TITLE: **Lessons Learned and Challenges Ahead for Cotton Genome Mapping**

DISCIPLINE: **Genomics and Biotechnology**

AUTHORS: M Ulloa (Corresponding author)
USDA-ARS, Western Integrated Cropping Systems Research Unit
17053 N. Shafter Ave.
Shafter, CA 93263
Phone: 661-746-8009
Fax: 661-746-1619
Email: mauricio.ulloa@ars.usda.gov

S. Saha
USDA-ARS, Genetics and Precision Agriculture Research Unit
810 Highway 12 E.
Mississippi State, MS 39762

J.Z. Yu
USDA-ARS, Crop Germplasm Research Unit
2765 F& B Rd.
College Station, TX 77845

J.N. Jenkins
USDA-ARS, Genetics and Precision Agriculture Research Unit
810 Highway 12 E.
Mississippi State, MS 39762

W.R. Meredith
USDA-ARS, Crop Genetics and Production Research Unit
Stoneville, MS 38776

R.J. Kohel
USDA-ARS, Crop Germplasm Research Unit
2881 F& B Rd.
College Station, TX 77845

ACKNOWLEDGMENTS: We would like to thank Dr. Dale Spurgeon, Todd Campbell and the anonymous reviewers for their suggestions to make this a better manuscript. We would also like to thank Sherry Ellberg and Annette Williams for their assistance in running gel electrophoresis.

DISCLAIMER: Mention of trade names or commercial products in this article is solely for the purpose of providing specific information and does not imply recommendation or endorsement by the U.S. Department of Agriculture or the University of California.

ABBREVIATIONS: RFLP, restriction fragment length polymorphism; SSR, simple sequence repeat; SNP, single nucleotide polymorphism; EST, expressed sequence tag; BAC, bacterial artificial chromosome; SSCP, single-strand conformation polymorphism.

Lessons Learned and Challenges Ahead for Cotton Genome Mapping

ABSTRACT

The goal of genetic linkage mapping is to discover genes for improving trait performance. The process of manipulating the cotton genome is complex because the cotton fibers (lint) used in textiles are derived from the seed trichomes (hairs) of four *Gossypium* species. In just over a decade, our understanding of the structure of the cotton genome has expanded tremendously. Prior to 1994, genetic linkage mapping in cotton was limited to linkage detection among morphological markers and the placement of those markers on chromosomes using cytogenetic stocks. Since then, approximately 27 major mapping efforts in *Gossypium* have emerged. In addition, PCR-based markers, such as microsatellites, have been placed on 26 cotton linkage groups/chromosomes. Over the past decade new tools have emerged (e.g., BAC physical mapping, differential display, microarrays, and SNPs), providing new avenues for gene discovery that circumvent some of the limitations of genetic linkage mapping. Global research efforts, such as the International Cotton Genome Initiative (ICGI) and Cotton Marker Database (CMD), have also provided opportunity for wide collaboration. However, difficulties have risen from the use of different electrophoretic systems (RFLP, gel systems, and dye-primers) for calling alleles on mapping populations. In addition, the need for standard nomenclature in naming molecular markers, as well as in identifying quantitative trait loci (QTL) in the cotton genome, continues to exist. Real gains in cotton improvement will require an unprecedented understanding of the molecular genetics of complex traits, and expanded genetic linkage maps that can be used to manipulate the cotton genome in ways that were previously inconceivable. Herein, we provided an overview of the progress, future, and challenges facing genetic mapping efforts in cotton.

Cotton (*Gossypium* spp.) is the leading natural textile fiber of the world and the world's sixth largest source of vegetable oil. In just over a decade, our understanding of the structural organization of the cotton genome has progressed tremendously (Ulloa et al., 2007). While much attention has been focused on genetic linkage maps and QTL analyses in the recent literature, the fundamental analytical process of genetic linkage analysis dates back to Mendel. In those days, geneticists could only study traits identifiable by discrete phenotypes that were controlled by one, or at most a few genes, and their interactions (Smith et al., 1999). However, the critical agronomic traits (e.g., yield, fiber quality, and most disease and pest resistance) are controlled by complex interactions of many genes. These genes are distributed across the genome and act to produce wide and continuous phenotypic variation.

To understand the complexities of genetic linkage mapping in cotton, we must first understand the evolution and structure of the cotton genome. The cotton fibers (lint) used in textiles are derived from the seed trichomes (hairs) of four *Gossypium* species: two Old World diploid species, *G. arboreum* L. and *G. herbaceum* L. (A genome), and two New World allotetraploid species, *G. barbadense* L. and *G. hirsutum* L. (AD genome). The singular feature of the cotton fiber that sets it apart from other plant seed trichomes is the fact that the cellulose microfibrils are laid down in a spiral arrangement. When the boll opens and the seed trichomes desiccate, they collapse into convoluted ribbons; this is what makes cotton fibers 'spinnable'. The most widely cultivated of the four cotton species is *G. hirsutum*, also known as Upland or Acala cotton. *G. barbadense* (Pima or Egyptian cotton) contributes the extra long staple lint that occupies a significant niche in the cotton market. In addition, Pima cotton is typically used as a parent to generate genetic populations and/or segregating families for genetic linkage mapping. *G. arboreum* and *G. herbaceum*, which are still cultivated in Africa and Asia, are important

because they are the only extant representatives of the lineage in which the complex seed trichome first evolved (Saunders, 1961). These two ancestral species are therefore critical experimental models for gene discovery. Also important are the New World D genome diploid species which have served as important sources of genes for lint quality even though none of the species produces commercial lint-fibers. Five of these species are adapted to the desert environments of Baja California, Mexico [*G. armourianum* Kearney, *G. harknessii* Brandegee, and *G. davidsonii* Kellogg] and NW mainland Mexico [*G. turneri* Fryxell and *G. thurberi* Todaro]. An additional seven species [*G.* sp. nov (Ulloa et al., 2006), *G. aridum* (Rose & Standley) Skovsted, *G. lobatum* Gentry, *G. laxum* Phillips, *G. schwendimanii* Fryx. & Koch, *G. gossypioides* (Ulbrich) Standley, and *G. trilobum* (Mociño & Sessé ex DC.) Skovsted,] are located in the Pacific coast states of Mexico and, with the exception of the last species, are arborescent in growth habit (Ulloa et al., 2006). Finally, *Gossypium raimondii* Ulbrich is endemic to Peru, while *G. klotzschianum* Andersson is found in the Galápagos Islands.

While the mapping of morphological traits continues (Endrizzi and Nelson, 1989; Percy 1999; Kohel and Bird, 2002), the arrival of molecular markers (RFLPs, RAPDs, AFLPs, SSRs, SNPs, etc.) has dramatically altered the utility and application of genetic linkage mapping in cotton. Although isozymes can rightly claim to be the first "molecular" markers, it was not until DNA sequence differences among organisms could be visualized that the true era of molecular markers began. The advantage of molecular markers is that they are phenotype neutral whereas morphological markers may be difficult to maintain through the breeding process or may have deleterious effects on other traits. Linkage analyses with morphological markers and correlations with chromosome behavior (cytogenetics) helped the development of the framework for mapping molecular markers. With a nearly unlimited pool of genetic markers, cotton

geneticists could construct linkage maps of entire genomes that could be used to dissect complex traits (Mei et al., 2004; Lin et al., 2005; Shen et al., 2005; Ulloa et al., 2005), locate genes (Rungis et al., 2002; Zhang and Stewart, 2004; Feng et al., 2005; Zhang et al., 2005), and compare the structure of related genomes (Brubaker et al., 1999; Rong et al., 2004; Desai et al., 2005; Rong et al., 2007).

The severe genetic bottlenecks that the cultivated cotton genomes passed through during domestication clearly limited the amount of genetic diversity available for plant selection. It is likely that many useful alleles are still to be found outside the current cultivated gene pools. Introgressing these traits requires wide or interspecific crosses. Genetic linkage maps of these interspecific populations will be invaluable in documenting and tracking the introgression of alien chromatin into donor genomes. The most common population type used for mapping the tetraploid cotton genome is the interspecific F_2. Other types of populations used include backcross (BC_1) families, recombinant inbred lines (RIL), and doubled haploid (DH) families (Ulloa et al., 2007).

The fundamental goal of genetic linkage mapping is gene discovery, or when the gene itself is yet to be located, the identification of a linked proxy that can be used in marker assisted selection. These are the primary tools breeders will use to reach their crop improvement objectives in the next decade. The sophisticated genetic linkage maps developed during the past decade are the result of our ability to look at DNA sequence differences directly with molecular markers, and to use powerful computers to analyze genetic linkages at scales that were previously impossible to analyze. Prior to 1994, genetic linkage mapping in cotton was limited to linkage detection among morphological markers and the placement of these morphological markers on chromosomes using cytogenetic stocks. In 1985, the diploid A-genome map

comprised 18 morphological markers in 7 linkage groups, and the tetraploid AD-genome map comprised approximately 80 markers distributed across 18 linkage groups (Endrizzi et al., 1985). Now, almost fully resolved molecular genetic maps for the diploid A genome and D genomes and the tetraploid AD genome are available (Rong et al., 2004; Frelichowski et al., 2005; Guo et al., 2007; Rong et al., 2007), and all 26 AD chromosomes or linkage groups (and their At and Dt homoeologs) have been linked to the cytological maps (Rong et al. 2004; Wang et al., 2006). In addition, PCR-based markers, such as microsatellites, haven been placed on the 26 cotton linkage groups/chromosomes (Frelichowski et al., 2005; Guo et al., 2007; http://www.cottonmarker.org/; previously http://www.mainlab.clemson.edu/cmd/). The purpose of this manuscript is to provide the reader with an overview of the past, present, and future of cotton genome research, with emphasis on structural genomics (genetic and QTL mapping).

MATERIAL AND METHODS

Mapping populations. Populations and genetic information in this study were published previously except for additional markers added to linkage group-chromosome 3 (Fig. 1; Park et al., 2005; Frelichowski et al., 2006) and additional electrophoresis evaluations done under agarose, acrylamide, single-strand conformation polymorphism (SSCP) gel systems, and using a DNA analyzer (see below).

