Genetic analysis of the transition from wild to domesticated cotton (*G. hirsutum*)

Corrinne E. Grover[1], Mi-Jeong Yoo[1,a], Michael A. Gore[2], David B. Harker[3,b], Robert L. Byers[3], Alexander E. Lipka[4], Guanjing Hu[1], Daojun Yuan[1], Justin L. Conover[1], Joshua A. Udall[3,c], Andrew H. Paterson[5], and Jonathan F. Wendel[1]

[1]Department of Ecology, Evolution and Organismal Biology, Iowa State University, Ames, IA 50011
[2]Plant Breeding and Genetics Section, School of Integrative Plant Science, Cornell University, Ithaca, NY 14853
[3]Plant and Wildlife Science Department, Brigham Young University, Provo, UT 84602
[4]Department of Crop Sciences, University of Illinois, Urbana, IL 61801
[5]Plant Genome Mapping Laboratory, University of Georgia, Athens, GA 30602
[a]current address: Department of Biology, University of Florida, Gainesville, FL 32611
[b]current address: Department of Dermatology, The University of Texas Southwestern Medical Center, Dallas, TX 75390
[c]current address: Southern Plains Agricultural Research Center, USDA-ARS, College Station, TX, 77845-4988

Corresponding Author: Jonathan F. Wendel (jfw@iastate.edu)

Data Availability: All data are available via fig**share** (https://figshare.com/projects/Genetic_analysis_of_the_transition_from_wild_to_domesticated_cotton_G_hirsutum_QTL/62693) and GitHub (https://github.com/Wendellab/QTL_TxMx).

Keywords: QTL, domestication, *Gossypium hirsutum*, cotton

Summary: An F_2 population between truly wild and domesticated cotton was used to identify QTL associated with selection under domestication. Multiple traits characterizing the domesticated cotton fiber were evaluated, and candidate genes are described for all traits. QTL are unevenly distributed between subgenomes of the domesticated polyploid, with many fiber QTL located on the genome derived from the D parent, which has no spinnable fiber, but a majority of QTL on the other subgenome. QTL are many (68) and environmentally labile. Together, with candidate gene analyses, the results suggest recruitment of many environmentally responsive factors during cotton domestication.

Abstract

The evolution and domestication of cotton is of great interest from both economic and evolutionary standpoints. Although many genetic and genomic resources have been generated for cotton, the genetic underpinnings of the transition from wild to domesticated cotton remain poorly known. Here we generated an intraspecific QTL mapping population specifically targeting the domesticated cotton fiber phenotype. We used 465 F_2 individuals derived from an intraspecific cross between the wild *Gossypium hirsutum* var. *yucatanense* (TX2094) and the elite cultivar *G. hirsutum* cv. Acala Maxxa, in two environments, to identify 68 QTL associated with phenotypic changes under domestication. These QTL average approximately 46 Mbp in size, and together represent 29% (647 Mbp) of the 2,260 Mbp genome. Although over 70% of QTL were recovered from the A-subgenome, many key fiber QTL were detected in the D-subgenome, which was derived from a species with unspinnable fiber. We found that many QTL are environmentally labile, with only 41% shared between the two environments, indicating that QTL associated with *G. hirsutum* domestication are genomically clustered but environmentally labile. Possible candidate genes were recovered and discussed in the context of the phenotype. We found some support for the previously noted biased recruitment under domestication of factors from one of the two co-resident genomes of allopolyploid cotton. We conclude that the evolutionary forces that shape intraspecific divergence and domestication in cotton are complex, and that phenotypic transformations likely involved multiple interacting and environmentally responsive factors.

Introduction

The cotton genus (*Gossypium*) represents the largest source of natural textile fiber worldwide. Although four species of cotton were independently domesticated, upland cotton (*G. hirsutum* L.) accounts for more than 90% of global cotton production. Native to the northern coast of the Yucatan peninsula in Mexico, *G. hirsutum* is now widely cultivated across the globe (Wendel and Albert 1992). Domestication of *G. hirsutum* occurred circa 5,000 years ago, producing many phenotypic changes common to plant domestication, including decreased plant stature, earlier flowering, and loss of seed dormancy. An additional primary target unique to cotton domestication was the single-celled epidermal trichomes (i.e., fibers) that cover the cotton seed. Cotton fiber morphology varies greatly in length, color, strength, and density among the myriad accessions that span the wild-to-domesticate continuum. As a species, *G. hirsutum* is highly diverse, both morphologically and ecologically, and has a correspondingly long and complex taxonomic history (Paul A. Fryxell 1979, 1976, 1968; P. A. Fryxell 1992) that includes the modern, cryptic inclusion of at least two distinct species (Wendel and Grover 2015; Gallagher et al.

2017). Truly wild forms of *G. hirsutum* (race *yucatanense*) occur as scattered populations in coastal regions of the semiarid tropical and subtropical zones of the Caribbean, northern South America, and Mesoamerica (Coppens d'Eeckenbrugge and Lacape 2014). These are distinguished from domesticated and feral forms by their short, coarse, brown fibers, as well as their sprawling growth habit, photoperiod sensitivity, and seed dormancy requirements, among others (Figure 1). Results from molecular marker analyses, including allozymes (Wendel and Albert 1992), restriction fragment length polymorphisms (RFLP) (Brubaker and Wendel 1994), simple sequence repeats (SSR) (Liu and Wendel 2002; Y. Zhang et al. 2011; Tyagi et al. 2014; Zhao et al. 2015; McCarty et al. 2018; Kaur, Tyagi, and Kuraparthy 2017), SNP arrays (Ai et al. 2017; Hinze et al. 2017; Cai et al. 2017), and next-generation sequencing (Ma et al. 2018; Fang, Wang, et al. 2017; Reddy et al. 2017) have quantified genetic diversity and aspects of population structure among wild, feral, and domesticated stocks of the species, as well as the allopolyploid origin of the species. Notably, the allopolyploid origin of *G. hirsutum* includes a diploid species with no spinnable fiber. That is, while the maternal progenitor of the allopolyploid lineage is derived from the African "A-genome" whose two extant species have been independently domesticated for fiber production, the paternal parent is derived from the fiberless Mesoamerican "D-genome" clade.

Recent advances have improved our understanding of the genetic changes targeted by humans during the several millennia of cotton domestication and improvement by evaluating gene expression differences that distinguish wild and domesticated cotton fiber, either globally or for a few key genes among accessions (Haigler et al. 2009; Bao et al. 2011; Tuttle et al. 2015; Argiriou et al. 2012; Kim et al. 2012). Genome-scale surveys have elucidated many of the genes that are differentially expressed between wild and domesticated cotton (Hovav, Udall, Chaudhary, et al. 2008; Chaudhary et al. 2009; Rapp et al. 2010; Nigam et al. 2014; Yoo and Wendel 2014), or among developmental stages of fiber development (Shi et al. 2006; Gou et al. 2007; Taliercio and Boykin 2007; Hovav, Udall, Chaudhary, et al. 2008; Hovav, Udall, Hovav, et al. 2008; Al-Ghazi et al. 2009; Rapp et al. 2010; Q. Q. Wang et al. 2010; Nigam et al. 2014; Yoo and Wendel 2014; Tuttle et al. 2015). These many studies indicate that domestication has dramatically altered the transcriptome of cotton fiber development, but to date the specific upstream variants and interacting partners responsible for these downstream developmental differences remain to be discovered.

From a genetic perspective, multiple independent quantitative trait loci (QTL) analyses have been performed to identify chromosomal regions contributing to phenotypic variation among various cotton genotypes. Most QTL analyses to date have focused either on crosses between modern cultivars of *G. hirsutum* or on crosses between cultivated forms of *G. hirsutum* with *G. barbadense*, another cultivated species which possesses superior fiber quality but with the limitations of lower yield and a narrower range of adaptation (Y. Hu et al. 2019; Chandnani et al. 2017; Fang, Wang, et al. 2017). Interspecific cotton crosses often generate negative genetic correlations between fiber quality and lint yield and suffer from F_2 breakdown (reviewed in (J. Zhang, Percy, and McCarty 2014)). Taken together, these numerous studies have reported more than 2,274 QTL (Said, Knapka, et al. 2015) pertaining to agronomically and economically important traits (e.g., plant architecture; biotic and abiotic stress resistance; fiber, boll, and seed quality and productivity). Several meta-analyses have attempted to identify possible QTL clusters and hotspots by uniting these QTL studies through a consensus map (Rong et al. 2007; Lacape et al. 2010; Said, Knapka, et al. 2015; Said, Song, et al. 2015); QTL cluster denotes a genomic region containing many QTL of sometime variable traits, whereas a QTL hotspots are clusters of QTL for a single trait (Said, Song, et al. 2015). These meta-analyses compiled QTL studies of both intraspecific *G. hirsutum* populations and interspecific *G. hirsutum* × *G. barbadense* populations, ultimately creating a QTL database from intraspecific and interspecific populations (Said, Knapka, et al. 2015). To date, QTL analyses have yielded multiple, sometimes conflicting, insights that are accession or environmentally dependent. Some aspects of fiber development, for example, are associated with QTL enrichment in the D-subgenome of polyploid cotton (C. Jiang et al. 1998; Rong et al. 2007; Lacape et al. 2005; Han et al. 2006; Qin et al. 2008; Said, Song, et al. 2015), which derives from a fiberless parent, but not all mapping populations reflect this bias (Lacape et al. 2010; Ulloa et al. 2005; C. Li et al. 2013). Likewise, QTL found in some environments and/or populations are not significant in similar, but non-identical, environments and in other mapping populations (Said, Knapka, et al. 2015; Said, Song, et al. 2015; Lacape et al. 2010). Some research does suggest that cotton fiber QTL are genomically clustered, yet with heterogeneous phenotypic effects (Rong et al. 2007; Lacape et al. 2010; Qin et al. 2008). Said et al. (Said et al. 2013; Said, Song, et al. 2015) showed that just as QTL clusters and hotspots exist for fiber quality, they also exist for other traits (e.g., yield, seed quality, leaf morphology, disease resistance), and these hotspots, while found on every chromosome, tend to concentrate in specific regions of the genome. In particular, comparisons between intraspecific and interspecific populations reveal common QTL clusters and hotspots, possibly indicative of shared genetic architecture among cultivars and between species (Said, Song, et al. 2015). While these QTL analyses have increased our understanding of the number and location of chromosomal regions that contribute to differences between cultivars and species, there remains a significant gap in our understanding of genes targeted during the initial domestication of cotton and their effects, which ultimately led to the development of modern cultivars.

Here we provide an evolutionary quantitative genetics perspective on the domestication of the dominant cultivated cotton species, *G. hirsutum*, through identification and characterization of QTL for traits that have played important roles during domestication. In contrast to previous studies, we utilize an *intraspecific* cross between a truly wild form of *G. hirsutum* (var. *yucatanense*, accession TX2094) and an elite cultivar (*G. hirsutum* cv. Acala Maxxa), to bracket the “before” and “after” phenotypic characteristics of the domestication process that played out over the last 5,000 years or so. Numerous domestication-related traits were characterized in both the parents and their segregating progeny in two environments, representing characters from several broader phenotypic categories: (1) plant architecture, (2) fruiting habit, (3) phenology, (4) flower, (5) seed, (6) fiber-length, (7) fiber quality, and (8) fiber color. We generated a SNP-based genetic linkage map to anchor each QTL to the *G. hirsutum* cotton reference genome (elite accession TM1; (Saski et al. 2017)) and identify plausible candidate genes for each trait. We show that the QTL associated with *G. hirsutum* domestication are both clustered and environmentally labile. Possible candidate genes were recovered and discussed for each trait. This study provides valuable insights into the genetic basis of cotton domestication and provides information that will assist in identifying cotton domestication genes and their functional effects on cotton biology.

Materials and Methods

Plant Materials and Phenotyping

A total of 465 F_2 individuals were derived from a cross between *Gossypium hirsutum* var. *yucatanense* accession TX2094 as the maternal parent (USDA GRIN accession PI 501501, collected by J. McD. Stewart) and the modern elite cultivar *G. hirsutum* cv. Acala Maxxa as the paternal parent. The *G. hirsutum* var. *yucatanense* accession was previously identified as being truly wild using both allozyme (Wendel and Albert 1992) and RFLP analysis (Brubaker and Wendel 1994), as well as by morphological evidence. To allow for the replication of alleles over time and space, these individuals were grown as two subpopulations (October 2009 to July 2010), with 231 plants located in a greenhouse at Iowa State University (Ames, Iowa), and the remaining 234 in a greenhouse at the U. S. Arid-Land Agricultural Research Center (Maricopa, Arizona); nine representatives of each parental accession were also grown in each greenhouse. At Iowa State, individual seeds were separately planted in 7.6 L (two gallon) containers containing 15:7:3:3 soil:sand:peat:perlite. Plants were grown under natural sunlight (10-11 hours of daylight) with daytime and nighttime temperatures of 25±2 and 20±2°C, respectively. Plants were fertilized twice a week with 125 ppm Nitrogen. In Maricopa, individual seeds were separately planted into 18.9 L (five gallon) pots containing moistened Sunshine Mix #1 (Sun Gro Horticulture Inc., Bellevue, WA) and perlite (4:1 ratio). Plants were grown under natural sunlight in a greenhouse with daytime and nighttime temperatures at 30±2 and 22±2°C, respectively. All Maricopa plants were fertilized every two-weeks with 20–20–20 (200 ppm N) Peters Professional plant nutrient solution. These two populations were subsequently evaluated for multiple traits in each of the following eight categories: (1) plant architecture, (2) fruiting habit, (3) phenology, (4) flower, (5) seed, (6) fiber length, (7) fiber quality, and (8) fiber color (Table 1). Traits were selected to cover the range of possible domestication phenotypes.

At 150 (±7) days after planting, 12 plant architecture traits were evaluated, which include plant height, fruiting branch length, branch angle, stem pubescence, and branch patterning (Table 1). Data were collected for branch angles at the intersection of 1st, 3rd and 5th sympodia (secondary axes) with the main stem; however, due to high variation in the data observed from the 1st and 3rd sympodia, only data from the 5th sympodium was considered further. In addition, the first node having a branch with red coloring was recorded in the Ames population only (Table 1). Stem pubescence was scored independently by two people using the five-grade (1–5) scale developed by Lee (1968) (J. A. Lee 1968), where 1 is fully pubescent; the average of the two scores was recorded.

Four categories of traits relating to phenology, flowering, and fruiting were also examined. Eleven phenological traits (Table 1) were recorded, and, for consistency between the two greenhouse populations, we hand-pollinated flowers for 30 days following the emergence of the first flower. Four floral traits were examined, including pollen color, the presence or absence of petal spot, average stigma distance, and the presence or absence of curly styles. For pollen color, there exists a gradient of color from cream to yellow; however, we restricted our classifications to the parental color codes, i.e., “cream” vs. “yellow” observed in Maxxa and TX2094, respectively. Upon maturation, several traits related to boll/seed development were also measured on harvested bolls, such as number of mature seeds, fuzzy seed weight, and average seeded cotton weight (Table 1).

Finally, 358 fiber samples harvested from the 465 F_2 plants were collected and sent to the Cotton Incorporated Textile Services Laboratory (Cotton Incorporated, Cary, NC) for analysis by the AFIS Pro system (Uster Technologies, Charlotte, NC), an industry standard for evaluating fiber length and other quality traits (Table 1). Fiber color was determined by a MiniScan XE Plus colorimeter (ver. 6.4, Hunter Associates Laboratory, Inc., Reston, VA), which measures color properties of L^*, a^*, and b^*. L^* is a lightness component, ranging from 0 to 100 (from dark to bright), and a^* (from green to red) and b^* (from blue to yellow) are chromatic components ranging from -120 to 120 (Yam and Papadakis 2004). Values were measured three times on the same fiber sample and averaged for each trait (i.e., mean L^*, mean a^*, and mean b^*).

Genotyping and Genetic Map Construction

A total of 384 KASPar-based SNP assays (277 co-dominant) were used to genotype 483 F_2 plants and parents (KBioscience Ltd., Hoddesdon, UK). SNP assays were designed as previously reported for *G. hirsutum* (Byers et al. 2012). Genomic DNA was extracted from leaf tissue using the Qiagen DNeasy

Plant Mini Kit (Qiagen, Stanford, CA, USA) and normalized to an approximate concentration of 60 ng/µL.

Specific Target Amplification (STA) PCR was used to pre-amplify the target region of genomic DNA containing the SNPs of interest, but without the discriminating SNP base in the primer sequence. The PCR conditions for this protocol included a 15-min denaturing period at 95°C followed by 14 two-step cycles: 15 s at 95°C followed by 4 min at 60°C. This effectively increased the concentration of the target DNA relative to the remaining DNA. The sample amplicons produced by the STA protocol were then genotyped using the Fluidigm 96.96 Dynamic Arrays genotyping EP1 System (San Francisco, CA). Each Fluidigm plate run included eight control samples: two Acala Maxxa, two TX2094, two pooled parental DNA (synthetic heterozygotes), and two no-template controls (NTC). These controls served as guideposts during the genotyping process. The STA amplicons and the SNP assays were loaded onto a Fluidigm 96.96 chip, where a touchdown PCR protocol on the Fluidigm FC1 thermal cycler (San Francisco, CA, USA) was used to allow the competing KASPar primers to amplify the appropriate SNP allele in each sample.

