 0782, an Upland Cotton Germplasm Line with Superior Fiber Quality. Journal of plant registrations 5:207-210

Bechere E, Auld DL, Cantrell RG, Hequet E, Krifa M, Misra S, Smith CW (2007) Registration of TTU 0774-3-3 and TTU 0808-1-6-1 Upland Cotton Germplasm Lines with Improved Fiber Length and Strength. J Plant Reg 1:58-59

Boopathi NM, Hoffmann LV (2016) Genetic Diversity, Erosion, and Population Structure in Cotton Genetic Resources. Genetic Diversity and Erosion in Plants. Springer, pp 409-438

Bradow JM, Davidonis GH (2000) Quantitation of fiber quality and the cotton production-processing interface: a physiologist's perspective. J Cotton Sci 4:34-64

Brown I, Smith CW, Auld D, Hague S, Hequet EF, Jones D (2012) Registration of TAM 94L-25-M24, TAM 94L-25-M25, and TAM 94L-25-M30 Mutant Upland Cotton Germplasm with Improved Fiber Length and Strength. Journal of Plant Registrations 6:195-199

Cai Y, Cui X, Rodgers J, Thibodeaux D, Martin V, Watson M, Pang S-S (2011) An investigation on different parameters used for characterizing short cotton fibers. Textile Research Journal 81:239-246

Cantrell R, Roberts C, Waddell C (2000) Registration of Acala 1517-99 Cotton. Crop Science 40:1200-1200

Clement J, Constable G, Stiller W, Liu S (2012) Negative associations still exist between yield and fibre quality in cotton breeding programs in Australia and USA. Field crops research 128:1-7

Clement JD, Constable GA, Stiller WN, Liu SM (2015) Early generation selection strategies for breeding better combinations of cotton yield and fibre quality. Field Crops Research 172:145-152

Constable G, Llewellyn D, Walford SA, Clement JD (2015) Cotton Breeding for Fiber Quality Improvement. In: Cruz VMV, Dierig DA (eds) Industrial Crops: Breeding for BioEnergy and Bioproducts. Springer New York, New York, NY, pp 191-232

Dong Z, Shi Y, Zhang J, Wang S, Li J, Liu A, Tang S, CHU P, YUAN Y-l (2009) Molecular marker-assisted selection and pyramiding breeding of major QTLs for cotton fiber length. Cotton Sci 21:279-283

Doudna JA, Charpentier E (2014) The new frontier of genome engineering with CRISPR-Cas9. Science 346:1258096

Endo T, Chiba B, Wagatsuma K, Saeki K, Ando T, Shomura A, Mizubayashi T, Ueda T, Yamamoto T, Nishio T (2016) Detection of QTLs for cold tolerance of rice cultivar 'Kuchum'and effect of QTL pyramiding. Theoretical and applied genetics 129:631-640

Fukuoka S, Saka N, Mizukami Y, Koga H, Yamanouchi U, Yoshioka Y, Hayashi N, Ebana K, Mizobuchi R, Yano M (2015) Gene pyramiding enhances durable blast disease resistance in rice. Scientific reports 5:7773

GUO X-h, CAI C-p, YUAN D-d, ZHANG R-s, XI J-l, GUO W-z (2016) Development and identification of Verticillium wilt-resistant upland cotton accessions by pyramiding QTL related to resistance. Journal of Integrative Agriculture 15:512-520

Hattori Y, Nagai K, Furukawa S, Song X-J, Kawano R, Sakakibara H, Wu J, Matsumoto T, Yoshimura A, Kitano H (2009) The ethylene response factors SNORKEL1 and SNORKEL2 allow rice to adapt to deep water. Nature 460:1026

Herring AD, Auld DL, Ethridge MD, Hequet EF, Bechere E, Green CJ, Cantrell RG (2004) Inheritance of fiber quality and lint yield in a chemically mutated population of cotton. Euphytica 136:333-339

Kong Depei QL, Zhang Xueyan, Liu Ji, Wang Peng, Li Fuguang (2017) Optimization of EMS Mutagenesis Condition and Screening of Mutants in Gossypium arboretum. L. Cotton Science 29:336-344

Lubbers EL, Chee PW (2009) The Worldwide Gene Pool of G. hirsutum and its Improvement. In: Paterson AH (ed) Genetics and Genomics of Cotton. Springer US, New York, NY, pp 23-52

Meredith WR (1984) Quantitative genetics. Cotton:131-150

Naoumkina M, Bechere E, Fang DD, Thyssen GN, Florane CB (2017) Genome-wide analysis of gene expression of EMS-induced short fiber mutant Ligon lintless-y (liy) in cotton (Gossypium hirsutum L.). Genomics 109:320-329

Pang Y, Chen K, Wang X, Wang W, Xu J, Ali J, Li Z (2017) Simultaneous Improvement and Genetic Dissection of Salt Tolerance of Rice (Oryza sativa L.) by Designed QTL Pyramiding. Frontiers in Plant Science 8

Patel JD, Wright RJ, Auld D, Chandnani R, Goff VH, Ingles J, Pierce GJ, Torres MJ, Paterson AH (2014) Alleles conferring improved fiber quality from EMS mutagenesis of elite cotton genotypes. Theoretical and Applied Genetics 127:821-830

Patel JD, Wright RJ, Chandnani R, Goff VH, Ingles J, Paterson AH (2016) EMS-mutated cotton populations suggest overlapping genetic control of trichome and lint fiber variation. Euphytica 208:597-608

Paterson AH, Boman RK, Brown SM, Chee PW (2004) Reducing the genetic vulnerability of cotton. Crop science 44:1900

Paterson AH, Wendel JF, Gundlach H, Guo H, Jenkins J, Jin D, Llewellyn D, Showmaker KC, Shu S, Udall J (2012) Repeated polyploidization of Gossypium genomes and the evolution of spinnable cotton fibres. Nature 492:423-427

Reinisch AJ, Dong JM, Brubaker CL, Stelly DM, Wendel JF, Paterson AH (1994) A detailed RFLP map of cotton, Gossypium hirsutum x Gossypium barbadense: chromosome organization and evolution in a disomic polyploid genome. Genetics 138:829

Shalem O, Sanjana NE, Zhang F (2015) High-throughput functional genomics using CRISPR-Cas9. Nature reviews Genetics 16:299

Smith CW (2003) Registration of TAM 94L-25 and TAM 94J-3 germplasm lines of upland cotton with improved fiber length. Crop science 43:742-744

Tanksley SD, McCouch SR (1997) Seed banks and molecular maps: unlocking genetic potential from the wild. Science 277:1063-1066

Thibodeaux D, Senter H, Knowlton J, Mcalister D, Cui X (2008) The impact of short fiber content on the quality of cotton ring spun yarn. J Cotton Sci 12:368-377

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

Ulloa M (2006) Heritability and correlations of agronomic and fiber traits in an okra-leaf upland cotton population. Crop science 46:1508-1514

Ulloa M, Meredith Jr WR (2000) Genetic linkage map and QTL analysis of agronomic and fiber quality traits in an intraspecific population. Journal of Cotton Science 4:161-170

Van der Sluijs M, Hunter L (1999) Neps in cotton lint. Textile Progress 28:1-47

Waghmare VN, Rong J, Rogers CJ, Bowers JE, Chee PW, Gannaway JR, Katageri I, Paterson AH (2016) Comparative transmission genetics of introgressed chromatin in Gossypium (cotton) polyploids. American journal of botany 103:719-729

Wangzhen G, Tianzhen Z, Yezhang D (2005) Molecular marker assisted selection and pyramiding of two QTLs for fiber strength in upland cotton. Acta Genetica Sinica 32:1275-1285

Yuan Y, Wang T, Shi Y, Shang H, Liu A, Li J, Gong J, Wang T, Gong W-k, Chen T, Li B (2014) Molecular marker-assisted selection and pyramiding effect of major QTLs for cotton fiber strength. New Biotechnology 31:S14

Zhang Z, Rong J, Waghmare VN, Chee PW, May OL, Wright RJ, Gannaway JR, Paterson AH (2011) QTL alleles for improved fiber quality from a wild Hawaiian cotton, Gossypium tomentosum. Theoretical and applied genetics 123:1075

Table 3.1- Superior fiber quality mutant lines selected from PATEL *et al.* (2014) for population development

Mutant id	Mean	Control mean	Significance	Background	Improvement in	% improv.
926	1.26	1.15	0.0001	TAM 94L25	LEN	9.00%
1903	1.3	1.18	0.0001	Acala 1517-99	LEN	9.60%
2888	1.25	1.15	0.0001	TAM 94L25	LEN	8.60%
3028	1.27	1.18	0.001	Acala 1517-99	LEN	7.20%
2925	8.68	5.78	0.0001	TAM 94L25	ELONG	50.00%
2877	3.94	4.83	0.0001	TAM 94L25	MIC	18.40%
1793	37.11	33.84	0.005	Acala 1517-99	STR	9.70%

Table 3.2- Crossing scheme of F1 hybrids to study effect of combination of different novel alleles on fiber traits

Pop id	Crosses between F1 hybrid	Fiber trait targeted	Mutant parental lines	Population size
Pop R	1903-1 X 2925-1 2877-2 X 1793-1	(LEN +ELON) X (MIC + STR)	Acala 1517-99 + TAM 94L25	95
Pop S	926-4 X 3028-2 2888-1 X 1903-3	LEN	Acala 1517-99 + TAM 94L25	94

Table 3.3- Parent-offspring regression estimates of heritability for seven cotton fiber traits across two mutant-containing populations

	Athens 14	**Tifton 14**	**Total_14**
MIC_A13	0.21	0.37	0.29
LEN_A13	0.46	0.42	0.44
UNIF_A13	0.28	0.3	0.29
STR_A13	0.36	0.39	0.37
ELON_A13	0.34	0.31	0.33
SFC_A13	0.37	0.46	0.41
Lint %_A13	0.21	0.3	0.25

Table 3.4-Correlations between seven cotton fiber traits in two mutant-containing populations

	MIC	**UHM**	**UI**	**STR**	**ELO**	**SFC %**
Pop R						
UHM	-0.24*					
UI	0	0.50*				
STR	0	0.55*	0.54*			
ELO	-0.10	-0.15	0.16	-0.08		
SFC	-0.03	-0.53*	-0.83*	-0.56*	-0.24*	
Lint %	0.14	-0.08	0.12	0.11	0.39*	-0.11
Pop S						
UHM	-0.50*					
UI	-0.10	0.47*				
STR	-0.21*	0.40*	0.44*			
ELO	0.12	-0.13	0.35*	0.10		
SFC	0.26*	-0.70*	-0.80*	-0.47*	-0.28*	
Lint %	0.39*	-0.22*	0.06	-0.19*	0.45*	0.03

* shows significance at $p < 0.0001$

Table 3.5- Variance components for seven cotton fiber traits across two mutant-containing populations

Fiber traits	Source	DF	SS	MS	F Value	% Contribution	DF	SS	MS	F Value	% Contribution
		Pop R					Pop S				
MIC	G	94	31.3	0.33	3.20*	33.6	93	44.4	0.48	4.27*	48.8
	E	2	17.6	8.8	84.40*	18.9	2	4.95	2.47	22.12*	5.4
	G*E	185	24.6	0.13	1.28	26.4	186	20.69	0.11	0.99	22.7
	Error	190	19.8	0.1			188	21.03	0.11		
Len	G	94	0.65	0.01	4.23*	50.5	93	0.78	0.01	4.01*	45.3
	E	2	0.04	0.02	11.89*	3.0	2	0.08	0.04	19.64*	4.8
	G*E	185	0.29	0	0.95	22.4	186	0.46	0	1.19	27.0
	Error	190	0.31	0			188	0.39	0		
UI	G	94	269	2.86	2.11*	32.8	93	290.26	3.12	2.04*	29.4
	E	2	51	25.5	18.79*	6.2	2	155.76	77.88	50.84*	15.8
	G*E	185	243	1.31	0.97	29.6	186	252.16	1.36	0.88	25.6
	Error	190	258	1.36			188	288	1.53		
STR	G	94	832	8.85	2.85*	41.6	93	635.8	6.84	2*	32.5
	E	2	52.9	26.5	8.52*	2.7	2	75.97	37.99	11.1*	3.9
	G*E	185	523	2.83	0.91	26.2	186	598.83	3.22	0.94	30.6
	Error	190	590	3.11			188	643.59	3.42		
ELON	G	94	77.8	0.83	4.89*	19.3	93	55.04	0.59	3.37*	21.5
	E	2	257	129	760*	64.0	2	136.31	68.15	389*	53.3
	G*E	185	34.9	0.19	1.11	8.7	186	31.4	0.17	0.96	12.3
	Error	190	32.2	0.17			188	32.98	0.18		
SFC	G	94	61.6	0.65	2.05*	31.1	93	82.49	0.89	2.5*	33.4
	E	2	16.2	8.09	25.35*	8.2	2	36.74	18.37	51.7*	14.9
	G*E	185	59.5	0.32	1.01	30.1	186	61.02	0.33	0.92	24.7
	Error	190	60.7	0.32			188	66.8	0.36		
Lint %	G	94	1012	10.8	1.35*	20.9	93	1370.1	14.73	2.62*	26.5
	E	2	1264	632	79.51*	26.1	2	1574.6	787.28	140*	30.5
	G*E	185	1058	5.63	0.71	21.9	186	1163.8	6.26	1.11	22.5
	Error	190	1502	7.95			188	1057.7	5.63		

Table 3.6 – Comparing of each fiber trait between two population, parental and check lines

Pop id	Trait	Pop mean	TAM	ACA	TXA	% improvement compare to TXA	GA 230	Fiber Max 832	Delta Pine 393	Total # of improved lines in	Best line	% improv to TXA	Difference in % to pop R
R	MIC	4.3	4.53*	4.54*	4.5*	4.5	4.55*	4.26	4.57*	23	R093	17.2	..
R	LEN	1.19	1.14*	1.12*	1.15*	3.4	1.17*	1.14*	1.13*	55	R001	10.5	..
R	UNIF	83.72	82.33*	82.11*	82.66*	1.3	83.6*	82.96*	82.69*	43	R073	4	..
R	STR	31.83	30.44*	29.15*	29.9*	6.4	30.48*	30.62*	30.33*	38	R090	16.6	..
R	ELON	5.36	4.48*	5.37	4.65*	15.3	4.96	5.11	5.28	0	R038	32	..
R	SFC %	7.7	8.57*	8.67*	8.44*	8.8	7.88*	8.35*	8.17*	59	R073	19.5	..
R	Lint %	36.59	38.93¥	38.01¥	39.11¥	-6.5	41.04¥	40.17¥	39.49¥	0	R010	5.5	..
S	MIC	4.17	4.53*	4.54*	4.5*	7.3	4.55*	4.26	4.57*	49	S79	19.4	-3*
S	LEN	1.24	1.13*	1.12*	1.15*	7.8	1.17*	1.14*	1.13*	78	S44	14.9	4.2*
S	UNIF	84.39	82.33*	82.11*	82.66*	2.1	83.6*	82.96*	82.69*	74	S27	4.3	0.8*
S	STR	33.72	30.44*	29.15*	29.9*	12.8	30.48*	30.62*	30.33*	87	S18	21.8	5.9*
S	ELON	4.89	4.48*	5.37¥	4.65	5.2	4.96	5.11	5.28¥	0	S25	21.3	-8.8¥
S	SFC %	7.28	8.57*	8.67*	8.44*	13.7	7.88*	8.35*	8.17*	87	S44	24.8	-5.5*
S	Lint %	36.09	38.93¥	38.01¥	39.11¥	-7.7	41.04¥	40.17¥	39.49¥	0	S24	2.2	-1.4¥

Parental or checks cells with "*" are significantly inferior to population average by $p<0.01$ and "¥" are significantly superior to population average by $p<0.01$

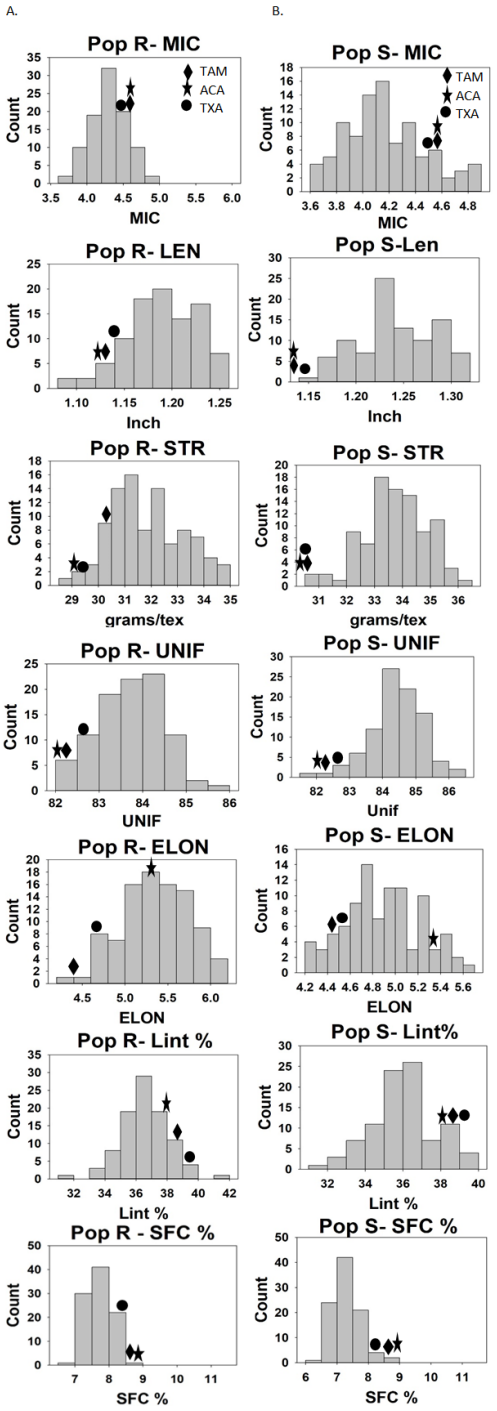


Figure 3.1- Distribution of genotypes in populations for different fiber traits. (A) pop R and (B)