In the first mapping population (Pop 1) previously reported (Shappley et al., 1998), 96 bulk-sampled plots from $F_{2.3}$ ('HS46' x 'MAR') families were used to isolate DNA for genetic mapping. DNA was also isolated from a second mapping population (Pop 2) composed of 119 bulk-sampled plots from $F_{2.3}$ (MD5678ne X 'Prema') families (Ulloa and Meredith, 2000), a third population (Pop 3) of 199 bulk-sampled plots from $F_{2.3}$ (HQ95-6 X MD51ne) families, and a fourth population (Pop 4) of 150 bulk-sampled plots from $F_{2.3}$ (119-5 X MD51ne) families

(Ulloa and Meredith, 2002). The complete set of RFLP molecular data (scored bands) from each population was used to develop the cotton joinmap (Ulloa et al., 2002; Ulloa et al., 2005).

Marker analysis. PCR amplification of molecular markers such as BNL, CIR, NAU, MUSB, MUCS, and MUSS was performed in a total volume of 15 µl containing 20 ng of template genomic DNA, 0.1 µM of each primer (forward and reverse), 1X PCR buffer, 0.2 mM dNTPs , and 1 U of Taq polymerase (Amplitaq, Applied Biosystem, Foster city, CA) with the following cycling profile: 1 cycle of 2 min. at 94°C, 10 cycles of 15 sec. at 94°C, 30 sec. at 60°C (step -0.5°C/cycle for cycles 2-10), and 1 min. at 72°C, 35 cycles of 15 sec. at 94°C, 30 sec. at 55°C, and 1 min. at 72°C.

PCR products were separated on a 3% super fine resolution (SFR^{TM}) agarose gel (Amresco, Solon, OH) containing 1X TBE at 80 volts for 4-5 hrs, and visualized by Alphaimager software (v. 5.5, Alpha Innotech Corporation, San Leandro, CA) after staining with ethidium bromide. Primer-pairs resulting in discrete PCR banding patterns denoting a molecular marker were scored.

For 10% acrylamide (Bio-Rad Laboratories, Hercules, CA) and single-strand conformation polymorphism (SSCP) gel systems (MDE^{TM} gel solution, Lonza Rockland Inc., Rockland, Maine), PCR products were separated on a Sequi-Gen GT Nucleic Acid Electrophoresis Cell (Bio-Rad Laboratories, Hercules, CA) containing 1X and 0.6X TBE at 3 Watts for 5 and 12 hrs, respectively. Staining was performed with a DNA Silver Staining kit (Amersham Biosciences Company, Piscataway, NJ) for visualization.

Fluorescent-labeled (RED Well Dye D3, D4) selected PCR primers were synthesized by Sigma-Aldrich (Sigma-Proligo, The Woodlands, TX). PCR amplification was performed on a total volume of 15 µL containing 2µL of DNA template (concentration 10 ng), 0.1 µM each of

forward and reverse primers, 1X PCR buffer, 3 mM of $MgCl_2$, 0.2 mM of dNTPs, and 0.5 U of Gold Taq polymerase (Amplitaq, Applied Biosystem, Foster City, CA) with cycling profile of 1 cycle of 10 min at 94°C (hot start); 10 cycles of 15 sec at 94°C, 30 sec at 60°C (step -0.5°C/cycle for cycles 2-10), and 1min at 72°C; 35 cycles of 15 sec at 94°C, 30 sec at 55°C, and 1 min at 72°C; final extension cycle of 6 min at 72 °C. PCR products were first checked and separated on 3% SFR™ agarose gel (Amresco, Solon, OH) containing 1X TBE at 90 V for 4-5 h, and visualized using Alphaimager v. 5.5 after staining with ethidium bromide. After visualization, SSR markers with discrete PCR banding patterns were again electrophoresed using a capillary DNA Analyzer system (CEQ 8000 Beckman Coulter Company, Fullerton, CA).

Construction of genetic linkage map. Genomic DNA was prepared from 183 recombinant inbred lines (RILs) generated from a cross between an Upland cotton (*Gh* TM1) and a superior fiber quality Pima (*Gb* 3-79) (Kohel et al., 2001; Kohel and Yu, 2004; Park et al., 2005; Frelichowski et al., 2006). A genetic linkage map was constructed by using the JoinMap 4.0 computer program (Van Ooijen, 2006). The Kosambi map function was used to convert recombination frequency to genetic map distance (centimorgan, cM). Most of the linkage groups were determined at LOD scores ≥ 6. Deviation from a 1:1 segregation ratio expected for RILs was tested by a χ^2 statistic for each marker.

QTL analysis of fiber quality-related traits. The detection of QTLs for bundle fiber strength (*Sf*), length at 2.5% (2.5% *Lf*), and fineness (*Ff*) were reported using RFLP, RAPD, and SSR markers (Kohel et al., 2001; Ulloa and Meredith, 2002; Park et al., 2005; Ulloa et al., 2005; Frelichowski et al., 2006). QTL analysis was conducted using MapQTL 4.0 (Van Ooijen and Maliepaard, 1996). Composite interval mapping (CIM) (Zeng, 1994) with the MQM module was performed in a stepwise forward regression procedure to detect putative QTL position using

a likelihood ratio (LR) test statistic. Empirical threshold values for LR were determined after 1000 permutation tests for all traits (Churchill and Doerge, 1994).

Single Nucleotide Polymorphism (SNP) markers. Herein, we use the example of the *GhEXPA6-1* gene family from An et al. (2007) to illustrate the problems and potential solution for SNP marker discovery in tetraploid cotton species. Specific primers were used to amplify genes of known sequence from several genotypes of tetraploid species and two diploid ancestral species (A and D genomes) of tetraploid cotton (At and Dt subgenomes). DNA PCR-amplicons were cloned and sequenced from individual plants to [1] identify the sequence variation in these DNA amplicons, [2] use sequences from different genotypes to construct a phylogenetic tree using PAUP software (Swofford, 1998), [3] discover putative SNPs from comparative analysis of the consensus sequences within individual phylogram clades (considered as putative loci), [4] discover the sub-genomic origin of the sequences based on the positions in the phylogenetic tree of the sequences with the diploid A and D ancestral species, and [5] identify the chromosomal location(s) of the SNP markers using the deletion method, thus confirming the true allelic nature of the putative SNP marker (An et al., 2006; Buriev et al., 2006a, 2006b; Hsu et al., 2007).

RESULTS AND DISCUSSION

RFLP assessment of QTLs on the genetic linkage joinmap from 4 populations.

Based on 111 shared RFLP loci, the percentage of common heterozygous loci between populations (Pop 1 – 4) varied from 9.0% to 41.0%. Identification of homologous RFLP loci in a linkage group reveals the approximate number of loci of cDNA-probes monomorphic in one population but polymorphic in the other. Knowledge of homology represents a means to greatly increase the informativeness of the genetic joinmap in the tetraploid Upland, *G. hirsutum*, cotton. The cDNA probes used to construct the joinmap detected multiple segregating loci. This level of

segregation has been observed in previous studies (Kianian and Quiros, 1992). We placed 188 QTLs from the four populations on 21 linkage groups of the joinmap. Eight linkage groups carried QTLs at the same locus or within of the linkage group from at least two of the four different populations. Linkage group 11, which is a part of chromosome 5, carried QTLs from two different populations close to each other or within neighboring regions of locus C115A1V. Locus C115A1V was polymorphic in two populations, and the heterozygote condition (family means with both alleles, one allele from each parent) had the highest values for lint yield (1074 kg ha^{-1}) and for fiber strength (234.0 kN m kg^{-1}, Table 1). Approximately 60% of the putative QTLs for fiber quality traits were placed on two cotton chromosomes. Chromosome 3, with approximately eleven linkage groups containing 31 detected putative QTLs from the four cotton populations, explained from 4.7 to 38.5% of the trait variation (Table 2). Chromosome 26, with approximately nine linkage groups containing 36 detected QTLs, explained from 3.4 to 44.6% of the trait variation (Table 3). For the most densely populated linkage groups with markers in the joinmap, groups 1 (part of chromosome 3) and 2 (part of chromosome 26), the order of RFLP loci changed moderately or drastically depending on the locus.

Based on the genome-specific linkage group/chromosome for the *G. hirsutum* tetraploid cotton, the At subgenome exhibited 64 QTLs from five chromosomes (3, 5, 9, 10, and 12). In contrast, the D subgenome exhibited 52 QTLs from three chromosomes (14, 20, and 26; Tables 1 - 3; Ulloa et al., 2005). This result agrees with previous studies, which indicated that fiber trait variability in *G. hirsutum* may predominately occur on certain linkage groups/chromosomes or may be clustered (Shappley et al., 1998; Ulloa et al., 2000; Ulloa et al., 2005; Rong et al., 2007). Characterization of the comparative organization of heterozygous RFLP loci in linkage groups from different populations increases the marker density of the Upland cotton joinmap and

facilitates both genetic and physical mapping applications. QTL examination of the joinmap for agronomic, fiber quality and physiological traits reveals highly recombined and gene-abundant regions in the cotton genome.

The RFLP marker system has the largest number of publicly available markers (Rong et al., 2004; Ulloa et al., 2005) but it is a marker system that is neither noted for high throughput nor ease of use as a tool in a breeding program to speed cultivar improvement The RFLP markers and the genetic linkage maps based on this type of marker will only remain useful as the RFLPs are converted into Polymerase Chain Reaction (PCR) based markers or other marker types.

Microsatellite or Simple Sequence Repeat (SSR) markers. This marker development effort within the cotton community continues. A recent advancement is the establishment of an *ad hoc* bioinformatics resource, the Cotton Marker Database (CMD; at http://www.cottonmarker.org, sponsored by Cotton Incorporated). At present CMD contains publication, sequence, primer, mapping and homology data for nine major cotton microsatellite projects, collectively representing 8,213 microsatellites. The collection of publicly available cotton SSR markers in a centralized, readily accessible and curated web-enabled database provides a more efficient utilization of microsatellite resources and will help to accelerate basic and applied research in molecular breeding (Blenda et al., 2006).

The International Cotton Genome Initiative (ICGI) was launched to facilitate the growth of cotton genomic knowledge and resources including, but not limited to, a saturated and fully integrated genetic and physical map of the cotton genome. A consensus linkage map is being developed from consolidated data generated by the cotton community, using a common set of framework markers such as microsatellites, or SSRs. Since 2005, great progress has been made

in the development of PCR based-markers and the mapping of these markers in a consensus map. Additional molecular markers have been mapped and placed on the 26 different cotton linkage groups/chromosomes, increasing the number of mapped loci in the existing CMD database to approximately 1,300. At the USDA-ARS at Shafter, CA, we continue to identify and map additional markers of previously unknown usefulness for genome analysis between Upland/Acala and Pima cottons. Polymorphisms (differences between different types of cottons) were discovered using a panel of cottons which included four different species (*G. hirsutum*, *G. barbadense*, *G. arboreum*, and *G. raimondii*). These markers were determined to be polymorphic because of changes in DNA organization at the chromosome level or because of single nucleotide polymorphisms (SNP). Fig. 1 shows the progress in marker saturation for chromosome 3.