Fluorescence intensity for each sample was measured with the EP1 reader (Fluidigm Corp, San Francisco, CA) and plotted on two axes. Some assays required more amplification in order to produce distinct clusters. For those that did not form distinct clusters during the initial analysis, an additional five cycles of PCR were performed on the plate and fluorescence intensity measured again until all assays produced sufficient resolution for cluster calling. Genotypic calls based on EP1 measurements were made using the Fluidigm SNP Genotyping Analysis program (Fluidigm 2011). All genotype calls were manually checked for accuracy and ambiguous data points that either failed to amplify and/or cluster near parental controls were scored as missing data. The final raw output for an individual chip included data from each of the multiple scans performed to ensure that the optimal amplification conditions for each assay was represented. The text output from genotyping was arranged to a compatible format for genetic mapping using Excel. Files are available at https://github.com/Wendellab/QTL_TxMx.

A genetic map based on the KASPar genotyping data was constructed using regression mapping as implemented in JoinMap4 (VAN Ooijen 2011). A LOD threshold of 5.0 was used and linkage distances were corrected with the Kosambi mapping function. Loci were excluded from the map if they failed to meet a Chi-Square test ($\alpha = 0.05$) for expected Mendelian ratios.

QTL analysis

For each location, the raw phenotypic values of each trait were evaluated for statistical outliers in SAS version 9.3 (SAS Institute 2012) by examination of Studentized deleted residuals (Kutner et al. 2004), which were obtained from a simple linear model fitted with fixed effects for the grand mean and a single randomly sampled, representative SNP marker. Next, the Box-Cox power transformation (Box and Cox 1964) procedure was used on the outlier-screened phenotypic values with the previous simple model to find the most appropriate transformation that corrected for unequal variances and non-normality of error terms. To identify the optimal lambda value for each trait, this procedure tested lambda values that ranged from −2 to +2 with increments of 0.05 using PROC TRANSREG in SAS version 9.3 (SAS Institute). QTL were detected with QTL Cartographer V1.17 (Basten, Weir, and Zeng 2005) using the composite interval mapping (CIM) method (Zeng 1993, 1994) with a window size of 10 cM and a 2 cM walk speed. The LOD thresholds used to identify QTL were determined using a permutation test (1000 repetitions, α= 0.05) (Churchill and Doerge 1994), and the confidence intervals were set as the map interval corresponding to one-LOD interval on either side of the LOD peak (Mangin, Goffinet, and Rebaï 1994). If the QTL were separated by a minimum distance of 20 cM, they were considered two different QTL (Ungerer et al. 2002). The linkage map showing the location of QTL (Figure 2) was generated by MapChart 2.2 (Voorrips 2002) and colorized in Adobe Photoshop Creative Suite 5 (Adobe). QTL nomenclature follows a method used in rice (McCouch et al. 1997), which starts with “q”, followed by an abbreviation of the trait name.

Candidate gene searches

Linkage groups were assigned to *G. hirsutum* chromosomes (Table 2) using molecular marker sequences as gmap (Wu and Nacu 2010; Wu and Watanabe 2005) queries against the published *G. hirsutum* cv TM-1 (Saski et al. 2017) genome (annotation gff version 1.1), using default values and permitting two possible paths (to accommodate homoeologs). A consensus of markers was used to identify the candidate chromosome for each linkage group, using the highest scoring path for each marker; however, when both paths were equally likely, both were used to derive the consensus. Candidate genes contained within the QTL confidence interval were identified by using the genomic coordinates of the first and last marker for each linkage group as a boundary, and subsequently intersecting the genomic boundaries of each linkage group with the genome annotation via bedtools 2 (Quinlan and Hall 2010). All scripts and parameters are available at https://github.com/Wendellab/QTL_TxMx.

Candidate genes were further screened for previously established expression differences in developing fibers (Bao, Hu et al, in prep), for putative transcription factors (Saski et al. 2017), and for non-silent SNPs between the parental accessions. For the latter, reads derived from *G. hirsutum* Maxxa (SRA:SRR617482) and *G. hirsutum* TX2094 (SRA:SRR3560138-3560140) were mapped against the TM-1 genome (Saski et al. 2017) and SNPs were annotated using the Best Practices pipeline of GATK (Van der Auwera et al. 2013). The resulting vcf files were processed with vcftools (Danecek et al. 2011) and SnpSift (Cingolani, Patel, et al. 2012) to (1) only recover sites with differences between *G. hirsutum* Maxxa and *G. hirsutum* TX2094, (2) remove sites with missing data, and (3) only recover SNPs where the wild *G. hirsutum* TX2094 shared the ancestral SNP with an outgroup species, *G. mustelinum* (SRA SRR6334743). The resulting 3.6 million SNPs were annotated with SnpEff (Cingolani, Platts, et al. 2012) for the putative effects of each change, and SnpSift was again used to restrict the final vcf to only those SNPs where an effect was annotated. The final set of genes with annotated effects was further limited to only those regions under a QTL. These genes were additionally classified (Table S1) as to whether they also: (1) exhibit differential expression; (2) are putative TFs; or (3) belong to a curated list of potentially fiber-relevant cotton genes, based on existing literature (*Cotton Fiber: Physics, Chemistry and Biology* 2018). The genes that fell within ± 5% QTL distance from the QTL peak were prioritized, where the genomic distance to the peak was calculated as: genomic location of the QTL peak ± (0.05 * genomic range of the QTL). The QTL peak was placed on the genome sequence by using the genomic QTL boundaries (determined above) to relate the number of cM to the amount of sequence in that same region (in base pairs). The +/- 5% range All program run information and relevant parameters are available at https://github.com/Wendellab/QTL_TxMx.

All data and scripts are available via GitHub (https://github.com/Wendellab/QTL_TxMx). Supplemental files are available at FigShare (https://figshare.com/projects/Genetic_analysis_of_the_transition_from_wild_to_domesticated_cotton_G_hirsutum_QTL/62693) and on GitHub. All other data, e.g., genomes and downloaded sequences are listed in the methods. Seed from the mapping population is available from the GRIN National Genetic Resources Program.

Results

Phenotypic variation

All of the traits investigated (Table 1) exhibited phenotypic variability between two parents, TX2094 and Maxxa, except Fruiting Branch Length of the 5th branch, Branch Angle of the 5th sympodium, and average number of flowers during a 30-day period (P-value < 0.05; Supplementary Table 1). In general, the phenotypes reflected the expected “domestication syndrome” in Maxxa, as represented by its: (1) reduced plant height; (2) fewer total nodes; (3) fewer nodes to first fruiting branch; (4) better fruiting habit (e.g., longer fruiting branches); (5) early flowering; (6) greater production of flowers, bolls, and seeds; and (7) enhanced fiber quantity and quality (Supplementary Table 1).

The F_2 plants displayed a wide range of phenotypic variability in two greenhouse environments, Ames and Maricopa. The northern latitude of Ames contributed to variability for traits reflective of a cooler, less-sunny environment compared to the F_2 plants grown in Maricopa, Arizona. That is, plants grown in Ames typically were taller, with shorter fruiting branch lengths and a greater number of nodes; however, these plants also exhibited a greater number of nodes to first fruiting branch, as well as a higher ratio of non-fruiting to fruiting branches. Interestingly, the Ames population also exhibited both later flowering and more flowers during a 30-day interval. The flowers themselves exhibited greater distance between stigma and style, and produced more seeds per boll with an overall lighter seed weight (per boll), indicative of smaller seed size. Other flower and fiber traits exhibited continuous variation in all the F_2 plants, from TX2094-like to Maxxa-like phenotypes; however, the two subpopulations were often statistically distinguishable. For example, 50 Fuzzy Seed weight (g) was 3.96 and 4.13 in Ames and Maricopa, respectively, which is significantly different ($\alpha = 0.05$). Observations such as these are unexpected under the null hypothesis that subpopulations should not be phenotypically distinct, and they likely reflect an interaction with the environment. Phenotypic measurements for parents and progeny are found in Supplemental Table 1).

Linkage map construction

KASPar-based SNP genotyping was used to construct a single genetic linkage map (total genetic length = 1952.1 cM) from the *G. hirsutum* F_2 populations using JoinMap (Stam 1993) to merge individual population maps. Of the 384 markers used for genotyping, 357 were successfully mapped to create 32 linkage groups (Table 2). Among those 384 originally targeted markers, 84 markers were homoeolog-specific by design (see Byers et al. 2012). To determine whether the homoeologous genome of these markers was specific and accurately identified, linkage groups with multiple homoeolog-diagnostic SNPs were examined for genome consensus. Seventy (83%) of the 84 assays that resided in linkage groups with at least one other homoeologous assay agreed with the consensus for the target genome. Twenty-three of the 32 linkage groups contained two or more homoeologous SNP assays, 21 (91%) of which were in complete agreement with their genome identification. The two incongruent linkage groups both consisted of only two homoeologous SNP assays. Thus, of the 32 linkage groups in the map, 29 (91%) of these can be putatively assigned a genome based on these predictive SNP assays. These assignments suggest that 16 linkage groups (#2, 3, 6, 8, 11, 12, 13, 18, 19, 20, 24, 25, 26, 29, 30, and 32) are representative of the D_T subgenome whereas 14 (#1, 4, 5, 9, 10, 14, 15, 16, 17, 21, 22, 23, 28, and 31) are representative of the A_T subgenome (here, the subscript $_T$ denotes “from tetraploid”). All map-based subgenome identifications agreed with the consensus for candidate gene/chromosome identification (see below). These linkage groups cover all 26 chromosomes in *G. hirsutum* genome (Table 2).

Identification of QTL and QTL clusters

A total of 68 QTL were detected from marker-trait analysis of the two independent populations (Figure 2). The QTL detected from the independent populations represented all phenotypic categories (33 QTL for 19 traits in the Ames population; 35 QTL for 15 traits in the Maricopa population). These QTL map to 15 linkage groups (14 chromosomes); 48 QTL mapped onto 8 chromosomes of A_T subgenome, while 20 QTL mapped onto 6 chromosomes of D_T subgenome (Supplementary Table 2). In general, these *G. hirsutum* chromosomes carry an average and median of 5 and 3.5 QTL respectively; however, four chromosomes (A09, A10, D08, D09) have only a single QTL each and one (A06) includes a remarkable 20 QTL (19 overlapping QTL, 63% of which are for fiber color; Figure 2). If QTL for the same trait class are merged, these 19 overlapping QTL on chromosome A06 reduce to a total of 5-8 distinct QTL, which is still above average (Figure 2). The top 45 QTL ($R^2 > 7\%$) are summarized in Table 3. A full listing of identified QTL, map and genomic information, and other relevant information is included in Supplementary Table 2 and is discussed in the context of phenotype (see below).

Candidate Gene identification

A total of 13,274 genes are predicted within the genomic range of the 68 QTL (Supplementary Table 2), representing approximately 20% of the predicted gene models for the *G. hirsutum* cv. TM1 genome (Saski et al. 2017). The genomic regions occupied by QTL average approximately 46 Mbp in size (median = 36 Mbp), for a total genomic length of approximately 647 Mbp or 29% of the total sequenced genome length of 2,260 Mbp (Supplementary Table 3). For each phenotype (e.g., plant architecture, fiber color, etc), between 968 - 4,329 distinct genes were recovered. Candidate genes for each phenotype are discussed below.

We further screened the 13,274 candidate genes for (1) genes with non-silent mutations in the domesticated Maxxa (using the outgroup polyploid species *G. mustelinum* to infer the ancestral state), to filter for possible functional differences at the protein level; (2) genes with expression differences between Maxxa and TX2094, to filter for genes that have been up- or down-regulated under domestication; (3) transcription factors; or (4) known cotton fiber genes of interest (see methods for details) (Supplementary Table 4). In general, more genes were found within the QTL boundaries for the A subgenome (9708 versus 4016 in D_T), in congruence with the proportion of the A subgenome covered by QTL (approximately 512 Mbp in A_T versus 136 in D_T).

From the genome-wide total of 34,870 genes that have one or more SNP between TX2094 and Maxxa, 87% (30,337 genes) are affected by at least one putatively non-silent mutation. Over half of these genes have SNPs that change the amino acid (19,195 genes), and slightly more than half have changes in the untranslated regions (UTR; 19,829) in an approximately 3:5 ratio favoring mutations in the 5′ UTR. These are slightly greater than the number of genes that have silent SNPs (39%; 13,579 genes). Only

2.6% of genes have a SNP that changes the start or stop (in an approximate 2:3 ratio, start:stop). Genome-wide, there exists no bias toward the A or D subgenome for any of the above categories. Of those 30,337 genes with non-silent TX2094 versus Maxxa SNP, 18% (6,134 genes) fall within a QTL in a ratio of approximately 2.5 A_T:1 D_T (4,444 genes in A_T versus 1,690 in D_T). This ratio is approximately equivalent to the overall representation of the genome under QTL, i.e., 3A_T:1D_T. Of the 6,134 genes with a non-silent SNP that occur under the QTL, 62% (3,807 genes) have predicted amino acid changes between TX2094 and Maxxa (2,761 A_T genes and 1,046 D_T) that could potentially be visible to selection (Table 4).

To further explore the candidate genes under the QTL, we also quantified the number of genes under QTL that exhibit differential expression (DGE) during fiber development (Bao, Hu, et al. in prep). Of the 5,168 genes differentially expressed between TX2094 and Maxxa (in either 10 or 20 dpa fiber; adjusted P-value < 0.005), approximately 20% (1,063 genes) are located under one of the QTL (Table 4). Between 7-11% of genes for each phenotypic group experienced DGE in the fiber stages surveyed (10 and 20 dpa). Interestingly, there appears to be little bias toward differential expression of genes under fiber-related QTL versus non-fiber QTL for these fiber-derived expression data. Differentially expressed genes that also contain nonsynonymous and/or UTR SNPs, account for about half of the DGE-QTL genes (553 genes), 350 of which have predicted amino acid changes.

Finally, we also considered two categories of genes of possible interest under the QTL: transcription factors (TF) and previously identified fiber-relevant genes (see methods). Seventy-nine putative TF (Saski et al. 2017) were predicted in the QTL regions (55A:24D), representing approximately 1% of the genes related to each trait. Of these 79 TF, 19 had putative amino acid changes. A single transcription factor under the QTL exhibited expression changes, i.e., Gohir.A08G064100, which is located under the fiber color QTL qCa-LG21-1M and qCL-LG21-2M (Supplementary Table 2). We also screened the QTL genes for a compilation of 703 genes mined from the fiber biology literature (see methods). Of these, 99 genes (~14%) were found under one or more QTL. Less than 1% of each phenotypic category was composed of genes derived from this list, aside from the fiber quality QTL, which had slightly more than 1% of genes derived from this list (1.3%). This slight bias toward the fiber quality QTL is expected, as the list of mined genes was initially biased toward fiber quality.

Plant architecture

Five QTL were detected for 3 of 12 traits related to plant architecture, 4 of which reside on chromosome A06 alone. These four include one QTL for fruiting branch length (FB2; Maricopa), and three for stem pubescence (SP; 2 from Ames, 1 from Maricopa). Particularly notable were the SP QTL located on ChrA06 (LG04), which explained 22.0-52.7% of the SP phenotypic variation. One QTL for plant height

(PH) was detected in the D_T-subgenome (D07; LG19) in Maricopa population, which explained 9.4% of the phenotypic variation (R^2) and showed additivity. For PH, the TX2094 allele contributes to increasing height, although the two parental alleles work additively (Table 3; Supplementary Table 2).

Homology search of markers associated with these QTL identified 2,078 candidate genes in the QTL regions for plant architecture (Supplementary Table 4) , 368 of which fell in a ± 5% window around the QTL peak (Supplementary Table 5) . For plant height (PH), the single QTL had 107 candidate genes in the ± 5% window surrounding the QTL peak, some of which have clear relevance to plant growth. These include, among others: Gohir.D07G161300, a phototropic-responsive NPH3 family protein (Christie et al. 2018); Gohir.D07G160100, a YUC8-like gene (Hentrich, Sánchez-Parra, et al. 2013); Gohir.D07G166500, an auxin-responsive family protein (Gallavotti 2013); and the tandem duplicates, Gohir.D07G167500 and Gohir.D07G167600, both of which encode ptuative far-red impaired responsive (FAR1) family proteins (W. Tang et al. 2013). Eleven of the genes contained within this narrowed window exhibit differential expression between TX2094 and Maxxa, including a QUASIMODO-like homolog (Gohir.D07G164900), which leads to a dwarf plant phenotype in Arabidopsis (Orfila et al. 2005).

Fruiting branch length (FB) also had a single QTL, under which 22 genes were predicted in the ± 5% window, many of which contain predicted SNP effects, either changing the amino acid or other non-silent mutations. A single gene experienced differential expression in the window, Gohir.A06G200200, a hydrolase-like protein; however, the relevance of this gene, or others in the narrowed window, to fruiting branch length is unclear.