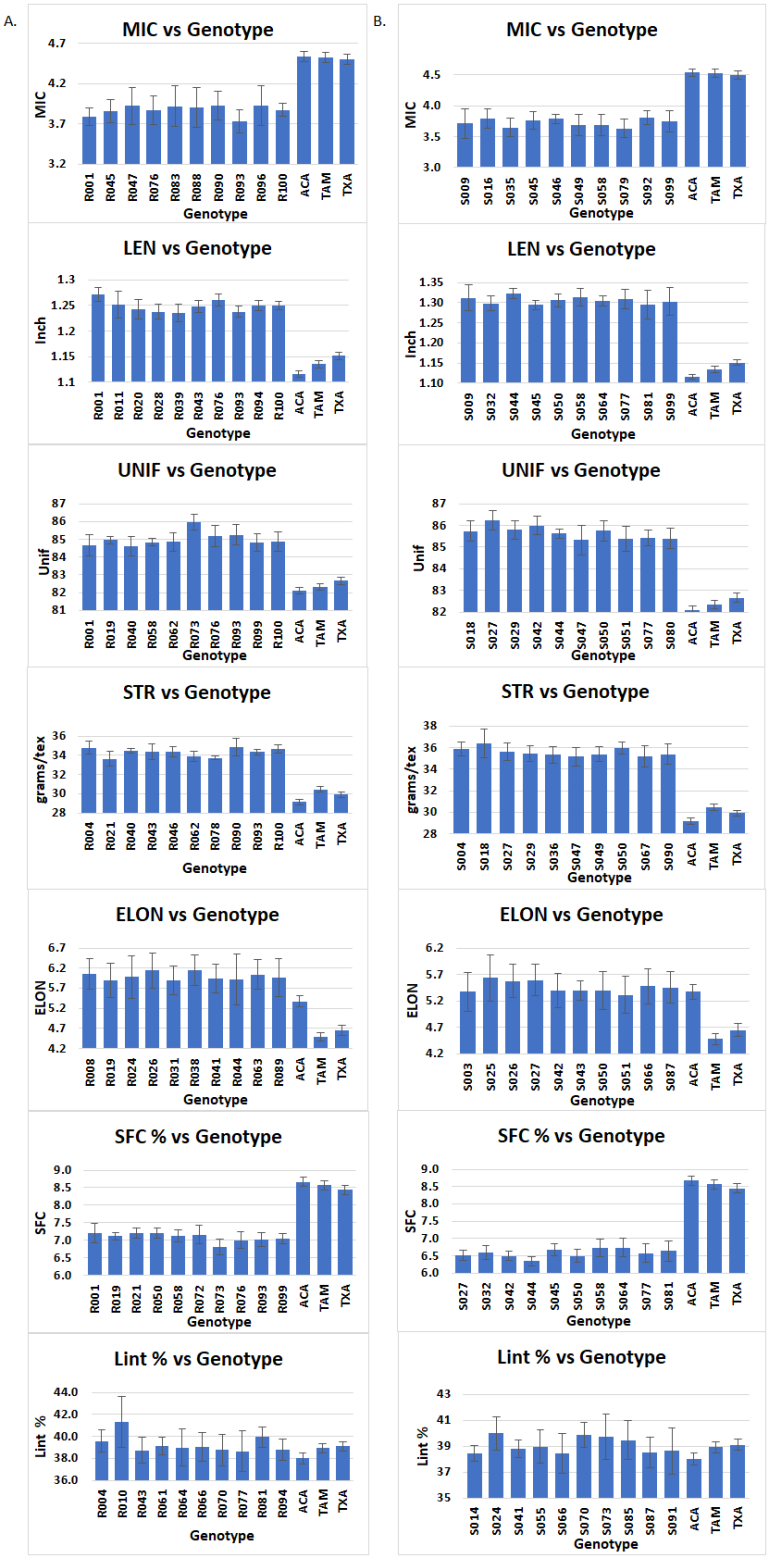


Figure 3.2- The 10 best lines for different fiber traits compared with parental lines. (A) pop R and (B) pop S

CHAPTER 4

A TERMINAL DELETION ON CHROMOSOME 18 CAUSES THE *Li*$_2$ SHORT FIBER PHENOTYPE AND EXEMPLIFIES THE COMPLEX NATURE OF COTTON FIBER QTLs

---

Patel, Jinesh[1]; Huang, Xianzhong [1,2]; Lin, Lifeng [1]; Das, Sayan[1]; Chandnani, Rahul [1]; Khanal, Sameer[1]; Adhikari, Jeevan[1]; Shehzad, Tariq [1]; Guo, Hui[1]; Roy, Eileen[1]; Rong, Junkang and Paterson, Andrew H[1]. To be submitted to Proceedings of the National Academy of Sciences (PNAS).

Abstract

Extreme elongation distinguishes about one-fourth of cotton seed epidermal cells as 'lint' fibers useful for the textile industry, from 'fuzz' fibers (< 5 mm). *Ligon lintless-2* (*Li_2*), a dominant mutant with no lint fiber but normal fuzz fiber, offers insight into pathways and mechanisms that differentiate spinnable cotton from its progenitors. A population of 1,545 F2 plants derived from a cross between Pima S-7 (*G. barbadense*) and an *Li_2* mutant line (*G. hirsutum*) and 144 DNA markers were used to fine map the *Li_2* region on chromosome 18. CISP15 was 0.4 cM from *Li_2* and two 'dominant' markers (i.e. with null alleles in the *Li_2* genotype) SSR7 and SSR18 showed complete linkage with *Li_2*. Non-random distribution of markers with null alleles suggests the *Li_2* phenotype to result from a 176-221 kb deletion of the terminal region of chromosome 18 that may have been masked in prior pooled-sample mapping strategies. The deletion includes at least 40 genes, among which 10 had annotations suggesting roles in fiber development and contained a total of 13 SNP alleles differentiating non-fiber and fiber producing *Gossypium* species, with 11 in coding regions of 8 genes inferred to cause amino acid changes. Based on qPCR at seven different fiber development stages, the 10 candidate genes in the putative *Li_2* deletion showed 5 broad categories of expression patterns. Two genes with a striking expression difference during fiber elongation stages show a VIGS induced *Li_2* like phenotype, each belonging to Glycosyltransferase Family 1. However, at least 7 of the 10 genes with annotations related to fiber development showed higher expression in wild-type compared to *Li_2* mutant during fiber development stages, consistent with the hypothesis that some cotton fiber (and perhaps other) QTLs comprise groups of closely-spaced genes that are functionally diverse but coordinately regulated.

Introduction

The ability of single cotton fiber cells to reach as much as 6 cm in length makes them the longest cell known in the plant kingdom (KIM AND TRIPLETT 2001) and confers their value to the textile industry. There are four major stages of cotton fiber development, i.e. fiber cell initiation, elongation, secondary wall biosynthesis, and maturation (BASRA AND MALIK 1984) After initiation, only about one fourth of cotton seed epidermal cells elongate and differentiate into spinnable lint fibers while the remainder become shorter fuzz fiber (BASRA AND MALIK 1984; TIWARI AND WILKINS 1995; KIM AND TRIPLETT 2001). Fiber elongation typically last for 15 to 20 days, during which the fiber expands at a rate of 2mm/day (LEE *et al.* 2007b; XU *et al.* 2008). Genes participating in regulation of H_2O_2 and reactive oxygen species (ROS), Ca2+, stress, brassinosteroids, water transportation, cell wall loosening, different secondary metabolic pathways, ethylene related signaling pathways, and pectin biosynthesis have been identified (HOVAV *et al.* 2008a; PANG *et al.* 2010; HAIGLER *et al.* 2012; LI *et al.* 2013; FANG *et al.* 2014; SHAN *et al.* 2014; TANG *et al.* 2014; YANG *et al.* 2014). Secondary cell wall synthesis overlaps fiber elongation, during which large amounts of cellulose are synthesized and deposited which ultimately leads to thickening of the cell wall to ~3-4 μm (MANSOOR AND PATERSON 2012). Cell wall thickening is necessary to impart fiber strength, which is an important component of cotton fiber quality (PANG *et al.* 2010). Expression of many genes related to transcription (*GhMYB1 and GhMYB7),* thaumatin-like protein *(GbTLP1),* cellulose synthase (*GhCesA4*), species-specific expansin (*GbEXPATR*), and leucine-rich repeat protein kinase (*LRR RLK*) have been detected during secondary cell wall synthesis (MUNIS *et al.* 2010; KIM *et al.* 2011; SUN *et al.* 2015; HUANG *et al.* 2016; ISLAM *et al.* 2016; LI *et al.* 2016c; FANG *et al.* 2017). The final step of fiber development is fiber maturation, which may last until 60 DPA. During this process, there is

accumulation of minerals and decline of water potential in fiber (JOHN AND KELLER 1996; MANSOOR AND PATERSON 2012). At the end of the process, approximately 500,000 elongated lint fibers will have been produced in a single fruit containing 30 to 35 seeds (HOVAV *et al.* 2008c).

Cotton mutants containing fiber anomalies are excellent tools to help in deciphering the complex process of fiber development. Several natural and artificial cotton mutants have been discovered, genetically mapped and/or used to identify and map other genes related to fiber development (RONG *et al.* 2005; HINCHLIFFE *et al.* 2011; PATEL *et al.* 2014; JIANG *et al.* 2015; HINCHLIFFE *et al.* 2016; MA *et al.* 2016; PATEL *et al.* 2016; THYSSEN *et al.* 2016; WAN *et al.* 2016; THYSSEN *et al.* 2017) and the process has been accelerated by the recent availability of cotton genome sequences (PATERSON *et al.* 2012; WANG *et al.* 2013; LI *et al.* 2014a; LI *et al.* 2015; ZHANG *et al.* 2015). GhMML3_A12, a MYBMIXTA-like transcription factor 3 /GhMYB25- like gene was identified as the cause of a naked seed mutant (N1), with Natural Antisense Transcripts (NATs) produced due to 3' antisense promoter activity causing production of 21–22 nt small RNAs and self-cleavage of the gene (GhMML3_A12) thus producing naked seeds by inhibiting fiber initiation (WAN *et al.* 2016). Three mutants with fuzzy seeds and short lint fiber have been identified in cotton, viz. Ligon lintless-1 (Li_1) (GRIFFEE AND LIGON 1929a), Ligon lintless-2 (Li_2) (NARBUTH AND KOHEL 1990; KOHEL *et al.* 1992) and Ligon lintless-like mutant (Lix) (CAI *et al.* 2013). While Li_1 and Li_x has significantly deformed leaves and stem, Li_2 plants show normal vegetative growth. Another recessive mutant Ligon lintless-3 (*li3*), controlling lint fiber initiation was identified in a *fl* mutant population (MA *et al.* 2016). The mutant genes Li_1, Li_2, *Lix*, and *li3* map to chromosomes 22 (D4), 18 (D13), 4 (A4), and 26 (D12), respectively (RONG *et al.* 2005; CAI *et al.* 2013; MA *et al.* 2016). A single nucleotide mutation

changing an amino acid of an actin gene, GhACT_LI1, disrupts organization of F-actin and cell morphology, thus causing a dwarf vegetative habit and lintless phenotype in *Li1* (THYSSEN *et al.* 2017). *Li₂* lies at the end of chromosome 18. Rong et al. (2005) located it 0.5 cM away from a RFLP marker A1552 and 1.8 cM from two co-segregating RFLP markers, Gate4BF10 and Coau1O05. Multiple studies have revealed the genetic position, differential gene expression and metabolite changes, but not yet identifying the mutation responsible for the *Li₂* phenotype (HINCHLIFFE *et al.* 2011; GILBERT *et al.* 2013; NAOUMKINA *et al.* 2013; NAOUMKINA *et al.* 2014; THYSSEN *et al.* 2014; NAOUMKINA *et al.* 2015).

We sought to identify the *Li₂* gene, by an approach integrating positional (genetic and physical mapping) information, evolutionary information, and differential gene expression, using Virus Induced Gene Silencing (VIGS) to evaluate candidate gene(s). The identification of 7 genes in the target region that appear to function in fiber development is consistent with a recent model (Paterson et al 2012) for cotton fiber QTLs, while two genes in which fiber length could be reduced by VIGS provide new insight into the identity of *Li₂* and the functions of other genes related to elongation of the longest single cell in the plant kingdom.

Materials and Methods

Plant material and population development

A F_2 mapping population was developed by an interspecific cross between an *Li₂* mutant (*G. hirsutum*, TM-1) and Pima S6 (*G. barbadense*), to obtain rich polymorphism of DNA markers. All F1 hybrids displayed the mutant phenotype, consistent with the dominant nature of *Li₂*. F1 plants were selfed and seed were collected, obtaining and planting total of 1,545 F2 individuals in greenhouse or field for this study. Noting that the trait sometimes shows variation within a plant, multiple bolls were observed to determine the *Li₂* trait of the F_2 plants.

Genetic marker development

Initial screening and identifying linked markers near *Li₂* used available SSR (Simple Sequence Repeat) and EST (Expressed Sequence Tag) markers from multiple published genetic linkage maps (PARK *et al.* 2005; RONG *et al.* 2005; GUO *et al.* 2007; RONG *et al.* 2007; YU *et al.* 2007; ZHANG *et al.* 2008; XIAO *et al.* 2009; ZHANG *et al.* 2009; WANG *et al.* 2013). From each map, polymorphic markers near the telomere of the long arm of chromosome 18 (SHAN *et al.* 2016) were mapped with respect to *Li₂*. As markers were exhausted, closely-linked markers were used to select BACs from a *G. raimondii* library that were sequenced, developing new markers from BAC sequences using CID (http://www.shrimp.ufscar.br/cid/index.php) (FREITAS *et al.* 2008) and SSR locator (DA MAIA *et al.* 2008). We also developed CISP (Conserved Intron Scanning Polymorphic) markers as described (FELTUS *et al.* 2006). A total of 144 markers from these different approaches were first tested in a small population to check polymorphism and proximity to *Li₂*. Seven markers with clear bands and which immediately flank *Li₂* were mapped in all 1,545 F2 individuals. Genetic maps were built using JoinMap 4.1 (VAN OOIJEN 2011).

Screening of BAC library

For chromosome walking, cotton BAC libraries were screened with overgo probes designed from genetic markers mapping closely to the *Li₂* phenotype, identifying 124 BACs that might be in the *Li₂* region (Table s1). To reduce the chances of false positive hits and to minimize sequencing redundant BACs, we used a *Gossypium raimondii* physical map (LIN *et al.* 2010), designing probes from tightly linked genetic markers and hybridizing them to *Gossypium raimondii* BAC library and corresponding contigs which led to contig ctg2409 on the physical map (LIN *et al.* 2010). We sequenced three BACS namely GR174O23, GR109E22, and GR174F21, from which markers developed from GR174F21 were more tightly linked with the

Li$_2$ phenotype than the others. With the availability of the *G. raimondii* genome sequence, we BLASTed 13,662 BAC end sequences (BESs) from the *G. raimondii* library (LIN *et al.* 2010) to identify BACs with one end sequence clearly anchored to the terminal region of chromosome D 13, but the other end unanchored. The goal was to find BACs that could be sequenced to extend the terminal region of chromosome 18 and get closer to *Li*$_2$. To determine how informative these BACs will be as compared to GR174F21, we performed DNA fingerprinting.

Validation of deletion

A total of 12 genes were selected from across the target region, with two between CISP15 and the centromere and ten between CISP15 and the telomere. Primers (20-24 bp with ~500 bp amplicons) were designed using Primer 3 (http://frodo.wi.mit.edu/). All amplicons had at least 5 SNPs between the At and Dt subgenomes of tetraploid genome. Three genotypes each from *Li*$_2$ homozygous and wild type along with 12 pair of primers were amplified using the thermal profile: 95˚C for 30 sec (denaturation), 55˚C for 30 sec (annealing), 72˚C for 1 min (extension) for a total of 36 cycles, followed by a terminal extension step at 72˚C for 10 min. PCR amplicons were separated in 10% SDS-PAGE gels. Amplicons were treated with exonuclease I and shrimp alkaline phosphatase enzymatic digestion to remove unused primers and dNTPs, then amplified with BigDye mixture at 95˚C for 5 min and then 95˚C for 10 sec, 50˚C for 5 sec, and 60˚C for 4 min for a total of 30 cycles. A clean up step was performed using Sephadex plates to remove excessive primers and dyes, and sequenced on an ABI3730 capillary sequencer. Sequence analysis was done using the Bioedit program (HALL 1999) and the Codon Code Aligner software (CodonCode Corporation, Dedham, Massachusetts, USA) to check deletion of the terminal end of chromosome 18 in *Li*$_2$ mutant lines compared to wild type lines. We extracted At and Dt

sequence from the *G. hirsutum* genome using a python script to compare SNPs in the sequenced regions and to validate absence of the Dt genomic region in Li_2 mutants (LI *et al.* 2015)

Mining for putative candidate gene(s)

A 221 kb candidate region was extracted from the *G. raimondii* sequence (PATERSON *et al.* 2012) using a python script, and scanned using FGENESH to determine mRNA and protein sequences of putative genes (SALAMOV AND SOLOVYEV 2000) using the protein sequences of these genes for Blastp (protein-protein blast) in the NCBI protein database to identify possible homologs and deduce possible functions. A total of 10 putative candidate genes were identified in the region which might serve in fiber elongation process.

Comparing gene sequences between different cotton genomes

A and AtDt (tetraploid) genome cottons produce elongated fibers whereas D and F genome cottons (APPLEQUIST *et al.* 2001; HOVAV *et al.* 2008a; CHAUDHARY *et al.* 2009a) do not. As reference sequence data from the D genome (*G. raimondii*) and resequencing data of the A (*G. herbaceum* L.), F (*Gossypium longicalyx*) and AtDt (*G. hirsutum* 'Acala Maxxa', *G. hirsutum* race yucatanense, and *G. mustelinum*) genomes were available (PATERSON *et al.* 2012), sequences of predicted genes were compared by using Burrows-Wheeler Aligner (BWA) (LI AND DURBIN 2009) to align resequencing data to the reference *G. raimondii* (PATERSON *et al.* 2012) and Sequence Alignment/Map (SAMtools) (LI *et al.* 2009) was used for SNP calling.

RNA isolation and quantitative reverse transcription-PCR

Different stages of cotton boll development i.e. fibers at initiation, elongation, and secondary cell wall thickening stages were used to examine the expression profiles of 10 genes using qRT-PCR. Tissue samples were collected at 3 (fiber + ovule mix), 6 (fiber +ovule mix), 10 (fiber), 12 (fiber), 15 (fiber), 21 (fiber) and 24 (fiber) days post anthesis (DPA). RNA was extracted using

the Purelink Plant RNA reagent kit, according to the manufacturer's instructions (Life Technologies Corporation, Grand Island, NY) with a DNA digestion step using turbo DNA-free kit (Applied Biosystems, Waltham, Massachusetts, USA). Quality of RNA was determined by electrophoresis in 1% agarose gels and concentration was determined in a NanoDrop 2000 spectrophotometer (NanoDrop Technologies Inc., Wilmington, DE). A total of 500 ng of mRNA was used to obtain first-strand cDNA using a SuperScript™ III Reverse Transcriptase cDNA Synthesis kit (Thermo Fisher Scientific) according to the manufacturer's instructions. qRT-PCR was carried out on a Roche LightCycler480 II instrument (Roche Diagnostics Ltd, Rotkreuz, Switzerland) in a 25-µl volume containing 10 ng of cDNA, 5 pM of each primer, and 25 µl of Fast SYBR Green Master Mixture (Thermo Fisher Scientific) according to the manufacturer's protocol. The PCR conditions were as follows: primary denaturation at 95 °C for 20 s followed by 40 amplification cycles of 3 s at 95 °C, and 30 s at 60 °C. Melting curve analysis was performed to ensure there was no primer-dimer formation. Three replicate assays were performed with independently isolated RNAs, and each PCR reaction was loaded in triplicate. Relative expression levels of each gene are presented using the $2^{-\Delta Ct}$ method (Livak *et al.*, 2001). The cotton *GhACTIN1*gene was used as the positive control.