Improved resolution and integration of cotton genome maps are desirable in order to reduce or close gaps in genome regions with very low or very high rates of recombination. Several methods have been suggested to increase map resolution. One strategy is to unite different genetic maps using common markers (Ulloa et al., 2002). Physical mapping is another valuable method of mapping genome regions with very high or very low rates of recombination (Yu et al., 2004; Ulloa et al., 2005; Xu et al., 2007; Rong et al., 2007) because such regions are difficult to resolve by segregation analysis. In regions of high rates of recombination, a large number of polymorphic markers are required to close gaps in the linkage map. Comprehensive structural, functional and comparative studies of any genome are increasingly dependent upon the availability of an anchored physical map which shows the order of all genetic components in correspondence to their chromosomal localization. Anchoring of phenotypic information (such as trait or QTL) onto the physical map requires its integration with the genetic map representing

the relative positions of the genes and/or markers on chromosomes. It is essential to develop a cotton physical map if we plan to clone QTL/genes. Chromosome walking to a gene may take several years if the physical distance of a QTL/gene is unknown.

Single-Nucleotide Polymorphism (SNP) markers. Currently, considerable attention is focused on the development of single-nucleotide polymorphism (SNP) marker systems, as they allow scientists to access sequence differences directly. In contrast, markers based on electrophoretic differences, including RFLP, AFLPs and SSRs can mask underlying genetic variation. However, the complex polyploid cotton genome will continue to complicate marker discovery and application. One of the primary forces impeding genetic improvement of cotton is the lack of information at the molecular level about genes that control important traits. Many of these genes are highly conserved both at the sequence and size level and are inappropriate for detection using a conventional PCR and agarose gel system.

SNPs are single-base variations in the genetic code usually represented as two, or sometimes three, different bases at a single position (Brookes, 1999). SNP markers have become an efficient tool in genome mapping, as they are reported to be the most abundant and polymorphic markers. SNPs are normally biallelic markers scattered across the whole genome and can also be associated with functional genes. The development of cost-effective techniques for finding SNPs is sometime challenged by the complexity of the cotton genome due to polyploidy. The disomic polyploids, such as cotton, usually contains two divergent "paralogous" copies of each gene (one per subgenome), which may complicate segregation analysis in genetic mapping. A group of physically linked SNPs that are co-inherited as a series of alleles is considered a haplotype. It is important to identify haplotypes that can distinguish allelic differences at a single locus in a polyploid system because of presence of duplicated loci.

Efficient discovery of SNP markers needs to distinguish differences between subgenomes (At and Dt) as well as allelic variants at a locus. The success of overcoming this problem will depend on the power of markers to detect polymorphism at a single locus.

Herein, we report a strategy to identify SNP markers in tetraploid cotton. We utilize the example of the *GhEXPA6-1* gene family from An et al. (2007) to illustrate this strategy. Original sequence information was collected from NCBI GenBank of the *GhEXPA6-1* gene from *G. hirsutum* cv Siokra 1-4 (Harmer et al., 2002). A gene specific primer pair was utilized to amplify genomic DNAs from individual plants of four tetraploid species, *G. hirsutum*, AD_1 [TM-1, HS46 and MARCABUCAG8US-1-88 (MAR)], *G. barbadense*, AD_2 (3-79), *G. tomentosum* Nuttall ex Seemann, AD_3, *G. mustelinum* Miers ex Watt, AD_4, and two diploid genome species, *G. arboreum*, A_2 and *G. raimondii* Ulbrich., D_5, respectively. The PCR products were cloned and sequenced using standard procedures. Multiple clones from individual amplified products and at least three identical sequences from each amplicon were used for the analysis to minimize error in sequencing. The sequences were used to construct phylogenetic trees. The putative ancestral and homoeologous relationships of the *GhEXPA6-1* gene family members among the genotypes were detected by phylogenetic grouping and comparison to the diploid ancestral A- and D-genome relative species with the sequences of tetraploid species of AD-genome cottons. Results showed the presence of two broad clades in the phylogram (Fig. 2; An et al., 2007).

The sequences from an individual clade of the phylogenetic tree were aligned and compared to detect the putative SNPs. The unique combination of SNPs in a sequence within a clade of the phylogram was used to detect the haplotype (Tables 4, 5). Each clade in the dendogram was considered as a putative locus and haplotype group. The hypothesis was based on the assumption that sequences at each locus will be more similar than sequences between the

loci. However, this strategy of identifying haplotype based on parsimonious analysis, without prior genetic knowledge, may separate significantly different alleles of a locus in two different clades. Therefore, it could sometimes misrepresent allelic differences as locus differences. Under such circumstances true SNPs may not be identified.

To further validate the true allelic nature of the SNP marker, we identified the chromosomal location of the SNP markers associated with the *GhEXPA6-1* gene using deletion analysis. Primers specific to the SNP markers were designed to anneal immediately upstream or downstream of the SNP site as the forward or reverse primer, respectively, so that the polymorphism could be detected by one-base extension technology with the ABI Prism SNaPshot^{TM} multiplex kit (Applied Biosystems, Foster City, CA). Several SNP markers associated with the *GhEXPA6-1* gene family members were delimited to the long arm of chromosome 3 by hypoaneuploid cytogenetic deficiency tests (An et al., 2007). This result further confirmed the true nature of the SNP markers because it would not be possible to detect a missing marker associated with a false SNP allele in the aneuploid plant. This strategy has the potential to overcome the problems in the discovery of SNP markers associated with the presence of duplicated loci in many polyploid species. The use of SNP markers will facilitate the association of candidate genes with complex traits and map-based cloning of these genes.

Difficulties arising from different systems and comparisons. In just over a decade, our understanding of the structural organization of the cotton genome has progressed tremendously. Prior to 1994, genetic linkage mapping in cotton was limited to linkage detection among morphological markers and the placement of those markers on chromosomes using cytogenetic stocks. Since then, approximately 27 major mapping efforts in *Gossypium* have emerged (Table 6). Twenty-two of those efforts have addressed mapping of the tetraploid

genome, and a few have addressed diploid maps. Those mapping efforts have increased the need for standard nomenclature of markers, as well as identified QTLs. In addition, use of different electrophoresis systems (RFLP, agarose, acrylamide, SSCP gels, and dye-primers) for calling alleles on mapping populations has created confusion when researchers have attempted to merge or join data from different mapping projects. Global research efforts, such as the International Cotton Genome Initiative (ICGI) and Cotton Marker Database (CMD) can facilitate wide collaboration not only in efforts to saturate and expand genetic maps, but also in the standardization of nomenclature of DNA markers and QTLs.

Lint yield and fiber quality are the two traits most important to cotton production. A number of fiber properties are used to measure or estimate fiber quality, namely, fineness, length, strength, micronaire (estimator of fiber fineness), and elongation (May, 1999; Ulloa et al., 2006). There is also a variety of different ways to estimate these fiber properties, which can confound the results of studies to associate QTLs with particular fiber properties. For example, measurements of micronaire can be confounded by the relationship between fiber fineness and the maturity of the fibers (May, 1999). In some cases, comparisons across QTL studies are not appropriate because it is not reasonable to combine different estimators (Rong et al., 2007). Another source of confusion is that the parents of each population could have different levels of marker and trait polymorphisms. Each population is a unique sample of the germplasm and a given set of QTLs will only be of direct value in a specific population.

Results of QTL comparisons from cotton populations (Ulloa et al., 2005; Rong et al., 2007) were consistent with results of studies in other crops (Beavis, 1994; Austin and Lee, 1996; Groh et al., 1998; Melchinger et al., 1998; Monforte and Tanksley, 2000). Consistency of QTLs in different environments and/or genetic backgrounds was found to be modest to low. Even

though consistencies of QTLs were found in some instances to be as low as 20% between four experiments, QTL examination for agronomic and fiber quality traits revealed highly recombined and gene-abundant regions on cotton chromosomes (Ulloa et al., 2005; Tables 1 – 3, 7). In addition, gene members of the lint fiber development network appear to be clustered with heterogeneous phenotypic effects (Ulloa et al., 2005; Rong et al., 2007).

The validation of QTLs for yield and fiber traits in the cotton crop remains to be confirmed. Beavis (1994), assessing QTLs in corn, suspected that the use of more powerful or modified analytical techniques would have a minor impact on analyses of polygenic traits because the most important factors are primarily functions of the experimental design, i.e., the numbers and types of progeny as well as the field plot designs that are used to evaluate traits. To realize the full potential of molecular markers in breeding for polygenic traits, it may be necessary to develop new marker-assisted breeding methods, rather than augmenting existing methods with marker-aided selection. The selection of QTL regions to transfer with marker-assisted selection (MAS) and/or to consider in a selection index should be based on QTL effects verified in an independent validation sample (Melchinger et al., 1998).

Future of cotton structural genomics. Fiber, biotic and abiotic stress EST single-gene sets using cDNA microarrays in standard and mutant (gene knockouts) germplasm will be used to develop genome-wide expression profiles of genes. The expression profiles will facilitate the classification and categorization of genes with similar expression patterns within a developmental framework, regardless of whether the gene function is known. These efforts are likely to be augmented by more extensive and innovative use of mutants. Mutagenized populations have been instrumental in defining gene function through genetic analysis and phenotypic screenings. There are about 145 mutant phenotypic markers that have been described

in allotetraploid cotton. Most of these mutants have originated spontaneously from genetic nurseries or were selected from exotic germplasm sources (Percy and Kohel, 1999). The application and utilization of these mutants is still very limited in cotton.