For stem pubescence, 239 genes were found in the ±5% windows surround the three QTL associated with stem pubescence. Gohir.A06G111500 encodes cellulose synthase 6, which is relevant in trichome development (Haigler et al. 2009; Betancur et al. 2010; Nixon et al. 2016); this gene was differentially expressed between the parents and may be important for this QTL. Also differentially expressed under the QTL is Gohir.A06G115200, related to the actin-like ATPase NHO1 (or GLI1) from *Arabidopsis thaliana*. The QTL also contains two trichome birefringence-like proteins, Gohir.A01G048200 and Gohir.A01G048300, the latter of which is differentially expressed; in *Arabidopsis,* members of this gene family are known to be involved in the synthesis and deposition of cellulose in secondary cell walls (Bischoff, Selbig, and Scheible 2010).

Fruiting habit and Phenology

Twelve QTL were detected for five traits related to fruiting habit and phenology (see Table 1), eleven of which reside in A_T subgenome. Four and three QTL were identified for Total Number of Nodes

(TN) and Plant Height-to-Total Number of Nodes Ratio (PH_by_TN), respectively, in the Ames and Maricopa populations. All QTL for PH_by_TN showed additivity, whereas only one exhibited additivity for TN (i.e., qTN-LG01-1A); the remaining three QTL exhibited partial- or over-dominance (Supplementary Table 2). Three QTL were identified for Total Number of Nodes at First Flower (TNFF) in the Ames population only, which exhibited either partial or over-dominance and explained 5.6-12.4% of the phenotypic variation. One QTL each was detected for Total Number of Non-Fruiting Branches (TNFB), and Number of Bolls (NB), again in Ames; both exhibited overdominance. For TNFB and NB, qTNFB-LG28-1A and qNB-LG13-1A explained 18.5% and 35.1% of the phenotypic variance, respectively. The QTL for NB was located in the D_T-subgenome and was the only QTL detected for phenology.

Homology search of QTL-associated markers recovered 3,266 candidate genes in the QTL intervals controlling fruiting habit and phenology. A single transcription factor, Gohir.D04G178400, was found in the single QTL for number of bolls (NB), as was a late embryogenesis abundant protein (Gohir.D04G180300); neither, however, exhibited DGE or non-silent SNPs. Similarly, the single QTL for TNFF revealed few genes exhibiting DGE or SNP changes (Supplementary Table 4).

Interestingly, the QTL for PH_by_TN included two profilin genes, Gohir.A06G068300 and Gohir.A06G068400. Members of the profilin gene family have been implicated in the parallel domestication events of all domesticated cotton species (Bao et al. 2011); however, the profilin members in these regions did not exhibit DGE in fibers or SNP differences. The three QTL regions for PH_by_TN also contains two genes showing homology to transcription factors (Gohir.A01G087000 and Gohir.A01G087100; similar to *Arabidopsis* WRKY and GRAS transcription factors, respectively), although neither is noted as a predicted transcription factor in the genome sequence (Saski et al. 2017). The QTL regions for PH_by_TN also contain at least eight cytochrome P450-like genes, a relatively large superfamily of genes with diverse metabolic roles (Mizutani 2012; Mizutani and Ohta 2010).

The QTL for the total number of nodes (TN) include two differentially expressed auxin efflux carrier family proteins, Gohir.A05G289500 and Gohir.A05G289600, and a CCR-related gene (Gohir.A05G297200); members of the latter gene family may be involved in lignin biosynthesis during development (Lauvergeat et al. 2001). Finally, the QTL also contains a differentially expressed SIS3-like homolog (Gohir.A01G143800); SIS3 is involved in the growth response to high concentrations of exogenous sugars (Y. Huang et al. 2010).

The genomic ranges for the number of nodes at first flower (NN) includes Gohir.A12G183300, a homolog of TEOSINTE BRANCHED1/CYCLOIDEA/PCF 15 (TCP15) near QTL qNN-LG01-1A (Supplementary Table 4). The homolog for this locus is known to be involved in determining plant height, internode length and leaf shape in *Arabidopsis* (Kieffer et al. 2011; Davière et al. 2014), maize

(Doebley, Stec, and Hubbard 1997; Studer and Doebley 2012), and tomato (Steiner et al. 2012). Another NN QTL, qNN-LG28-1A, is proximal to a homolog of SPL2 (Gohir.A01G154600), which is involved in shoot maturation and the transition to flowering (Shikata et al. 2009).

Flower

Eight QTL were identified for three floral traits, which individually explain 5.2-60.4% of the phenotypic variation and most of which exhibited varying degrees of dominance. Two QTL were detected for Average Stigma Distance (SD) in the Maricopa population only, one in each polyploid subgenome (Chr A05 and Chr D09). A single QTL was identified for Curly Style (CS) in the Ames population only, with the curly allele originating from TX2094. Five QTL were detected for Pollen Color (PC) on two A and one D chromosomes (A05, A10, and D04); presence of TX2094 alleles generated more yellow pollen (Supplementary Table S2).

Candidate gene search revealed 1,926 genes in the QTL intervals for floral traits. The single QTL for curly style (CS; qCS.LG01.1A) yielded no notable genes, which may indicate a need for further exploration or to expand the search to the entire QTL window. Indeed, expanding the window for candidate gene search revealed a DVL-homolog (Gohir.A12G170400). Members of the DVL gene family affect multiple organs in *Arabidopsis*, including conferring phenotypic changes in both fruit and inflorescence (Wen, Lease, and Walker 2004). The QTL for average stigma distance revealed no notable candidate genes. For pollen color, the candidate gene Gohir.A05G284900 is a MYB33-like homolog; in *Arabidopsis*, this gene is involved in anther development (Millar and Gubler 2005).

Seed

Six QTL were identified in the Maricopa population only for two of nine seed-related traits, i.e., Fuzzy Seed Weight (FSW) and Seed Weight (SW). Three QTL were identified for each trait, all of which exhibited either partial or over-dominance (5 and 1 QTL, respectively). Five of the QTL reside on A_T subgenome chromosomes (two on A01 and three on A07), and explain 3.6–10.4% of the phenotypic variation in seed-related traits. A single FSW QTL was located on chromosome D13.

QTL for these seed traits contained 3,306 candidate genes. For the seed weight QTL regions, these include Gohir.A07G189000 (UDP-D-glucose/UDP-D-galactose 4-epimerase 3) and several FASCICLIN-like arabinogalactans (FLA) including Gohir.A07G192300, a FLA2-like gene. Both Gohir.A07G189000 and Gohir.A07G192300 exhibit up-regulation in domesticated (versus wild) cottons (YOO AND WENDEL 2014) and have *Arabidopsis* homologs that function in cell wall biosynthesis. Also included in the ± 5% QTL region is a Pfifferling (PFI)-like homolog (Gohir.D13G119200), which functions in seed (embryo) development in *Arabidopsis* (Steinborn et al. 2002), and an expansion (EXPA5)-like homolog (Gohir.A07G205900), which may act to mediate cell wall expansion (Y. Lee, Choi, and Kende 2001; Shcherban et al. 1995).

Fiber length

Fiber-related characteristics were among the obvious phenotypic targets during domestication of cotton. Not surprisingly, therefore, 37 QTL were detected for fiber-related traits (i.e., length, color, and measures of quality), of which 15 were for fiber length. As observed in some other populations, a majority of these (73% or 11 QTL) were located in the subgenome (D_T) derived from the parental diploid (D genome) that has short, unspinnable fiber. These QTL were dispersed over 5 of the 13 D_T chromosomes and 2 of the 13 A_T-derived chromosomes, individually explaining from 4.7 to 12.9% of the phenotypic variation. Despite having far fewer QTL, the A_T-subgenome exhibited QTL for four of the six length traits evaluated, including L(n) CV [%], L(w) [in], L5% (n) [in], and UQL(w) [in]. None of the A_T-subgenome QTL were in the top 5 fiber-length related QTL, explaining at most 9.36% of the trait variation. Conversely, QTL for all six fiber length traits examined were found on D_T-subgenome chromosomes (Supplementary Table S2), and nearly half of these QTL individually explain over 10% of the phenotypic variation (R^2) for their categories (5 out of 11 D_T QTL representing 5 out of 6 length traits). Three of the top five QTL for length traits were detected on chromosome D07 and most (4) of the top five showed partial- or over-dominance. Interestingly, all three D07 QTL were from different length categories (L(n) CV [%], L(n) [in], L(w) CV [%]).

Candidate gene searches for fiber length QTL revealed several possibilities, including five cellulose synthase-like genes, all of which clustered on chromosome D11 (Gohir.D11G245500 - Gohir.D11G245900). The middle gene in the cluster, Gohir.D11G245700, exhibited both amino acid changes and differential gene expression between wild and domesticated *G. hirsutum*, supporting a possible role in fiber domestication. These D_{T}-derived cellulose synthase genes were associated with the QTL for UQL_(w)_[in], qUQL-LG06-1M, which is associated with “Upper Quantile Length by Weight” and explains 6% of the variation. Interestingly, an additional cellulose synthase-like gene (Gohir.A08G144300) was also differentially expressed between wild and domesticated cotton; however, this gene was not contained within a fiber length QTL, but was rather found associated with the fiber color QTL qCL-LG21-2M (mean L^*). Similarly, several genes typically associated with flavonoid production (e.g. flavanone 3-hydroxylase and two chalcone-flavanone isomerase) were found within the fiber length QTL for L(w)_[in] rather than the QTL for fiber color where they would be expected to influence the brown coloration found in wild fibers.

Additional candidate genes were found, including a single gene on A06 and several interspersed on D07. On chromosome A06, a beta tubulin-like gene (Gohir.A06G128600) is found within both the "Mean Length by Weight" QTL qLw-LG04-1M and the "5% Length by Number" QTL qL5n-LG04-1M. Beta tubulin genes are relevant to cell wall development because they orient the cellulose microfibrils (Spokevicius et al. 2007), a major component of secondary cell walls. Chromosome D07 also contains a beta tubulin-like gene (Gohir.D07G122700); however, this beta tubulin-like candidate is associated with the "Mean Length by Number" QTL qLn-LG19-1A. A second candidate for "Mean Length by Number" was also detected within QTL qLn-LG19-1A on D07. This candidate is a putative FASCICLIN(FLA12)-like protein (Gohir.D07G121700), whose *Arabidopsis* homolog affects cell wall architecture and composition, including cellulose content (MacMillan et al. 2010). Three additional candidates were found on D07, including two candidates on for the "Coefficient of Variation of the Length by Number" QTL qLnCV-LG19-1A. Namely, a RAB GTPase homolog (Gohir.D07G105100) (Lunn et al. 2013) and a dynamin(DL1)-like protein (Gohir.D07G118700) (Collings et al. 2008) were detected within this QTL region. In *Arabidopsis*, these genes influence the cell wall composition (both) and cellular expansion (DL1). Notably, the DL1-like candidate exhibits differential expression between wild and domesticated cotton fiber. A final candidate is found for the QTL qLwCV-LG19-1A (trait = "Coefficient of Variation of the Length by Weight"), Gohir.D07G152700, whose closest homolog is the YABBY1 transcription factor in *Arabidopsis* that is exclusively expressed in trichomes (Schliep et al. 2010). This candidate gene also exhibits an amino acid change between wild and domesticated cotton.

Fiber color

Fiber color is an important agronomic trait conferred by the accumulation of flavonoids in mature fibers (Tuttle et al. 2015; Xiao et al. 2014, 2007; Feng et al. 2013; T. Li et al. 2012; Hua et al. 2007). Nineteen QTL were detected for the three fiber color traits evaluated: mean L^* (bright/dark), mean a^* (green/red), and mean b^* (blue/yellow). Many of these overlapped and were therefore aggregated into two broad QTL on chromosomes A06 and A08, each encompassing more than 50 cM (52 and 57 cM, respectively). The QTL on chromosome A06 were typically of major effect, individually explaining from 24.3 to 64.5% of the phenotypic variation, whereas those on chromosome A08 typically explained less than 10% of the variation (from 3.3 to 12.0%). Five candidate genes were recovered from the QTL on chromosome A06, four of which are hydroxycinnamoyl-CoA shikimate/quinate hydroxycinnamoyl transferase (HCT)-like genes (i.e., Gohir.A06G192300, Gohir.A06G193100, Gohir.A06G193400, Gohir.A06G193500), whose *Arabidopsis thaliana* homologs mediate the balance of plant growth and flavonoid (i.e., flavonol and anthocyanin) biosynthesis (Besseau et al. 2007). Two flavin-binding monooxygenase family (YUCCA)-

like proteins were found within the color QTL detected here. The first YUCCA-like gene is the remaining candidate on A06,Gohir.A06G138000, and the second is one of the two candidates detected on chromosome A08 (i.e., Gohir.A08G192500). *Arabidopsis* homologs of the YUCCA family function in the production of auxin (Hentrich, Sánchez-Parra, et al. 2013; Hentrich, Böttcher, et al. 2013), a key regulator of plant development that may also be involved in the regulation of flavonol synthesis (Lewis et al. 2011). The remaining candidate from chromosome A08 encodes chalcone-flavanone isomerase family-like protein, which also functions in flavonoid biosynthesis in *Arabidopsis* (W. Jiang et al. 2015).

Interestingly, both chromosomes A06 and A08 have multiple loci that are predicted defense-related proteins (23 and 6 on A06 and A08, respectively) within the QTL (Supplementary Table 4). While not predicted to function in color, it is notable that (1) flavonoids can function in plant defense (Treutter 2006) and (2) genes conferring chemical defense compounds may cluster genomically (Takos and Rook 2012). Therefore, while these genes may not directly affect fiber color, selection for defense or for color may have had indirect effects on color. Likewise, the broad QTL on chromosome A08, while associated with color and not overlapping other fiber QTL, contains several genes that may influence fiber morphology. These include a cellulose synthase-like gene (Gohir.A08G144300), a myb 66-like gene (S. Wang et al. 2004) (Gohir.A08G168600), and a beta-6 tubulin-like gene (Ji et al. 2002) (Gohir.A08G182500). Cellulose synthase genes are commonly implicated in fiber morphogenesis, and the latter two are known to function in fiber initiation and elongation, respectively (Haigler et al. 2009; Betancur et al. 2010). The fiber color QTL on chromosome A06 also contains a gene known to influence fiber elongation (Shi et al. 2006) (Gohir.A06G152100). In this case, however, the QTL overlaps with other QTL associated with fiber length and stem pubescence, as would be expected for a gene known to influence trichome development.

Other fiber qualities

While a total of 14 “other” measures of fiber quality were evaluated (Table 1), only two traits related to short fiber content produced QTL. Both traits were located on chromosome D07, with “Short Fiber Content by Number” (SFC (n) [%]) and “Short Fiber Content by Weight” (SFC (w) [%]) revealing one and two QTL, respectively. Notably, one of the SFC (w) [%] QTL was entirely contained within the QTL for SFC (n) [%]. Several candidates affecting cell wall composition and synthesis were found within these two regions. These include two candidates located within 5% of a fiber length QTL peak, i.e., YABBY1 and FLA12 (see above), and several located within the complete fiber length QTL regions. Distinctly within the three “other fiber” QTL were two subtilisin protease-like candidates (i.e., Gohir.D07G151000 and Gohir.D07G159000), each associated with a different short fiber content QTL. Subtilisin proteases have been associated with cell wall composition in *Arabidopsis thaliana*, specifically the mucilage content of cell walls (Rautengarten et al. 2008). Another candidate, Gohir.D07G150700, is a galacturonosyltransferase 1 (GAUT1)-like gene, whose *Arabidopsis thaliana* homologs influence cell wall composition by controlling pectin biosynthesis (Atmodjo et al. 2011). Two actin-related genes were also found in the QTL region for "other" fiber qualities. Gohir.D07G154900 is an actin depolymerizing factor, which affects the elongation pattern of cotton fibers (Augustine et al. 2008; Q.-S. Huang et al. 2014), and Gohir.D07G159100 is an exostosin (MUR3)-like gene, whose *Arabidopsis* orthologs are involved in actin organization and the synthesis of cell wall components (Reiter, Chapple, and Somerville 1997).

Comparison of putative QTL between subpopulations, between subgenomes, and among chromosomes

The F_2 seed derived from a single cross between *G. hirsutum* accessions TX2094 and Maxxa were planted in two different greenhouse environments, in Maricopa, AZ and Ames, IA (see methods). The 68 total QTL detected were nearly evenly divided between the two subpopulations, with Maricopa recovering two more QTL than Ames (35 versus 33, respectively). While the number of QTL recovered in each subpopulation was similar, less than half (14 QTL, or 41%) were shared between the two locations. Likewise, while both populations detected QTL on a similar number of chromosomes (11 and 10 in Maricopa and Ames, respectively), approximately one-third of those chromosomes had QTL from only one population. On average, the QTL detected in Ames had a more narrow range, both overall (11.98 versus 23.14 cM) and when only considering shared QTL regions (18.95 versus 28.33 cM). Slight and opposing subgenome biases were found for the chromosomes recovered from each subpopulation, with Ames recovering QTL on 7 A_T and 3 D_T chromosomes, whereas Maricopa recovered QTL on 5 A_T and 6 D_T chromosomes.