Virus-induced gene silencing (VIGS) assay

Through PCR, from cDNA we amplified 351 to 599 bp fragments for all 10 putative candidate genes in the suspected deleted part of chromosome 18. These fragments were cloned and inserted into the VIGS binary vector pYL156 (pTRV2) that was double digested with either EcoRI KpnI, EcoRI XhoI or KpnI XhoI. The vectors pTRV1(pYL192) which is a helper plasmid, and cloned pTRV2 vectors were introduced into *Agrobacterium* strain GV3101 through electroporation (Bio-Rad Gene Pulser II). For the VIGS assay, the transformed *Agrobacterium*

colonies containing pTRV1 and cloned pTRV2 were inoculated into 5 mL of LB medium containing kanamycin (50 μg/mL) and gentamycin (25 μg/mL) and grown overnight at 28 °C in an Innova 4080 shaking at a speed of 80 rpm. The next day, the cultures were transferred into a flask containing 50 ml LB media with kanamycin (50 μg/mL) and gentamycin (25 μg/mL) antibiotic concentration plus 10 mM MES (2-(4 morpholino)-ethane sulfonic acid) and 20 μM acetosyringone and left to grow overnight at 28 °C with shaking at 80 rpm (GAO *et al.* 2011). The culture was resuspended using infiltration buffer (10 mM MgCl2, 10 mM MES and 200 μM acetosyringone). Final OD 600 of the culture was adjusted to 1.5. Seven-day old seedlings with fully expanded cotyledons but no visible true leaf were infiltrated with solution made from mixtures of pTRV1 and pTRV2 vectors in ratio of 1:1. Infected plants along with controls were grown in growth chambers under 15 hour light/ 9 hour dark cycles with light intensity of 120 μE m^{-2} S^{-1} and temperature of 24 °C.

Results

Phenotyping of *Li₂* and segregation of the *Li₂* region

The *Li₂* mutant has seeds with no lint fiber but normal fuzz fiber. To determine whether a plant was *Li₂* or wild type, multiple open bolls were inspected on each plant (Figure 1). Some plants showed both *Li₂* and wild type bolls (Figure 1), in which case the plant was considered *Li₂*. Much literature has shown the *Li₂* mutant phenotype to be genetically dominant and governed by a single locus, thus a segregation ratio of 3 lintless to 1 linted individual(s) was expected (NARBUTH AND KOHEL 1990; KOHEL *et al.* 1992; RONG *et al.* 2005). However, among 1,545 F2 plants we found 1091 to be lintless and 454 wild type, a 2.4:1 ratio suggesting enrichment of the wild type. As shown in Table 1, a similar pattern was also observed for DNA marker genotypes

in the *Li₂* region. Such segregation distortion may be due to presence of favorable alleles in this region from *G. barbadense*, or to a deleterious effect of the *Li₂* mutant (from *G. hirsutum*).

Fine mapping of the *Li₂* locus

Initial mapping of the *Li₂* locus in 135 individuals showed 12 markers to all be toward the centromeric side of the gene, i.e., lacking flanking markers on the telomeric side that would permit identification of recombinants in the interval containing the gene. To try to develop markers on the telomeric side of *Li₂*, we designed overgo probes using genetic sequences of markers closest to the gene (Gr_ea17f11, A1552, Gh.fbr.sw02661 and COAU2K07), hybridized them to *Gossypium raimondii* BACs, and sequenced the BACs, namely GR174F21, GR109E22, and GR174O23 (figure S1A). Based on alignment and segregation of markers developed from these BACs with *Li₂*, it appeared that GR109E22 and GR174O23 overlapped each other, but there was a gap between GR109E22 and GR174F21, with the latter being closest to *Li₂*. Also, no new BACs were identified that were distal to GR174F21, suggesting that it is at the end of the chromosome. With the availability of the *G. raimondii* sequence (Paterson et al. 2012), the gap between GR174F21 and GR109E22 was determined to be 26.7 kb. The remaining sequence at the terminal end of chromosome D13 (chromosome 18 in *G. hirsutum*) from GR174F21 was determined to be 37.6 kb, including 14.5 kb of undefined base pairs (N instead of A, C, G and T). We BLASTed 13,662 BAC end sequences from the *G. raimondii* library (LIN *et al.* 2010) to seek BACs with one end in the terminal region of chromosome D13, finding GR102N11 and GR006L12, but DNA fingerprinting showed that sequencing these BACs would not improve coverage of the region (figure 1B). GR174F21 was later found to extend from 60,413,412 to 60,534,221bp of chromosome D13 of *G. hirsutum*, with just 77 bp beyond it to the end of the

chromosome (ZHANG *et al.* 2015). Thus, GR174F21 was the last informative BAC sequence for the terminal region of chromosome D13.

For fine mapping, four markers NAU2980, NAU3447, NAU3827 and NAU3223 covering the *Li₂* region were used along with new markers developed from BAC GR109E22 (CISP15) and GR174F21 (SSR 7 and SSR18). A total of 1,545 F2 plants were genotyped. Based on the genetic map CISP15 was 0.4 cM away from *Li₂* while SSR7 and SSR18 co-segregated with *Li₂* (figure 2). Both co-segregating markers were dominant and in repulsion phase with *Li₂*, somewhat limiting their value.

Deletion of a chromosome segment appears responsible for the *Li₂* phenotype

To try to mitigate the constraint that SSR7 and SSR18 from BAC GR174F21 were genetically dominant, two more SSRs (CISP39 and gr174F21_4.4) were found to be polymorphic and map near *Li₂* and previously identified SSRs. However, while the majority of SSRs in other regions had been co-dominant, these (like SSR7 and 18) were also dominant, all four only amplifying *G. barbadense* alleles. This observation suggested the hypothesis that a segmental deletion near the chromosome 18 terminus may account for the *Li₂* mutant, with its genetic dominance due to haploinsufficiency, i.e. a single functional allele does not produce enough gene product to confer a wild-type phenotype. To further investigate the hypothesis of a deletion in the *Li₂* region of chromosome 18, we sequenced additional segments in this region from the wild type (*G. barbadense)* and homozygous *Li₂* mutants (*G. hirsutum*). The D-genome sequence was mined for genic sequences beyond the last co-dominant marker, i.e. CISP15 towards the telomere. Although precautions were taken while designing primers and running PCR, only 5 gene primer sets amplified single bands useful for further study. One (Gorai.013G268200) was between CISP15 and the centromere; and four, namely

Gorai.013G269200, Gorai.013G270000, Gorai.013G270800, and Gorai.013G271400 were between CISP15 and the telomere. As *Li₂* lies on a chromosome of the Dt genome, amplicon sequence in the hypothetical deletion region from *Li₂* homozygous mutants should only show sequence from the At genome while wild type plants should show sequences from both At and Dt genomes. Sequencing of Gorai.013G268200 which is between the centromere and co-dominant marker CISP 15, suggested presence of both genome (At and Dt) segments in mutant and wild type. Furthermore, due to stringency of the primers of Gorai.013G269200 only Dt was amplified in *Li₂* and wild type, suggesting no deletion in *Li₂* mutant till this point. Amplicons of three genes, Gorai.013G270000, Gorai.013G270800, and Gorai.013G271400 were missing Dt segments in *Li₂* homozygous lines, with only genotypes corresponding to the At genome present in the mutants whereas both At and Dt genotypes were present in the wild type (Figure 2). Thus, three consecutive primer pairs fail to amplify a chromosome 18 (Dt) locus from the *G. hirsutum* genotype carrying *Li₂*, while Gorai.013G268200 and Gorai.013G269200 (closer to CISP15 than the other three) amplify both At and Dt loci, supporting the hypothesis that a terminal deletion is responsible for the *Li₂* phenotype. The results suggest the deletion breakpoint to be between Gorai.013G269200 and Gorai.013G270000, with at least 176 kb and perhaps as much as 221 kb missing from the terminal end of chromosome D13 in *Li₂* mutants.

Candidate *Li₂* gene sequences

The putative 176-221 kb deletion of chromosome 18 in the *Li₂* mutants spanned a total of 40 genes, among which 10 had annotations suggesting function as cell wall proteins and cell wall enzymes, or associated with secondary metabolism, hormonal regulation, and post-transcriptional modification. To further narrow down the list we compared coding sequences (CDS) and untranslated regions (5’ and 3’ UTRs) of the 10 genes between non-fiber producing

(D and F genomes) and fiber producing (A, At, and Dt genome) taxa. We found 13 SNP alleles differentiating non-fiber and fiber producing groups, with 11 in coding regions of 8 genes inferred to cause amino acid changes and potentially affect gene function (Table 2).

Differential expression of candidate gene during fiber development based on Rt-qPCR

Based on qPCR at seven different fiber development stages (Figure 4), the 10 candidate genes in the putative *Li₂* deletion showed 5 broad categories of expression patterns. Two genes (*GhUBE11-D1a* and *GhC4H*) were both expressed primarily at 3 and 6 DPA, with substantially diminished expression later, and with only small differences between *Li₂* and wild type. Four genes (*GhUGT87A1-D1a*, *GhUGT87A2*, *GhUGT87A1-D1b*, *GhETO1*) were expressed at low levels at 3 DPA, with much higher levels at 6 and 12 DPA (and higher in wild type than *Li₂*), diminishing at 15 DPA with wild type maintaining higher expression than *Li₂* for *GhUGT87A1-D1a* and *GhETO1* (and noting that the transition from 6 to 10 DPA is confounded with sampling of ovule versus fiber tissue so is not directly comparable). Two genes (*GhE1310*, *GhEXPA8*) paralleled the general expression pattern of the second group, but with striking enrichment of expression in *Li₂* at 15 DPA. *GhIRX7* showed *Li₂* enriched expression early in development (3-6 DPA) but wild-type enriched expression late (21, 24 DPA). *GhUBE11-D1b*, a ubiquitin-activating enzyme, was enriched in wild-type at all developmental stages except 10 DPA.

Virus-Induced Gene Silencing (VIGS)

To further investigate the 10 candidate genes in the putative *Li₂* deletion, we cloned fragments of each gene, ranging in size from 351 to 599 bp. Due to constraints on growth chamber space, we performed an initial VIGS experiment in the green house with monitored temperature and humidity, finding significant reduction in fiber length of plants treated with TRV2:UGT87A1-D1a and TRV2:UGT87A2. In a follow up study in a growth chamber, multiple plants treated

with TRV2:UGT87A1-D1a showed 32 to 40% reduction in fiber length compared to control lines, with 17 to 23% reduction in fiber length for plants treated with TRV2:UGT87A2 (Figure 5). Despite the use of a growth chamber with controlled environmental conditions validated in prior work (GAO *et al.* 2011), VIGS phenotypes varied somewhat between different plants and within the same plants, as has been reported previously in cotton (WAN *et al.* 2016; ANDRES *et al.* 2017). Similar problems were also seen in plants treated with the control vector TRV2:GRCLA1.

Phylogenetic analysis

Sequences of *UGT87A1* and *UGT87A2* genes were extracted from genome sequences of *G. raimondii* (PATERSON *et al.* 2012), *G. arboreum* (LI *et al.* 2014a), *G. hirsutum* (LI *et al.* 2015) and resequencing data for *G. longicalyx* (PATERSON *et al.* 2012). Protein sequences were predicted using FGENESH and employed in phylogenetic analysis performed using MrBayes v3.2 (KUMAR *et al.* 2016b) with the JTT model of protein evolution (JONES *et al.* 1992). The analysis was run for 5,000,000 generations with sample frequency every 100 generations. The resulting tree was visualized in MEGA7 (KUMAR *et al.* 2016b). *GhUGT87A1_D1a* and *GhUGT87A2* appear more closely related to each other than either is to *GhUGT87A1_D1b*. Further protein sequence alignment using MUSCLE (EDGAR 2004) showed truncation of *GhUGT87A2* in the At subgenome at the C-terminal end, with partial deletion (16 of 44 amino acid) of a highly conserved region known as PSPG (plant secondary product glycosyltransferase) (Figure 6) that plays a role in transferring sugar molecules from donor to acceptor molecules (OSMANI *et al.* 2009; TERASAKA *et al.* 2012). Thus, deletion in PSPG can affect normal transfer of sugar molecules by *GhUGT87A2* of the At subgenome, causing partial or complete loss of function.

Discussion

Investigation of the causal agent(s) of the cotton *Li₂* mutant phenotype, conferring seeds with no lint fiber but normal fuzz fiber, provides an early example in support of the hypothesis that groups of closely-spaced genes that are functionally diverse but coordinately regulated may be of central importance to cotton fiber development (PATERSON *et al.* 2012). Cotton fiber development comprises a long series of carefully coordinated processes, including initiation, elongation, secondary wall deposition and maturation to form the longest single cell known in the plant kingdom, and involves much of the transcriptome (KIM AND TRIPLETT 2001). Mutants in fiber initiation, elongation, and secondary cellulose synthesis (KIM AND TRIPLETT 2001; RONG *et al.* 2005; CAI *et al.* 2013) provide excellent tools for dissecting these processes, with both fundamental and commercial importance.

Two prior studies had mapped *Li₂* to a different position than ours. Hinchliffe *et al.* (2011) mapped the locus between marker DPL0547 and DPL0922, with EST-derived SSR NAU3991 completely linked with the *Li₂* locus. Based on mapping and differential gene expression data, they suggested that a putative gene producing plectin-related protein might be responsible for the *Li₂* phenotype. Using 'super bulked segregant analysis' sequencing (sBSAseq) and a larger F2 population of the same pedigree used by Hinchliffe *et al.* (2011), Thyssen et al. (2014) identified two putative *Li₂* candidate genes, TIP (aquaporin) and ZnF (C2H2-type zinc finger family protein) although finding no changes in the coding sequences of any genes in the *Li₂* vicinity. Based on a *G. hirsutum* sequence (ZHANG *et al.* 2015), the *Li₂* candidate (Gh_D13G2394) identified by Thyssen et al. (2014) was between and nearly equidistant from our markers CISP15 (260.7 kb) and NAU2980 (270 kb). In our study, recombination between these markers, additional markers in the deletion region and the *Li₂* phenotype placed the *Li₂* gene in the

deletion region rather than the region suggested by Thyssen et al. (2014). We found a total of 22 recombinations between these markers and *Li₂* always segregated with marker CISP 15, a finding that would be extremely improbable if the *Li₂* gene was where Thyssen et al. (2014) suggest.

The strategy used by Thyssen et al. (2014) of bulking 100 segregants based on *Li₂* phenotype, could mix *Li₂* homozygotes and heterozygotes, masking SNPs in the deletion region that we suggest to be critical to the *Li₂* phenotype since heterozygotes will have at least one wild type allele. Also, diploid *G. raimondii* was the best available reference genome when they conducted their study, which might have differed somewhat from the tetraploid genome and diminished the accuracy of mapping the region. A similar complication was observed in mapping the *Li1* gene using diploid *G. arboreum* and *G. raimondii* for aligning SNPs (THYSSEN *et al.* 2015; THYSSEN *et al.* 2017). While two nearby regions of a chromosome could each include different genes conferring a similar phenotype, for example seed shattering in sorghum involving *SpWRKY* and *YABBY* loci which are only 300 kb apart (LIN *et al.* 2012; TANG *et al.* 2013), this scenario is rare.

Haploinsufficiency, such as we postulate to account for the genetic dominance of the mutant *Li₂* allele, has been reported in plants, albeit rarely (MEINKE 2013; YUAN *et al.* 2014a). Our observations corroborate prior reports of abnormal expression of the *Li₂* phenotype and incomplete penetrance of the *Li₂* gene (AN *et al.* 2010), resembling haploinsufficiency and incomplete penetrance associated with segmental deletion in multiple human studies (KÖHN *et al.* 2009; KLAASSEN *et al.* 2013; TODARELLO *et al.* 2014; EL KHATTABI *et al.* 2015; KLAASSEN *et al.* 2016). The scarcity of haploinsufficiency in plants might be due to a greater degree of tolerance of plant genomes than others for changes of gene dosage, as plant lineages have incurred genome duplications more frequently than any other taxa known (PATERSON *et al.*

2010). Genes with similar function or connected genes tend to cluster in genomes (NEI 2003; YI *et al.* 2007). Deletion of part of a chromosome containing connected genes might have a haploinsufficient effect as one copy of multiple genes involved in a biosynthetic pathway is removed, causing disruption in the process (THOMAS *et al.* 2006). QTL clusters for single or multiple fiber traits have been identified in cotton (RONG *et al.* 2007; LI *et al.* 2016a; ZHANG *et al.* 2016b) and related to closely-spaced genes that are functionally diverse but coordinately regulated (PATERSON *et al.* 2012).

The failure of fibers to elongate properly in the absence of two wild-type Li_2 alleles suggests that the Dt genome of tetraploid cotton may be differentiated from its diploid progenitor by a neomorphic mutation. While diploid D-genome cottons do not produce lint fibers, QTL studies have shown many fiber-quality related QTLs to locate in Dt genomic regions of tetraploid cottons (RONG *et al.* 2007; ZHAO *et al.* 2012; WANG *et al.* 2013). The fact that loss of even a single wild-type Li_2 allele causes extremely short fiber suggests that homologs of these genes, including homoeologs in the At genome, are functionally diverged and not able to compensate for this loss. Having a partially functional homologous gene/s may cause abnormal expression of the wild phenotype and incomplete penetrance such as we observe. A careful study of gene expression using RNAseq in plants with such atypical phenotypes might clarify such an anomaly.

Based on gene expression and comparative genomics results, we found seven genes namely, *GhIRX7_D*, *GhETO1_D*, *GhUBE11_Db*, *GhUGT87A1_Da*, *GhUGT87A2_D*, *GhUGT87A1_Db* and *GhEXPA8* that might have roles in fiber development. Employing VIGS, we could see impact of knocking down of genes *GhUGT87A1_Da* and *GhUGT87A2_D* on fiber length, confirming their importance in fiber elongation.

GhIRX7_D belongs to the glycosyltransferase family 47 (ZHONG AND YE 2003; GESHI *et al.* 2010). Disrupting of *IRX7* or *FRA8 (FRAGILE FIBER 8)* causes reduction in secondary wall thickness, decline in amounts of cellulose and xylan, collapse of xylem vessels, decrease in stem strength, and dwarf phenotype (BROWN *et al.* 2005; ZHONG *et al.* 2005; BROWN *et al.* 2007; LEE *et al.* 2007a). Reduced expression during secondary wall synthesis (Figure 4) of *GhIRX7_D* in *Li*$_2$ mutants might also affect secondary wall biosynthesis, contributing to the phenotype at later stages of fiber development.