The use of chemical and radiation mutagens to induce genetic variation in cotton has been of interest to some breeders and researchers, but its full utilization has been limited. Early efforts at chemical mutagenesis of cotton were no doubt impeded to a large degree by the large polyploid genome and high oil content of the seed. However, Auld et al. (1998) recently reported the first exhaustive and successful attempt at chemical (EMS) mutagenesis in tetraploid cotton. Mutant families in cotton species can exhibit a range of phenotypes, providing a valuable resource for identifying mutant alleles and determining the role of genes on fiber yield and quality. In addition, interest in the use of mutants has been renewed recently with the introduction of targeted induced local lesions in genomes (TILLING). TILLING relies on the detection of base-pair mismatches in the formation of heteroduplexes from heteroallelic DNA (McCallum et al., 2000a, b; Colbert et al., 2001). Mutant alleles, identified on the basis of single nucleotide polymorphisms (SNPs), also provide an immediate molecular marker that can be applied to mapping populations to map loci that may not be identified with more traditional DNA markers.

Linkage disequilibrium (LD) or association mapping is an alternative paradigm for mapping of markers and genes. Recent studies on the extent of LD have been performed for the human (Reich et al. 2001), *Drosophila* (Langley et al., 2000), and maize genomes (Remington et al., 2001). Cotton researchers are now poised to make significant advances in functional analysis of the cotton genome. These advances will aid in developing a functionally-anchored genetic map, and in association analyses to link genes to phenotypes and QTLs via marker development.

Radiation hybrid mapping (RHM) uses radiation treatment to increase chromosome rearrangement and recombination events. This technique also offers the advantage of very high rates of polymorphism between donor and recipient cell lines. RHM has been used extensively to map the genomes of human and certain animal species (Cox et al., 1993; Walter et al., 1994). Only recently has RHM been used in cotton. Gao (2003) developed a novel approach for RHM based on wide-cross *in vivo* hybridization, demonstrating that the RH-wide-cross whole-genome (WWRH) approach can be used to map the cotton genome as a complement to traditional linkage mapping. The major limitation of RHM or WWRH is that they map only one chromosome at a time. Another approach to high resolution mapping is with the development and use of chromosome-specific recombinant inbred lines. An example is that being developed by chromosome substitution lines from *G. barbardense* 3-79 into *G. hirsutum* TM-1 for agronomic and fiber trait studies (Saha et al., 2004). These new genomic resources will provide additional approaches for the improvement of cotton that overcome the limitation of mapping only one chromosome at a time.

In cotton as in other crops, SNPs are being developed as the ultimate DNA marker. Their importance has increased recently with the availability of DNA sequence databases and comparative analyses of sequence diversity (http://snp.cshl.org). Compared to other markers, SNPs are more abundant. For example, in humans a SNP is found every 1000-2000 bp (Sachidanandam et al., 2001), and in *Arabidopsis* every 3300 bp (Drenkard et al., 2000). These types of markers are easily subjected to high-throughput automation. Because of their high density in the genome, SNPs can provide a marker that is tightly linked to a gene or even located within the gene of interest (An et al., 2007; Hsu et al., 2007). Therefore, these markers provide an opportunity to tag functional genes with complex QTLs. This feature has renewed interest in

so-called association studies based on linkage disequilibrium (LD) to identify either marker tags for genes of medical or agronomic importance, or causal mutations responsible for a specific phenotype in other organisms (Templeton et al., 2000; Thornsberry et al., 2001).

The cotton community has indicated that 5,000 DNA informative/mapping PCR markers will be necessary to provide the requisite tools for genome analysis, evaluation of germplasm collections, and marker-assisted selection, and these will soon be available. At present, there are approximately 8,213 microsatellite markers in cotton that are available to the public (BNL, CIR, CM, JESPR, NAU, MGHES, MUSB, MUCS, MUSS, and TMB; http://www.cottonmarker.org/), and they continue to accumulate. The challenge for the cotton community in the next decade is to develop a fully integrated genetic and physical map of the cotton genome that incorporates all expressed genes. Using bacterial artificial chromosome (BAC) libraries, Dong et al. (2001) constructed from the Upland cotton genetic standard TM-1 (Kohel et al., 1970) 100,000 fingerprints that have been assembled and manually edited into 6,000 physical contigs (Yu et al., 2004; Xu et al., 2007). Four thousand cotton EST unigenes and DNA markers are currently anchored on the draft physical map of TM-1 contigs to identify gene-rich islands of the Upland cotton chromosomes. Once such an integrated genome map is complete with high genome coverage it will provide the research tool required by the next generation of cotton breeders, geneticists, and molecular biologists. The BAC corresponding to a diagnostic DNA marker can be used directly for map-based cloning of a gene or QTL when the marker is mapped to a locus closely linked to the gene or QTL of interest.

While it is often said that progress in cotton genomics is lagging behind that in other agronomic crops, the past decade has seen substantial growth and development, conceptually and technically. Undoubtedly, this tremendous progress will continue. In addition, one of the issues

that was stressed in the last two ICGI Research Conferences (India 2004 and Brazil 2006) was the selection of a genome for sequencing. Most modern productive cotton cultivars are tetraploids [*G. hirsutum* (AD_1) and *G. barbadense* (AD_2)], both with two morphologically distinct subgenomes (At and Dt). The ideal situation would be to sequence all four of these genomes. However, limited resources will probably force the cotton community to sequence only one of these genomes. The abundance of sequence information resulting from this effort will increase the knowledge of structural and functional genomics of the cotton crop. This new knowledge will greatly benefit the research community, and eventually producers and consumers.

The next phase of cotton genome mapping will be characterized by multi-disciplinary and multi-institutional collaborative initiatives that combine expertise in classical and molecular genetics, molecular biology, structural and functional genomics, biotechnology and bioinformatics to develop publicly-available genomic tools and resources. The research priorities of these collaborative initiatives will be directed by the needs of the broader cotton research community, and these efforts will continue to facilitate and advance the development of integrated, molecular consensus maps for marker-assisted breeding and the development of bioinformatic tools with user-friendly interfaces for the deposition and retrieval of data from public web-sites.

REFERENCES

Altaf, M.K., J.M. Stewart, M.K. Wajahatullah, J. Zhang, and R.G. Cantrell. 1997. Molecular and morphological genetics of a trispecies F_2 population of cotton. p. 448-452. *In* Proc. Beltwide Cotton Conf., Natl. Cotton Counc. of Amer., Memphis, TN.

An, C., S. Saha, J.N. Jenkins, B. Scheffler, T.A. Wilkins, and D.M. Stelly. 2006. Chromosome assignment of six *alpha-expansin* genes using SNP markers in cotton. p. 737. *In* Proc. Beltwide Cotton Conf., Natl. Cotton Counc. of Amer., Memphis, TN.

An, C., S. Saha, J.N. Jenkins, B. Scheffler, T.A. Wilkins, and D.M. Stelly. 2007. Transcriptome profiling, sequence characterization, and SNP marker based chromosomal assignment of the *EXPANSIN* genes in cotton. Mol. Genet. Genom. [Online]. Available at http://www.springerlink.com. D0I 10.1007/s00438-007-0270-9.

Auld, D.L., M.D. Ethridge, J.K. Dever, and P.D. Dotray. 1998. Chemical mutagenesis as a tool in cotton improvement. p. 550-552. *In* Proc. Beltwide Cotton Conf., Natl. Cotton Counc. of Amer., Memphis, TN.

Austin, D.F., and M. Lee. 1996. Comparative mapping in $F_{2.3}$ and $F_{6.7}$ generations of quantitative trait loci for grain yield and yield components in maize. Theor. Appl. Genet. 92: 817-826.

Beavis, W.D. 1994. The power and deceit of QTL experiments: lessons from comparative QTL studies. p. 250-266. *In* D.B.Wilkinson (ed.) Proc 49th Annu. Corn Sorghum Res. Conf., 7-8 Dec. 1994, Chicago, IL, Am. Seed Trade Assoc., Washington, DC.

Blenda, A., J. Scheffler, B. Scheffler, M. Palmer, J.-M. Lacape, J.Z. Yu, C. Jesudurai, S. Jung, S. Muthukumar, P. Yellambalase, S. Ficklin, M. Staton, R. Eshelman, M. Ulloa**,** S. Saha, B. Burr, S. Liu, T. Zhang, D. Fang, A. Pepper, S. Kumpatla, J. Jacobs, J. Tomkins, R. Cantrell, and D. Main. 2006. CMD: A Cotton Microsatellite Database resource for *Gossypium* genomics. BMC Genomics 7: 132.

Brookes, A.J. 1999. The essence of SNPs. Gene 234: 177-186.

Brubaker, C.L., and A.H.D. Brown. 2003. The use of multiple alien chromosome addition aneuploids facilitates genetic linkage mapping of the *Gossypium* G genome. Genome 46: 774-791.

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

Buriev, Z.T., S. Saha, I. Abdurakhmonov, J.N. Jenkins, A. Abdukarimov, B. Scheffler, D.M. Stelly, and M.J.E. Wubben. 2006a. Evidence of clustering for the *MIC-3* gene family associated with nematode resistance in *Gossypium* spp. p. 162. *In* Proc. ASA-CSSA-SSSA Annual Meeting, Indianapolis, IN.

Buriev, Z., S. Saha, I. Y. Abdurakhmonov, J.N. Jenkins, A. Abdukarimov, B.E. Scheffler, D.M. Stelly, and M.J.E. Wubben. 2006b. A strategy for developing SNP markers for the *MIC-3* root-specific gene family associated with nematode resistance in *Gossypium* spp. pp. 74. *In* Proc. 3rd International Cotton Genome Initiative (ICGI) Conference, Brasilia, Brazil.

Churchill, G.A., and R.W. Doerge. 1994. Empirical threshold values for quantitative trait mapping. Genetics 138: 963-971.

Colbert, T., B.J. Till, R. Tompa, S. Reynolds, M.N. Steine, A.T. Yeung, C.M. McCallum, L. Comai, and S. Henikoff. 2001. High-throughput screening for induced point mutations. Plant Physiol. 126: 480-484.

Cox, D.R., M. Burmeister, E.R. Price, S. Kim, and R.M. Myers. 1993. Radiation hybrid mapping: a somatic cell genetic method for constructing high-resolution maps of mammalian chromosomes. Science 250: 245-250.

Desai, A. 2005. Comparison of marker order synteny and floral trait evolution among diploid and tetraploid genomes of *Gossypium*. M.S. Thesis, Univ. of Georgia, Athens, GA.

Dong, J., R.J. Kohel, H.-B. Zhang, and J.Z. Yu. 2001. Bacterial artificial chromosome (BAC) libraries constructed from the genetic standard of Upland cottons. *In* Proc. 8th Intl. Conf. on Plant and Animal Genome Research. San Diego, CA.