The QTL shared between the Ames and Maricopa populations were most frequently associated with fiber color (9 out of 14 shared QTL), with the remaining five influencing fiber length (2 shared QTL), flower/pollen color (2 QTL), and plant architecture/stem pubescence (1 QTL). Half of the shared QTL were located on chromosome A06, in overlapping QTL spanning 48 and 52 cM in Ames and Maricopa, respectively (Figure 2). Fiber length QTL, however, were frequently unshared (6 in Ames and 5 in Maricopa), often located on different chromosomes and conferring different traits. The unshared fiber length QTL from both subpopulations were primarily on D_T subgenome chromosomes (i.e., D04, D07, D11, and A09 for the A_T QTL; D08, D13, and A06 for the D_T QTL) and spanned the phenotypic traits encompassed by fiber length. Notably, three trait categories had no shared QTL, i.e, fiber quality, plant architecture, and seed. The sole three "other fiber quality" QTL were found in Ames only, the six seed

QTL were found in Maricopa only, and the 12 fruiting habit QTL were split between Ames (7 QTL) and Maricopa (5 QTL) (Supplementary Table S2).

The distribution and total length of QTL also varied between the two polyploid subgenomes. Over 70% of detected QTL (48 of 68 QTL) are found on A_T subgenome chromosomes, 21 from the Ames subpopulation and 27 from the Maricopa subpopulation. Most of the A_T-derived QTL (62.5%) are for fiber color (19) or fruiting habit (11), meaning that most of the QTL that overlap between the two subpopulations (11 out of 14) are located on A_T chromosomes. For most of the categories (i.e., fiber color, fruiting habit, plant architecture, and seed), the A_T subgenome showed more QTL ranging from a 4:1 bias in the plant architecture categories to the fiber color category, which solely contained A_T-derived QTL. Both the mean and median length of A_T derived QTL is slightly larger than D_T derived (average 19 versus 14 cM, respectively, for mean, and 18 versus 12 cM for median), likely due to the large color QTL on chromosome A06. Interestingly, the fiber length and quality categories contained only D_T-derived QTL, together accounting for 70% of the QTL from D_T subgenome chromosomes. This observation is congruent with some previous research that has suggested D-genome recruitment during fiber domestication.

Discussion

QTL lability and the complex genetic architecture of cotton domestication phenotypes

The molecular underpinnings of the domesticated cotton fiber phenotype is of substantial interest from both evolutionary and economic standpoints. Because a cotton “fiber” is a single-celled, highly exaggerated structure, it provides a unique model for the evolutionary and developmental transformations that are possible in a single cell. Economically, cotton fibers are central to a multi-billion dollar and globally vital industry, one that has a vested interest in manipulating the genetics of domesticated fiber. Consequently, myriad studies have attempted to reveal the key players in fiber development. The results of these experiments and analyses have been diverse and often in conflict, underscoring the complex nature of cotton fiber biology and also the diverse suite of populations that have variously been employed. Comparison between the present research and previously generated QTL suffers from this same complexity. Many of the phenotypic traits evaluated here have been evaluated in other crosses and under different conditions, as summarized in the Cotton QTL Database v. 2.3 (Said, Knapka, et al. 2015) and CottonGen (Yu et al. 2014). As noted by others, QTL results of an individual study (such as the one presented here) are frequently incongruent with QTL results from other crosses grown under different conditions (Said, Knapka, et al. 2015; Lacape et al. 2010; Rong et al. 2007; Said, Song, et al. 2015). This observation is clear from our results alone, where less than half of the QTL were shared across two similar environments. When extended to previous QTL results, even our most robust shared QTL (*i.e.*, fiber color, chromosome A06) exhibit more complicated inheritance; *i.e.*, the Cotton QTL Database lists 62 QTL for fiber color spread across 21 of the 26 cotton chromosomes whereas we detect a single chromosome for both environments. A notable difference between ours and previous studies, however, is that ours was designed to capture the array of changes that characterize the transformation of the truly wild form of *G. hirsutum* into the modern elite cultivars that presently comprise the modern annualized crop plant. This cross should capture the major differences between wild and domesticated forms of *G. hirsutum*, whereas previous research has focused on differences between either (1) elite lines of the independently domesticated species *G. hirsutum* and *G. barbadense* (i.e., Pima cotton), or (2) between *G. hirsutum* landraces and/or elite cultivars, which reflect differences in improvement rather than those accompanying initial domestication.

Notwithstanding these substantive differences among studies, both the results presented here and earlier indicate that the genetic architecture underlying fiber morphology and development (among other domestication phenotypes) is complex and is responsive to environmental conditions. Consequently, uncovering QTL represent an important yet insufficient step in disentangling the genetic underpinnings of fiber development and cotton domestication. The complex interactions among genes important to understanding the QTL recovered remain to be elucidated, but many important enabling tools for such analyses have been developed. For example, gene coexpression network analyses can reveal modules of interconnected genes involved in key traits, as shown for cottonseed (G. Hu et al. 2016) and fiber (Gallagher et al. in prep), using the comparative context of wild versus domesticated *G. hirsutum*. In these examples, domestication appears to have increased the coordinated expression among genes and gene modules relevant to domesticated phenotypes. Ongoing research on *cis*/*trans* regulatory differences between wild and domesticated *G. hirsutum* (Bao, Hu, et al. in prep) suggests that changes in both *cis* and *trans* regulation have occurred during domestication; notably, the latter are frequently associated with environmental responsiveness and therefore may contribute to the environmental variability of QTL as reported here.

Multiple sources of information can narrow candidate gene identification

A primary goal of QTL analyses is to uncover the genomic basis of phenotypic differences. In many cases, QTL regions encompass a large region of the genome, and hence contain many genes. Here, each individual QTL recovered between 62 and 2,080 genes (average = 540), resulting in 968 - 4,329 possible candidate genes for each phenotype (Supplementary Table 2). In the present analysis, we narrow the candidate genes to those found within ± 5% cM of the QTL peak and focus on those genes with secondary evidence, i.e., DGE, amino acid changes, transcription factors, and/or those with relevant functions in related species. The genes mentioned here as candidates, while not exhaustive, represent strong candidates for the their respective phenotypes. The strength of these candidates, however, is limited by the information available. For the fiber QTL, we were able to leverage existing expression information for the accessions used in the QTL mapping cross, which provides additional evidence supporting individual genes as candidates. A caveat, however, is that since the expression sampling was completed for an independent project and QTL are often environmentally labile, genes exhibiting differential expression (or lack thereof) in the dataset used here may not represent the expression patterns that would be observed in the individuals used in the initial QTL cross and grown under the conditions of the QTL subpopulations. Furthermore, differential expression data were only available for two timepoints during fiber development, albeit key timepoints (Haigler et al. 2012). Future QTL research may be improved by integrating multiple data types from the outset, including expression from tissues relevant to the phenotypes evaluated for each parent grown in each environment; however, the results of the present were improved (for the fiber phenotype) by considering the data available.

Implications for domestication and future prospects

Domestication is a complex process involving a multiplicity of traits and the coordinated alteration of gene expression for numerous genes, for all but the simplest of traits (Olsen and Wendel 2013a, [b] 2013; Meyer and Purugganan 2013). With respect to cotton, a large number of QTL analyses have been conducted, specifically focused on economically valuable fiber characteristics, with some interest in other agronomically important phenotypes. These analyses have used either different species (Lacape et al. 2010; Said, Song, et al. 2015; C. Jiang et al. 1998; Lacape et al. 2005; Said, Knapka, et al. 2015; B. Wang, Zhuang, et al. 2017; Mei et al. 2004; Rong et al. 2007; B. Wang et al. 2016; B. Wang, Draye, et al. 2017; Draye et al. 2005; P. Chee et al. 2005; P. W. Chee et al. 2005; A. H. Paterson et al. 2003) or different cultivated lines of the same species (Jamshed et al. 2016; Shang et al. 2015; Qin et al. 2008; Lin et al. 2009; Tan et al. 2015; C. Li et al. 2013, 2012; S. Tang et al. 2015; Ulloa et al. 2005; Z.-S. Zhang et al. 2005; Shang et al. 2016; Tan et al. 2018; Shen et al. 2006; H. Wang et al. 2015) to provide perspectives on the genetic control of various traits. While each contributes to our multi-dimensional understanding of the controls on phenotypes, (1) it is not immediately clear that interspecies QTL are useful in cotton breeding programs (Jamshed et al. 2016; Shang et al. 2015; Lin et al. 2009), and (2) inter-cultivar or inter-line crosses provide a limited perspective on the underlying genetic architecture leading to modern elite lines. The present QTL analysis was designed specifically to reveal the genetic architecture underlying the morphological transformation from wild to domesticated upland cotton, *G. hirsutum*. Like many of existing QTL analyses in cotton, our cross, while having allelic replication only in two environments, also demonstrates that the genomic differences that underlie many wild vs. cultivated characteristics are environmentally variable. Although the parents of the cross were phenotypically similar in both environments, less than half of the QTL were conserved across the two subpopulations. This variability is likely due to pleiotropic and environmentally labile regulatory factors and genetic interactions (Metzger et al. 2016; Signor and Nuzhdin 2018; Wittkopp, Haerum, and Clark 2004; Rhoné et al. 2017; Coolon et al. 2014; Chen, Nolte, and Schlötterer 2015) playing a role in divergence between wild and domesticated species. This complexity is also increased by the allopolyploid nature of cotton, whose subgenomes evolved in isolation for 5-10 million years but now are reunited in a common nucleus, where they have coexisted for 1-2 million years. It is notable that, congruent with other QTL analyses, we find important fiber related QTL on the subgenome derived from the parent with the much shorter, inferior fiber (D genome). The involvement of the D-genome in the evolution of transgressive fiber phenotypes has been noted in multiple analyses, including for QTL (C. Jiang et al. 1998; Rong et al. 2007; Lacape et al. 2005; Han et al. 2006; Said, Song, et al. 2015; Qin et al. 2008), expression (Hovav, Chaudhary, et al. 2008; Yoo and Wendel 2014; Fang, Guan, and Zhang 2017; T. Zhang et al. 2015), and in selective genomic sweeps (Fang, Gong, et al. 2017; Fang, Wang, et al. 2017), yet the underlying genetic basis for this phenomenon remains unclear. Further work using advanced populations in which individual QTL have been isolated in isogenic backgrounds, combined with a multi-omics or systems biology perspective, is one promising approach to developing a fuller understanding of cotton biology as well as the domestication process.

Acknowledgements

The authors would like to thank Gwen Coyle, Kristen Cox, Joel Gilley, Anna Tuchin, and Kara Grupp for experimental assistance. The authors acknowledge computational support and assistance from the Iowa State University ResearchIT Unit (http://researchit.las.iastate.edu/). Funding was provided by the National Science Foundation Plant Genome Program, the USDA-ARS, and by Cotton Incorporated.

References

Ai, Xiantao, Yajun Liang, Junduo Wang, Juyun Zheng, Zhaolong Gong, Jiangping Guo, Xueyuan Li, and Yanying Qu. 2017. "Genetic Diversity and Structure of Elite Cotton Germplasm (Gossypium Hirsutum L.) Using Genome-Wide SNP Data." *Genetica* 145 (4-5): 409–16.

Al-Ghazi, Yves, Stéphane Bourot, Tony Arioli, Elizabeth S. Dennis, and Danny 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 Physiology* 50 (7): 1364–81.

Argiriou, Anagnostis, Apostolos Kalivas, Georgios Michailidis, and Athanasios Tsaftaris. 2012.

"Characterization of PROFILIN Genes from Allotetraploid (Gossypium Hirsutum) Cotton and Its Diploid Progenitors and Expression Analysis in Cotton Genotypes Differing in Fiber Characteristics." *Molecular Biology Reports* 39 (4): 3523–32.

Atmodjo, Melani A., Yumiko Sakuragi, Xiang Zhu, Amy J. Burrell, Sushree S. Mohanty, James A. Atwood 3rd, Ron Orlando, Henrik V. Scheller, and Debra Mohnen. 2011. "Galacturonosyltransferase (GAUT)1 and GAUT7 Are the Core of a Plant Cell Wall Pectin Biosynthetic Homogalacturonan:galacturonosyltransferase Complex." *Proceedings of the National Academy of Sciences of the United States of America* 108 (50): 20225–30.

Augustine, Robert C., Luis Vidali, Ken P. Kleinman, and Magdalena Bezanilla. 2008. "Actin Depolymerizing Factor Is Essential for Viability in Plants, and Its Phosphoregulation Is Important for Tip Growth." *The Plant Journal: For Cell and Molecular Biology* 54 (5): 863–75.

Bao, Ying, Guanjing Hu, Lex E. Flagel, Armel Salmon, Magdalena Bezanilla, Andrew H. Paterson, Zining Wang, and Jonathan F. Wendel. 2011. "Parallel up-Regulation of the Profilin Gene Family Following Independent Domestication of Diploid and Allopolyploid Cotton (Gossypium)." *Proceedings of the National Academy of Sciences of the United States of America* 108 (52): 21152–57.

Basten, Christopher, Bruce Weir, and Zhao-Bang Zeng. 2005. "QTL Cartographer, Version 1.17," April. https://www.researchgate.net/publication/228705502_QTL_Cartographer_version_117.

Besseau, Sébastien, Laurent Hoffmann, Pierrette Geoffroy, Catherine Lapierre, Brigitte Pollet, and Michel Legrand. 2007. "Flavonoid Accumulation in Arabidopsis Repressed in Lignin Synthesis Affects Auxin Transport and Plant Growth." *The Plant Cell* 19 (1): 148–62.

Betancur, Lissete, Bir Singh, Ryan A. Rapp, Jonathan F. Wendel, M. David Marks, Alison W. Roberts, and Candace H. Haigler. 2010. "Phylogenetically Distinct Cellulose Synthase Genes Support Secondary Wall Thickening in Arabidopsis Shoot Trichomes and Cotton Fiber." *Journal of Integrative Plant Biology* 52 (2): 205–20.

Bischoff, Volker, Joachim Selbig, and Wolf-Rüdiger Scheible. 2010. "Involvement of TBL/DUF231 Proteins into Cell Wall Biology." *Plant Signaling & Behavior* 5 (8): 1057–59.

Box, G. E. P., and D. R. Cox. 1964. "An Analysis of Transformations." *Journal of the Royal Statistical Society. Series B, Statistical Methodology* 26 (2): 211–43.

Brubaker, Curt L., and Jonathan F. Wendel. 1994. "Reevaluating the Origin of Domesticated Cotton (Gossypium Hirsutum; Malvaceae) Using Nuclear Restriction Fragment Length Polymorphisms (RFLPs)." *American Journal of Botany* 81 (10): 1309–26.

Byers, Robert L., David B. Harker, Scott M. Yourstone, Peter J. Maughan, and Joshua A. Udall. 2012. "Development and Mapping of SNP Assays in Allotetraploid Cotton." *TAG. Theoretical and Applied Genetics. Theoretische Und Angewandte Genetik* 124 (7): 1201–14.

Cai, Caiping, Guozhong Zhu, Tianzhen Zhang, and Wangzhen Guo. 2017. "High-Density 80 K SNP Array Is a Powerful Tool for Genotyping G. Hirsutum Accessions and Genome Analysis." *BMC Genomics* 18 (1): 654.

Chandnani, Rahul, Zhengsheng Zhang, Jinesh D. Patel, Jeevan Adhikari, Sameer Khanal, Daohua He, Nino Brown, Peng W. Chee, and Andrew H. Paterson. 2017. "Comparative Genetic Variation of Fiber Quality Traits in Reciprocal Advanced Backcross Populations." *Euphytica/ Netherlands Journal of Plant Breeding* 213 (11): 241.

Chaudhary, Bhupendra, Ran Hovav, Lex Flagel, Ron Mittler, and Jonathan 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 (August): 378.

Chee, Peng, Xavier Draye, Chun-Xiao Jiang, Laura Decanini, Terrye A. Delmonte, Robert Bredhauer, C. Wayne Smith, and Andrew H. Paterson. 2005. "Molecular Dissection of Interspecific Variation between Gossypium Hirsutum and Gossypium Barbadense (cotton) by a Backcross-Self Approach: I. Fiber Elongation." *TAG. Theoretical and Applied Genetics. Theoretische Und Angewandte Genetik* 111 (4): 757–63.

Chee, Peng W., Xavier Draye, Chun-Xiao Jiang, Laura Decanini, Terrie A. Delmonte, Robert Bredhauer,

C. Wayne Smith, and Andrew H. Paterson. 2005. “Molecular Dissection of Phenotypic Variation between Gossypium Hirsutum and Gossypium Barbadense (cotton) by a Backcross-Self Approach: III. Fiber Length.” *TAG. Theoretical and Applied Genetics. Theoretische Und Angewandte Genetik* 111 (4): 772–81.

Chen, Jun, Viola Nolte, and Christian Schlötterer. 2015. “Temperature Stress Mediates Decanalization and Dominance of Gene Expression in Drosophila Melanogaster.” *PLoS Genetics* 11 (2): e1004883.

Christie, John M., Noriyuki Suetsugu, Stuart Sullivan, and Masamitsu Wada. 2018. “Shining Light on the Function of NPH3/RPT2-Like Proteins in Phototropin Signaling.” *Plant Physiology* 176 (2): 1015–24.

Churchill, G. A., and R. W. Doerge. 1994. “Empirical Threshold Values for Quantitative Trait Mapping.” *Genetics* 138 (3): 963–71.