GhEXPA8 belongs to the expansin superfamily, wall-loosening proteins which help with cell expansion (SAMPEDRO AND COSGROVE 2005). Multiple studies have discovered the role of expansin genes during fiber development, especially during fiber elongation (ORFORD AND TIMMIS 1998; JI *et al.* 2003; SHI *et al.* 2006; SHAN *et al.* 2014; BAJWA *et al.* 2015). The result suggests that *GhEXPA8* is important for fiber length and any reduction in its expression during fiber elongation (as seen in *Li*$_2$ mutant) might have repercussions.

GhETO1_D encodes an ethylene-overproduction protein. In Arabidopsis, *ETO1* controls the rate of ethylene synthesis by negatively regulating 1-aminocyclopropane-1-carboxylic acid synthase (ACS) (WANG *et al.* 2004; YOSHIDA *et al.* 2005; CHRISTIANS *et al.* 2009)**.** Ethylene plays an important role in fiber elongation but excessive production might curb fiber development (SHI *et al.* 2006). A steep increase of tenfold in ethylene concentration is seen in mutants with loss-of-function *eto1* mutations causing shorter seedlings, smaller leaves, and reduced sizes of roots, petioles and inflorescences (CHAE *et al.* 2003; WANG *et al.* 2004; CHRISTIANS *et al.* 2009; PANG *et al.* 2010; LUO *et al.* 2014)**.** Application of abscisic acid (ABA) enormously inhibited root growth in *eto1* mutants by promoting ethylene biosynthesis (LUO *et al.* 2014). Elevated levels of ABA have been observed during fiber development of *Li*$_2$ mutant

lines, which coupled with reduced expression of *GhETO1_D* during fiber elongation might have contributed to the *Li₂* phenotype (CHEN *et al.* 1997; GILBERT *et al.* 2013). *GhUBE11_Db* encodes Ubiquitin-activating enzyme E1 that catalyzes the first step of three consecutive enzymatic cascades in a ubiquitination reaction. Interestingly, regulation of ethylene biosynthesis through ubiquitin-mediated protein degradation is assisted by ethylene-overproduction protein (ETO1) (LYZENGA AND STONE 2012). This suggests that multiple genes in the same biological process have been affected by deletion in *Li₂* mutants.

GhUGT87A1_Da, *GhUGT87A2_D* and *GhUGT87A1_Db* belong to Glycosyltransferase (GT) Family 1, the largest GT family (YONEKURA-SAKAKIBARA AND HANADA 2011; HUANG *et al.* 2015). *UGT87A1* and *UGT87A2* are nearly identical, and may have similar function and redundant effects. *UGT87A2* in *Arabidopsis* has been involved in ascorbic acid homeostasis, cell wall biosynthesis, controlling H_2O_2 level (ROS scavenging activity), delaying flowering, enhancing germination, root growth and prevention of cell damage during stress (SAINT PAUL 2010; WANG *et al.* 2012; LI *et al.* 2016b). Regulation of H_2O_2 or ROS is required to avoid arresting the fiber elongation process, as higher levels of H_2O_2 could trigger secondary cell wall biosynthesis (LI *et al.* 2007; CHAUDHARY *et al.* 2008; HOVAV *et al.* 2008a; CHAUDHARY *et al.* 2009a; GUO *et al.* 2016). Hovav *et al.* (2008) found elevated concentration of H_2O_2 in *G. herbaceum* (A genome) and *G. longicalyx* (F genome) during early fiber elongation, but the latter was not able to reduce the H_2O_2 level thus reducing fiber elongation, enhancing stress conditions and triggering secondary cell wall biosynthesis. Deletion of these genes might have increased H_2O_2 or ROS accumulation, permitting premature halting of fiber elongation and causing the *Li₂* phenotype. Such higher concentration of ROS and higher expression of stress related genes have been reported in *Li₂* mutant (HINCHLIFFE *et al.* 2011; NAOUMKINA *et al.* 2013). Further, better

germination was observed in *87A2OE* than *ugt87a2* knock out mutant line and wild type (LI *et al.* 2016b). We also observed segregation distortion in our population with significant reduction of plants having homozygous *G. hirsutum* alleles (homozygous Li_2 mutant allele), suggesting some role of these genes in reproductive organs, seed viability and germination.

While the cotton Li_2 mutant phenotype is essentially discrete and tentatively due to two specific genes, we posit that the putative deletion conferring this phenotype may exemplify consequences of allele substitution at many cotton (and perhaps other) QTLs, comprising groups of closely-spaced genes that are functionally diverse but coordinately regulated (PATERSON *et al.* 2012).

Reference

Abdel-Salam, M., M. Negm and C. S. Ardabb, 2009 The Egyptian cotton, current constraints and future opportunities. Textile Industries Holding Co., Modern Press-Alexandria-Egypt.

Ahloowalia, B., and M. Maluszynski, 2001 Induced mutations–A new paradigm in plant breeding. Euphytica 118**:** 167-173.

Ai, X., Y. Liang, J. Wang, J. Zheng, Z. Gong *et al.*, 2017 Genetic diversity and structure of elite cotton germplasm (Gossypium hirsutum L.) using genome-wide SNP data. Genetica**:** 1-8.

Al-Ghazi, Y., S. Bourot, T. Arioli, E. S. Dennis and D. J. Llewellyn, 2009 Transcript profiling during fiber development identifies pathways in secondary metabolism and cell wall structure that may contribute to cotton fiber quality. Plant Cell Physiol 2009.

Alonso, J. M., A. N. Stepanova, T. J. Leisse, C. J. Kim, H. Chen *et al.*, 2003 Genome-wide insertional mutagenesis of Arabidopsis thaliana. Science 301**:** 653-657.

An, C., J. N. Jenkins, J. C. McCarty and S. Saha, 2010 Atypical Ligon lintless-2 phenotype in cotton. J Cotton Sci 14.

An, G., D.-H. Jeong, K.-H. Jung and S. Lee, 2005 Reverse Genetic Approaches for Functional Genomics of Rice. Plant Molecular Biology 59**:** 111-123.

Andres, R. J., V. Coneva, M. H. Frank, J. R. Tuttle, L. F. Samayoa *et al.*, 2017 Modifications to a LATE MERISTEM IDENTITY1 gene are responsible for the major leaf shapes of Upland cotton (Gossypium hirsutum L.). Proceedings of the National Academy of Sciences 114**:** E57-E66.

Applequist, W. L., R. Cronn and J. F. Wendel, 2001 Comparative development of fiber in wild and cultivated cotton. Evolution & Development 3**:** 3-17.

Aslam, U., H. M. Cheema, S. Ahmad, I. A. Khan, W. Malik *et al.*, 2016 COTIP: Cotton TILLING Platform, a resource for plant improvement and reverse genetic studies. Frontiers in plant science 7.

Atkinson, N. J., and P. E. Urwin, 2012 The interaction of plant biotic and abiotic stresses: from genes to the field. Journal of Experimental Botany 63**:** 3523-3543.

Auld, D., E. Bechere, M. Ethridge, W. Becker, E. Hequet *et al.*, 2000 Registration of TTU 202-1107-B and TTU 271-2155-C mutant germplasm lines of upland cotton with improved fiber quality. Crop science 40**:** 1835-1835.

Auld, D., M. Ethridge, J. Dever and P. Dotray, 1998 Chemical mutagenesis as a tool in cotton improvement. Beltwide Cotton Conferences (USA).

Auld, D., M. Heikkinen, D. Erickson, J. Sernyk and J. Romero, 1992 Rapeseed mutants with reduced levels of polyunsaturated fatty acids and increased levels of oleic acid. Crop science 32**:** 657-662.

Auld, D., G. G. Light, M. Fokar, E. Bechere and R. D. Allen, 2009 Mutagenesis Systems for Genetic Analysis of Gossypium. Genetics and Genomics of Cotton**:** 1-18.

Auld, D. L., 2000 Registration of TTU 202-1107-B and TTU 271-2155-C mutant germplasm lines of upland cotton with improved fiber quality. Crop science 40**:** 1835-1836.

Awan, M. A., C. Konzak, J. Rutger and R. Nilan, 1980 Mutagenic effects of sodium azide in rice. Crop Science 20**:** 663-668.

Bajwa, K. S., A. A. Shahid, A. Q. Rao, A. Bashir, A. Aftab *et al.*, 2015 Stable transformation and expression of GhEXPA8 fiber expansin gene to improve fiber length and micronaire value in cotton. Frontiers in Plant Science 6**:** 838.

Balcells, L., J. Swinburne and G. Coupland, 1991 Transposons as tools for the isolation of plant genes. Trends in Biotechnology 9**:** 31-37.

Barb, J. G., J. E. Bowers, S. Renaut, J. I. Rey, S. J. Knapp *et al.*, 2014 Chromosomal evolution and patterns of introgression in Helianthus. Genetics 197**:** 969-979.

Basra, A. S., and C. P. Malik, 1984 Development of the cotton fiber. Int. Rev. Cytol 89**:** 65-113.

Beasley, C., and E. Egli, 1977 Fiber production in vitro from a conditional fiberless mutant of cotton. Developmental biology 57**:** 234-237.

Bechere, E., D. Auld, R. Cantrell, E. Hequet, M. Krifa *et al.*, 2007a Registration of TTU 0774-3-3 and TTU 0808-1-6-1 upland cotton germplasm lines with improved fiber length and strength. Journal of plant registrations 1**:** 58-59.

Bechere, E., D. Auld and E. Hequet, 2009a Development of ‘naked-tufted’seed coat mutants for potential use in cotton production. Euphytica 167**:** 333-339.

Bechere, E., D. Auld, M. Krifa, C. W. Smith and R. Cantrell, 2011 Registration of TTU 0782, an Upland Cotton Germplasm Line with Superior Fiber Quality. Journal of plant registrations 5**:** 207-210.

Bechere, E., D. L. Auld, R. G. Cantrell, E. Hequet, M. Krifa *et al.*, 2007b Registration of TTU 0774-3-3 and TTU 0808-1-6-1 Upland Cotton Germplasm Lines with Improved Fiber Length and Strength. J. Plant Reg. 1**:** 58-59.

Bechere, E., D. L. Auld, P. Dotray and H. Kebede, 2010 Registration of Four Upland Cotton (Gossypium hirsutum L.) Genetic Stock Mutants with Tolerance to Imazamox.

Bechere, E., D. L. Auld, P. A. Dotray, L. V. Gilbert and H. Kebede, 2009b Imazamox Tolerance in Mutation-Derived Lines of Upland Cotton.

Boopathi, N. M., and L. V. Hoffmann, 2016 Genetic Diversity, Erosion, and Population Structure in Cotton Genetic Resources, pp. 409-438 in *Genetic Diversity and Erosion in Plants*. Springer.

Bradow, J. M., and G. H. Davidonis, 2000 Quantitation of fiber quality and the cotton production-processing interface: a physiologist's perspective. J. Cotton Sci 4**:** 34-64.

Brown, D. M., F. Goubet, V. W. Wong, R. Goodacre, E. Stephens *et al.*, 2007 Comparison of five xylan synthesis mutants reveals new insight into the mechanisms of xylan synthesis. The Plant Journal 52**:** 1154-1168.

Brown, D. M., L. A. Zeef, J. Ellis, R. Goodacre and S. R. Turner, 2005 Identification of novel genes in Arabidopsis involved in secondary cell wall formation using expression profiling and reverse genetics. The Plant Cell 17**:** 2281-2295.

Brown, I., C. W. Smith, D. Auld, S. Hague, E. F. Hequet *et al.*, 2012 Registration of TAM 94L-25-M24, TAM 94L-25-M25, and TAM 94L-25-M30 Mutant Upland Cotton Germplasm with Improved Fiber Length and Strength. Journal of Plant Registrations 6**:** 195-199.

Brubaker, C. L., F. Bourland and J. F. Wendel, 1999 The origin and domestication of cotton. Cotton: Origin, history, technology, and production. John Wiley & Sons, New York**:** 3-31.

Brubaker, C. L., and J. F. Wendel, 1994 Reevaluating the origin of domesticated cotton (Gossypium hirsutum; Malvaceae) using nuclear restriction fragment length polymorphisms (RFLPs). American journal of botany**:** 1309-1326.

Cai, C., X. Tong, F. Liu, F. Lv, H. Wang *et al.*, 2013 Discovery and identification of a novel Ligon lintless-like mutant (Lix) similar to the Ligon lintless (Li1) in allotetraploid cotton. Theoretical and Applied Genetics 126**:** 963-970.

Cai, Y., X. Cui, J. Rodgers, D. Thibodeaux, V. Martin *et al.*, 2011 An investigation on different parameters used for characterizing short cotton fibers. Textile Research Journal 81**:** 239-246.

Cantrell, R., C. Roberts and C. Waddell, 2000 Registration of Acala 1517-99 Cotton. Crop Science 40**:** 1200-1200.

Chae, H. S., F. Faure and J. J. Kieber, 2003 The eto1, eto2, and eto3 Mutations and Cytokinin Treatment Increase Ethylene Biosynthesis in Arabidopsis by Increasing the Stability of ACS Protein. The Plant Cell 15**:** 545-559.

Chapagain, A. K., A. Y. Hoekstra, H. H. G. Savenije and R. Gautam, 2006 The water footprint of cotton consumption: An assessment of the impact of worldwide consumption of cotton products on the water resources in the cotton producing countries. Ecological Economics 60**:** 186-203.

Chaudhary, B., R. Hovav, L. Flagel, R. Mittler and J. F. Wendel, 2009a Parallel expression evolution of oxidative stress-related genes in fiber from wild and domesticated diploid and polyploid cotton (Gossypium). BMC genomics 10**:** 378.

Chaudhary, B., R. Hovav, L. Flagel, R. Mittler and J. F. Wendel, 2009b Parallel expression evolution of oxidative stress-related genes in fiber from wild and domesticated diploid and polyploid cotton (Gossypium). BMC Genomics 10.

Chaudhary, B., R. Hovav, R. Rapp, N. Verma, J. A. Udall *et al.*, 2008 Global analysis of gene expression in cotton fibers from wild and domesticated Gossypium barbadense. Evolution & development 10**:** 567-582.

Chee, P., E. Lubbers, O. May, J. Gannaway and A. H. Paterson, 2004 Changes in genetic diversity of the U.S. Upland cotton, pp. in *Beltwide Cotton Conference*. National Cotton Council, San Antonio, TX.

Chen, J.-G., X.-M. Du, X. Zhou and H.-Y. Zhao, 1997 Levels of cytokinins in the ovules of cotton mutants with altered fiber development. J Plant Growth Regul 16.

Chen, L., Z.-S. Zhang, M.-C. Hu, W. Wang, J. Zhang *et al.*, 2008 Genetic linkage map construction and QTL mapping for yield and fiber quality in upland cotton (Gossypium hirsutum L.). Acta Agron Sin 34**:** 1199-1205.

Chen, Z., K. Feng, C. E. Grover, P. Li, F. Liu *et al.*, 2016 Chloroplast DNA structural variation, phylogeny, and age of divergence among diploid cotton species. PloS one 11**:** e0157183.

Chen, Z. J., and X. Guan, 2011 Auxin boost for cotton. Nat Biotech 29**:** 407-409.

Chen, Z. J., B. E. Scheffler, E. Dennis, B. A. Triplett, T. Zhang *et al.*, 2007 Toward sequencing cotton (Gossypium) genomes. Plant physiology 145**:** 1303-1310.

Chopra, V., 2005 Mutagenesis: Investigating the process and processing the outcome for crop improvement. CURRENT SCIENCE-BANGALORE- 89**:** 353.

Christians, M. J., D. J. Gingerich, M. Hansen, B. M. Binder, J. J. Kieber *et al.*, 2009 The BTB ubiquitin ligases ETO1, EOL1 and EOL2 act collectively to regulate ethylene biosynthesis in Arabidopsis by controlling type-2 ACC synthase levels. The Plant Journal 57**:** 332-345.

Clement, J., G. Constable, W. Stiller and S. Liu, 2012 Negative associations still exist between yield and fibre quality in cotton breeding programs in Australia and USA. Field crops research 128**:** 1-7.

Clement, J. D., G. A. Constable, W. N. Stiller and S. M. Liu, 2015 Early generation selection strategies for breeding better combinations of cotton yield and fibre quality. Field Crops Research 172**:** 145-152.

Constable, G., D. Llewellyn, S. A. Walford and J. D. Clement, 2015 Cotton Breeding for Fiber Quality Improvement, pp. 191-232 in *Industrial Crops: Breeding for BioEnergy and Bioproducts*, edited by V. M. V. Cruz and D. A. Dierig. Springer New York, New York, NY.

Cotton, F. T. U., 1998 COTTON IMPROVEMENT.

Cui, X., T. A. Calamari, K. Q. Robert, J. B. Price and M. D. Watson, 2003 Measuring the short fiber content of cotton. Textile research journal 73**:** 891-895.

Culp, T., D. Harrell and T. Kerr, 1979 Some genetic implications in the transfer of high fiber strength genes to upland cotton. Crop Science 19**:** 481-484.

da Maia, L. C., D. A. Palmieri, V. Q. de Souza, M. M. Kopp, F. I. F. de Carvalho *et al.*, 2008 SSR Locator: Tool for Simple Sequence Repeat Discovery Integrated with Primer Design and PCR Simulation. International Journal of Plant Genomics 2008**:** 412696.

Davey, J. W., P. A. Hohenlohe, P. D. Etter, J. Q. Boone, J. M. Catchen *et al.*, 2011 Genome-wide genetic marker discovery and genotyping using next-generation sequencing. Nature reviews. Genetics 12**:** 499.

De Serres, F. J., and A. Hollaender, 2012 *Chemical mutagens: Principles and methods for their detection*. Springer Science & Business Media.

Dong, Z., Y. Shi, J. Zhang, S. Wang, J. Li *et al.*, 2009 Molecular marker-assisted selection and pyramiding breeding of major QTLs for cotton fiber length. Cotton Sci 21**:** 279-283.

Doudna, J. A., and E. Charpentier, 2014 The new frontier of genome engineering with CRISPR-Cas9. Science 346**:** 1258096.

Edgar, R. C., 2004 MUSCLE: multiple sequence alignment with high accuracy and high throughput. Nucleic acids research 32**:** 1792-1797.

El Khattabi, L., F. Guimiot, E. Pipiras, J. Andrieux, C. Baumann *et al.*, 2015 Incomplete penetrance and phenotypic variability of 6q16 deletions including SIM1. Eur J Hum Genet 23**:** 1010-1018.

Endrizzi, J., E. Turcotte and R. Kohel, 1985 Genetics, cytogenetics, and evolution of Gossypium. Adv. Genet 23**:** 271-375.

Fang, L., R. Tian, X. Li, J. Chen, S. Wang *et al.*, 2014 Cotton fiber elongation network revealed by expression profiling of longer fiber lines introgressed with different Gossypium barbadense chromosome segments. BMC Genomics 15**:** 838.