Drenkard, E., B.G. Richter, S. Rozen, L.M. Stutius, N.A. Angell, M. Mindrinos, R.J. Cho, P.J. Oefner, R.W. Davis, and F.M. Ausubel. 2000. A simple procedure for the analysis of single nucleotide polymorphisms facilitates map-based cloning in Arabidopsis. Plant Physiol. 124: 1483-1492.

Endrizzi, J.E., and R. Nelson. 1989. Linkage analysis and arm location of the open bud (*Ob1*) and yellow petal (*Y2*) loci in chromosome 18 of cotton. Genome 32: 1041-1043.

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

Feng, C.D., M.D. Stewart, and J.F. Zhang. 2005. STS markers linked to the *Rf1* fertility restorer gene of cotton. Theor. Appl. Genet. 110: 237-243.

Frelichowski, J.E., M.B. Palmer, D. Main, J.P. Tomkins, R.G. Cantrell, D.M. Stelly, J.Z. Yu, R.J. Kohel, and M. Ulloa. 2006. Cotton genome mapping with new microsatellites from Acala 'Maxxa' BAC-ends. Mol. Gen. Genom. 275: 479-491.

Frelichowski, J.E., M. Ulloa, M. Palmer, D. Main, J.P. Tomkins, D.M. Stelly, R.G. Cantrell, R.J. Kohel, and J.Z. Yu. 2005. Genetic, physical and QTL mapping assessments of BAC-end derived microsatellite markers developed from cotton (*Gossypium. hirsutum* L.) Acala 'Maxxa'. p. 225. *In* Proc. 13th Intl. Conf. on Plant and Animal Genome Research. San Diego, CA.

Gao, W. 2003. Wide-cross whole-genome radiation hybrid (WWRH) mapping and identification of cold-responsive genes using oligo-gene microarray analysis in cotton. Ph.D. diss. Texas A&M Univ., College Station, TX.

Gill, K.S., B.S. Bill, T.R. Endo, and T. Taylor. 1996. Identification and high density mapping of gene-rich regions in chromosome group 1 of wheat. Genetics 144: 1883-1891.

Groh, S., D. Gonzalez-de-leon, M.M. Khairallah, C. Jiang, D. Bergvinson, M. Bohn, D.A. Hoisington, and A.E. Melchinger. 1998. QTL mapping in tropical maize: III. Genomic regions for resistance to *Diatraea* spp. and associated traits in two RIL populations. Crop Sci. 38: 1062-1072.

Guo, W., C. Cai, C. Wang, Z. Han, X. Song, K. Wang, X. Niu, C. Wang, K. Lu, B. Shi, and T. Zhang. 2007. A microsatellite-based, gene-rich linkage map reveals genome structure, function and evolution in *Gossypium*. Genetics 176: 527-541.

Harmer, S.E., S. J. Orford, and J.N. Timmis. 2002. Characterisation of six *α-expansin* genes in *Gossypium hirsutum* (Upland cotton). Mol. Genet. Genom. 268: 1-9.

Hsu, C.-Y., C. An, S. Saha, Din-Pow Ma, J.N. Jenkins, B. Scheffler, and D.M. Stelly. 2007. Molecular and SNP characterization of two genome specific transcription factor genes GhMyb8 and GhMyb10 in cotton species. Euphytica [Online]. Available at http://www.springer.com/east/home. DOI 10.1007/s10681-007-9485-4.

Jiang, C., R.J. Wright, S. S. Woo, T.A. DelMonte, and A.H. Paterson. 2000. QTL analysis of leaf morphology in tetraploid *Gossypium* (cotton). Theor. Appl. Genet. 100: 409-418.

Karaca, M., S. Saha, J. Jenkins, A. Zipf, R. Kohel, and D.M. Stelly. 2002. Simple sequence markers (SSR) linked to the Ligon Lintless (*Li1*) mutant in cotton. J. Hered. 93: 221-224.

Kianian, S.F., and C.F. Quiros. 1992. Generation of *Brasica oleracea* composite RFLP map: linkage arrangements among various populations and evolutionary implications. Theor. Appl. Genet. 84: 544-554.

Kohel, R.J., and L.S. Bird. 2002. Inheritance and linkage analysis of the yellow pulvinus mutant of cotton. J. Cotton Sci. 6: 115-118.

Kohel, R.J., T.R. Richmond, and C.F. Lewis. 1970. Texas Marker-1. A description of a genetic standard for *Gossypium hirsutum* L. Crop Sci. 10: 670-671.

Kohel, R.J., and J.Z. Yu. 2004. Permanent recombinant inbred mapping population for cotton genome research. *In* Proc. 2nd International Cotton Genome Initiative (ICGI) Conf., Hyderabad, India.

Kohel, R.J., J.Z. Yu, Y.-H. Park, and G.R. Lazo. 2001. Molecular mapping and characterization of traits controlling fiber quality in cotton. Euphytica 121: 163-172.

Lacape, J.-M., T.-B. Nguyen, B. Courtis, J.-L. Belot, M. Giband, J.-P. Gourlot, G. Gawryziak, S. Roques, and B. Hau. 2005. QTL Analysis of cotton fiber quality using multiple *Gossypium hirsutum* x *Gossypium barbadense* backcross generations. Crop Sci. 45: 123-140.

Lacape, J.-M., T.B. Nguyen, S. Thibivilliers, B. Bojinov, B. Courtois, R.G. Cantrell, B. Burr, and B. Hau. 2003. A combined RFLP-SSR-AFLP map of tetraploid cotton based on a *Gossypium hirsutum* x *Gossypium barbadense* backcross population. Genome 46: 612-626.

Langley, C.H., B.P. Lazzaro, W. Phillips, E. Heikkinen, and J.M. Braverman. 2000. Linkage disequilibria and the site frequency spectra in the su(s) and su(w(a)) regions of the *Drosophila melanogaster* X chromosome. Genetics 156: 1837-1852.

Lin, Z., D. He, X. Zhang, Y. Nie, X. Guo, C. Feng, and J.M. Stewart. 2005. Linkage map construction and mapping QTL for cotton fiber quality using SRAP, SSR and RAPD. Plant Breed. 124: 180-187.

May, O.L. 1999. Genetic variation in fiber quality. p. 183-230. *In* A.S Basra (ed.) Cotton fibers developmental biology, quality improvement, and textile processing. The Haworth Press, Inc., Binghamton, NY.

McCallum, C.M., L. Comai, E.A. Greene, and S. Henikoff. 2000a. Targeted induced local lesions in genomes (TILLING) for plant functional genomics. Plant Physiol. 123: 439-442.

McCallum, C.M., L. Comai, E.A. Greene, and S. Henikoff. 2000b. Targeted screening for induced mutations. Nat. Biotechnol. 18: 455-457.

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*). Theor. Appl. Genet. 108: 280-291.

Melchinger, A.E., H.F Utz, and C.C. Schon. 1998. Quantitative trait locus (QTL) mapping using different testers and independent population samples in maize reveals low power of QTL detection and large bias in estimates of QTL effects. Genetics 149: 383-403.

Monforte, A.J. and S.D. Tanksley. 2000. Fine mapping of a quantitative trait locus (QTL) from *Lycopersicon hirsutum* chromosome 1 affecting fruit characteristics and agronomic traits: breaking linkage among QTLs affecting different traits and dissection of heterosis for yield. Theor. Appl. Genet. 100: 471-479.

Nguyen, T.-B., M. Giband, P. Brottier, A.-M. Risterucci, and J.-M. Lacape. 2004. Wide coverage of the tetraploid cotton genome using newly developed microsatellite markers. Theor. Appl. Genet. 109: 167-175.

Park, Y.-H., M.S. Alabady, M. Ulloa, B. Sickler, T.A. Wilkins, J.Z. Yu, D.M. Stelly, R.J. Kohel, O.M. El-Shihy, and R.G. Cantrell. 2005. Genetic Mapping of New Cotton Fiber Loci using EST-derived microsatellites in an interspecific recombinant inbred line cotton population. Mol. Genet. Genom. 274: 428-441.

Paterson, A.H., Y. Saranga, M. Menz, C. Jiang, and R.J. Wright. 2003. QTL analysis of genotype x environmental interactions affecting cotton fiber quality. Theor. Appl. Genet. 106: 384-396.

Percy, R.G. 1999. Inheritance of cytoplasmic-virescent *cyt-v* and dense-glanding *dg* mutants in American Pima cotton. Crop Sci. 39: 372-374.

Percy, R.G., and R.J. Kohel. 1999. Qualitative Genetics. p. 319-360. *In* C.W. Smith and J. Cothren (eds.) Cotton: Origin, History, Technology, and Production. Wiley, New York, NY.

Petes, T.D. 2001. Meiotic recombination hot spots and cold spots. Nat. Rev. Genet. 2: 360-369.

Reich, D.E., M. Cargill, S. Bolk, J. Ireland, P.C. Sabeti, D.J. Richter, T. Lavery, R. Kouyoumjian, S.F. Farhadian, R. Ward, and E.S. Lander. 2001. Linkage disequilibrium in the human genome. Nature 411: 199-204.

Reinisch, A.J., J.M. Dong, C.L. Brubaker, J.F. Wendel, and A.H. Paterson. 1994. A detailed RFLP map of cotton, *Gossypium hirsutum x Gossypium barbadense*: chromosome organization and evolution in a disomic polyploid genome. Genetics 138: 829-847.

Remington, D.L., J.M. Thornsberry, Y. Matsuoka, L.M. Wilson, S.R. Whitt, J. Doeblay, S. Kresovich, M.M. Goodman, and E.S. Buckler. 2001. Structure of linkage disequilibrium and phenotypic associations in the maize genome. Proc. Nat. Acad. Sci. USA 98: 11479-11484.

Rong, J.-K., C. Abbey, J.E. Bowers, C.L. Brubaker, C. Chang, P.W. Chee, T.A. Delmonte, X.L. Ding, J.J. Garza, B.S. Marler, C.-H. Park, G.J. Pierce, K.M. Rainey, V.K. Rastogi, S.R. Schulze, N.L. Trolinder, J.F. Wendel, T.A. Wilkins, T.D. Williams-Coplin, R.A. Wing, R.J. Wright, X. Zhao, L. Zhu, and A.H. Paterson. 2004. A 3347-locus genetic recombination map of sequence-tagged sites reveals features of genome organization, transmission and evolution of cotton (*Gossypium*). Genetics 166: 389-417.