Cingolani, Pablo, Viral M. Patel, Melissa Coon, Tung Nguyen, Susan J. Land, Douglas M. Ruden, and Xiangyi Lu. 2012. “Using Drosophila Melanogaster as a Model for Genotoxic Chemical Mutational Studies with a New Program, SnpSift.” *Frontiers in Genetics* 3 (March): 35.

Cingolani, Pablo, Adrian Platts, Le Lily Wang, Melissa Coon, Tung Nguyen, Luan Wang, Susan J. Land, Xiangyi Lu, and Douglas M. Ruden. 2012. “A Program for Annotating and Predicting the Effects of Single Nucleotide Polymorphisms, SnpEff: SNPs in the Genome of Drosophila Melanogaster Strain w1118; Iso-2; Iso-3.” *Fly* 6 (2): 80–92.

Collings, David A., Leigh K. Gebbie, Paul A. Howles, Ursula A. Hurley, Rosemary J. Birch, Ann H. Cork, Charles H. Hocart, Tony Arioli, and Richard E. Williamson. 2008. “Arabidopsis Dynamin-like Protein DRP1A: A Null Mutant with Widespread Defects in Endocytosis, Cellulose Synthesis, Cytokinesis, and Cell Expansion.” *Journal of Experimental Botany* 59 (2): 361–76.

Coolon, Joseph D., C. Joel McManus, Kraig R. Stevenson, Brenton R. Graveley, and Patricia J. Wittkopp. 2014. “Tempo and Mode of Regulatory Evolution in Drosophila.” *Genome Research* 24 (5): 797–808.

Coppens d’Eeckenbrugge, Geo, and Jean-Marc Lacape. 2014. “Distribution and Differentiation of Wild, Feral, and Cultivated Populations of Perennial Upland Cotton (Gossypium Hirsutum L.) in Mesoamerica and the Caribbean.” *PloS One* 9 (9): e107458.

Cotton Fiber: Physics, Chemistry and Biology. 2018. 1st ed. Springer International Publishing.

Danecek, Petr, Adam Auton, Goncalo Abecasis, Cornelis A. Albers, Eric Banks, Mark A. DePristo, Robert E. Handsaker, et al. 2011. “The Variant Call Format and VCFtools.” *Bioinformatics* 27 (15): 2156–58.

Davière, Jean-Michel, Michael Wild, Thomas Regnault, Nicolas Baumberger, Herfried Eisler, Pascal Genschik, and Patrick Achard. 2014. “Class I TCP-DELLA Interactions in Inflorescence Shoot Apex Determine Plant Height.” *Current Biology: CB* 24 (16): 1923–28.

Doebley, J., A. Stec, and L. Hubbard. 1997. “The Evolution of Apical Dominance in Maize.” *Nature* 386 (6624): 485–88.

Draye, Xavier, Peng Chee, Chun-Xiao Jiang, Laura Decanini, Terrye A. Delmonte, Robert Bredhauer, C. Wayne Smith, and Andrew H. Paterson. 2005. “Molecular Dissection of Interspecific Variation between Gossypium Hirsutum and G. Barbadense (cotton) by a Backcross-Self Approach: II. Fiber Fineness.” *TAG. Theoretical and Applied Genetics. Theoretische Und Angewandte Genetik* 111 (4): 764–71.

Fang, Lei, Hao Gong, Yan Hu, Chunxiao Liu, Baoliang Zhou, Tao Huang, Yangkun Wang, et al. 2017. “Genomic Insights into Divergence and Dual Domestication of Cultivated Allotetraploid Cottons.” *Genome Biology* 18 (1): 33.

Fang, Lei, Xueying Guan, and Tianzhen Zhang. 2017. “Asymmetric Evolution and Domestication in Allotetraploid Cotton (Gossypium Hirsutum L.).” *The Crop Journal* 5 (2): 159–65.

Fang, Lei, Qiong Wang, Yan Hu, Yinhua Jia, Jiedan Chen, Bingliang Liu, Zhiyuan Zhang, et al. 2017. “Genomic Analyses in Cotton Identify Signatures of Selection and Loci Associated with Fiber Quality and Yield Traits.” *Nature Genetics* 49 (7): 1089–98.

Feng, Hongjie, Xinhui Tian, Yongchang Liu, Yanjun Li, Xinyu Zhang, Brian Joseph Jones, Yuqiang Sun, and Jie Sun. 2013. “Analysis of Flavonoids and the Flavonoid Structural Genes in Brown Fiber of Upland Cotton.” *PloS One* 8 (3): e58820.

Fluidigm. 2011. “SNP Genotyping.”

Fryxell, P. A. 1992. “A Revised Taxonomic Interpretation of Gossypium L (Malvaceae).” *Rheeda* 2: 108–65.

Fryxell, Paul A. 1968. “A Redefinition of the Tribe Gossypieae.” *Botanical Gazette* 129 (4): 296–308.

———. 1976. “A Nomenclator of Gossypium: The Botanical Name of Cotton.” Technical Bulletins 158582. United States Department of Agriculture, Economic Research Service. https://ideas.repec.org/p/ags/uerstb/158582.html.

———. 1979. *Natural History of the Cotton Tribe*. 1st edition. Texas A&M University Press.

Gallagher, Joseph P., Corrinne E. Grover, Kristen Rex, Matthew Moran, and Jonathan F. Wendel. 2017. “A New Species of Cotton from Wake Atoll, Gossypium Stephensii (Malvaceae).” *Systematic Botany* 42 (1): 115–23.

Gallavotti, Andrea. 2013. “The Role of Auxin in Shaping Shoot Architecture.” *Journal of Experimental Botany* 64 (9): 2593–2608.

Gou, Jin-Ying, Ling-Jian Wang, Shuang-Ping Chen, Wen-Li Hu, and Xiao-Ya Chen. 2007. “Gene Expression and Metabolite Profiles of Cotton Fiber during Cell Elongation and Secondary Cell Wall Synthesis.” *Cell Research* 17 (5): 422–34.

Haigler, Candace H., Lissete Betancur, Michael R. Stiff, and John R. Tuttle. 2012. “Cotton Fiber: A Powerful Single-Cell Model for Cell Wall and Cellulose Research.” *Frontiers in Plant Science* 3 (May): 104.

Haigler, Candace H., Bir Singh, Guirong Wang, and Deshui Zhang. 2009. “Genomics of Cotton Fiber Secondary Wall Deposition and Cellulose Biogenesis.” In *Genetics and Genomics of Cotton*, edited by Andrew H. Paterson, 385–417. New York, NY: Springer US.

Han, Zhiguo, Changbiao Wang, Xianliang Song, Wangzhen Guo, Jinying Gou, Chunhong Li, Xiaoya Chen, and Tianzhen Zhang. 2006. “Characteristics, Development and Mapping of Gossypium Hirsutum Derived EST-SSRs in Allotetraploid Cotton.” *TAG. Theoretical and Applied Genetics. Theoretische Und Angewandte Genetik* 112 (3): 430–39.

Hentrich, Mathias, Christine Böttcher, Petra Düchting, Youfa Cheng, Yunde Zhao, Oliver Berkowitz, Josette Masle, Joaquín Medina, and Stephan Pollmann. 2013. “The Jasmonic Acid Signaling Pathway Is Linked to Auxin Homeostasis through the Modulation of YUCCA8 and YUCCA9 Gene Expression.” *The Plant Journal: For Cell and Molecular Biology* 74 (4): 626–37.

Hentrich, Mathias, Beatriz Sánchez-Parra, Marta-Marina Pérez Alonso, Víctor Carrasco Loba, Laura Carrillo, Jesus Vicente-Carbajosa, Joaquín Medina, and Stephan Pollmann. 2013. “YUCCA8 and YUCCA9 Overexpression Reveals a Link between Auxin Signaling and Lignification through the Induction of Ethylene Biosynthesis.” *Plant Signaling & Behavior* 8 (11): e26363.

Hinze, Lori L., Amanda M. Hulse-Kemp, Iain W. Wilson, Qian-Hao Zhu, Danny J. Llewellyn, Jen M. Taylor, Andrew Spriggs, et al. 2017. “Diversity Analysis of Cotton (Gossypium Hirsutum L.) Germplasm Using the CottonSNP63K Array.” *BMC Plant Biology* 17 (1): 37.

Hovav, Ran, Bhupendra Chaudhary, Joshua A. Udall, Lex Flagel, and Jonathan F. Wendel. 2008. “Parallel Domestication, Convergent Evolution and Duplicated Gene Recruitment in Allopolyploid Cotton.” *Genetics* 179 (3): 1725–33.

Hovav, Ran, Joshua A. Udall, Bhupendra Chaudhary, Ryan Rapp, Lex Flagel, and Jonathan F. Wendel. 2008. “Partitioned Expression of Duplicated Genes during Development and Evolution of a Single Cell in a Polyploid Plant.” *Proceedings of the National Academy of Sciences of the United States of America* 105 (16): 6191–95.

Hovav, Ran, Joshua A. Udall, Einat Hovav, Ryan Rapp, Lex Flagel, and Jonathan F. Wendel. 2008. “A Majority of Cotton Genes Are Expressed in Single-Celled Fiber.” *Planta* 227 (2): 319–29.

Huang, Quan-Sheng, Xun-Ji Chen, Jian-Ping Li, Xiao-Yan Hao, Guo Chen, Zumuremu, and Lin Shao. 2014. “The Phosphorylation of an Actin Depolymerizing Factor by a Calcium-Dependent Protein Kinase Regulates Cotton Fiber Elongation.” *Acta Physiologiae Plantarum / Polish Academy of*

Sciences, Committee of Plant Physiology Genetics and Breeding 36 (10): 2627–36.

Huang, Yadong, Chun Yao Li, Donna L. Pattison, William M. Gray, Sungjin Park, and Susan I. Gibson. 2010. "SUGAR-INSENSITIVE3, a RING E3 Ligase, Is a New Player in Plant Sugar Response." *Plant Physiology* 152 (4): 1889–1900.

Hua, Shuijin, Xuede Wang, Shuna Yuan, Mingyan Shao, Xiangqian Zhao, Shuijin Zhu, and Lixi Jiang. 2007. "Characterization of Pigmentation and Cellulose Synthesis in Colored Cotton Fibers." *Crop Science* 47: 1540–46.

Hu, Guanjing, Ran Hovav, Corrinne E. Grover, Adi Faigenboim-Doron, Noa Kadmon, Justin T. Page, Joshua A. Udall, and Jonathan F. Wendel. 2016. "Evolutionary Conservation and Divergence of Gene Coexpression Networks in Gossypium (Cotton) Seeds." *Genome Biology and Evolution* 8 (12): 3765–83.

Hu, Yan, Jiedan Chen, Lei Fang, Zhiyuan Zhang, Wei Ma, Yongchao Niu, Longzhen Ju, et al. 2019. "Gossypium Barbadense and Gossypium Hirsutum Genomes Provide Insights into the Origin and Evolution of Allotetraploid Cotton." *Nature Genetics* 51 (4): 739–48.

Jamshed, Muhammad, Fei Jia, Juwu Gong, Koffi Kibalou Palanga, Yuzhen Shi, Junwen Li, Haihong Shang, et al. 2016. "Identification of Stable Quantitative Trait Loci (QTLs) for Fiber Quality Traits across Multiple Environments in Gossypium Hirsutum Recombinant Inbred Line Population." *BMC Genomics* 17 (March): 197.

Jiang, C., 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 of the United States of America* 95 (8): 4419–24.

Jiang, Wenbo, Qinggang Yin, Ranran Wu, Guangshun Zheng, Jinyue Liu, Richard A. Dixon, and Yongzhen Pang. 2015. "Role of a Chalcone Isomerase-like Protein in Flavonoid Biosynthesis in Arabidopsis Thaliana." *Journal of Experimental Botany* 66 (22): 7165–79.

Ji, Shengjian, Yingchun Lu, Jun Li, Gang Wei, Xiujuan Liang, and Yuxian Zhu. 2002. "A Beta-Tubulin-like cDNA Expressed Specifically in Elongating Cotton Fibers Induces Longitudinal Growth of Fission Yeast." *Biochemical and Biophysical Research Communications* 296 (5): 1245–50.

Kaur, Baljinder, Priyanka Tyagi, and Vasu Kuraparthy. 2017. "Genetic Diversity and Population Structure in the Landrace Accessions of Gossypium Hirsutum." *Crop Science* 57: 2457–70.

Kieffer, Martin, Vera Master, Richard Waites, and Brendan Davies. 2011. "TCP14 and TCP15 Affect Internode Length and Leaf Shape in Arabidopsis." *The Plant Journal: For Cell and Molecular Biology* 68 (1): 147–58.

Kim, Hee Jin, Barbara A. Triplett, Hong-Bin Zhang, Mi-Kyung Lee, Doug J. Hinchliffe, Ping Li, and David D. Fang. 2012. "Cloning and Characterization of Homeologous Cellulose Synthase Catalytic Subunit 2 Genes from Allotetraploid Cotton (Gossypium Hirsutum L.)." *Gene* 494 (2): 181–89.

Kutner, Michael, Christopher Nachtsheim, John Neter, and William Li. 2004. *Applied Linear Statistical Models*. 5th edition. McGraw-Hill/Irwin.

Lacape, Jean-Marc, Danny Llewellyn, John Jacobs, Tony Arioli, David Becker, Steve Calhoun, Yves Al-Ghazi, 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 (June): 132.

Lacape, Jean-Marc, Trung-Bieu Nguyen, Brigitte Courtois, Jean-Louis Belot, Marc Giband, Jean-Paul Gourlot, Gerard Gawryziak, Sandrine Roques, and Bernard Hau. 2005. "QTL Analysis of Cotton Fiber Quality Using Multiple Gossypium Hirsutum Gossypium Barbadense Backcross Generations." *Published in Crop Sci.* 45: 123–40.

Lauvergeat, V., C. Lacomme, E. Lacombe, E. Lasserre, D. Roby, and J. Grima-Pettenati. 2001. "Two Cinnamoyl-CoA Reductase (CCR) Genes from Arabidopsis Thaliana Are Differentially Expressed during Development and in Response to Infection with Pathogenic Bacteria." *Phytochemistry* 57 (7): 1187–95.

Lee, J. A. 1968. "Genetical Studies Concerning the Distribution of Trichomes on the Leaves of GOSSYPIUM HIRSUTUM L." *Genetics* 60 (3): 567–75.

Lee, Y., D. Choi, and H. Kende. 2001. "Expansins: Ever-Expanding Numbers and Functions." *Current*

Opinion in Plant Biology 4 (6): 527–32.

Lewis, Daniel R., Melissa V. Ramirez, Nathan D. Miller, Prashanthi Vallabhaneni, W. Keith Ray, Richard F. Helm, Brenda S. J. Winkel, and Gloria K. Muday. 2011. "Auxin and Ethylene Induce Flavonol Accumulation through Distinct Transcriptional Networks." *Plant Physiology* 156 (1): 144–64.

Li, Chengqi, Changbiao Wang, Na Dong, Xiaoyun Wang, Haihong Zhao, Richard Converse, Zhe Xia, Rui Wang, and Qinglian Wang. 2012. "QTL Detection for Node of First Fruiting Branch and Its Height in Upland Cotton (Gossypium Hirsutum L.)." *Euphytica/ Netherlands Journal of Plant Breeding* 188 (3): 441–51.

Li, Chengqi, Xiaoyun Wang, Na Dong, Haihong Zhao, Zhe Xia, Rui Wang, Richard L. Converse, and Qinglian Wang. 2013. "QTL Analysis for Early-Maturing Traits in Cotton Using Two Upland Cotton (*Gossypium Hirsutum* L.) Crosses." *Breeding Science* 63 (2): 154–63.

Lin, Zhongxu, Yanxin Zhang, Xianlong Zhang, and Xiaoping Guo. 2009. "A High-Density Integrative Linkage Map for Gossypium Hirsutum." *Euphytica/ Netherlands Journal of Plant Breeding* 166 (1): 35–45.

Li, Tingchun, Honghong Fan, Zhengpeng Li, Jun Wei, Yi Lin, and Yongping Cai. 2012. "The Accumulation of Pigment in Fiber Related to Proanthocyanidins Synthesis for Brown Cotton." *Acta Physiologiae Plantarum / Polish Academy of Sciences, Committee of Plant Physiology Genetics and Breeding* 34 (2): 813–18.

Liu, Bao, and Jonathan F. Wendel. 2002. "Intersimple Sequence Repeat (ISSR) Polymorphisms as a Genetic Marker System in Cotton: TECHNICAL NOTE." *Molecular Ecology Notes* 1 (3): 205–8.

Lunn, Daniel, Sanyasi R. Gaddipati, Gregory A. Tucker, and Grantley W. Lycett. 2013. "Null Mutants of Individual RABA Genes Impact the Proportion of Different Cell Wall Components in Stem Tissue of Arabidopsis Thaliana." *PloS One* 8 (10): e75724.

MacMillan, Colleen P., Shawn D. Mansfield, Zbigniew H. Stachurski, Rob Evans, and Simon G. Southerton. 2010. "Fasciclin-like Arabinogalactan Proteins: Specialization for Stem Biomechanics and Cell Wall Architecture in Arabidopsis and Eucalyptus: FLAs Specialized for Stem Biomechanics and Cell Walls." *The Plant Journal: For Cell and Molecular Biology* 62 (4): 689–703.