Fang, X., X. Liu, X. Wang, W. Wang, D. Liu *et al.*, 2017 Fine-mapping qFS07.1 controlling fiber strength in upland cotton (Gossypium hirsutum L.). Theoretical and Applied Genetics**:** 1-12.

Feltus, F. A., H. P. Singh, H. C. Lohithaswa, S. R. Schulze, T. D. Silva *et al.*, 2006 A comparative genomics strategy for targeted discovery of single-nucleotide polymorphisms and conserved-noncoding sequences in orphan crops. Plant Physiol 140**:** 1183-1191.

Freitas, P. D., D. S. Martins and P. M. Galetti, 2008 CID: a rapid and efficient bioinformatic tool for the detection of SSRs from genomic libraries. Molecular Ecology Resources 8**:** 107-108.

Gallagher, J. P., C. E. Grover, K. Rex, M. Moran and J. F. Wendel, 2017 A New Species of Cotton from Wake Atoll, Gossypium stephensii (Malvaceae). Systematic Botany 42**:** 115-123.

Gao, W., Z. J. Chen, Z. Y. John, R. J. Kohel, J. E. Womack *et al.*, 2006 Wide-cross whole-genome radiation hybrid mapping of the cotton (Gossypium barbadense L.) genome. Molecular genetics and genomics 275**:** 105-113.

Gao, X., R. C. Britt Jr, L. Shan and P. He, 2011 Agrobacterium-mediated virus-induced gene silencing assay in cotton. JoVE (Journal of Visualized Experiments)**:** e2938-e2938.

Geshi, N., J. Harholt, Y. Sakuragi, J. Krüger Jensen and H. V. Scheller, 2010 Glycosyltransferases of the GT47 Family, pp. 265-283 in *Annual Plant Reviews*. Wiley-Blackwell.

Gilbert, M. K., J. M. Bland, J. M. Shockey, H. Cao, D. J. Hinchliffe *et al.*, 2013 A transcript profiling approach reveals an abscisic acid-specific glycosyltransferase (UGT73C14)

induced in developing fiber of Ligon lintless-2 mutant of cotton (Gossypium hirsutum L.). PloS one 8**:** e75268.

Gore, M. A., D. D. Fang, J. A. Poland, J. Zhang, R. G. Percy *et al.*, 2014 Linkage map construction and quantitative trait locus analysis of agronomic and fiber quality traits in cotton. The Plant Genome 7.

Greene, E. A., C. A. Codomo, N. E. Taylor, J. G. Henikoff, B. J. Till *et al.*, 2003 Spectrum of chemically induced mutations from a large-scale reverse-genetic screen in Arabidopsis. Genetics 164**:** 731-740.

Gregorio, G., D. Senadhira, R. Mendoza, N. Manigbas, J. Roxas *et al.*, 2002 Progress in breeding for salinity tolerance and associated abiotic stresses in rice. Field Crops Research 76**:** 91-101.

Griffee, F., and L. Ligon, 1929a Occurrence of "lintless" cotton plants and inheritance of character "lintless" *Journal of the American Society of Agronomy* 21**:** 711-717.

Griffee, F., and L. Ligon, 1929b Occurrence of "lintless" cotton plants and the inheritance of the character "lintless. J Am Soc Agron 21**:** 711-717.

Grover, C. E., H. Kim, R. A. Wing, A. H. Paterson and J. F. Wendel, 2007 Microcolinearity and genome evolution in the AdhA region of diploid and polyploid cotton (Gossypium). The Plant Journal 50**:** 995-1006.

Guan, X., M. Pang, G. Nah, X. Shi, W. Ye *et al.*, 2014 miR828 and miR858 regulate homoeologous MYB2 gene functions in Arabidopsis trichome and cotton fibre development. Nature communications 5**:** 3050.

Guo, K., X. Du, L. Tu, W. Tang, P. Wang *et al.*, 2016 Fibre elongation requires normal redox homeostasis modulated by cytosolic ascorbate peroxidase in cotton (Gossypium hirsutum). Journal of Experimental Botany 67**:** 3289-3301.

Guo, W., C. Cai, C. Wang, Z. Han, X. Song *et al.*, 2007 A microsatellite-based, gene-rich linkage map reveals genome structure, function, and evolution in Gossypium. Genetics**:** genetics.107.070375.

Haigler, C., 2007 Substrate supply for cellulose synthesis and its stress sensitivity in the cotton fiber. Cellulose: Molecular and Structural Biology**:** 147-168.

Haigler, C., 2010 Physiological and anatomical factors determining fiber structure and utility. Physiology of cotton**:** 33-47.

Haigler, C., L. Betancur, M. Stiff and J. Tuttle, 2012 Cotton fiber: a powerful single-cell model for cell wall and cellulose research. Frontiers in Plant Science 3.

Haigler, C., B. Singh, G. Wang and D. Zhang, 2009 Genomics of cotton fiber secondary wall deposition and cellulose biogenesis. Genetics and Genomics of Cotton**:** 1-33.

Hall, T. A., 1999 BioEdit: a user-friendly biological sequence alignment editor and analysis program for Windows 95/98/NT, pp. 95-98 in *Nucleic acids symposium series*.

Harland, S., 1929 The work of the genetics department of the Cotton Research Station, Trinidad. Emp Cotton Grow Rev 6**:** 304-314.

Hatfield, P. M., M. M. Gosink, T. B. Carpenter and R. D. Vierstra, 1997 The ubiquitin-activating enzyme (E1) gene family in Arabidopsis thaliana. The Plant Journal 11**:** 213-226.

Herring, A. D., D. L. Auld, M. D. Ethridge, E. F. Hequet, E. Bechere *et al.*, 2004 Inheritance of fiber quality and lint yield in a chemically mutated population of cotton. Euphytica 136**:** 333-339.

Hinchliffe, D. J., B. D. Condon, G. Thyssen, M. Naoumkina, C. A. Madison *et al.*, 2016 The GhTT2_A07 gene is linked to the brown colour and natural flame retardancy phenotypes of Lc1 cotton (Gossypium hirsutum L.) fibres. Journal of experimental botany 67**:** 5461-5471.

Hinchliffe, D. J., R. B. Turley, M. Naoumkina, H. J. Kim, Y. Tang *et al.*, 2011 A combined functional and structural genomics approach identified an EST-SSR marker with complete linkage to the Ligon lintless-2 genetic locus in cotton (Gossypium hirsutum L.). BMC genomics 12**:** 445.

Ho, M.-H., S. Saha, J. N. Jenkins and D.-P. Ma, 2010 Characterization and Promoter Analysis of a Cotton RING-Type Ubiquitin Ligase (E3) Gene. Molecular Biotechnology 46**:** 140-148.

Hovav, R., J. A. Udall, B. Chaudhary, E. Hovav, L. Flagel *et al.*, 2008a The Evolution of Spinnable Cotton Fiber Entailed Prolonged Development and a Novel Metabolism. PLoS Genet 4**:** e25.

Hovav, R., J. A. Udall, B. Chaudhary, E. Hovav, L. Flagel *et al.*, 2008b The evolution of spinnable cotton fiber entailed prolonged development and a novel metabolism. PLoS Genet 4.

Hovav, R., J. A. Udall, E. Hovav, R. Rapp, L. Flagel *et al.*, 2008c A majority of cotton genes are expressed in single-celled fiber. Planta 227**:** 319-329.

Hu, H., X. He, L. Tu, L. Zhu, S. Zhu *et al.*, 2016 GhJAZ2 negatively regulates cotton fiber initiation by interacting with the R2R3-MYB transcription factor GhMYB25-like. The Plant Journal 88**:** 921-935.

Huang, G.-Q., S.-Y. Gong, W.-L. Xu, W. Li, P. Li *et al.*, 2013 A fasciclin-like arabinogalactan protein, GhFLA1, is involved in fiber initiation and elongation of cotton. Plant physiology 161**:** 1278-1290.

Huang, J., F. Chen, S. Wu, J. Li and W. Xu, 2016 Cotton GhMYB7 is predominantly expressed in developing fibers and regulates secondary cell wall biosynthesis in transgenic Arabidopsis. Science China Life Sciences 59**:** 194-205.

Huang, J., C. Pang, S. Fan, M. Song, J. Yu *et al.*, 2015 Genome-wide analysis of the family 1 glycosyltransferases in cotton. Molecular Genetics and Genomics 290**:** 1805-1818.

Hulse-Kemp, A. M., J. Lemm, J. Plieske, H. Ashrafi, R. Buyyarapu *et al.*, 2015 Development of a 63K SNP Array for Cotton and High-Density Mapping of Intra-and Inter-Specific Populations of Gossypium spp. G3: Genes, Genomes, Genetics**:** g3. 115.018416.

Hussein, H., F. Al Enani and M. El Moghazi, 1982 Histological and morphological characteristics of a glandless cotton mutant induced with sodium azide. Egypt. J. Genet. Cytol 11**:** 167-174.

Hutchinson, J., 1951 Intra-specific differentiation in Gossypium hirsutum. Heredity 5**:** 161-193.

Iqbal, M., O. Reddy, K. El-Zik and A. Pepper, 2001 A genetic bottleneck in the'evolution under domestication'of upland cotton Gossypium hirsutum L. examined using DNA fingerprinting. TAG Theoretical and Applied Genetics 103**:** 547-554.

Islam, M. S., L. Zeng, G. N. Thyssen, C. D. Delhom, H. J. Kim *et al.*, 2016 Mapping by sequencing in cotton (Gossypium hirsutum) line MD52ne identified candidate genes for fiber strength and its related quality attributes. Theoretical and Applied Genetics 129**:** 1071-1086.

Jeon, J. S., S. Lee, K. H. Jung, S. H. Jun, D. H. Jeong *et al.*, 2000 T-DNA insertional mutagenesis for functional genomics in rice. The Plant Journal 22**:** 561-570.

Jeong, D.-H., S. An, H.-G. Kang, S. Moon, J.-J. Han *et al.*, 2002 T-DNA insertional mutagenesis for activation tagging in rice. Plant physiology 130**:** 1636-1644.

Ji, S. J., Y. C. Lu, J. X. Feng, G. Wei, J. Li *et al.*, 2003 Isolation and analyses of genes preferentially expressed during early cotton fiber development by subtractive PCR and cDNA array. Nucleic Acids Res 31.

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

Jiang, Y., M. Ding, Y. Cao, F. Yang, H. Zhang *et al.*, 2015 Genetic fine mapping and candidate gene analysis of the Gossypium hirsutum Ligon lintless-1 (Li1) mutant on chromosome 22(D). Molecular Genetics and Genomics 290**:** 2199-2211.

John, M. E., and G. Keller, 1996 Metabolic pathway engineering in cotton: Biosynthesis of polyhydroxybutyrate in fiber cells. Proceedings of the National Academy of Sciences of the United States of America 93**:** 12768-12773.

John, Z. Y., R. J. Kohel, D. D. Fang, J. Cho, A. Van Deynze *et al.*, 2012 A high-density simple sequence repeat and single nucleotide polymorphism genetic map of the tetraploid cotton genome. G3: Genes, Genomes, Genetics 2**:** 43-58.

Jones, D. T., W. R. Taylor and J. M. Thornton, 1992 The rapid generation of mutation data matrices from protein sequences. Bioinformatics 8**:** 275-282.

Juan, H., F. ShuLi, S. MeiZhen, P. ChaoYou, W. HengLing *et al.*, 2015 Cloning and function analysis of uridine diphosphate glycosyltransferase gene GhUGT85O1 in cotton (Gossypium hirsutum). Journal of Agricultural Biotechnology 23**:** 701-710.

Kearney, T. H., and G. J. Harrison, 1927 Inheritance of smooth seeds in cotton. J Agric Res 35**:** 193-217.

Kelly, B., N. Abidi, D. Ethridge and E. F. Hequet, 2015 Fiber to fabric. Cotton**:** 665-744.

Kelly, C. M., E. F. Hequet and J. K. Dever, 2012 Interpretation of AFIS and HVI fiber property measurements in breeding for cotton fiber quality improvement. J Cotton Sci 16**:** 1-16.

Kim, C., H. Guo, W. Kong, R. Chandnani, L.-S. Shuang *et al.*, 2016 Application of genotyping by sequencing technology to a variety of crop breeding programs. Plant Science 242**:** 14-22.

Kim, H. J., N. Murai, D. D. Fang and B. A. Triplett, 2011 Functional analysis of Gossypium hirsutum cellulose synthase catalytic subunit 4 promoter in transgenic Arabidopsis and cotton tissues. Plant science 180**:** 323-332.

Kim, H. J., and B. A. Triplett, 2001 Cotton fiber growth in planta and in vitro. Models for plant cell elongation and cell wall biogenesis. Plant Physiology 127**:** 1361-1366.

Kim, H. J., and B. A. Triplett, 2004 Characterization of GhRac1 GTPase expressed in developing cotton (Gossypium hirsutum L.) fibers. Biochimica et Biophysica Acta (BBA)-Gene Structure and Expression 1679**:** 214-221.

Klaassen, P., S. Duijff, H. Swanenburg de Veye, F. Beemer, G. Sinnema *et al.*, 2016 Explaining the variable penetrance of CNVs: Parental intelligence modulates expression of intellectual impairment caused by the 22q11. 2 deletion. American Journal of Medical Genetics Part B: Neuropsychiatric Genetics 171**:** 790-796.

Klaassen, P., S. Duijff, H. Swanenburg de Veye, J. Vorstman, F. Beemer *et al.*, 2013 Behavior in preschool children with the 22q11. 2 deletion syndrome. American Journal of Medical Genetics Part A 161**:** 94-101.

Kodym, A., and R. Afza, 2003 Physical and chemical mutagenesis. Plant functional genomics**:** 189-203.

Kohel, R. J., E. V. Narbuth and C. R. Benedict, 1992 Fiber Development of Ligon Lintless-2 Mutant of Cotton. Crop Science 32**:** 733-735.

Köhn, L., S. J. Bowne, L. S. Sullivan, S. P. Daiger, M. S. Burstedt *et al.*, 2009 Breakpoint characterization of a novel∼ 59 kb genomic deletion on 19q13. 42 in autosomal-dominant retinitis pigmentosa with incomplete penetrance. European Journal of Human Genetics 17**:** 651-655.

Kong Depei, Q. L., Zhang Xueyan, Liu Ji, Wang Peng, Li Fuguang, 2017 Optimization of EMS Mutagenesis Condition and Screening of Mutants in Gossypium arboretum. L. Cotton Science 29**:** 336-344.

Krifa, M., and M. D. Ethridge, 2006 Compact spinning effect on cotton yarn quality: Interactions with fiber characteristics. Textile Research Journal 76**:** 388-399.

Krishnan, A., E. Guiderdoni, G. An, C. H. Yue-ie, C.-d. Han *et al.*, 2009 Mutant resources in rice for functional genomics of the grasses. Plant physiology 149**:** 165-170.

Kumar, P., Y. He, R. Singh, R. F. Davis, H. Guo *et al.*, 2016a Fine mapping and identification of candidate genes for a QTL affecting Meloidogyne incognita reproduction in Upland cotton. BMC Genomics 17**:** 567.

Kumar, S., G. Stecher and K. Tamura, 2016b MEGA7: Molecular Evolutionary Genetics Analysis version 7.0 for bigger datasets. Molecular biology and evolution**:** msw054.

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

Lacape, J.-M., T.-B. Nguyen, B. Courtois, J.-L. Belot, M. Giband *et al.*, 2005 QTL Analysis of Cotton Fiber Quality Using Multiple × Backcross Generations. Crop Sci. 45**:** 123-140.

Lee, C., R. Zhong, E. A. Richardson, D. S. Himmelsbach, B. T. McPhail *et al.*, 2007a The PARVUS gene is expressed in cells undergoing secondary wall thickening and is essential for glucuronoxylan biosynthesis. Plant and cell physiology 48**:** 1659-1672.

Lee, J. J., A. W. Woodward and Z. J. Chen, 2007b Gene expression changes and early events in cotton fibre development. Annals of Botany 100**:** 1391-1401.

Li, C., Y. Dong, T. Zhao, L. Li, C. Li *et al.*, 2016a Genome-Wide SNP Linkage Mapping and QTL Analysis for Fiber Quality and Yield Traits in the Upland Cotton Recombinant Inbred Lines Population. Frontiers in Plant Science 7**:** 1356.

Li, D.-D., X.-M. Ruan, J. Zhang, Y.-J. Wu, X.-L. Wang *et al.*, 2013 Cotton plasma membrane intrinsic protein 2s (PIP2s) selectively interact to regulate their water channel activities and are required for fibre development. New Phytologist 199**:** 695-707.

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

Li, F., G. Fan, K. Wang, F. Sun, Y. Yuan *et al.*, 2014a Genome sequence of the cultivated cotton Gossypium arboreum. Nature genetics 46**:** 567-572.

Li, H., and R. Durbin, 2009 Fast and accurate short read alignment with Burrows–Wheeler transform. Bioinformatics 25**:** 1754-1760.

Li, H., B. Handsaker, A. Wysoker, T. Fennell, J. Ruan *et al.*, 2009 The sequence alignment/map format and SAMtools. Bioinformatics 25**:** 2078-2079.

Li, H. B., Y. M. Qin, Y. Pang, W. Q. Song, W. Q. Mei *et al.*, 2007 A cotton ascorbate peroxidase is involved in hydrogen peroxide homeostasis during fibre cell development. New Phytol 175.

Li, L., J. Huang, L. Qin, Y. Huang, W. Zeng *et al.*, 2014b Two cotton fiber-associated glycosyltransferases, GhGT43A1 and GhGT43C1, function in hemicellulose glucuronoxylan biosynthesis during plant development. Physiologia Plantarum 152**:** 367-379.

Li, P., Y. j. Li, B. Wang, H. m. Yu, Q. Li *et al.*, 2016b The Arabidopsis UGT87A2, a stress-inducible family 1 glycosyltransferase, is involved in the plant adaptation to abiotic stresses. Physiologia Plantarum.

Li, Y., L. Tu, F. A. Pettolino, S. Ji, J. Hao *et al.*, 2016c GbEXPATR, a species-specific expansin, enhances cotton fibre elongation through cell wall restructuring. Plant biotechnology journal 14**:** 951-963.

Lin, L., G. J. Pierce, J. E. Bowers, J. C. Estill, R. O. Compton *et al.*, 2010 A draft physical map of a D-genome cotton species (Gossypium raimondii). BMC genomics 11**:** 395.

Lin, Z., X. Li, L. M. Shannon, C.-T. Yeh, M. L. Wang *et al.*, 2012 Parallel domestication of the Shattering1 genes in cereals. Nature genetics 44**:** 720-724.

Liszkay, A., E. van der Zalm and P. Schopfer, 2004 Production of reactive oxygen intermediates (O2−, H2O2, and OH) by maize roots and their role in wall loosening and elongation growth. Plant Physiology 136**:** 3114-3123.

Liu, D., J. Zhang, X. Liu, W. Wang, D. Liu *et al.*, 2016 Fine mapping and RNA-Seq unravels candidate genes for a major QTL controlling multiple fiber quality traits at the T1 region in upland cotton. BMC Genomics 17**:** 295.