Rong, J.-K., F. A. Feltus, V.N. Waghmare, G.J. Pierce, P.W. Chee, X. Draye, Y. Saranga, R.J. Wright, T.A. Wilkins, O.L. May, C.W. Smith, J.R. Gannaway, J.F. Wendel, and A.H. Paterson. 2007. Meta-analysis of polyploid cotton QTLs shows unequal contributions of subgenomes to complex network of genes and gene clusters implicated in lint fiber development. Genetics [Online]. Available at http://www.genetics.org/. 10.1534/genetics.107.074518.

Rungis, D., D. Llewellyn, E.S. Dennis, and B.R. Lyon. 2002. Investigation of the chromosomal location of the bacterial blight resistance gene present in an Australian cotton (*Gossypium hirsutum* L.) cultivar. Aust. J. Agri. Res. 53: 551-560.

Sachidanandam, R., D. Weissman, S.C. Schmidt, J.M. Kakol, L.D. Stein, J.C. Mullikin, B.J. Mortimore, D.L. Willey, S.E. Hunt, C.G. Cole, P.C. Coggill, C.M. Rice, Z.M. Ning, J. Rogers, D.R. Bentley, P.Y. Kwok, E.R. Mardis, R.T. Yeh, B. Schultz, L. Cook, R. Davenport, M. Dante, L. Fulton, L. Hillier, R.H. Waterston, J.D. McPherson, B. Gilman, S. Schaffner, W.J. Van Etten, D. Reich, J. Higgins, M.J. Daly, B. Blumenstiel, J. Baldwin, N.S. Stange-Thomann, M.C. Zody, L. Linton, E.S. Lander, and D. Altschuler. 2001. A map of human genome sequence variation containing 1.42 million single nucleotide polymorphisms. Nature 409: 928-933.

Saha, S., J. Wu, J.N. Jenkins, J.C. McCarty**,** Jr., O.A. Gutierrez, D.M. Stelly, R.G. Percy, and D.A. Raska. 2004. Effect of Chromosome substitutions from *Gossypium barbadense* L. 3-79 into *G*. *hirsutum* L. TM-1 on agronomic and fiber traits. J. Cotton Sci. 8: 162-169.

Saunders, J.H. 1961. The Wild Species of *Gossypium* and their Evolutionary History. Oxford Univ. Press, London.

Saranga, Y., M. Menz, C. Jiang, R.J. Wright, D. Yakir, and A.H. Paterson. 2001. Genomic dissection of genotype x environment interactions conferring adaptation of cotton to arid conditions. Genome Res. 11: 1988-1995.

Shappley, Z.W., J.N. Jenkins, W.R. Meredith, and J.C. McCarty**,** Jr. 1998. An RFLP linkage map of Upland cotton, *Gossypium hirsutum* L. Theor. Appl. Genet. 97: 756-761.

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

Shen, X., W. Guo, X. Zhu, Y. Yuan, J.Z. Yu, R.J. Kohel, and T. Zhang. 2005. Molecular mapping of QTLs for qualities in three diverse lines in Upland cotton using SSR markers. Mol. Breed. 15: 169-181.

Smith, C.W., R.G. Cantrell, H.S. Moser, and S.R. Oakley. 1999. History of cultivar development in the United States. p. 99-171. *In* C.W. Smith and J. Cothren (ed.) Cotton: Origin, History, Technology, and Production. John Wiley & Sons, New York, NY.

Swofford, D.L. 1998. PAUP*. Phylogenetic analysis using parsimony, Version 4b.10. Sinauer Associates, Sunderland, Massachusetts.

Templeton, A.R., K.M. Weiss, D.A. Nickerson, E. Boerwinkle, and C.F. Sing. 2000. Cladistic structure within the human lipoprotein lipase gene and its implications for phenotypic association studies. Genetics 156: 1259-1275.

Thornsberry, J.M., M.M. Goodman, J. Doebley, S. Kresovich, D. Nielsen, and E.S. Buckler. 2001. Dwarf8 polymorphisms associate with variation in flowering time. Nature Genet. 28: 286-289.

Ulloa, M., C. Brubaker, and P. Chee. 2007. Cotton. p. 1-49. *In* C. Kole (ed.) Genome Mapping & Molecular Breeding. Vol. 6: Technical Crops. Springer, New York, NY.

Ulloa, M., R.G. Cantrell, R. Percy, Z. Lu, and E. Zeiger. 2000. QTL Analysis of stomatal conductance and relationship to lint yield in an interspecific cotton. J. Cotton Sci. 4: 10-18.

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

Ulloa, M., and W.R. Meredith, Jr. 2002. QTL Examination on a RFLP joinmap from four intraspecific cotton populations. p. 247. *In* Proc. 10th Intl. Conf. on Plant and Animal Genome Research. San Diego, CA.

Ulloa, M., W.R. Meredith**,** Jr., Z. W. Shappley, and A.L. Kahler. 2002. RFLP genetic linkage maps from four F2.3 populations and a joinmap of *Gossypium hirsutum* L. Theor. Appl. Genet. 104: 200-208.

Ulloa, M., S. Saha, J.N. Jenkins, W.R. Meredith**,** Jr., J. C. McCarty, and M.D. Stelly. 2005. Chromosomal assignment of RFLP linkage groups harboring important QTLs on an intraspecific cotton (*Gossypium hirsutum* L.) joinmap. J. Hered. 96: 132-144.

Ulloa, M., J.M. Stewart, C.E. Garcia, A.A. Godoy. M.N. Gaytan**,** and N. Acosta. 2006. Cotton genetic resources in the western states of Mexico; *In situ* conservation status and germplasm collection for *ex situ* preservation. Genetic Resources and Crop Evol. 53: 653-668.

Van Ooijen, J.W. 2006. JoinMap 4, Software for calculation of genetic linkage maps in experimental populations. Kyazma B.V., Wageningen, Netherlands.

Van Ooijen, J.W., and C. Maliepaard. 1996. MapQTL™ Version 3.0, Sotfware for the calculation of QTL positions on genetic maps. Plant Research International, Wageningen, Netherlands.

Waghmare, V.N., J. Rong, C.J. Rogers, G.J. Pierce, J.F. Wendel, and A.H. Paterson. 2005. Genetic mapping of a cross between *Gossypium hirsutum* (cotton) and the Hawaiian endemic, *Gossypium tomentosum*. Theor. Appl. Genet. 111: 665-676.

Walter, M.A., D.J. Spillet, P. Thomas, J. Weissienbach, and P.N. Goodfellow. 1994. A method for constructing radiation hybrid maps of whole genomes. Nature Genet. 7: 22-28.

Wang, K., X. Song, Z. Han, W. Guo, J.Y. Yu, J. Sun, J. Pan, R.J. Kohel, and T. Zhang. 2006. Complete assignment of the chromosomes of *Gossypium hirsutum* L. by translocation and fluorescence in situ hybridization mapping. Theor. Appl. Genet. 113: 73-80.

Wright, R., P. Thaxton, K. El-Zik, and A.H. Paterson. 1998. D-subgenome bias of *Xcm* resistance genes in tetraploid *Gossypium* (cotton) suggest that polyploidy formation has created novel avenues for evolution. Genetics 149: 1978-1996.

Wright, R., P. Thaxton, A.H. Paterson, and K. El-Zik. 1999. Molecular mapping of genes affecting pubescence of cotton. J. Hered. 90: 215-219.

Xu, Z., J.Z. Yu, J. Cho, L. Covelenda, J. Dong, G. Song, L. Zhang, M. Atkins, J. Yu, K. Ding, M. Ulloa, J. Tomkins, D.M. Stelly, H.-B. Zhang, and R.J. Kohel. 2007. An integrated BAC/BIBAC-based physical and genetic map of the cotton genome. p. 49. *In* Proc. 15th Intl. Conf. on Plant and Animal Genome Research. San Diego, CA.

Yu, J.Z., R.J. Kohel, H.-B.Zhang, D.M. Stelly, Z. Xu, J. Dong, L. Covaleda, M.-K. Lee, P. Cui, G.R. Lazo, P. Gupta, and K. Ding. 2004. Toward an integrative physical and genetic map of the cultivated allotetraploid cotton genome. pp. 42. *In* Proc. 12th Intl. Conf. on Plant and Animal Genome Research. San Diego, CA.

Yu, J., P. Yong-Ha, G.R. Lazo, and R.J. Kohel. 1998. Molecular mapping of the cotton genome: QTL analysis of fiber quality properties. p. 485. *In* Proc. Beltwide Cotton Conf., Natl. Cotton Counc. of Amer., Memphis, TN.

Zeng, Z.-B. 1994. Precision mapping of quantitative trait loci. Genetics 136: 1457-1468.

Zhang, J., and J.M. Stewart. 2004. Identification of molecular markers linked to the fertility restorer genes for CMS-D_8 in cotton (*Gossypium hirsutum* L.). Crop Sci. 44: 1209-1217.

Zhang, J.F., J.M. Stewart, and T.H. Wang. 2005. Linkage analysis between gametophytic restorer *Rf(2)* gene and genetic markers in cotton. Crop Sci. 45: 147-156.

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

Table 1. QTLs placed on linkage group No. 11 (Chromosome No. 5) of the genetic RFLP joinmap developed from four Upland F_2-derived F_3 cotton populations with LOD score > 2.0 affecting agronomic and fiber quality traits.

Cotton Trait		Pop[z]	Locus	Group/Chrom	LOD[y]	Mean values of Cotton Traits				
						Female[x]	Heterozygote[x]	Male[x]	SD[w] ±	%Exp-Var[v]
Lint Yield	(kg ha^{-1})	M (2)	C115A1V	13 / 5	2.8	992.9	1074.7	1001.3	10.9	9.0
		M (AVG)	C115A1V	13 / 5	2.1	992.5	1040.1	990.4	9.4	6.1
First Pick		M (2)	C115A1V	13 / 5	3.2	918.6	1005.8	928.8	11.4	9.0
		M (AVG)	C115A1V	13 / 5	2.0	746.6	789.7	731.9	9.9	6.1
Seed Weight	(g)	H (6)	C115A1V	9 / 5	4.2	0.093	0.098	0.094	0.0	19.7
		H (9)	C115A1V	9 / 5	2.3	0.102	0.104	0.101	0.0	9.7
Boll Weight	(g)	H (6)	C115A1V	9 / 5	2.2	4.8	5.0	4.8	1.7	9.5
50% Fiber Span Length	(mm)	H (AVG)	C115A1V	9 / 5	2.0	14.48	14.73	14.48	0.0	5.5
Fiber Strength	(kN m kg^{-1})	H (6)	C115A1V	9 / 5	2.0	220.5	228.5	224.2	3.2	8.5
		H (9)	C115A1V	9 / 5	2.0	228.5	234.0	232.2	3.2	4.2

[z]Populations H = HQ95-6 x MD51ne, and M = 119-5 x MD51ne, and (1), (2), (6), (9) and (AVG) = Quantitative data collected from field 1, 2, 6, 9, and average from all fields.