Mangin, B., B. Goffinet, and A. Rebaï. 1994. "Constructing Confidence Intervals for QTL Location." *Genetics* 138 (4): 1301–8.

Ma, Zhiying, Shoupu He, Xingfen Wang, Junling Sun, Yan Zhang, Guiyin Zhang, Liqiang Wu, et al. 2018. "Resequencing a Core Collection of Upland Cotton Identifies Genomic Variation and Loci Influencing Fiber Quality and Yield." *Nature Genetics* 50 (6): 803–13.

McCarty, Jack C., Dewayne D. Deng, Johnie N. Jenkins, and Lige Geng. 2018. "Genetic Diversity of Day-Neutral Converted Landrace Gossypium Hirsutum L. Accessions." *Euphytica/ Netherlands Journal of Plant Breeding* 214 (10): 173.

McCouch, S. R., Y. G. Cho, P. E. Yano, M. Blinstrub, H. Morishima, and T. Kinosita. 1997. "Report on QTL Nomenclature." *Rice Genetics Newsletter* 14: 11–13.

Mei, M., N. H. Syed, W. Gao, P. M. Thaxton, C. W. Smith, D. M. Stelly, and Z. J. Chen. 2004. "Genetic Mapping and QTL Analysis of Fiber-Related Traits in Cotton (Gossypium)." *TAG. Theoretical and Applied Genetics. Theoretische Und Angewandte Genetik* 108 (2): 280–91.

Metzger, Brian P. H., Fabien Duveau, David C. Yuan, Stephen Tryban, Bing Yang, and Patricia J. Wittkopp. 2016. "Contrasting Frequencies and Effects of Cis- and Trans-Regulatory Mutations Affecting Gene Expression." *Molecular Biology and Evolution* 33 (5): 1131–46.

Meyer, Rachel S., and Michael D. Purugganan. 2013. "Evolution of Crop Species: Genetics of Domestication and Diversification." *Nature Reviews. Genetics* 14 (12): 840–52.

Millar, Anthony A., and Frank Gubler. 2005. "The Arabidopsis GAMYB-like Genes, MYB33 and MYB65, Are microRNA-Regulated Genes That Redundantly Facilitate Anther Development." *The Plant Cell* 17 (3): 705–21.

Mizutani, Masaharu. 2012. "Impacts of Diversification of Cytochrome P450 on Plant Metabolism."

Biological & Pharmaceutical Bulletin 35 (6): 824–32.

Mizutani, Masaharu, and Daisaku Ohta. 2010. “Diversification of P450 Genes during Land Plant Evolution.” *Annual Review of Plant Biology* 61: 291–315.

Nigam, Deepti, Palacharla Kavita, Rajiv K. Tripathi, Alok Ranjan, Ridhi Goel, Mehar Asif, Alpika Shukla, Gunjita Singh, Debashis Rana, and Samir V. Sawant. 2014. “Transcriptome Dynamics during Fibre Development in Contrasting Genotypes of Gossypium Hirsutum L.” *Plant Biotechnology Journal* 12 (2): 204–18.

Nixon, B. Tracy, Katayoun Mansouri, Abhishek Singh, Juan Du, Jonathan K. Davis, Jung-Goo Lee, Erin Slabaugh, et al. 2016. “Comparative Structural and Computational Analysis Supports Eighteen Cellulose Synthases in the Plant Cellulose Synthesis Complex.” *Scientific Reports* 6 (June): 28696.

Olsen, Kenneth M., and Jonathan F. Wendel. 2013a. “A Bountiful Harvest: Genomic Insights into Crop Domestication Phenotypes.” *Annual Review of Plant Biology* 64 (February): 47–70.

———. 2013b. “Crop Plants as Models for Understanding Plant Adaptation and Diversification.” *Frontiers in Plant Science* 4 (August): 290.

Orfila, Caroline, Susanne Oxenbøll Sørensen, Jesper Harholt, Naomi Geshi, Hazel Crombie, Hoai-Nam Truong, J. S. Grant Reid, J. Paul Knox, and Henrik Vibe Scheller. 2005. “QUASIMODO1 Is Expressed in Vascular Tissue of Arabidopsis Thaliana Inflorescence Stems, and Affects Homogalacturonan and Xylan Biosynthesis.” *Planta* 222 (4): 613–22.

Paterson, A. H., Y. Saranga, M. Menz, C-X Jiang, and R. Wright. 2003. “QTL Analysis of Genotype × Environment Interactions Affecting Cotton Fiber Quality.” *TAG. Theoretical and Applied Genetics. Theoretische Und Angewandte Genetik* 106 (3): 384–96.

Qin, Hongde, Wangzhen Guo, Yuan-Ming Zhang, and Tianzhen Zhang. 2008. “QTL Mapping of Yield and Fiber Traits Based on a Four-Way Cross Population in Gossypium Hirsutum L.” *TAG. Theoretical and Applied Genetics. Theoretische Und Angewandte Genetik* 117 (6): 883–94.

Quinlan, Aaron R., and Ira M. Hall. 2010. “BEDTools: A Flexible Suite of Utilities for Comparing Genomic Features.” *Bioinformatics* 26 (6): 841–42.

Rapp, Ryan A., Candace H. Haigler, Lex Flagel, Ran H. Hovav, Joshua A. Udall, and Jonathan F. Wendel. 2010. “Gene Expression in Developing Fibres of Upland Cotton (Gossypium Hirsutum L.) Was Massively Altered by Domestication.” *BMC Biology* 8 (November): 139.

Rautengarten, Carsten, Björn Usadel, Lutz Neumetzler, Jürgen Hartmann, Dirk Büssis, and Thomas Altmann. 2008. “A Subtilisin-like Serine Protease Essential for Mucilage Release from Arabidopsis Seed Coats.” *The Plant Journal: For Cell and Molecular Biology* 54 (3): 466–80.

Reddy, Umesh K., Padma Nimmakayala, Venkata Lakshmi Abburi, C. V. C. M. Reddy, Thangasamy Saminathan, Richard G. Percy, John Z. Yu, et al. 2017. “Genome-Wide Divergence, Haplotype Distribution and Population Demographic Histories for Gossypium Hirsutum and Gossypium Barbadense as Revealed by Genome-Anchored SNPs.” *Scientific Reports* 7 (January): 41285.

Reiter, W. D., C. Chapple, and C. R. Somerville. 1997. “Mutants of Arabidopsis Thaliana with Altered Cell Wall Polysaccharide Composition.” *The Plant Journal: For Cell and Molecular Biology* 12 (2): 335–45.

Rhoné, Bénédicte, Cédric Mariac, Marie Couderc, Cécile Berthouly-Salazar, Issaka Salia Ousseini, and Yves Vigouroux. 2017. “No Excess of Cis-Regulatory Variation Associated with Intraspecific Selection in Wild Pearl Millet (Cenchrus Americanus).” *Genome Biology and Evolution* 9 (2): 388–97.

Rong, Junkang, F. Alex Feltus, Vijay N. Waghmare, Gary J. Pierce, Peng W. Chee, Xavier Draye, Yehoshua Saranga, 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 (4): 2577–88.

Said, Joseph I., Joseph A. Knapka, Mingzhou Song, and Jinfa Zhang. 2015. “Cotton QTLdb: A Cotton QTL Database for QTL Analysis, Visualization, and Comparison between Gossypium Hirsutum and G. Hirsutum × G. Barbadense Populations.” *Molecular Genetics and Genomics: MGG* 290 (4): 1615–25.

Said, Joseph I., Zhongxu Lin, Xianlong Zhang, Mingzhou Song, and Jinfa 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 (1): 776.

Said, Joseph I., Mingzhou Song, Hantao Wang, Zhongxu Lin, Xianlong Zhang, David D. Fang, and Jinfa Zhang. 2015. "A Comparative Meta-Analysis of QTL between Intraspecific Gossypium Hirsutum and Interspecific G. Hirsutum × G. Barbadense Populations." *Molecular Genetics and Genomics: MGG* 290 (3): 1003–25.

Saski, Christopher A., Brian E. Scheffler, Amanda M. Hulse-Kemp, Bo Liu, Qingxin Song, Atsumi Ando, David M. Stelly, et al. 2017. "Sub Genome Anchored Physical Frameworks of the Allotetraploid Upland Cotton (Gossypium Hirsutum L.) Genome, and an Approach toward Reference-Grade Assemblies of Polyploids." *Scientific Reports* 7 (1): 15274.

Schliep, Martin, Berit Ebert, Ulrike Simon-Rosin, Daniela Zoeller, and Joachim Fisahn. 2010. "Quantitative Expression Analysis of Selected Transcription Factors in Pavement, Basal and Trichome Cells of Mature Leaves from Arabidopsis Thaliana." *Protoplasma* 241 (1-4): 29–36.

Shang, Lianguang, Qingzhi Liang, Yumei Wang, Xiaocui Wang, Kunbo Wang, Abdugheni Abduweli, Lingling Ma, Shihu Cai, and Jinping Hua. 2015. "Identification of Stable QTLs Controlling Fiber Traits Properties in Multi-Environment Using Recombinant Inbred Lines in Upland Cotton (Gossypium Hirsutum L.)." *Euphytica/ Netherlands Journal of Plant Breeding* 205 (3): 877–88.

Shang, Lianguang, Yumei Wang, Xiaocui Wang, Fang Liu, Abdugheni Abduweli, Shihu Cai, Yuhua Li, Lingling Ma, Kunbo Wang, and Jinping Hua. 2016. "Genetic Analysis and QTL Detection on Fiber Traits Using Two Recombinant Inbred Lines and Their Backcross Populations in Upland Cotton." *G3* 6 (9): 2717–24.

Shcherban, T. Y., J. Shi, D. M. Durachko, M. J. Guiltinan, S. J. McQueen-Mason, M. Shieh, and D. J. Cosgrove. 1995. "Molecular Cloning and Sequence Analysis of Expansins--a Highly Conserved, Multigene Family of Proteins That Mediate Cell Wall Extension in Plants." *Proceedings of the National Academy of Sciences of the United States of America* 92 (20): 9245–49.

Shen, Xinlian, Tianzhen Zhang, Wangzhen Guo, Xiefei Zhu, and Xiaoyang Zhang. 2006. "Mapping Fiber and Yield QTLs with Main, Epistatic, and QTL × Environment Interaction Effects in Recombinant Inbred Lines of Upland Cotton." *Crop Science* 46: 61–66.

Shikata, Masahito, Tomotsugu Koyama, Nobutaka Mitsuda, and Masaru Ohme-Takagi. 2009. "Arabidopsis SBP-Box Genes SPL10, SPL11 and SPL2 Control Morphological Change in Association with Shoot Maturation in the Reproductive Phase." *Plant & Cell Physiology* 50 (12): 2133–45.

Shi, Yong-Hui, Sheng-Wei Zhu, Xi-Zeng Mao, Jian-Xun Feng, Yong-Mei Qin, Liang Zhang, Jing Cheng, Li-Ping Wei, Zhi-Yong Wang, and Yu-Xian Zhu. 2006. "Transcriptome Profiling, Molecular Biological, and Physiological Studies Reveal a Major Role for Ethylene in Cotton Fiber Cell Elongation." *The Plant Cell* 18 (3): 651–64.

Signor, Sarah A., and Sergey V. Nuzhdin. 2018. "The Evolution of Gene Expression in Cis and Trans." *Trends in Genetics: TIG* 34 (7): 532–44.

Spokevicius, Antanas V., Simon G. Southerton, Colleen P. MacMillan, Deyou Qiu, Siming Gan, Josquin F. G. Tibbits, Gavin F. Moran, and Gerd Bossinger. 2007. "β-Tubulin Affects Cellulose Microfibril Orientation in Plant Secondary Fibre Cell Walls: β-Tubulin Affects MFA." *The Plant Journal: For Cell and Molecular Biology* 51 (4): 717–26.

Stam, Piet. 1993. "Construction of Integrated Genetic Linkage Maps by Means of a New Computer Package: Join Map." *The Plant Journal: For Cell and Molecular Biology* 3 (5): 739–44.

Steinborn, Katharina, Christoph Maulbetsch, Bianca Priester, Susanne Trautmann, Tobias Pacher, Bernd Geiges, Frank Küttner, et al. 2002. "The Arabidopsis PILZ Group Genes Encode Tubulin-Folding Cofactor Orthologs Required for Cell Division but Not Cell Growth." *Genes & Development* 16 (8): 959–71.

Steiner, Evyatar, Osnat Yanai, Idan Efroni, Naomi Ori, Yuval Eshed, and David Weiss. 2012. "Class I

TCPs Modulate Cytokinin-Induced Branching and Meristematic Activity in Tomato.” *Plant Signaling & Behavior* 7 (7): 807–10.

Studer, Anthony J., and John F. Doebley. 2012. “Evidence for a Natural Allelic Series at the Maize Domestication Locus Teosinte branched1.” *Genetics* 191 (3): 951–58.

Takos, Adam M., and Fred Rook. 2012. “Why Biosynthetic Genes for Chemical Defense Compounds Cluster.” *Trends in Plant Science* 17 (7): 383–88.

Taliercio, Earl W., and Deborah Boykin. 2007. “Analysis of Gene Expression in Cotton Fiber Initials.” *BMC Plant Biology* 7 (May): 22.

Tang, Shiyi, Zhonghua Teng, Tengfei Zhai, Xiaomei Fang, Fang Liu, Dajun Liu, Jian Zhang, et al. 2015. “Construction of Genetic Map and QTL Analysis of Fiber Quality Traits for Upland Cotton (Gossypium Hirsutum L.).” *Euphytica/ Netherlands Journal of Plant Breeding* 201 (2): 195–213.

Tang, Weijiang, Qiang Ji, Yongping Huang, Zhimin Jiang, Manzhu Bao, Haiyang Wang, and Rongcheng Lin. 2013. “FAR-RED ELONGATED HYPOCOTYL3 and FAR-RED IMPAIRED RESPONSE1 Transcription Factors Integrate Light and Abscisic Acid Signaling in Arabidopsis.” *Plant Physiology* 163 (2): 857–66.

Tan, Zhaoyun, Xiaomei Fang, Shiyi Tang, Jian Zhang, Dajun Liu, Zhonghua Teng, Ling Li, et al. 2015. “Genetic Map and QTL Controlling Fiber Quality Traits in Upland Cotton (Gossypium Hirsutum L.).” *Euphytica/ Netherlands Journal of Plant Breeding* 203 (3): 615–28.

Tan, Zhaoyun, Zhiqin Zhang, Xujing Sun, Qianqian Li, Ying Sun, Peng Yang, Wenwen Wang, et al. 2018. “Genetic Map Construction and Fiber Quality QTL Mapping Using the CottonSNP80K Array in Upland Cotton.” *Frontiers in Plant Science* 9 (February): 225.

Treutter, Dieter. 2006. “Significance of Flavonoids in Plant Resistance: A Review.” *Environmental Chemistry Letters* 4 (3): 147.

Tuttle, John R., Gyoungju Nah, Mary V. Duke, Danny C. Alexander, Xueying Guan, Qingxin Song, Z. Jeffrey Chen, Brian E. Scheffler, and Candace H. Haigler. 2015. “Metabolomic and Transcriptomic Insights into How Cotton Fiber Transitions to Secondary Wall Synthesis, Represses Lignification, and Prolongs Elongation.” *BMC Genomics* 16 (June): 477.

Tyagi, Priyanka, Michael A. Gore, Daryl T. Bowman, B. Todd Campbell, Joshua A. Udall, and Vasu Kuraparthy. 2014. “Genetic Diversity and Population Structure in the US Upland Cotton (Gossypium Hirsutum L.).” *TAG. Theoretical and Applied Genetics. Theoretische Und Angewandte Genetik* 127 (2): 283–95.

Ulloa, M., S. Saha, J. N. Jenkins, W. R. Meredith Jr, J. C. McCarty Jr, and D. M. Stelly. 2005. “Chromosomal Assignment of RFLP Linkage Groups Harboring Important QTLs on an Intraspecific Cotton (Gossypium Hirsutum L.) Joinmap.” *The Journal of Heredity* 96 (2): 132–44.

Ungerer, Mark C., Solveig S. Halldorsdottir, Jennifer L. Modliszewski, Trudy F. C. Mackay, and Michael D. Purugganan. 2002. “Quantitative Trait Loci for Inflorescence Development in Arabidopsis Thaliana.” *Genetics* 160 (3): 1133–51.

Van der Auwera, Geraldine A., Mauricio O. Carneiro, Christopher Hartl, Ryan Poplin, Guillermo del Angel, Ami Levy-Moonshine, Tadeusz Jordan, et al. 2013. “From FastQ Data to High-Confidence Variant Calls: The Genome Analysis Toolkit Best Practices Pipeline” 43 (1): 11.10.1–11.10.33.

VAN Ooijen, J. W. 2011. “Multipoint Maximum Likelihood Mapping in a Full-Sib Family of an Outbreeding Species.” *Genetics Research* 93 (5): 343–49.

Voorrips, R. E. 2002. “MapChart: Software for the Graphical Presentation of Linkage Maps and QTLs.” *Journal of Heredity*. https://doi.org/10.1093/jhered/93.1.77.