Lubbers, E., P. Chee, J. Gannaway, R. Wright, K. El-Zik *et al.*, 2004 Levels and patterns of genetic diversity in upland cotton, pp. in *Plant and Animal Genome XII Conference*, San Diego, CA.

Lubbers, E., S. Walker, L. May and P. Chee, 2006 Breeding cultivars and germplasm with enhanced yield and quality. In P. Roberts et al. (ed.) 2005 Georgia Cotton Research and Extension Reports. UGA/CPES Research – Extension Publication No. 6, University of Georgia, Athens, GA**:** 136-152.

Lubbers, E. L., and P. W. Chee, 2009 The Worldwide Gene Pool of G. hirsutum and its Improvement, pp. 23-52 in *Genetics and Genomics of Cotton*, edited by A. H. Paterson. Springer US, New York, NY.

Luo, X., Z. Chen, J. Gao and Z. Gong, 2014 Abscisic acid inhibits root growth in Arabidopsis through ethylene biosynthesis. The Plant Journal 79**:** 44-55.

Lyzenga, W. J., and S. L. Stone, 2012 Regulation of ethylene biosynthesis through protein degradation. Plant signaling & behavior 7**:** 1438-1442.

Ma, Q.-F., C.-H. Wu, M. Wu, W.-F. Pei, X.-L. Li *et al.*, 2016 Integrative transcriptome, proteome, phosphoproteome and genetic mapping reveals new aspects in a fiberless mutant of cotton. Scientific reports 6.

Machado, A., Y. Wu, Y. Yang, D. J. Llewellyn and E. S. Dennis, 2009 The MYB transcription factor GhMYB25 regulates early fibre and trichome development. The Plant Journal 59**:** 52-62.

Maluszynski, M., B. S. Ahloowalia and B. Sigurbjörnsson, 1995 Application of in vivo and in vitro mutation techniques for crop improvement. Euphytica 85**:** 303-315.

Mamoon Rehman, H., M. Amjad Nawaz, L. Bao, Z. Hussain Shah, J.-M. Lee *et al.*, 2016 Genome-wide analysis of Family-1 UDP-glycosyltransferases in soybean confirms their abundance and varied expression during seed development. Journal of Plant Physiology 206**:** 87-97.

Mansoor, S., and A. H. Paterson, 2012 Genomes for jeans: cotton genomics for engineering superior fiber. Trends in Biotechnology 30**:** 521-527.

May, O. L., 2002 Quality improvement of upland cotton (Gossypium hirsutum L.). Journal of crop production 5**:** 371-394.

Mba, C., R. Afza, S. Bado and S. M. Jain, 2010 Induced mutagenesis in plants using physical and chemical agents. Plant cell culture: essential methods 20**:** 111-130.

McCallum, C. M., L. Comai, E. A. Greene and S. Henikoff, 2000 Targeting induced locallesions in genomes (TILLING) for plant functional genomics. Plant physiology 123**:** 439-442.

McLendon, C. A., 1912 *Mendelian inheritance in cotton hybrids*. Georgia Experiment Station.

Meena, A. K., M. Ramesh, C. Nagaraju and B. L. Kumhar, 2017 A Review of QTL Mapping in Cotton: Molecular Markers, Mapping Populations and Statistical Methods. Int. J. Curr. Microbiol. App. Sci 6**:** 3057-3080.

Mehetre, S., and M. Thombre, 1983 Fibre properties of x-ray induced glandless mutants in American cotton. Journal of Maharashtra Agricultural Universities 8**:** 189-190.

Mei, M., N. Syed, W. Gao, P. Thaxton, C. Smith *et al.*, 2004 Genetic mapping and QTL analysis of fiber-related traits in cotton (Gossypium). TAG Theoretical and Applied Genetics 108**:** 280-291.

Mei, W., Y. Qin, W. Song, J. Li and Y. Zhu, 2009 Cotton GhPOX1 encoding plant class III peroxidase may be responsible for the high level of reactive oxygen species production that is related to cotton fiber elongation. J Genet Genomics 36.

Meinke, D. W., 2013 A survey of dominant mutations in Arabidopsis thaliana. Trends in Plant Science 18**:** 84-91.

Meredith, W., 1994 Genetics and management factors influencing textile fiber quality, pp. 256–261.

Meredith, W. R., 1984 Quantitative genetics. Cotton**:** 131-150.

Moon, J., G. Parry and M. Estelle, 2004 The Ubiquitin-Proteasome Pathway and Plant Development. The Plant Cell 16**:** 3181-3195.

Munis, M. F., L. Tu, F. Deng, J. Tan, L. Xu *et al.*, 2010 A thaumatin-like protein gene involved in cotton fiber secondary cell wall development enhances resistance against Verticillium dahliae and other stresses in transgenic tobacco. Biochem Biophys Res Commun 393**:** 38-44.

Naoumkina, M., E. Bechere, D. D. Fang, G. N. Thyssen and C. B. Florane, 2017 Genome-wide analysis of gene expression of EMS-induced short fiber mutant Ligon lintless-y (liy) in cotton (Gossypium hirsutum L.). Genomics 109**:** 320-329.

Naoumkina, M., D. J. Hinchliffe, R. B. Turley, J. M. Bland and D. D. Fang, 2013 Integrated metabolomics and genomics analysis provides new insights into the fiber elongation process in Ligon lintless-2 mutant cotton (Gossypium hirsutum L.). BMC Genomics 14.

Naoumkina, M., G. Thyssen, D. D. Fang, D. J. Hinchliffe, C. Florane *et al.*, 2014 The Li 2 mutation results in reduced subgenome expression bias in elongating fibers of allotetraploid cotton (Gossypium hirsutum L.). PLoS One 9**:** e90830.

Naoumkina, M., G. N. Thyssen and D. D. Fang, 2015 RNA-seq analysis of short fiber mutants Ligon-lintless-1 (Li 1) and–2 (Li 2) revealed important role of aquaporins in cotton (Gossypium hirsutum L.) fiber elongation. BMC plant biology 15**:** 65.

Narbuth, E. V., and R. J. Kohel, 1990 Inheritance and linkage analysis of a new fiber mutant in cotton. Journal of Heredity 81**:** 131-133.

Nei, M., 2003 Genome evolution: let's stick together. Heredity 90.

Ngematov, M., V. Kovalenko, V. Shumnyi and K. Asrorov, 1975 Induction of cytoplasmicmale sterility in cotton by the method of radiation mutagenesis. Soviet Genetics 11**:** 1593–1595.

Oosterhuis, D. M., and J. Jernstedt, 1999 Morphology and anatomy of the cotton plant. Cotton: origin, history, technology, and production**:** 175-206.

Orford, S. J., and J. N. Timmis, 1998 Specific expression of an expansin gene during elongation of cotton fibres. Biochimica et Biophysica Acta (BBA)-Gene Structure and Expression 1398**:** 342-346.

Osmani, S. A., S. Bak and B. L. Møller, 2009 Substrate specificity of plant UDP-dependent glycosyltransferases predicted from crystal structures and homology modeling. Phytochemistry 70**:** 325-347.

Palle, S. R., L. M. Campbell, D. Pandeya, L. Puckhaber, L. K. Tollack *et al.*, 2013 RNAi-mediated Ultra-low gossypol cottonseed trait: performance of transgenic lines under field conditions. Plant biotechnology journal 11**:** 296-304.

Pang, C. Y., H. Wang, Y. Pang, C. Xu, Y. Jiao *et al.*, 2010 Comparative proteomics indicates that biosynthesis of pectic precursors is important for cotton fiber and Arabidopsis root hair elongation. Mol Cell Proteomics 9.

Park, Y. H., M. S. Alabady, M. Ulloa, B. Sickler, T. A. Wilkins *et al.*, 2005 Genetic mapping of new cotton fiber loci using EST-derived microsatellites in an interspecific recombinant inbred line cotton population. Mol Genet Genomics 274**:** 428-441.

Patel, J. D., R. J. Wright, D. Auld, R. Chandnani, V. H. Goff *et al.*, 2014 Alleles conferring improved fiber quality from EMS mutagenesis of elite cotton genotypes. Theoretical and Applied Genetics 127**:** 821-830.

Patel, J. D., R. J. Wright, R. Chandnani, V. H. Goff, J. Ingles *et al.*, 2016 EMS-mutated cotton populations suggest overlapping genetic control of trichome and lint fiber variation. Euphytica 208**:** 597-608.

Paterson, A., Y. Saranga, M. Menz, C. X. Jiang and R. Wright, 2003 QTL analysis of genotype× environment interactions affecting cotton fiber quality. TAG Theoretical and Applied Genetics 106**:** 384-396.

Paterson, A. H., R. K. Boman, S. M. Brown and P. W. Chee, 2004 Reducing the genetic vulnerability of cotton. Crop science 44**:** 1900.

Paterson, A. H., Michael Freeling, H. Tang and X. Wang, 2010 Insights from the Comparison of Plant Genome Sequences. Annual Review of Plant Biology 61**:** 349-372.

Paterson, A. H., J. F. Wendel, H. Gundlach, H. Guo, J. Jenkins *et al.*, 2012 Repeated polyploidization of Gossypium genomes and the evolution of spinnable cotton fibres. Nature 492**:** 423-427.

Pathirana, R., 2011 Plant mutation breeding in agriculture. Plant sciences reviews**:** 107-126.

Percival, A., J. Wendel and J. Stewart, 1999 Taxonomy and germplasm resources. Cotton: origin, history, technology, and production**:** 33–63.

Percy, R., B. Hendon, E. Bechere and D. Auld, 2015 Qualitative genetics and utilization of mutants. Cotton**:** 155-186.

Pickersgill, B., 2007 Domestication of Plants in the Americas: Insights from Mendelian and Molecular Genetics. Annals of Botany 100**:** 925-940.

Pickett, J., 2007 UBE1, you're not alone. Nat Rev Mol Cell Biol 8**:** 599-599.

Potikha, T. S., C. C. Collins, D. I. Johnson, D. P. Delmer and A. Levine, 1999 The involvement of hydrogen peroxide in the differentiation of secondary walls in cotton fibers. Plant physiology 119**:** 849-858.

Qin, H., W. Guo, Y.-M. Zhang and T. Zhang, 2008a QTL mapping of yield and fiber traits based on a four-way cross population in *Gossypium hirsutum L.* TAG Theoretical and Applied Genetics 117**:** 883-894.

Qin, Y.-M., C.-Y. Hu and Y.-X. Zhu, 2008b The ascorbate peroxidase regulated by H2O2 and ethylene is involved in cotton fiber cell elongation by modulating ROS homeostasis. Plant signaling & behavior 3**:** 194-196.

Rathore, K. S., S. Sundaram, G. Sunilkumar, L. M. Campbell, L. Puckhaber *et al.*, 2012 Ultra-low gossypol cottonseed: generational stability of the seed-specific, RNAi-mediated phenotype and resumption of terpenoid profile following seed germination. Plant biotechnology journal 10**:** 174-183.

Raut, R., H. Jain and R. Panwar, 1971 RADIATION-INDUCED PHOTO-INSENSITIVE MUTANTS IN COTTON, pp. Indian Agricultural Research Inst., New Delhi.

Reinisch, A. J., J. M. Dong, C. L. Brubaker, D. M. Stelly, J. F. Wendel *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**:** 829.

Renny-Byfield, S., J. T. Page, J. A. Udall, W. S. Sanders, D. G. Peterson *et al.*, 2016 Independent domestication of two Old World cotton species. Genome biology and evolution 8**:** 1940-1947.

Rong, J., C. Abbey, J. E. Bowers, C. L. Brubaker, C. Chang *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., F. A. Feltus, V. N. Waghmare, G. J. Pierce, P. W. Chee *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.

Rong, J. K., G. J. Pierce, V. N. Waghmare, C. A. J. Rogers, A. Desai *et al.*, 2005 Genetic mapping and comparative analysis of seven mutants related to seed fiber development in cotton. Theoretical and Applied Genetics 111**:** 1137-1146.

Ruan, Y.-L., D. J. Llewellyn and R. T. Furbank, 2003 Suppression of sucrose synthase gene expression represses cotton fiber cell initiation, elongation, and seed development. The Plant Cell 15**:** 952-964.

Said, J. I., Z. Lin, X. Zhang, M. Song and J. Zhang, 2013 A comprehensive meta QTL analysis for fiber quality, yield, yield related and morphological traits, drought tolerance, and disease resistance in tetraploid cotton. BMC genomics 14**:** 776.

Said, J. I., M. Song, H. Wang, Z. Lin, X. Zhang *et al.*, 2015 A comparative meta-analysis of QTL between intraspecific Gossypium hirsutum and interspecific G. hirsutum × G. barbadense populations. Molecular Genetics and Genomics 290**:** 1003-1025.

Saint Paul, V. v., 2010 Stress inducible glycosyltransferases in Arabidopsis thaliana and their impact on plant metabolism and defense mechanisms, pp. lmu.

Salamov, A. A., and V. V. Solovyev, 2000 Ab initio gene finding in Drosophila genomic DNA. Genome research 10**:** 516-522.

Sampedro, J., and D. J. Cosgrove, 2005 The expansin superfamily. Genome Biology 6**:** 242.

Sasser, P., 1981 Basics of high volume instruments for fiber testing, pp. in *Proceedings-Beltwide Cotton Production Research Conferences*.

Seelanan, T., A. Schnabel and J. F. Wendel, 1997 Congruence and consensus in the cotton tribe (Malvaceae). Systematic Botany**:** 259-290.

Shalem, O., N. E. Sanjana and F. Zhang, 2015 High-throughput functional genomics using CRISPR-Cas9. Nature reviews. Genetics 16**:** 299.

Shamsuzzaman, K., M. Hamid, M. Azad, M. Hussain and M. Majid, 2003 Varietal improvement of cotton (Gossypium hirsutum) through mutation breeding. Improvement of new and traditional industrial crops by induced mutations and related biotechnology**:** 78.

Shan, C.-M., X.-X. Shangguan, B. Zhao, X.-F. Zhang, L.-m. Chao *et al.*, 2014 Control of cotton fibre elongation by a homeodomain transcription factor GhHOX3. Nature communications 5.

Shan, W., Y. Jiang, J. Han and K. Wang, 2016 Comprehensive cytological characterization of the Gossypium hirsutum genome based on the development of a set of chromosome cytological markers. The Crop Journal 4**:** 256-265.

Shappley, Z., J. Jenkins, C. Watson, A. Kahler and W. Meredith, 1996 Establishment of molecular markers and linkage groups in two F2 populations of upland cotton. Theoretical and applied genetics 92**:** 915-919.

Shappley, Z., J. N. Jenkins, J. Zhu and J. C. McCarty Jr, 1998 Quantitative trait loci associated with agronomic and fiber traits of upland cotton.

Shen, X., W. Guo, Q. Lu, X. Zhu, Y. Yuan *et al.*, 2007 Genetic mapping of quantitative trait loci for fiber quality and yield trait by RIL approach in Upland cotton. Euphytica 155**:** 371-380.

Shen, X., T. Zhang, W. Guo, X. Zhu and X. Zhang, 2006 Mapping Fiber and Yield QTLs with Main, Epistatic, and QTL × Environment Interaction Effects in Recombinant Inbred Lines of Upland Cotton. Crop Sci. 46**:** 61-66.

Shi, Y. H., S. W. Zhu, X. Z. Mao, J. X. Feng, Y. M. Qin *et al.*, 2006 Transcriptome profiling, molecular biological, and physiological studies reveal a major role for ethylene in cotton fiber cell elongation. Plant Cell 18.

Shoemaker, D., 1909 A study of leaf characters in cotton hybrids. Journal of Heredity**:** 116-118.

Shofner, F. M., Y. Chu and D. Thibodeaux, 1990 An overview of the advanced fiber information system, pp. 173-181 in *Proc. Int. Cotton Conf., Faserinstitut, Bremen, Germany*.

Smith, C. W., 2003 Registration of TAM 94L-25 and TAM 94J-3 germplasm lines of upland cotton with improved fiber length. Crop science 43**:** 742-744.

Stewart, J., 1995 Potential for crop improvement with exotic germplasm and genetic engineering, pp. 313-327 in *Proceeding of the world cotton research conference-I, Brishbane, Australia, February 14-17, Melbourne, 1995*.

Stewart, J. M., L. A. Craven, C. Brubaker and J. F. Wendel, 2015 Gossypium anapoides (Malvaceae), a new species from Western Australia. Novon 23**:** 447-451.

Suh, M., and P. Sasser, 1996 The technological and economic impact of high volume instrument (HVI) systems on the cotton and cotton textile industries. Journal of the Textile Institute 87**:** 43-59.

Sun, X., S. Y. Gong, X. Y. Nie, Y. Li, W. Li *et al.*, 2015 A R2R3-MYB transcription factor that is specifically expressed in cotton (Gossypium hirsutum) fibers affects secondary cell wall biosynthesis and deposition in transgenic Arabidopsis. Physiologia plantarum 154**:** 420-432.

Sunilkumar, G., L. M. Campbell, L. Puckhaber, R. D. Stipanovic and K. S. Rathore, 2006 Engineering cottonseed for use in human nutrition by tissue-specific reduction of toxic gossypol. Proceedings of the National Academy of Sciences 103**:** 18054-18059.

Suo, J. F., X. O. Liang, L. Pu, Y. S. Zhang and Y. B. Xue, 2003 Identification of GhMYB109 encoding a R2R3 MYB transcription factor that expressed specifically in fiber initials and elongating fibers of cotton (Gossypium Hirsutum L.). Biochimica Et Biophysica Acta-Gene Structure and Expression 1630**:** 25-34.

Tai, F. J., X. L. Wang, W. L. Xu and X. B. Li, 2008 Characterization and expression analysis of two cotton genes encoding putative UDP-Glycosyltransferases. Molecular Biology 42**:** 44-51.

Talebi, A. B., A. B. Talebi and B. Shahrokhifar, 2012 Ethyl methane sulphonate (EMS) induced mutagenesis in Malaysian rice (cv. MR219) for lethal dose determination. American Journal of Plant Sciences 3**:** 1661.

Tang, H., H. E. Cuevas, S. Das, U. U. Sezen, C. Zhou *et al.*, 2013 Seed shattering in a wild sorghum is conferred by a locus unrelated to domestication. Proceedings of the National Academy of Sciences 110**:** 15824-15829.

Tang, W., L. Tu, X. Yang, J. Tan, F. Deng *et al.*, 2014 The calcium sensor GhCaM7 promotes cotton fiber elongation by modulating reactive oxygen species (ROS) production. New Phytologist 202**:** 509-520.

Tanksley, S. D., and S. R. McCouch, 1997 Seed banks and molecular maps: unlocking genetic potential from the wild. Science 277**:** 1063-1066.

Terasaka, K., Y. Mizutani, A. Nagatsu and H. Mizukami, 2012 In situ UDP-glucose regeneration unravels diverse functions of plant secondary product glycosyltransferases. FEBS Letters 586**:** 4344-4350.