[y]LOD = The LOD scores for each QTL were derived from the likelihood ratio test statistic of QTL Cartographer.

[x]The mean values for each parent and the heterozygote, from the informative fragments, were calculated by MapQTL program.

[w]SD = Standard deviation calculated by MapQTL.

[v]%Exp = The trait percentage variation explained by the QTL, calculated by MapQTL.

Table 2. QTLs placed on linkage groups part of chromosome No. 3 of the genetic RFLP joinmap developed from four Upland F_2-derived F_3 cotton populations with LOD score > 2.0 affecting agronomic and fiber quality traits.

Cotton Trait	Pop[z]	Locus	Group /Joingrp	LOD[y]	Mean values of Cotton Traits			SD[w] ±	%Exp-Var[v]
					Female[x]	Heterozygote[x]	Male[x]		
Lint Percentage (%)	M (2)	F14D2I	7 / 3	2.1	35.5	34.5	34.7	1.00	10.0
	M (AVG)	F14D2I	7 / 3	2.8	35.8	34.6	34.8	1.00	16.6
	H (6)	C56E1bI	19 / 3	2.7	38.8	37.9	39.6	0.93	38.5
	H (AVG)	C56E1bI	19 / 3	2.1	36.9	36.1	37.6	0.93	32.4
	H (6)	C122B1I	22 / 3	2.1	38.7	38.1	39.0	1.02	11.7
	H (AVG)	C122B1I	22 / 3	2.0	36.9	36.4	37.0	1.01	6.6
Seed Weight (g)	M (1)	C104A1aI	3 / 3	4.6	9.3	9.8	9.7	0.7	11.4
	M (2)	C104A1aI	3 / 3	3.4	9.3	9.7	9.7	0.7	8.2
	M (AVG)	C104A1aI	3 / 3	4.5	9.3	9.8	9.7	0.7	10.8
Fiber Strength (kN m kg^{-1})	M (1)	C104A1b	3 / 3	2.0	226.7	213.5	213.5	3.1	16.5
	M (AVG)	C104A1aI	3 / 3	2.2	221.2	218.1	215.0	2.9	22.0
	M (1)	C70D2bV	7 / 3	4.5	218.7	213.3	232.9	1.7	30.1
	H (1)	C87A4VI	22 / 3	2.7	218.1	209.1	213.4	3.1	9.2
	H (1)	C114A6V	22 / 3	2.2	216.6	209.7	213.6	3.2	5.7
Fiber Elongation (%)	U (AVG)	C58D3cI	9 / 3	6.4	8.0	7.78	7.39	0.47	31.6
	H (1)	C56E1bI	19 / 3	2.0	7.8	7.6	8.2	0.64	24.4
	H (6)	F2E6I	4 / 3	2.3	8.0	8.0	7.6	0.75	5.0
	H (AVG)	F2E6I	4 / 3	3.1	7.9	7.8	7.6	0.60	6.4
50% Fiber Span Length (mm)	M (2)	C104A1aI	3 / 3	2.2	14.46	14.67	14.74	0.0	9.4
	M (2)	C108C5V	7 / 3	2.3	14.54	14.70	14.45	0.0	4.7
	H (AVG)	C44D2I	17 / 3	2.5	14.55	14.64	14.53	0.0	7.5
	H (AVG)	C56E1bI	19 / 3	2.3	14.55	14.71	14.45	0.0	17.5
	H (AVG)	C122B1I	22 / 3	2.4	14.56	14.65	14.48	0.0	7.5

2.5% Fiber Span Length	**M (1)**	**C104A1aI**	**3 / 3**	**6.5**	**27.88**	**28.56**	**28.75**	**0.20**	**15.8**
	M (2)	**C104A1aI**	**3 / 3**	**5.0**	**28.23**	**28.90**	**29.07**	**0.30**	**11.7**
	M (AVG)	**C104A1aI**	**3 / 3**	**6.7**	**28.06**	**28.73**	**28.90**	**0.20**	**15.8**
	H (1)	**C34E3V**	**19 / 3**	**2.4**	**29.42**	**29.24**	**28.81**	**0.75**	**7.2**
	H (6)	**C34E3V**	**19 / 3**	**2.1**	**30.17**	**30.18**	**29.60**	**1.04**	**7.7**
	H (9)	**C34E3V**	**19 / 3**	**2.8**	**31.40**	**30.93**	**30.66**	**0.75**	**9.7**
	H (AVG)	**C34E3V**	**19 / 3**	**2.7**	**30.00**	**29.86**	**29.42**	**0.84**	**8.1**
	H (9)	**C114A6V**	**22 / 3**	**2.4**	**31.06**	**30.95**	**30.44**	**1.45**	**9.3**
Micronaire (unit)	**M (1)**	**C104A1aI**	**3 / 3**	**2.5**	**49**	**48**	**47**	**1**	**7.5**
	M (2)	**C104A1aI**	**3 / 3**	**4.6**	**50**	**48**	**47**	**1**	**11.5**
	M (AVG)	**C104A1aI**	**3 / 3**	**4.0**	**49**	**47**	**47**	**1**	**9.2**
Fiber Maturity (%)	**U (AVG)**	**C70D2aV**	**7 / 3**	**2.5**	**83.5**	**85.4**	**86.1**	**1.8**	**33.6**
Fiber Perimeter (µm)	**U (AVG)**	**C70D2aV**	**7 / 3**	**3.0**	**49.7**	**48.3**	**48.3**	**1.7**	**9.3**

[z]Populations: U = MD5678ne x Prema, H = HQ95-6 x MD51ne, and M = 119-5 x MD51ne, and (1), (2), (6), (9) and (AVG) = Quantitative data collected from field 1, 2, 6, 9, and average from all fields.

[y]LOD = The LOD scores for each QTL were derived from the likelihood ratio test statistic of QTL Cartographer.

[x]The mean values for each parent and the heterozygote, from the informative fragments, were calculated by MapQTL program.

[w]SD = Standard deviation calculated by MapQTL.

[v]%Exp = The trait percentage variation explained by the QTL, calculated by MapQTL.

Table 3. QTLs placed on linkage groups part of chromosome 26 of the genetic RFLP joinmap developed from four Upland F_2-derived F_3 cotton populations with LOD score > 2.0 affecting agronomic and fiber quality traits.

Cotton Trait	Pop[z]	Locus	Group / Chromo	LOD[y]	Mean values of Cotton Traits			SD[w] ±	%Exp-Var[v]
					Female[x]	Heterozygote[x]	Male[x]		
Lint Yield (kg ha^{-1})	M (1)	C66C3V	6 / 26	2.4	1006.8	957.4	1033.4	9.5	10.1
	M (1)	C61A1I	9 / 26	2.3	995.9	1011.4	948.1	9.6	6.2
Lint Percentage (%)	U (AVG)	C80F3bI	1 / 26	2.8	38.3	38.4	37.0	1.32	19.6
	U (AVG)	C42B1bI	1 / 26	2.8	38.0	38.2	36.6	1.30	4.8
	M (2)	C58A1V	6 / 26	2.3	34.4	35.1	35.2	1.00	7.2
Seed Weight (g)	U (AVG)	C88C2I	1 / 26	2.4	0.1039	0.1081	0.1119	0.0075	3.8
	U (AVG)	C81F4bI	3 / 26	2.3	0.1009	0.1050	0.1098	0.0078	11.2
Fiber Strength (kN m kg^{-1})	U (AVG)	C38A5bI	1 / 26	3.1	217.5	223.1	231.4	3.4	24.6
	U (AVG)	C42B1bI	1 / 26	2.5	217.6	223.4	232.2	3.4	10.6
	H (1)	C60B1I	24 / 26	2.2	211.0	209.8	216.4	3.2	6.9
Fiber Elongation (%)	U (AVG)	C50C1bI	3 / 26	2.8	8.30	7.85	7.55	0.52	3.4
	U (AVG)	C81F4bI	3 / 26	3.8	8.40	7.71	7.52	0.48	11.9
50% Fiber Span Length (mm)	U (AVG)	C50C1bI	3 / 26	2.5	14.22	14.48	14.48	0.01	10.4
	M (2)	C58A1V	6 / 26	3.1	14.81	14.55	14.47	0.00	9.0
	M (2)	C50D1V	6 / 26	3.1	14.81	14.55	14.45	0.00	8.8
	M (AVG)	C50D1V	6 / 26	2.4	14.60	14.44	14.37	0.00	6.8
2.5% Fiber Span Length	U (AVG)	C38A5bI	1 / 26	2.5	28.25	29.01	29.87	1.15	11.5
	U (AVG)	C64C4V	15 / 26	2.2	29.37	27.73	29.75	0.65	44.6
Micronaire (unit)	U (AVG)	C50C6bV	1 / 26	3.2	46	44	43	2.6	9.1
	U (AVG)	C38A5bI	1 / 26	3.2	47	44	43	2.3	21.7
	U (AVG)	C42B1bI	1 / 26	4.2	47	44	42	2.4	19.3
	U (AVG)	C1A5I	1 / 26	3.3	45	44	42	2.7	6.2
Fiber Maturity (%)	U (AVG)	C5F4I	3 / 26	2.3	83.3	86.2	86.8	3.4	15.2

Fiber Perimeter (μm)	**U (AVG)**	**C42B1bI**	**1 / 26**	**2.9**	**49.8**	**49.4**	**47.8**	**1.9**	**11.3**
	U (AVG)	**C80A5bV**	**5 / 26**	**2.3**	**49.8**	**48.6**	**48.9**	**1.8**	**5.8**

[z]Populations: U = MD5678ne x Prema, H = HQ95-6 x MD51ne, and M = 119-5 x MD51ne, and (1), (2), (6), (9) and (AVG) = Quantitative data collected from field 1, 2, 6, 9, and average from all fields.