Wang, Baohua, Xavier Draye, Zhengsheng Zhang, Zhimin Zhuang, O. Lloyd May, Andrew H. Paterson, and Peng W. Chee. 2016. “Advanced Backcross Quantitative Trait Locus Analysis of Fiber Elongation in a Cross between Gossypium Hirsutum and G. Mustelinum.” *Crop Science* 56: 1760–68.

Wang, Baohua, Xavier Draye, Zhimin Zhuang, Zhengsheng Zhang, Min Liu, Edward L. Lubbers, Don Jones, O. Lloyd May, Andrew H. Paterson, and Peng W. Chee. 2017. “QTL Analysis of Cotton Fiber Length in Advanced Backcross Populations Derived from a Cross between Gossypium

Hirsutum and G. Mustelinum.” *TAG. Theoretical and Applied Genetics. Theoretische Und Angewandte Genetik* 130 (6): 1297–1308.

Wang, Baohua, Zhimin Zhuang, Zhengsheng Zhang, Xavier Draye, Lan-Shuan Shuang, Tariq Shehzad, Edward L. Lubbers, et al. 2017. “Advanced Backcross QTL Analysis of Fiber Strength and Fineness in a Cross between Gossypium Hirsutum and G. Mustelinum.” *Frontiers in Plant Science* 8 (October): 1848.

Wang, Hantao, Cong Huang, Huanle Guo, Ximei Li, Wenxia Zhao, Baosheng Dai, Zhenhua Yan, and Zhongxu Lin. 2015. “QTL Mapping for Fiber and Yield Traits in Upland Cotton under Multiple Environments.” *PloS One* 10 (6): e0130742.

Wang, Qin Qin, Fei Liu, Xu Sheng Chen, Xiao Jie Ma, Hou Qing Zeng, and Zhi Min Yang. 2010. “Transcriptome Profiling of Early Developing Cotton Fiber by Deep-Sequencing Reveals Significantly Differential Expression of Genes in a Fuzzless/lintless Mutant.” *Genomics* 96 (6): 369–76.

Wang, Shui, Jia-Wei Wang, Nan Yu, Chun-Hong Li, Bin Luo, Jin-Ying Gou, Ling-Jian Wang, and Xiao-Ya Chen. 2004. “Control of Plant Trichome Development by a Cotton Fiber MYB Gene.” *The Plant Cell* 16 (9): 2323–34.

Wendel, Jonathan F., and Victor 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 (1): 115–43.

Wendel, Jonathan F., and Corrinne E. Grover. 2015. “Taxonomy and Evolution of the Cotton Genus, Gossypium.” In *Cotton*, 25–44. Agronomy Monograph. Madison, WI: American Society of Agronomy, Inc., Crop Science Society of America, Inc., and Soil Science Society of America, Inc.

Wen, Jiangqi, Kevin A. Lease, and John C. Walker. 2004. “DVL, a Novel Class of Small Polypeptides: Overexpression Alters Arabidopsis Development.” *The Plant Journal: For Cell and Molecular Biology* 37 (5): 668–77.

Wittkopp, Patricia J., Belinda K. Haerum, and Andrew G. Clark. 2004. “Evolutionary Changes in Cis and Trans Gene Regulation.” *Nature* 430 (6995): 85–88.

Wu, Thomas D., and Serban Nacu. 2010. “Fast and SNP-Tolerant Detection of Complex Variants and Splicing in Short Reads.” *Bioinformatics* 26 (7): 873–81.

Wu, Thomas D., and Colin K. Watanabe. 2005. “GMAP: A Genomic Mapping and Alignment Program for mRNA and EST Sequences.” *Bioinformatics* 21 (9): 1859–75.

Xiao, Yue-Hua, Qian Yan, Hui Ding, Ming Luo, Lei Hou, Mi Zhang, Dan Yao, et al. 2014. “Transcriptome and Biochemical Analyses Revealed a Detailed Proanthocyanidin Biosynthesis Pathway in Brown Cotton Fiber.” *PloS One* 9 (1): e86344.

Xiao, Yue-Hua, Zheng-Sheng Zhang, Meng-Hui Yin, Ming Luo, Xian-Bi Li, Lei Hou, and Yan Pei. 2007. “Cotton Flavonoid Structural Genes Related to the Pigmentation in Brown Fibers.” *Biochemical and Biophysical Research Communications* 358 (1): 73–78.

Yam, Kit L., and Spyridon E. Papadakis. 2004. “A Simple Digital Imaging Method for Measuring and Analyzing Color of Food Surfaces.” *Journal of Food Engineering* 61 (1): 137–42.

Yoo, Mi-Jeong, and Jonathan F. Wendel. 2014. “Comparative Evolutionary and Developmental Dynamics of the Cotton (Gossypium Hirsutum) Fiber Transcriptome.” *PLoS Genetics* 10 (1): e1004073.

Yu, Jing, Sook Jung, Chun-Huai Cheng, Stephen P. Ficklin, Taein Lee, Ping Zheng, Don Jones, Richard G. Percy, and Dorrie Main. 2014. “CottonGen: A Genomics, Genetics and Breeding Database for Cotton Research.” *Nucleic Acids Research* 42 (Database issue): D1229–36.

Zeng, Z. B. 1993. “Theoretical Basis for Separation of Multiple Linked Gene Effects in Mapping Quantitative Trait Loci.” *Proceedings of the National Academy of Sciences of the United States of America* 90 (23): 10972–76.

———. 1994. “Precision Mapping of Quantitative Trait Loci.” *Genetics* 136 (4): 1457–68.

Zhang, Jinfa, Richard G. Percy, and Jack C. McCarty. 2014. “Introgression Genetics and Breeding between Upland and Pima Cotton: A Review.” *Euphytica/ Netherlands Journal of Plant Breeding*

198 (1): 1–12.

Zhang, Tianzhen, Yan Hu, Wenkai Jiang, Lei Fang, Xueying Guan, Jiedan Chen, Jinbo Zhang, et al. 2015. “Sequencing of Allotetraploid Cotton (Gossypium Hirsutum L. Acc. TM-1) Provides a Resource for Fiber Improvement.” *Nature Biotechnology* 33 (5): 531–37.

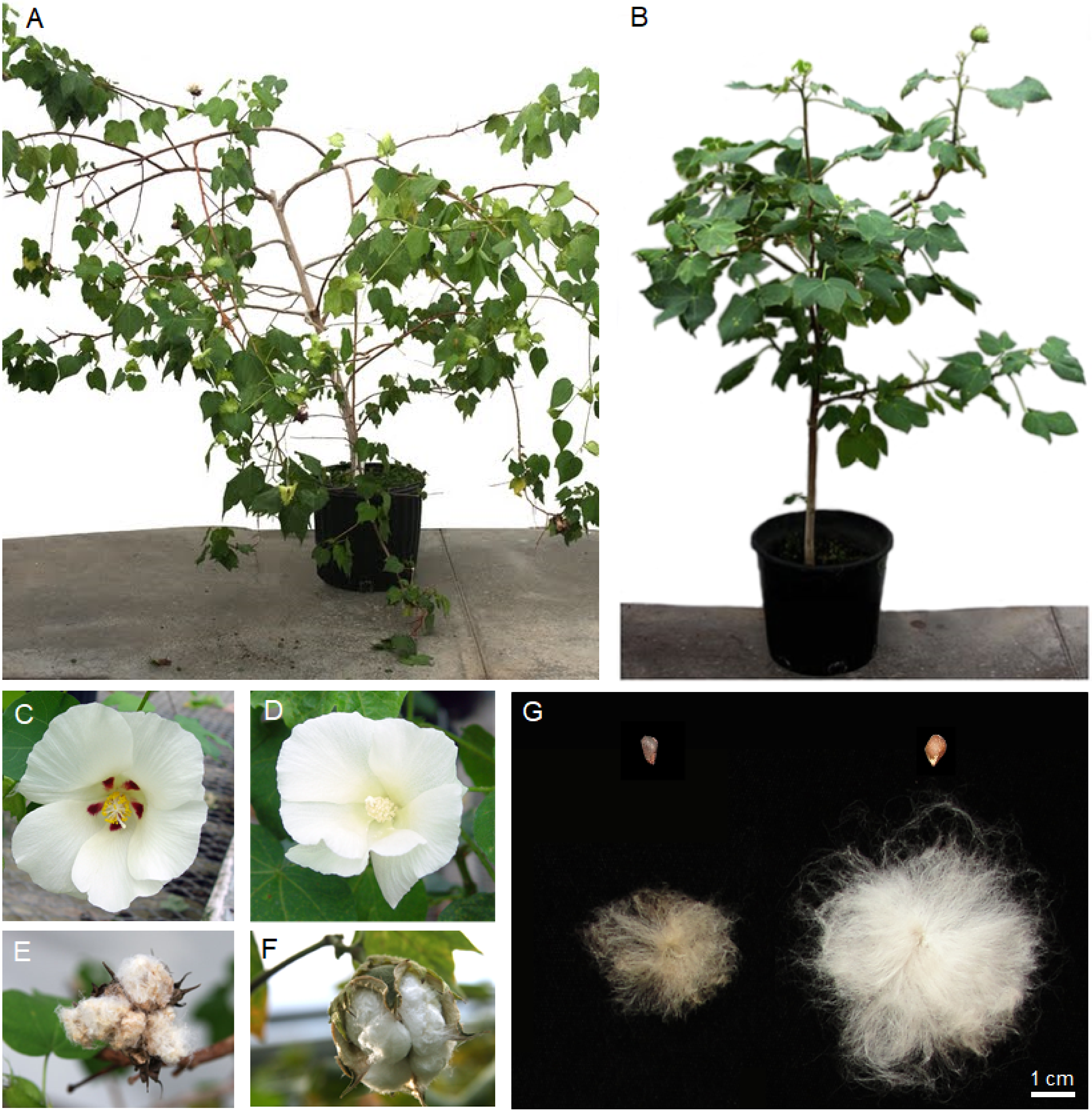
Zhang, Y., X. F. Wang, Z. K. Li, G. Y. Zhang, and Z. Y. Ma. 2011. “Assessing Genetic Diversity of Cotton Cultivars Using Genomic and Newly Developed Expressed Sequence Tag-Derived Microsatellite Markers.” *Genetics and Molecular Research: GMR* 10 (3): 1462–70.

Zhang, Zheng-Sheng, Yue-Hua Xiao, Ming Luo, Xian-Bi Li, Xiao-Ying Luo, Lei Hou, De-Mou Li, and Yan Pei. 2005. “Construction of a Genetic Linkage Map and QTL Analysis of Fiber-Related Traits in Upland Cotton (Gossypium Hirsutum L.).” *Euphytica/ Netherlands Journal of Plant Breeding* 144 (1): 91–99.

Zhao, Yunlei, Hongmei Wang, Wei Chen, Yunhai Li, Haiyan Gong, Xiaohui Sang, Feichao Huo, and Fanchang Zeng. 2015. “Genetic Diversity and Population Structure of Elite Cotton (Gossypium Hirsutum L.) Germplasm Revealed by SSR Markers.” *Plant Systematics and Evolution = Entwicklungsgeschichte Und Systematik Der Pflanzen* 301 (1): 327–36.

Figure 1. Morphological differentiation between *G. hirsutum* var. *yucatanense* TX2094 and *G. hirsutum* cv. Acala Maxxa. (A) Adult plant of TX2094, wild; (B) Adult plant of Acala Maxxa, domesticated; (C) TX2094 flower; (D) Acala Maxxa flower; (E) Open boll of TX2094; (F) Open boll of Acala Maxxa; (G) Ginned seed of TX2094 (top left) and Acala Maxxa (top right), and fiber of TX2094 (bottom left) and Acala Maxxa (bottom right). Photo credit: Kara Grupp & Mi-Jeong Yoo

Figure 2. Genetic linkage map that includes identified QTL associated with cotton fiber traits evaluated here, as generated by MapChart 2.2 (Voorrips 2002). While all chromosomes were recovered for the linkage map, only those linkage groups/chromosomes containing QTL are depicted here. QTL nomenclature follows that first used in rice (McCouch et al. 1997), which starts with “q”, followed by an abbreviation of the trait name. QTL are colored by trait category.


A
B
C
D
E
F
G
1 cm

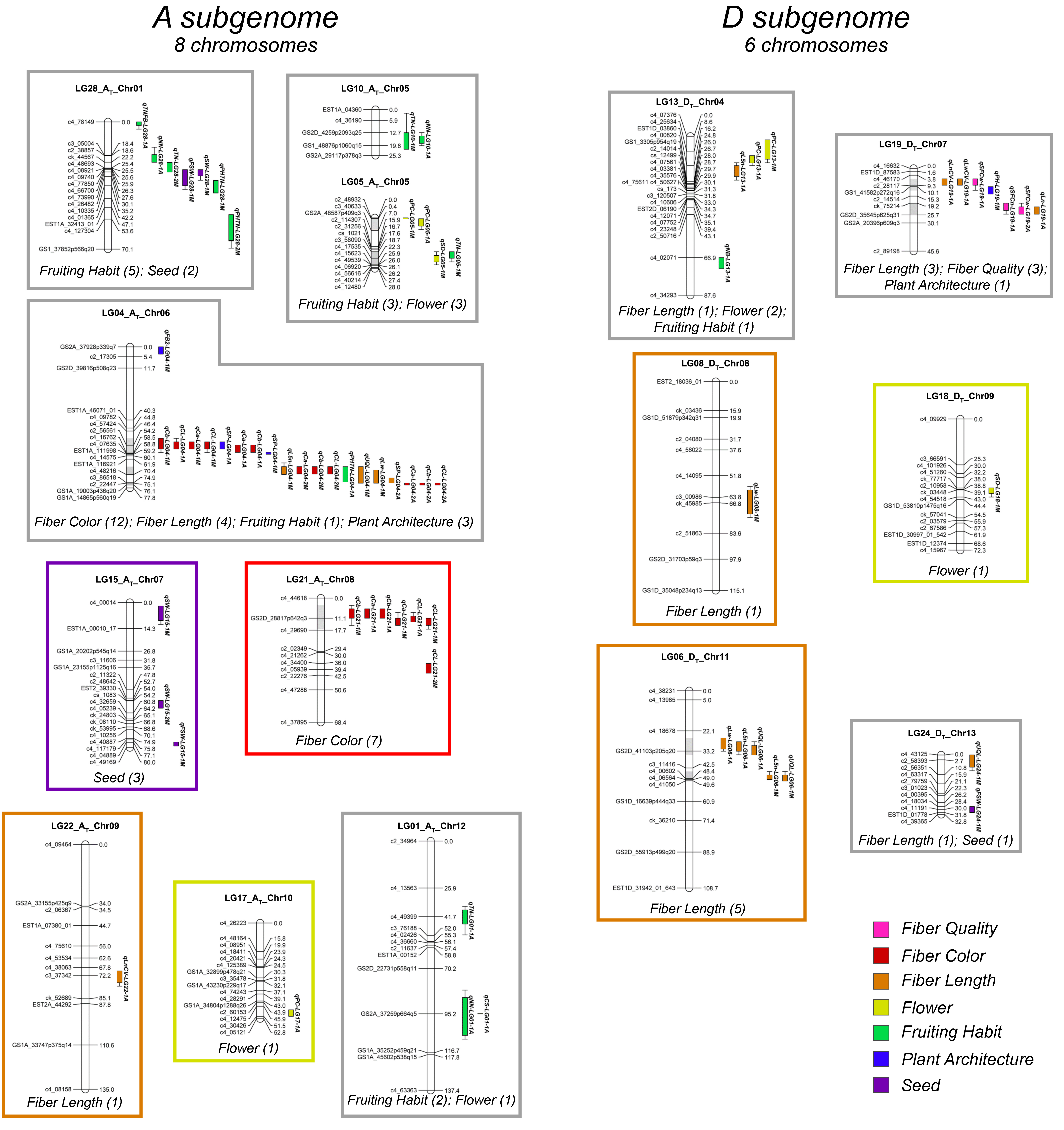

A subgenome
8 chromosomes
D subgenome
6 chromosomes
LG28_AT_Chr01
Fruiting Habit (5); Seed (2)
LG10_AT_Chr05
LG05_AT_Chr05
Fruiting Habit (3); Flower (3)
LG04_AT_Chr06
Fiber Color (12); Fiber Length (4); Fruiting Habit (1); Plant Architecture (3)
LG15_AT_Chr07
Seed (3)
LG21_AT_Chr08
Fiber Color (7)
LG22_AT_Chr09
Fiber Length (1)
LG17_AT_Chr10
Flower (1)
LG01_AT_Chr12
Fruiting Habit (2); Flower (1)
LG13_DT_Chr04
Fiber Length (1); Flower (2);
Fruiting Habit (1)
LG19_DT_Chr07
Fiber Length (3); Fiber Quality (3);
Plant Architecture (1)
LG08_DT_Chr08
Fiber Length (1)
LG18_DT_Chr09
Flower (1)
LG06_DT_Chr11
Fiber Length (5)
LG24_DT_Chr13
Fiber Length (1); Seed (1)
Fiber Quality
Fiber Color
Fiber Length
Flower
Fruiting Habit
Plant Architecture
Seed

Table 1. List of traits measured in this study. For detailed information on identified QTL, refer to Table 2.