Thibodeaux, D., H. Senter, J. Knowlton, D. Mcalister and X. Cui, 2008 The impact of short fiber content on the quality of cotton ring spun yarn. J Cotton Sci 12**:** 368-377.

Thomas, B. C., B. Pedersen and M. Freeling, 2006 Following tetraploidy in an Arabidopsis ancestor, genes were removed preferentially from one homeolog leaving clusters enriched in dose-sensitive genes. Genome research 16**:** 934-946.

Thyssen, G. N., D. D. Fang, R. B. Turley, C. Florane, P. Li *et al.*, 2014 Next generation genetic mapping of the Ligon-lintless-2 (Li 2) locus in upland cotton (Gossypium hirsutum L.). Theoretical and applied genetics 127**:** 2183-2192.

Thyssen, G. N., D. D. Fang, R. B. Turley, C. Florane, P. Li *et al.*, 2015 Mapping-by-sequencing of Ligon-lintless-1 (Li1) reveals a cluster of neighboring genes with correlated expression in developing fibers of Upland cotton (Gossypium hirsutum L.). Theoretical and Applied Genetics 128**:** 1703-1712.

Thyssen, G. N., D. D. Fang, R. B. Turley, C. B. Florane, P. Li *et al.*, 2017 A Gly65Val substitution in an actin, GhACT_LI1, disrupts cell polarity and F-actin organization resulting in dwarf, lintless cotton plants. The Plant Journal.

Thyssen, G. N., D. D. Fang, L. Zeng, X. Song, C. D. Delhom *et al.*, 2016 The Immature fiber mutant phenotype of cotton (Gossypium hirsutum) is linked to a 22-bp frame-shift deletion in a mitochondria targeted pentatricopeptide repeat gene. G3: Genes| Genomes| Genetics 6**:** 1627-1633.

Tiwari, S. C., and T. A. Wilkins, 1995 Cotton (*Gossypium hirsutum*) seed trichomes expand via diffuse growing mechanism. Canadian Journal of Botany 73**:** 746-757.

Todarello, G., N. Feng, B. S. Kolachana, C. Li, R. Vakkalanka *et al.*, 2014 Incomplete penetrance of NRXN1 deletions in families with schizophrenia. Schizophrenia research 155**:** 1-7.

Tyagi, P., M. A. Gore, D. T. Bowman, B. T. Campbell, J. A. Udall *et al.*, 2014 Genetic diversity and population structure in the US Upland cotton (Gossypium hirsutum L.). Theoretical and Applied Genetics 127**:** 283-295.

Ulloa, M., 2006 Heritability and correlations of agronomic and fiber traits in an okra-leaf upland cotton population. Crop science 46**:** 1508-1514.

Ulloa, M., and W. R. Meredith Jr, 2000 Genetic linkage map and QTL analysis of agronomic and fiber quality traits in an intraspecific population. Journal of Cotton Science 4**:** 161-170.

Van der Sluijs, M., and L. Hunter, 1999 Neps in cotton lint. Textile Progress 28**:** 1-47.

Van Esbroeck, G., and D. T. Bowman, 1998 Cotton germplasm diversity and its importance to cultivar development. Journal of Cotton Science.

Van Ooijen, J., 2011 Multipoint maximum likelihood mapping in a full-sib family of an outbreeding species. Genetics research 93**:** 343-349.

Viot, C., 2017 Evolution and domestication of diploid cultivated cottons: molecular genetics and agronomic evidence.

Waghmare, V. N., J. Rong, C. J. Rogers, J. E. Bowers, P. W. Chee *et al.*, 2016 Comparative transmission genetics of introgressed chromatin in Gossypium (cotton) polyploids. American journal of botany 103**:** 719-729.

Wakeham, H., 1955 Cotton fiber length distribution—an important quality factor. Textile Research Journal 25**:** 422-429.

Walford, S.-A., Y. Wu, D. J. Llewellyn and E. S. Dennis, 2012 Epidermal cell differentiation in cotton mediated by the homeodomain leucine zipper gene, GhHD-1. The Plant Journal 71**:** 464-478.

Wan, Q., X. Guan, N. Yang, H. Wu, M. Pan *et al.*, 2016 Small interfering RNAs from bidirectional transcripts of GhMML3_A12 regulate cotton fiber development. New Phytologist.

Wan, Q., Z. Zhang, M. Hu, L. Chen, D. Liu *et al.*, 2007 T1 locus in cotton is the candidate gene affecting lint percentage, fiber quality and spiny bollworm (Earias spp.) resistance. Euphytica 158**:** 241-247.

Wang, B., S. H. Jin, H. Q. Hu, Y. G. Sun, Y. W. Wang *et al.*, 2012 UGT87A2, an Arabidopsis glycosyltransferase, regulates flowering time via FLOWERING LOCUS C. New Phytologist 194**:** 666-675.

Wang, K. L.-C., H. Yoshida, C. Lurin and J. R. Ecker, 2004 Regulation of ethylene gas biosynthesis by the Arabidopsis ETO1 protein. Nature 428**:** 945-950.

Wang, Z., D. Zhang, X. Wang, X. Tan, H. Guo *et al.*, 2013 A whole-genome DNA marker map for cotton based on the D-genome sequence of Gossypium raimondii L. G3: Genes| Genomes| Genetics 3**:** 1759-1767.

Wangzhen, G., Z. Tianzhen and D. Yezhang, 2005 Molecular marker assisted selection and pyramiding of two QTLs for fiber strength in upland cotton. Acta Genetica Sinica 32**:** 1275-1285.

Waters, W. T., J. Phillips and L. A. Fiori, 1966 The Effect of Fiber-Bundle Elongation of Medium Staple Cottons on Processing Performance and Y arn Properties. Textile Research Journal 36**:** 1004-1012.

Wendel , J., and R. Cronn, 2003 Polyploidy and the evolutionary history

of cotton. Adv Agron 78**:** 139–186.

Wendel, J. F., and V. A. Albert, 1992 Phylogenetics of the Cotton Genus (Gossypium): Character-State Weighted Parsimony Analysis of Chloroplast-DNA Restriction Site Data and Its Systematic and Biogeographic Implications. Systematic Botany 17**:** 115-143.

Wendel, J. F., C. Brubaker, I. Alvarez, R. Cronn and J. M. Stewart, 2009 Evolution and natural history of the cotton genus, pp. 3-22 in *Genetics and genomics of cotton*. Springer.

Wendel, J. F., C. L. Brubaker and A. E. Percival, 1992 Genetic diversity in Gossypium hirsutum and the origin of upland cotton. American Journal of Botany**:** 1291-1310.

Wendel, J. F., and R. C. Cronn, 2003 Polyploidy and the evolutionary history of cotton. Advances in agronomy 78**:** 139-186.

Wendel, J. F., and C. E. Grover, 2015 Taxonomy and evolution of the cotton genus, Gossypium. Cotton**:** 25-44.

Xiao, J., K. Wu, D. Fang, D. M. Stelly, J. Yu *et al.*, 2009 New SSR Markers for Use in Cotton (*Gossypium* spp.*)* Improvement. The Journal of Cotton Science 13**:** 75-157.

Xu, P., J. Gao, Z. Cao, P. W. Chee, Q. Guo *et al.*, 2017 Fine mapping and candidate gene analysis of qFL-chr1, a fiber length QTL in cotton. Theoretical and Applied Genetics 130**:** 1309-1319.

Xu, Z., R. J. Kohel, G. Song, J. Cho, M. Alabady *et al.*, 2008 Gene-rich islands for fiber development in the cotton genome. Genomics 92**:** 173-183.

Yang, S., and S. Gordon, A study on cotton fibre elongation measurement.

Yang, Z., C. Zhang, X. Yang, K. Liu, Z. Wu *et al.*, 2014 PAG1, a cotton brassinosteroid catabolism gene, modulates fiber elongation. New Phytologist 203**:** 437-448.

Yi, G., S.-H. Sze and M. R. Thon, 2007 Identifying clusters of functionally related genes in genomes. Bioinformatics 23**:** 1053-1060.

Yonekura-Sakakibara, K., and K. Hanada, 2011 An evolutionary view of functional diversity in family 1 glycosyltransferases. The Plant Journal 66**:** 182-193.

Yoshida, H., M. Nagata, K. Saito, K. L. Wang and J. R. Ecker, 2005 Arabidopsis ETO1 specifically interacts with and negatively regulates type 2 1-aminocyclopropane-1-carboxylate synthases. BMC Plant Biology 5**:** 14.

Yu, J. W., S. X. Yu, C. R. Lu, W. Wang, S. L. Fan *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**:** 716-724.

Yu, Y., D. Yuan, S. Liang, X. Li, X. Wang *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.

Yuan, L., Y. Dou, S. F. Kianian, C. Zhang and D. R. Holding, 2014a Deletion Mutagenesis Identifies a Haploinsufficient Role for γ-Zein in opaque2 Endosperm Modification. Plant Physiology 164**:** 119-130.

Yuan, Y., T. Wang, Y. Shi, H. Shang, A. Liu *et al.*, 2014b Molecular marker-assisted selection and pyramiding effect of major QTLs for cotton fiber strength. New Biotechnology 31**:** S14.

Zhang, B., and J.-Y. Liu, 2016 Cotton cytosolic pyruvate kinase GhPK6 participates in fast fiber elongation regulation in a ROS-mediated manner. Planta 244**:** 915-926.

Zhang, F., X. Jin, L. Wang, S. Li, S. Wu *et al.*, 2016a GhFAnnxA affects fiber elongation and secondary cell wall biosynthesis associated with Ca2+ influx, ROS homeostasis and actin filament reorganization. Plant Physiology.

Zhang, J., and R. Percy, 2007 Improving Upland cotton by introducing desirable genes from Pima cotton, pp. 10-14 in *World Cotton Research Conf.*

Zhang, S.-W., X.-F. Zhu, L.-C. Feng, X. Gao, B. Yang *et al.*, 2016b Mapping of fiber quality QTLs reveals useful variation and footprints of cotton domestication using introgression lines. Scientific Reports 6**:** 31954.

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

Zhang, T., Y. Yuan, J. Yu, W. Guo and R. J. Kohel, 2003a Molecular tagging of a major QTL for fiber strength in Upland cotton and its marker-assisted selection. Theoretical and Applied Genetics 106**:** 262-268.

Zhang, X.-D., J. N. Jenkins, F. E. Callahan, R. G. Creech, Y. Si *et al.*, 2003b Molecular cloning, differential expression, and functional characterization of a family of class I ubiquitin-conjugating enzyme (E2) genes in cotton (Gossypium). Biochimica et Biophysica Acta (BBA) - Gene Structure and Expression 1625**:** 269-279.

Zhang, Y. X., Z. X. Lin, Q. Z. Xia, M. J. Zhang and X. L. Zhang, 2008 Characteristics and analysis of simple sequence repeats in the cotton genome based on a linkage map constructed from a BC1 population between Gossypium hirsutum and G-barbadense. Genome 51**:** 534-546.

Zhang, Z., J. Rong, V. N. Waghmare, P. W. Chee, O. L. May *et al.*, 2011 QTL alleles for improved fiber quality from a wild Hawaiian cotton, Gossypium tomentosum. Theoretical and applied genetics 123**:** 1075.

Zhang, Z., Y.-L. Ruan, N. Zhou, F. Wang, X. Guan *et al.*, 2017 Suppressing a Putative Sterol Carrier Gene Reduces Plasmodesmal Permeability and Activates Sucrose Transporter Genes during Cotton Fiber Elongation. The Plant Cell.

Zhang, Z. S., M. C. Hu, J. Zhang, D. J. Liu, J. Zheng *et al.*, 2009 Construction of a comprehensive PCR-based marker linkage map and QTL mapping for fiber quality traits in upland cotton (Gossypium hirsutum L.). Molecular Breeding 24**:** 49-61.

Zhao, L., L. Yuanda, C. Caiping, T. Xiangchao, C. Xiangdong *et al.*, 2012 Toward allotetraploid cotton genome assembly: integration of a high-density molecular genetic linkage map with DNA sequence information. BMC genomics 13**:** 539.

Zhao, Y., H. Wang, W. Chen, Y. Li, H. Gong *et al.*, 2015 Genetic diversity and population structure of elite cotton (Gossypium hirsutum L.) germplasm revealed by SSR markers. Plant systematics and evolution 301**:** 327-336.

Zhong, R., M. J. Peña, G.-K. Zhou, C. J. Nairn, A. Wood-Jones *et al.*, 2005 Arabidopsis Fragile Fiber8, Which Encodes a Putative Glucuronyltransferase, Is Essential for Normal Secondary Wall Synthesis. The Plant Cell 17**:** 3390-3408.

Zhong, R., and Z.-H. Ye, 2003 Unraveling the functions of glycosyltransferase family 47 in plants. Trends in Plant Science 8**:** 565-568.

Table 4.1 Segregation distortion of DNA markers in the ***Li₂*** region

	CISP15			**NAU2980**			**NAU3827**		
Genotype	observed	expected	Chi test	observed	expected	Chi test	observed	Expected	Chi test
GH/GH	284	380.00	2.02E-08	285	382.75	2.16E-08	288	383.50	7.56E-08
GH/GB	796	780.00		808	765.50		814	767.00	
GB/GB	440	380.00		438	382.75		432	383.50	
Total	1,520			1,531			1,534		

Table 4.2 – Properties of non-synonymous SNPs differentiating fiber and non-fiber producing *Gossypium* species

GENE ID (*G. ramondii*)	POSITION	STRAND	Non-Fiber	Fiber	SNP location	Non-Fiber aa	Fiber aa	GENE ID (*G. hirsutum*)	Annotated function
Gorai.013G269400	58104353	-	A	T	CDS	F	Y	Gh_D13G2434	glucuronoxylan glucuronosyltransferase IRX7
Gorai.013G270200	58142691	+	G	A	CDS	G	R	Gh_D13G2442	ethylene-overproduction protein 1-like
Gorai.013G270200	58143864	+	C	A	CDS	L	I	Gh_D13G2442	ethylene-overproduction protein 1-like
Gorai.013G270200	58144018	+	G	A	CDS	G	D	Gh_D13G2442	ethylene-overproduction protein 1-like
Gorai.013G270200	58144068	+	G	A	CDS	V	I	Gh_D13G2442	ethylene-overproduction protein 1-like
Gorai.013G270300	58144807	-	A	G	3-UTR	.	.	Gh_D13G2443	ubiquitin-activating enzyme E1 1-like
Gorai.013G270300	58144988	-	T	G	3-UTR	.	.	Gh_D13G2443	ubiquitin-activating enzyme E1 1-like
Gorai.013G270300	58145428	-	A	G	CDS	L	L	Gh_D13G2443	ubiquitin-activating enzyme E1 1-like
Gorai.013G270400	58154218	-	T	A	CDS	Q	H	Gh_D13G2444	ubiquitin-activating enzyme E1 1-like, transcript variant X2
Gorai.013G270800	58179089	+	A	G	CDS	E	G	Gh_D13G2448	glucan endo-1,3-beta-glucosidase 9-like
Gorai.013G270900	58181681	+	C	T	CDS	P	L	Gh_D13G2449	UDP-glycosyltransferase 87A1-like
Gorai.013G271000	58184415	+	C	G	CDS	R	G	Gh_D13G2450	UDP-glycosyltransferase 87A2-like, transcript variant X2
Gorai.013G271000	58185533	+	C	T	3-UTR	.	.	Gh_D13G2450	UDP-glycosyltransferase 87A2-like, transcript variant X2
Gorai.013G271100	58187286	+	T	A	CDS	I	N	Gh_D13G2452	UDP-glycosyltransferase 87A1-like
Gorai.013G271700	58235904	+	C	T	5-UTR	.	.	Gh_D13G2458	trans-cinnamate 4-monooxygenase-like
Gorai.013G272000	58243663	+	A	G	CDS	R	G	Gh_D13G2460	expansin-A8-like

a. b. c.

d.

Figure 4.1- Phenotype of seeds containing the *Li₂* mutant allele, and homozygous WT allele (a) the *Li₂* mutant allele (b) homozygous WT allele (c) a branch showing atypical expression of the *Li₂* phenotype (two bolls of *Li₂* phenotype on side with the center with normal lint or wild type phenotype); and (d) atypical expression of *Li₂* phenotype within a boll. The extreme shows normal expression of *Li₂* or wild type phenotype in a boll while central two shows abnormal phenotype in a boll of *Li₂* plant

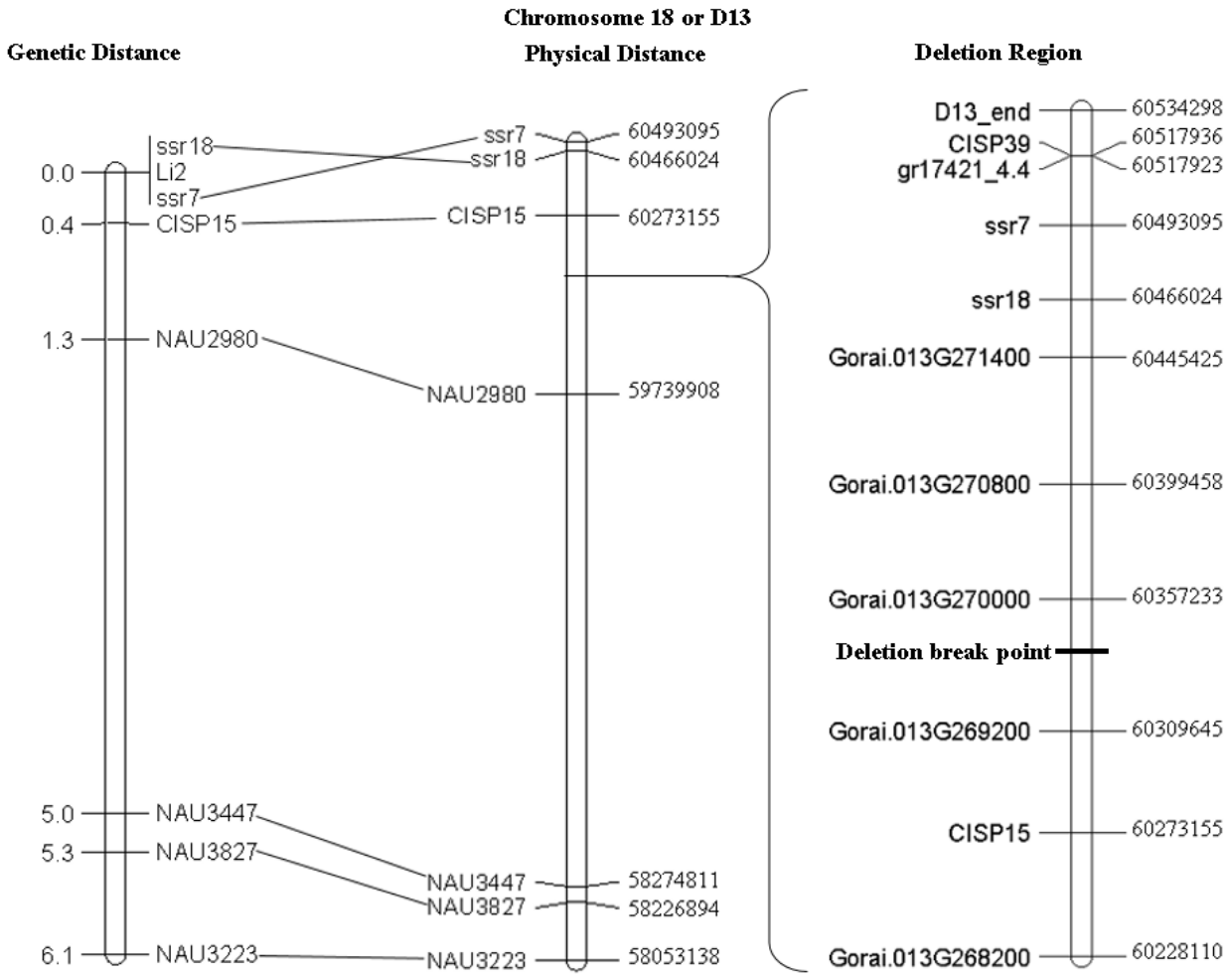
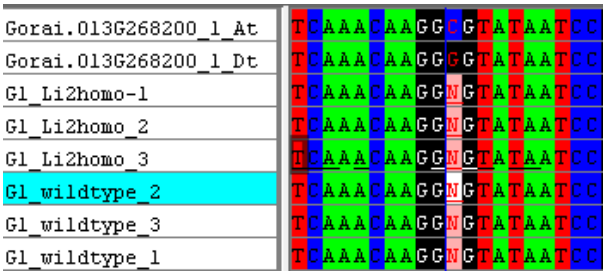
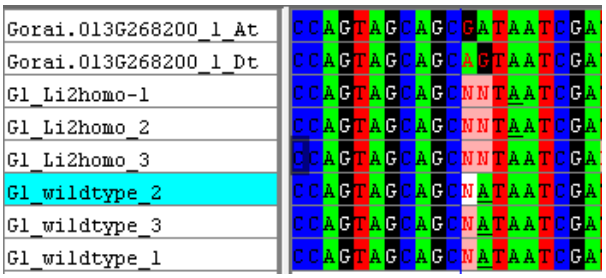


Figure 4.2- Genetic and physical map of the *Li₂* locus
Genetic distance is in cM and physical distance is in bp. An approximate deletion break point is shown. The physical map indicates that the markers from the deletion region (ssr7 and ssr18) are completely linked with the *Li₂* phenotype. Gorai.013Gxxxxxx represents genes from which amplicons were sequenced to investigate the deletion theory and find approximate break points.

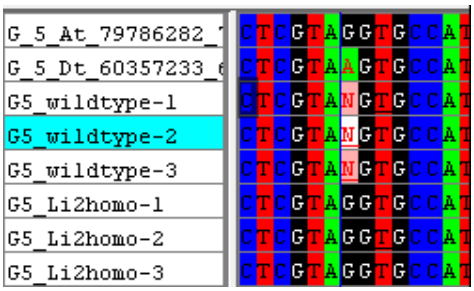
a.



b.



c.



d.

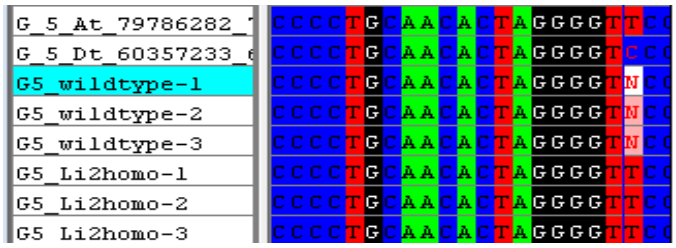


Figure 4.3 -Validating deletion of the terminal end of chromosome 18 in *Li₂* homozygous
Eight sequences are shown in these figures. The sequences are At subgenome, Dt subgenome, three biological replicates of wild type and three biological replicates of *Li₂* homozygous lines.
(a) and (b) are representation of the sequence form Gorai.013G268200. We can see both alleles from At and Dt sub genome are present in wild type and *Li₂* homozygous lines.
(c) and (d) are representation of the sequences from Gorai.013G270000. Here we can see both alleles from At and Dt sub genome are present in wild type and but allele from Dt subgenome is *Li₂* homozygous lines.

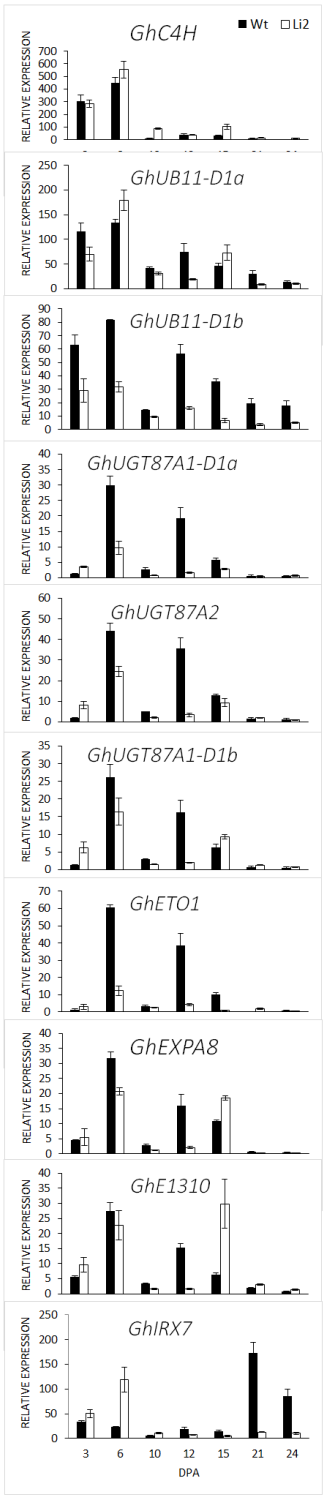
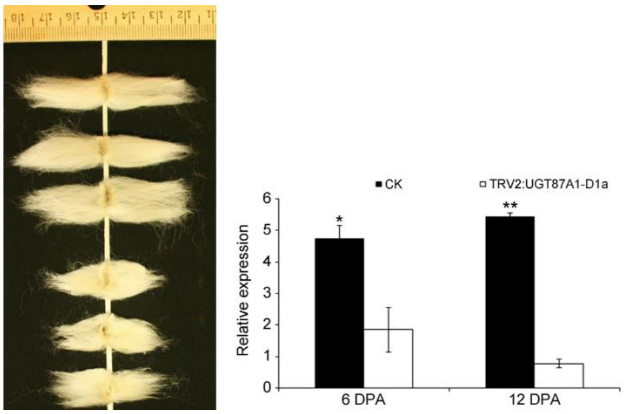
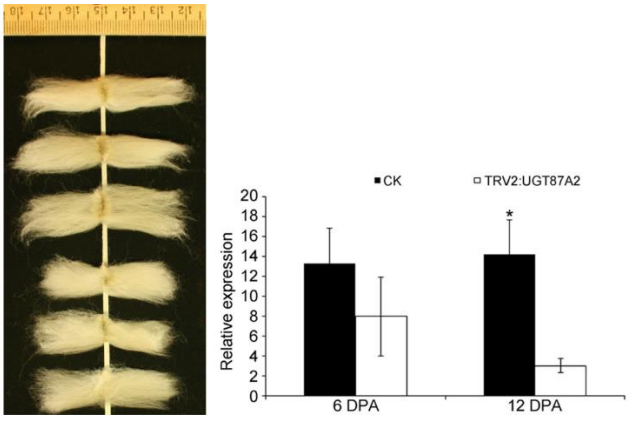


Figure 4.4 – Gene expression analysis using RT-qPCR for candidate genes in fiber tissues from wild-type and mutant plants. X-axis represents Days post-anthesis (DPA). Fiber and ovule tissues were mixed for 3 and 6 DPA, for remaining DPA (10,12,15,21 and 24) only fiber tissues were used for the study. Error bars represents standard error (SE) from three biological replicates.

a.



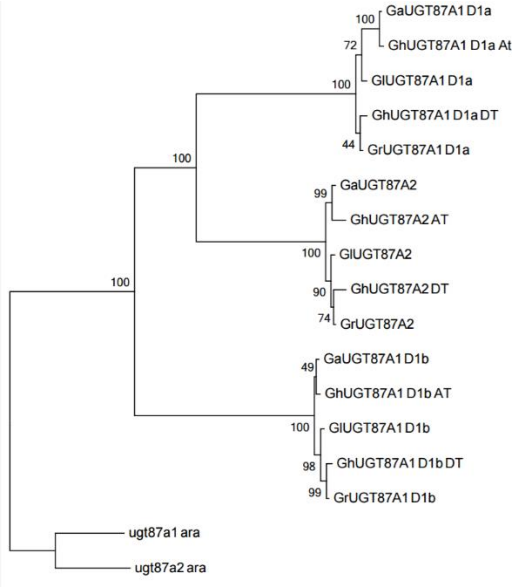
b.



c.

Figure 4.5 – Reduction of fiber length by Virus Induced Gene Silencing (VIGS)
(a) Representative plants treated with TRV2:GRCLA1 as a visual marker for verifying efficiency of viral infection. Fiber samples from plants infected by TRV2:UGT87A1-D1a (b) and TRV2:UGT87A2 (c) and relative expression of the genes at different fiber development stages (6 and 12 DPA)

a.



b.

```
ugt87a2_ara          QGSFLSVSEAQMEEIVKGLRESGVRFLWVARGGELKLKEALEGSLGVVVSWCDQLRVLCH
GaUGT87A1-D1b        MGSFLSVSNVQTEEITAGLQDSGVRFLLVAREGSWKLKHG-CGGEGLVVEWCDQLRVLCH
GhUGT87A1-D1b_AT     MGSFLSVSNVQMEEITAGLQDSGVRFLLVAREGSWKLKHG-YGGEGLVVEWCDQLRVLCH
GlUGT87A1-D1b        MGSFLSVSNVQMEEITAGLQDSGVRFLLVAREGSWKLKHG-CGGEGLVVEWCDQLRVLCH
GhUGT87A1-D1b_DT     MGSFLSVSNVQMEEITAGLQDSGVRFLLVAREGSWKLKHG-YGGEGLVVEWCDQLRVLCH
GrUGT87A1-D1b        MGSFLSVSNVQMEEITAGLQDSGVRFLLVAREGSWKLKHG-YGGEGLVVEWCDQLRVLCH
GhUGT87A2_DT         LGSFLSVSNTQMDEIVAGLQISGVRYLWVARGEASRLKDR-CGDMGLVVEWCDQLKVLCH
GrUGT87A2            LGSFLSVSNTQMDEIVAGLQISGVRYLWVARGEASRLKDR-CGDMGLVVEWCDQLKVLCH
GlUGT87A2            LGSFLSVSNTQMDEIVAGLQISGVRYLWVARGEASRLKDR-CGDMGLVVEWCDQLKVLCH
GhUGT87A2_AT         LGSFLSVSNTQMDEIVAGLQISGVRYLWVARGEASRLKDR-CGDMGLVVEWCDQLKVLCH
GaUGT87A2            LGSFLSVSNTQMDEIVGGLQISGVRYLWVARGEASRLKDR-CGDMGLVVEWCDQLKVLCH
GaUGT87A1-D1a        LGSFLSVSNDQMDEIAAGLQDSGVPYLWVARGETSWPTEK-GSEMGFVVEWCDQLKVLCY
GhUGT87A1-D1a_At     LGSFLSVSNDQMDEIAAGLQDSGVPYLWVARGETSWPTEK-GSEMGFVVEWCDQLKVLCY
GlUGT87A1-D1a        LGSFLSVSNDQMDEIAAGLQDSGVPYLWVARGGTSRPTEK-GGEMGLVVEWCDQLKVLCH
GhUGT87A1-D1a_DT     LGSFLSVSNDQMDEIAAGLQDSGVLYLWVARGGTSRLTEK-GSEMGLVVEWCDQLKVLCH
GrUGT87A1-D1a        LGSFLSVSNDQMDEIAAGLQDSGVLYLWVARGGTSRLTEK-GSEMGLVVEWCDQLKVLCH
                      *******: * :**. *:. :** :: ***    . .    .  *.**.*****.***:

ugt87a1_ara          AAIGGFWTHCGYNSTLEGICSGVPLLTFPVFWDQFLNAKMIVEEWRVGMGIERKKQMELL
ugt87a2_ara          KAVGGFWTHCGFNSTLEGIYSGVPMLAFPLFWDQKLNAKMIVEDWRVGMRIERTKKNELL
GaUGT87A1-D1b        PSIGGFWSHCGWNSVKEGIFAGTPFLTFPLFADQNLNSKLIVEDWKIGWRIKKHQ----F
GhUGT87A1-D1b_AT     PSIGGFWSHCGWNSVKEGIFAGTPFLTFPLFADQNLNSKLIVEDWKIGWRTKKHQ----F
GlUGT87A1-D1b        PSIGGFWSHCGWNSVKEGIFAGTPFLTFPLFADQNLNSKLIVEDWKIGWRIKKHQ----F
GhUGT87A1-D1b_DT     PSIGGFWSHCGWNSVKEGIFAGTPFLTFPLFADQNLNSKLIVEDWKIGWRIKKHQ----F
GrUGT87A1-D1b        PSIGGFWSHCGWNSVKEGIFAGTPFLTFPLFADQNLNSKLIVEDWKIGWRIKKHQ----F
GhUGT87A2_DT         PSVGGFWTHCGWNSILEAVFAGVPMLTFPLFLDQDTNSRQIVKDWGNGWRVNDAVKAEKL
GrUGT87A2            PSVGGFWTHCGWNSILEAVFAGVPMLTFPLFLDQDTNSRQIVEDWGNGWRVNDAVKAEKL
GlUGT87A2            PSVGGFWTHCGWNSILEAVFAGVPMLTFPLFLDQDTNSRQIVEDWGNGWRVNDAVKAEKL
GhUGT87A2_AT         PSVWGFWMHCGWNSILDV----------------------RIV-----------------
GaUGT87A2            PSVGGFWTHCGWNSILEAVFAGVPMLTFPLFLDQDTNSRQIVEDWGNGWRVNDAVKAEKL
GaUGT87A1-D1a        SSIGGFLTHCGWNSILEAIYAGIPMLTFPILFDQAPNSKQIVDDWKIGWRLKEEQKDGSL
GhUGT87A1-D1a_At     SSIGGFLTHCGWNSILEAIYAGIPMLTFPILFDQAPNSKQIVDDWKIGWRLKEEQKDGSL
GlUGT87A1-D1a        SSIGGFLTHCGWNSILEAIYAGITMLTFPILFDQAPNSKQIVDDWKIGWRLKEQQKDGSL
GhUGT87A1-D1a_DT     SSIGGFLTHCGWNSILEAIYAGIPMLTFPILFDQAPNSKQIVDDWKIGWRLKEQQKDGRL
GrUGT87A1-D1a        SSIGGFLTHCGWNSILEAIYAGIPMLTFPILFDQAPNSKQIVDDWKIGWRLKEQQKDGRL
                      :: **  ***:**  :                    **
```

Figure 4.6 – Phylogenetic analysis and protein sequence alignment
(a) Bayesian inference phylogenetic analysis was performed in MrBayes3.2. The consensus tree shows that *GhUGT87A1_D1a* and *GhUGT87A2* are more closely related than *GhUGT87A1*_D1b (b) Protein sequence alignment performed using MUSCLE indicates deletion of 16 amino acid in highly conserved PSPG (plant secondary product glycosyltransferase) region of 44-amino acids, shown in red boxes

Chapter 5

SUMMARY

Cotton faces two inter-related challenges -- to identify and more effectively manipulate genes participating in complex fiber development that results in the primary economic product of cotton, while broadening the genetic diversity available to select for improved fiber quality. Here we showed how mutants can be used to decipher genes involved in fiber development and increase genetic diversity in cotton.

Studies to understand genetics and identify the causal mutation of the *Li*$_2$ phenotype have been going on for almost three decades. Due to the subtelomeric location of the *Li*$_2$ locus, it was difficult to identify two flanking markers. Numerous studies on fine mapping, gene expression and metabolomics have been conducted but none had identified the causal mutation of *Li*$_2$ -- indeed, the nature of the locus shows that some studies were flawed and could not have identified the gene. Using genetic markers and a large F_2 population, we inferred that the causal mutation for the *Li*$_2$ phenotype is a terminal deletion of around 176 Kb of chromosome 18. We further identified seven candidate genes in the deletion region that could participate in fiber elongation, two of which individually show mutant phenotypes resembling *Li*$_2$.

De novo mutants are an important part of crop improvement programs, exemplified by those that were central to the "Green Revolution". Here we evaluated ten different populations derived from 12 previously-identified mutant lines affecting various fiber quality components, to see if the improvement of fiber quality was stable, transferable and whether stacking of multiple mutant alleles could improve multiple fiber quality attributes. Multiple lines within each

population were significantly improved for the fiber trait for which the population was developed. Multiple lines were also improved for multiple fiber traits for which they were crossed, suggesting allele pyramiding from mutant lines does improve multiple fiber quality attributes. Remarkably, one population developed from mutants to improve fiber length showed 75% of lines improved for four or more fiber traits and 45% improved for five or more fiber traits. Thus, improvement of one fiber trait can simultaneously improve other fiber traits as well.

The most improved lines from each population for micronaire, fiber length, fiber strength, uniformity index, fiber elongation, and lint percent were O013 (31.7%), P058 (16.1%), K92 (22.4%), Q47 (4.1%), N068 (45.8%), and O014 (13.9%), respectively. Such genotypes should be tested for stability of the phenotype and estimation of their yield potential before releasing them as new germplasm. Genotype S50 of pop S showed striking improvement of 14.1% in MIC, 13.5% in LEN, 3.8% in UNIF, 20.1% in STR, 16.1% in ELON and 23% compared to TXA but had significant lower Lint %. Such line could be directly used in a breeding program to develop fine, long and strong fiber. Genotype S32 showed improvement of 12.6% for MIC, 12.7% in LEN, 2.8% in UNIF, 17.4% in STR, 11.8% in ELON and 21.8% in SFC and no difference in Lint % with respect to TXA. Similarly, S46 showed improvement of 15.9% for MIC, 12.1% in LEN, 2.3% in UNIF, 13.2% in STR and 18% in SFC and no difference in ELON and Lint % with respect to TXA. Such lines should be tested for yield trial and could be release as germplasm with improvement in multiple fiber traits.

Each of the two studies here provide new evidence in support of the recent hypothesis that development of the cotton fiber, a complex organ that is the longest single cell known in the plant kingdom, involves groups of closely-spaced genes that are functionally diverse but coordinately regulated (PATERSON *et al.* 2012).