Category	Trait
Plant architecture (12)	Plant Height (PH); Fruiting Branch Length for 1st, 3rd and 5th branches (FB1, FB2, FB3); Plant Height-to-Fruiting Branch Length Ratio (PH/FB1, PH/FB2, PH/FB3); Branch Angle of 5th sympodium; Node with red branch†; Average Stem Pubescence (SP)
Fruiting habit (7)	Total Number of Nodes (TN); Plant Height-to-Total Number of Nodes Ratio (PH/TN); Nodes to First Fruiting Branch (NF); Total Number of Non-Fruiting Branches (TNFB); Total Number of Fruiting Branches (TFB); Newly produced Nodes during 30-day interval†; Number of Fruiting Branch after 30-day Interval†
Phenology (11)	Days to First Flower (FF); Total Number of Nodes at FF (TNFF)†; Nodes to Fruiting Branch at FF†; Number of Fruiting Branch at FF†; Total Number of Flowers during 30-day interval; Average Number of Flowers/Day; Green bolls retained after 30 days + 4 week interval§; Open Bolls Retained after 30 Days + 4 Week Interval§; Green Bolls Retained after 30 Days + 4 Week Interval†; Number of bolls at 1st day of 30-day interval (NB)†; Number of bolls at 30th day of 30-day interval†
Flower (4)	Pollen Color (Yellow/Cream) (PC); Petal Spot (Present/Absent) (PS); Average Stigma Distance (mm) (SD); Curly style (Present/Absent) (CS)†
Seed (9)	50 Fuzzy Seed weight (g) (FSW); 50 Seed weight (g) (SW); Average Number of Mature Seeds (5 Bolls); Average Seeded cotton Weight (5 Bolls) (BW); Average Number of Locules (5 Bolls); Average Boll Weight (5 bolls)†; Average Weight of Locules (5 Bolls) †
Fiber length (7)	Mean Length by Number (L(n) [in]); Coefficient of Variation of the Length by Number (L(n) CV [%]); Mean Length by Weight (L(w) [in]); Coefficient of Variation of the Length by Weight (L(w) CV [%]); 2.5% Length by Number (L2.5% (n) [in]); 5% Length by Number (L5% (n) [in]); Upper Quantile Length by Weight (UQL(w) [in])
Fiber color (3)	mean L^*; mean a^*; mean b^*

Other fiber qualities (14)	Number of dust particles per gram (Dust count by g); Fineness (Fine [mTex]); Immature Fiber Content (IFC [%]);Maturity Ratio (Mat Ratio); Nep size (µm); Neps per gram; Seed Coat Nep size (SCN Size (µm)); Seed Coat Nep Count per gram (SCN (Count by g)); Short Fiber Content by Number (SFC (n) [%]);Short Fiber Content by Weight (SFC (w) [%]); Total Count per gram; Number of Trash particles per gram (Trash count by g); Trash Size [um]; Visible Foreign Matter (VFM (%))

L^* is a lightness component, ranging from 0 to 100 (from dark to bright), and a^* (from green to red) and b^* (from blue to yellow) are chromatic components for ranging from -120 to 120 (YAM AND PAPADAKIS 2004)

[†] Traits were measured in Ames (IA) population only.

[§] Traits were measured in Maricopa (AZ) population only.

Table 2. Subgenome location of linkage group based on linkage map and genomically mapped markers. The

	Linkage group	# markers	Subgenome	G. arboreum	G. raimondii
LG01	AZ1/IA1,2	14	At	Chr12	
LG02	AZ2/IA3	8	Dt		Chr08
LG03	AZ4/IA37	2	Dt		Chr08
LG04	AZ5/IA6	17	At	Chr06	
LG05	AZ6/IA16	14	At	Chr05	
LG06	AZ7/IA25	12	Dt		Chr07
LG07	AZ8/IA26	3	At	Chr11	
LG08	AZ9/IA4	11	Dt		Chr04
LG09	AZ10/IA7	17	At	Chr03	
LG10	AZ11/IA34	5	At	Chr05	
LG11	AZ12/IA27-29	13	Dt		Chr09
LG12	AZ13/IA10	9	Dt		Chr02
LG13	AZ14/IA14	22	Dt		Chr12
LG14	AZ15/IA17,18	18	At	Chr11	
LG15	AZ16,17/IA22	19	At	Chr07	
LG16	AZ18/IA8	17	At	Chr13	
LG17	AZ19/IA15	16	At	Chr10	
LG18	AZ20/IA23	15	Dt		Chr06
LG19	AZ21/IA20	10	Dt		Chr01
LG20	AZ22/IA21	3	Dt		Chr01
LG21	AZ23/IA19	11	At	Chr08	
LG22	AZ24/IA5	12	At	Chr09	
LG23	AZ25/IA11,12	11	At	Chr03	
LG24	AZ26/IA9	11	Dt		Chr13
LG25	AZ27/IA33	9	Dt		Chr05
LG26	AZ28/IA36	7	Dt		Chr03
LG27	AZ29/IA13	2	Dt		Chr05
LG28	AZ30/IA24-38	16	At	Chr01 or Chr02	
LG29	AZ31/IA31	9	Dt		Chr10
LG30	AZ32/IA30	12	Dt		Chr11
LG31	AZ33/IA32	10	At	Chr06	
LG32	AZ34/IA35	2	Dt		Chr13

number of markers used to identify the chromosomes is listed. Start and end show the position in the corr

G. hirsutum At	G. hirsutum Dt	start	end
Chr12		785,478	100,081,291
	Chr12	22,239,698	51,613,748
	Chr12	61,838,133	62,384,281
Chr06		121,378,477	11,844,977
Chr05		93,493,253	32,455,072
	Chr11	7,839,868	70,028,342
Chr11		4,372,224	1,912,510
	Chr08	2,309,559	69,752,567
Chr03		7,756,446	101,465,028
Chr05		12,447,798	16,801,699
	Chr05	63,764,922	2,523,538
	Chr01	18,196,452	62,288,071
	Chr04	3,602,330	56,438,619
Chr11		10,951,928	109,622,090
Chr07		1,830,647	88,161,909
Chr13		96,773,839	3,404,007
Chr10		106,114,806	6,056,379
	Chr09	40,676,425	1,234,789
	Chr07	48,192,627	5,155,281
	Chr07	55,033,970	55,696,830
Chr08		117,527,995	2,877,637
Chr09		2,580,082	79,333,697
Chr02		326,615	84,855,995
	Chr13	58,413,766	11,655,805
	Chr02	12,742,894	61,010,408
	Chr03	6,483,364	50,173,362
	Chr13	62,947,661	62,948,929
Chr01		100,276,884	4,271,138
	Chr06	57,362,695	64,706,897
	Chr10	62,552,661	13,976,894
Chr04		807,278	75,498,219
	Chr13	1,182,162	1,182,460

esponding G. hirsutum cv. TM-1 subgenome.

Table 3. QTL detected for traits by CIM in the F_2 populations. Only QTL with R^2 ≥7% is presented here. The full list of QTL is found in Supplementary Table 2.

Trait[a]	Lambda value	Chr[b]	QTL name[c]	Flanking markers	Position (cM)	LOD	*A*[d]	*D*[e]	D/|A|[f]	GA[g]	R^2 (%)[h]
PH	0.60	D07	qPH-LG19-1M	c4_16632-GS2A_20396p609q3	12.09	31.09	-4.7780	-0.5251	-0.1099	A	9.41
TN	2.00	A12	qTN-LG01-1A	c4_49399-c3_76188	41.71	21.33	-82.3682	109.6526	1.3313	OD	9.90
	0.90	A01	qTN-LG28-2M	c4_78149-EST1A_32413_01	25.57	33.72	-1.1893	-0.5353	-0.4501	PD	7.34
TNFF	0.00	A12	qNN-LG01-1A	GS2A_37259p664q5-GS1A_35252p459q21	92.17	20.35	-0.0629	-0.0574	-0.9120	D	12.39
TNFB	1.00	A01	qTNFB-LG28-1A	c4_78149-c3_05004	0.01	23.97	-1.2089	-2.5326	-2.0949	OD	18.47
SP	1.15	A06	qSP-LG04-1A	c4_57424-c4_48216	54.21	192.49	1.5459	0.1243	0.0804	A	36.90
	0.65	A06	qSP-LG04-1M	GS2D_39816p508q23-GS1A_14865p560q19	58.77	309.96	0.6390	0.1244	0.1946	A	52.70
	1.15	A06	qSP-LG04-2A	c3_86518-GS1A_14865p560q19	74.42	86.34	1.2357	0.2887	0.2336	PD	21.95
NB	0.50	D04	qNB-LG13-1A	c4_02071-c4_34293	68.91	25.79	-0.4971	1.6910	3.4017	OD	35.11
CS	-2.00	A12	qCS-LG01-1A	GS2A_2514p880q5	95.21	273.85	0.3325	0.3844	1.1559	D	60.45
PC	2.00	A05	qPC-LG05-1A	c2_48932-GS2A_48587p409q3	2.45	26.88	-0.2962	0.3294	1.1122	D	10.34

	2.00	A05	qPC-LG05-1M	c2_48932-GS2A_48587p409q3	0.01	35.37	-0.3439	0.1242	0.3612	PD	9.51
	2.00	D04	qPC-LG13-1A	c4_25634-c4_00820	12.63	34.64	-0.3181	0.3563	1.1202	D	11.35
	2.00	D04	qPC-LG13-1M	c4_07376-c2_14014	6.01	46.06	-0.3521	0.3280	0.9317	D	14.13
	2.00	A10	qPC-LG17-1A	c4_12475-c4_30426	49.94	22.17	-0.2487	0.3128	1.2580	OD	9.49
FSW	-0.05	A07	qFSW-LG15-1M	c2_48642-c4_49169	79.14	34.43	-0.0024	-0.0006	-0.2411	PD	7.32
SW	1.00	A07	qSW-LG15-1M	c4_00014-EST1A_00010_17	6.01	24.52	0.1300	-0.2185	-1.6807	OD	8.96
	1.00	A07	qSW-LG15-2M	c2_11322-c4_49169	56.23	44.76	0.2146	-0.0483	-0.2253	PD	10.34
SFC(n) [%]	0.15	D07	qSFCn-LG19-1A	c2_28117-GS2D_35645p625q31	21.2	35.63	0.0364	0.0159	0.4364	PD	13.42
SFC(w) [%]	-0.15	D07	qSFCw-LG19-1A	c2_28117	7.82	22.84	-0.0092	-0.0170	-1.8560	OD	8.32
	-0.15	D07	qSFCw-LG19-2A	GS2D_35645p625q31	23.2	24.45	-0.0173	-0.0118	-0.6845	PD	11.38
L(n) [in]	2.00	D07	qLn-LG19-1A	GS2D_35645p625q31	23.2	26.05	-0.0568	-0.0299	-0.5270	PD	11.02
L(n) CV [%]	-0.30	D07	qLnCV-LG19-1A	c2_28117-GS2D_35645p625q31	7.82	38.44	-0.0043	-0.0061	-1.4294	OD	12.90

L(w) [in]	2.00	A06	qLw-LG04-1M	c2_56561-GS1A_14865p560q19	74.42	42.84	0.0608	0.0025	0.0411	A	9.36
	1.50	D11	qLw-LG06-1A	GS2D_41103p205q20	30.11	23.14	0.0466	0.0029	0.0623	A	7.54
L(w) CV [%]	-1.40	D07	qLwCV-LG19-1A	c2_28117-c2_14514	7.82	26.47	-0.0003	-0.0004	-1.4838	OD	10.48
L5% (n) [in]	2.00	A06	qL5n-LG04-1M	c4_09782-GS1A_14865p560q19	67.96	38.02	0.0977	0.0588	0.6022	PD	9.26
	1.00	D11	qL5n-LG06-1A	c4_18678-c4_00602	30.11	42.30	0.0578	0.0119	0.2059	PD	12.71
UQL(w) [in]	1.00	A06	qUQL-LG04-1M	c4_48216-GS1A_14865p560q19	74.42	36.75	0.0324	0.0102	0.3142	PD	7.16
	0.85	D11	qUQL-LG06-1A	c4_18678-c4_00602	32.11	40.58	0.0393	0.0031	0.0782	A	10.48
mean *a**	0.75	A08	qCa-LG21-1M	c4_44618-c4_47288	13.1	109.27	-0.4582	-0.0905	-0.1975	A	12.01
	0.75	A06	qCa-LG04-1M	EST1A_46071_01-c4_16762	54.21	125.25	-0.8050	-0.0785	-0.0975	A	29.60
	0.75	A06	qCa-LG04-2M	c4_16762-GS1A_14865p560q19	67.96	386.83	-1.1620	0.1323	0.1139	A	61.41
	0.80	A06	qCa-LG04-1A	EST1A_46071_01-c4_07635	56.21	170.71	-1.0662	0.2848	0.2671	PD	40.77
	0.80	A06	qCa-LG04-2A	c4_07635-GS1A_14865p560q19	76.14	286.00	-1.2368	0.2860	0.2313	PD	52.27

mean *b**	1.40	A06	qCb-LG04-1M	EST1A_46071_01-c4_16762	54.21	129.23	-13.9090	-0.3856	-0.0277	A	31.49
	1.40	A06	qCb-LG04-2M	c4_16762-GS1A_14865p560q19	67.96	400.09	-20.2542	2.6334	0.1300	A	64.49
	1.30	A06	qCb-LG04-1A	EST1A_46071_01-EST1A_111998	56.21	177.12	-11.2812	3.1356	0.2779	PD	41.55
	1.30	A06	qCb-LG04-2A	EST1A_111998-GS1A_14865p560q19	76.14	283.88	-12.8132	2.9755	0.2322	PD	51.01
	1.40	A08	qCb-LG21-1M	c4_44618-c4_34400	8.01	93.80	-6.6039	-0.6124	-0.0927	A	9.02
mean *L**	1.15	A06	qCL-LG04-2M	c4_16762-GS1A_14865p560q19	67.96	242.51	13.8931	-0.9291	-0.0669	A	49.55
	1.15	A06	qCL-LG04-1M	EST1A_46071_01-c4_16762	54.21	100.26	9.3721	1.4697	0.1568	A	24.26
	0.60	A06	qCL-LG04-2A	c4_07635-GS1A_14865p560q19	76.14	256.19	0.7360	-0.0997	-0.1355	A	49.05
	0.60	A06	qCL-LG04-1A	EST1A_46071_01-c4_07635	54.21	168.46	0.6272	-0.0974	-0.1553	A	36.88
	1.15	A08	qCL-LG21-1M	c4_44618-c4_34400	13.1	71.45	5.7489	0.5838	0.1016	A	11.24

---

[a] Trait information is provided in Table 1[*PH* Plant Height; *TN* Total Number of Nodes; *TNFF* Total Number of Nodes at First day of Flowering; *TNFB* Total Number of Non-Fruiting Branches; *SP* Average Stem Pubescence; *NB* Number of bolls at 1st day of 30-day interval; *CS* Curly style (Present/Absent); *PS* Petal Spot (Present/Absent); *PC* Pollen Color (Yellow/Cream); *FSW* 50 Fuzzy Seed weight (g) (FSW); *SW* 50 Seed weight (g); *Fine [mTex]* Fiber Fineness; *Mat Ratio* Maturity Ratio; *SFC(n) [%]*Short Fiber Content by Number; *SFC(w) [%]* Short Fiber Content by Weight; *L(n) [in]* Mean Length by Number; *L(n) CV [%]* Coefficient of Variation of the Length by Number; *L(w) [in]* Mean Length by Weight; *L(w) CV [%]* (L(w) [in]); *L5% (n) [in]* 5% Length by Number; *UQL(w) [in]* Upper Quantile Length by Weight (UQL(w) [in]); *mean a*, mean b**, *mean L** Fiber Color]

[b] The corresponding genome of *G. hirsutum* TM1 was indicated. A and D represents A and D subgenome, respectively.

[c] QTL name is provided as follows: the first two to four letters excluding "q" indicate trait name, following by linkage group (LG). The last letter indicates population where QTL was detected; *A* Ames, IA; M Maricopa, AZ.

[d] *A* additive effects. Positive/negative signs indicate alleles from Maxxa/TX2094 increase trait value, respectively.

[e] *D* dominant effects. Positive/negative signs indicate the effects increasing/decreasing trait value over the population mean

[f] *D/|A|* dominance/|additive|

[g] *GA* gene action; *A* additive (|d/a| = 0-0.2); *PD* partial dominance (|d/a| = 0.21-0.8); *D* dominance (|d/a| = 0.81-1.2); *OD* overdominance (|d/a| >1.2)

[h] Percentage of phenotypic variation explained by the marker genotype at the corresponding marker

Table 4: Number of genes in any QTL, or for QTL related to a specific trait, that also exhibit additional c
cotton

	Total	genes with non-silent changes *	genes with non-synonymous changes	differentially expressed **	transcription factors
All QTL	13,724	6,134	3,807	NA	79
Architecture	2,078	957	594	NA	13
Fiber Color	3,087	1,298	816	236	16
Fiber Length	4,329	1,930	1,178	345	27
Fiber Quality (other)	968	452	266	106	8
Flower	1,926	889	555	NA	10
Fruiting Habit	3,266	1,542	982	NA	12
Seed	3,306	1,513	921	NA	23

* includes start/stop adjustments and SNPs in UTR

** DGE only applies to fiber-related traits differences between wild and domesticated

known cotton genes
99
12
11
34
13
10
10
9