[y]LOD = The LOD scores for each QTL were derived from the likelihood ratio test statistic of QTL Cartographer.

[x]The mean values for each parent and the heterozygote, from the informative fragments, were calculated by MapQTL program.

[w]SD = Standard deviation calculated by MapQTL.

[v]%Exp = The trait percentage variation explained by the QTL, calculated by MapQTL.

Table 4. GhEXPA6-1 A genome haplotypes from four different species: *Gossypium arboreum* (A2), *G. barbadense* (Gb), *G. tomentosum* (Gt) HS46 and MARCABUCAG8US-1-88 (MAR) (*G. hirsutum*).

Haplotype	Position										Genotype
	57	68	85	225	267	297	443	445	466	553	
I	T	T	G	G	C	G	G	C	C	T	A2
II	C	C	A	A	C	A	C	G	T	G	Gb
III	T	T	G	A	T	G	C	G	C	G	Gt
IV	T	T	G	A	C	G	C	G	C	G	HS46 and Mar

Table 5. GhEXPA6-1 D genome haplotypes from five different species: 3-79 (*Gossypium barbadense*), *G. raimondii* (D5), *G. mustelinum* (Gm), HS46 and TM-1 (*G. hirsutum*), and *G. tomentosum* (Gt).

Haplotype	Position								Genotype
	57	73	85	92	145	152	327	388	
I	C	A	G	G	A	G	C	A	3-79
II	T	G	G	G	A	T	C	A	D5
III	C	G	G	G	G	T	C	A	Gm
IV	C	G	A	A	A	T	T	T	HS46 and TM-1
V	C	G	A	G	A	T	T	A	Gt

Table 6. Summary of published molecular genetic linkage maps for *Gossypium* species.

Mapping populations	No. of loci	No. of linkage groups	Length (cM)	QTL No.	Reference
57 *G. hirsutum* x *G. barbadense* F_2 progeny	683 RFLP loci	41	4675		Reinisch et al. (1994)
[*G. arboreum* x *G. trilobum*] x *G. hirsutum* F_2 progeny	216 loci --194 AFLP --19 RAPD --3 morphological markers	11	522		Altaf et al. (1997)
96 *G. hirsutum* [HS46 x MARCABUSCAG8US-1-88] $F_{2.3}$ progeny	120 RFLP loci	31	865	100	Shappley et al. (1998)
150-199 *G. barbadense* [PS 7] x *G. hirsutum* [Empire B2b6, B2, B3, and 'S295'] F_2 progeny	xxxx	48 - 49	921 - 2101	8 4	Wright et al. (1998) Wright et al. (1999)
58 *G. herbaceum* [A_2 (=A_1)-97] x *G. arboreum* [$A_2$47] F_2 progeny	161 loci --6 isozyme --155 RFLP	18	856		Brubaker et al. (1999)
62 *G. trilobum* [*s.n.*] x *G. raimondii* [*s.n.*] F_2 progeny	306 RFLP loci	17	1486		Brubaker et al. (1999)
119 *G. hirsutum* [MD5678ne x Prema] $F_{2.3}$ progeny	81 RFLP loci	17	701		Ulloa and Meredith (2000)
118 *G. hirsutum* [NM24016 x TM1] $F_{2.3}$ progeny	199 loci --125 RAPDS --68 SSRs	28	1058	2	Ulloa et al. (2000)
180 *G. hirsutum* [Mutant x 'Seberry'] F_2 progeny	xxxxx	26	3664	21	Jiang et al. (2000)
171 *G. hirsutum* [TM-1] x *G. barbadense* [3-79] F_2 progeny	355 loci --216 RFLP --139 RAPD	50	4766	13	Kohel et al. (2001) Yu et al. (1998)

58 *G. hirsutum* [TM-1] x *G. barbadense* [Hai 7124] haploid and doubled-haploid progeny	**489 loci**	**43**	**3315**		**Zhang et al. (2003)**
Genetic Joinmap	**283 RFLP loci; 1 morphological marker**	**47** **47**	**1503** **1503**	**92**	**Ulloa et al. (2002)** **Ulloa et al. (2005)**
<u>Population A</u>: 96 *G. hirsutum* [HS46 x MARCABUSCAG8US-1-88] $F_{2.3}$ progeny	**[A: 120 RFLP loci]**	**[A: 31]**	**[A: 865]**	**100**	**Shappley et al. (1998)**
<u>Population B:</u> 119 *G. hirsutum* [MD5678 x Prema] $F_{2.3}$ progeny	**[B: 81 RFLP loci]**	**[B: 17]**	**[B: 701]**	**26**	**Ulloa and Meredith (2000)**
<u>Population C</u>: 199 *G. hirsutum* [HQ95-6 x MD51ne] $F_{2.3}$ progeny	**[C: 82 RFLP loci; 1 morphological marker]**	**[C: 24]**	**[C: 830]**	**18**	**Ulloa and Meredith (2002)**
<u>Population D</u>: 155 *G. hirsutum* [119-5 subokra x MD51ne] $F_{2.3}$ progeny (Ulloa 2002)	**[D: 56 RFLP loci]**	**[D: 16]**	**[D: 520]**	**19**	**Ulloa and Meredith (2002)**
96 *G. hirsutum* [TM-1 x *Li1* (Ligon lintless)] F_2 progeny	**23 loci** **--22 SSR** **--1 morphological marker**	**8**	**218**	**1[c]**	**Karaca et al. (2002)**
214 'Siv'on' x F-177 F_2 progeny	**253 RFLP loci**	**13**	**4675** **Reinsch et al. (1994)**	**33** **79**	**Saranga et al. (2001)** **Paterson et al. (2003)**

94 *G. nelsonii* [Gos-5024] x *G. australe* [Gos-5005] F_2 progeny	**389 AFLP loci** **--213 *G. australe*** **--176 *G. nelsonii***	**13 linkage group assemblages** **--17 *G. australe*** **--21 *G. nelsonii***	**- 931** **- 773**		**Brubaker and Brown (2003)**
75 *G. hirsutum* [Guazuncho 2] x {*G. hirsutum* [Guazuncho 2] x *G. barbadense* [VH8-4602] } BC_1 progeny	**888 loci** **--465 AFLP** **--229 SSR** **--192 RFLP** **--2 morphological markers**	**37** **26**	**4397** **5519**	**80**	**Lacape et al. (2003)** **Lacape et al. (2005)** **Nguyen et al. (2004)**
94 *G. hirsutum* [Acala 44] x *G. barbadense* [Pima S-7] F_2 progeny	**392 loci** **--333 AFLP** **--12 RFLP** **--47 SSR**	**42**	**3287**	**7**	**Mei et al. (2004)**
82 *G. hirsutum* [TMS-22] x *G. tomentosum* [WT936] F_2 progeny	**589 RFLP loci**	**82**	**4259**	**11**	**Waghmare et al. (2005)**
57 *G. hirsutum* [race 'palmeri'] x *G. barbadense* [K101] F_2 progeny	**2584 STS loci**	**26**	**4448**		**Rong et al. (2004)**
62 *G. trilobum* x *G. raimondii* F_2 progeny	**763 STS loci**	**13**	**1493**		**Rong et al. (2004)**
69 *G. hirsutum* [Handan 208] x *G. barbadense* [Pima 90] F_2 progeny	**566 loci** **--205 SSR**	**41**	**5141**		**Lin et al. (2005)**
163 *G. hirsutum* [*G. anomalum* introgression line 7235 x TM-1] F_2 and RIL progeny	**86 SSR loci** **156 loci**	**21**	**667** **1024**	**37** **28**	**Shen et al. (2005)** **Zhang et al. (2003)** **Shen et al. (2007)**

169 *G. hirsutum* [HS427-10 x TM-1] F_2 progeny	**56 SSR loci**	**17**	**558**	**23**	**Shen et al. (2005)**
142 *G. hirsutum* [PD6992 x SM3] F_2 progeny	**73 SSR loci**	**22**	**588**	**11**	**Shen et al. (2005)**
167 *G. arboreum* [SMA4; PI529740] x *G. herbaceum* [A_1-97; PI529670] F_2 progeny	**275 RFLP loci**	**12**	**1147**		**Desai (2005)**
183 *G. hirsutum* [TM-1] x *G. barbadense* [3-79] recombinant inbred lines (RIL) progeny	**224 SSR loci**	**40**	**1277**	**9**	**Park et al. (2005)**
183 *G. hirsutum* [TM-1] x *G. barbadense* [3-79] recombinant inbred lines (RIL) progeny	**407 SSR loci**	**43**	**2126**	**25**	**Frelichowski et al. (2006)**

Table 7. Assessment of Fiber Quality (L = length, S = strength, and F = fineness) QTLs from 10 interspecific and eight intraspecific mapping populations, representing approximately 421 QTLs on genetic linkage maps. Information obtained from Ulloa et al. (2007).

At subgenome	Interspecific Populations	Intraspecific Populations	Dt subgenome	Interspecific Populations	Intraspecific Populations
Chromosome/LG	At - subgenome	At - subgenome	Chromosome/LG	Dt - subgenome	Dt - subgenome
1	S F		15	S S S S F	
2	L L F F		14	L S S F	L S F F
3	L L S S S F F F F	L L S S F	17	S F	
4	L S F		22	L S S F	
5	L S F F F		D08 / 19	F	
6	L F F		25	L L S S F F	L L
7		L L L S F	16	L S S F F	L S F
A02 / 8	L S F	S S	D03 / 24	L S F F	L S F
9	L L S F F	L S F	23	L S S F	L L S S F
10	L F	L L S S F F	20	L S S S F	L S F
A03 / 11	L S F	S F	D02 / 21	L S S S F	L
12	L L	L S F	26	L S	L S F F
A01 / 13	L L S S F		18	L L L S S S F F	L

FIGURE CAPTIONS

Figure 1. Linkage group/chromosome No. 3 from the At subgenome coverage from 2005 (Frelichowski et al. 2006) to 2007 (presented herein). The cotton genetic linkage map was constructed from 183 recombinant inbred lines (RILs) from the interspecific cross *Gh cv*. TM1 x *Gb acc*. Pima 3-79 with a total of 650 loci.

Figure 2. Phylogram results of An et al. (2007) showing the separation of sequences into two distinct clades of A and D sub-genome of tetraploid sequences. SNP markers associated with GhEXPA6-1 gene

Cotton Chromosome No. 3

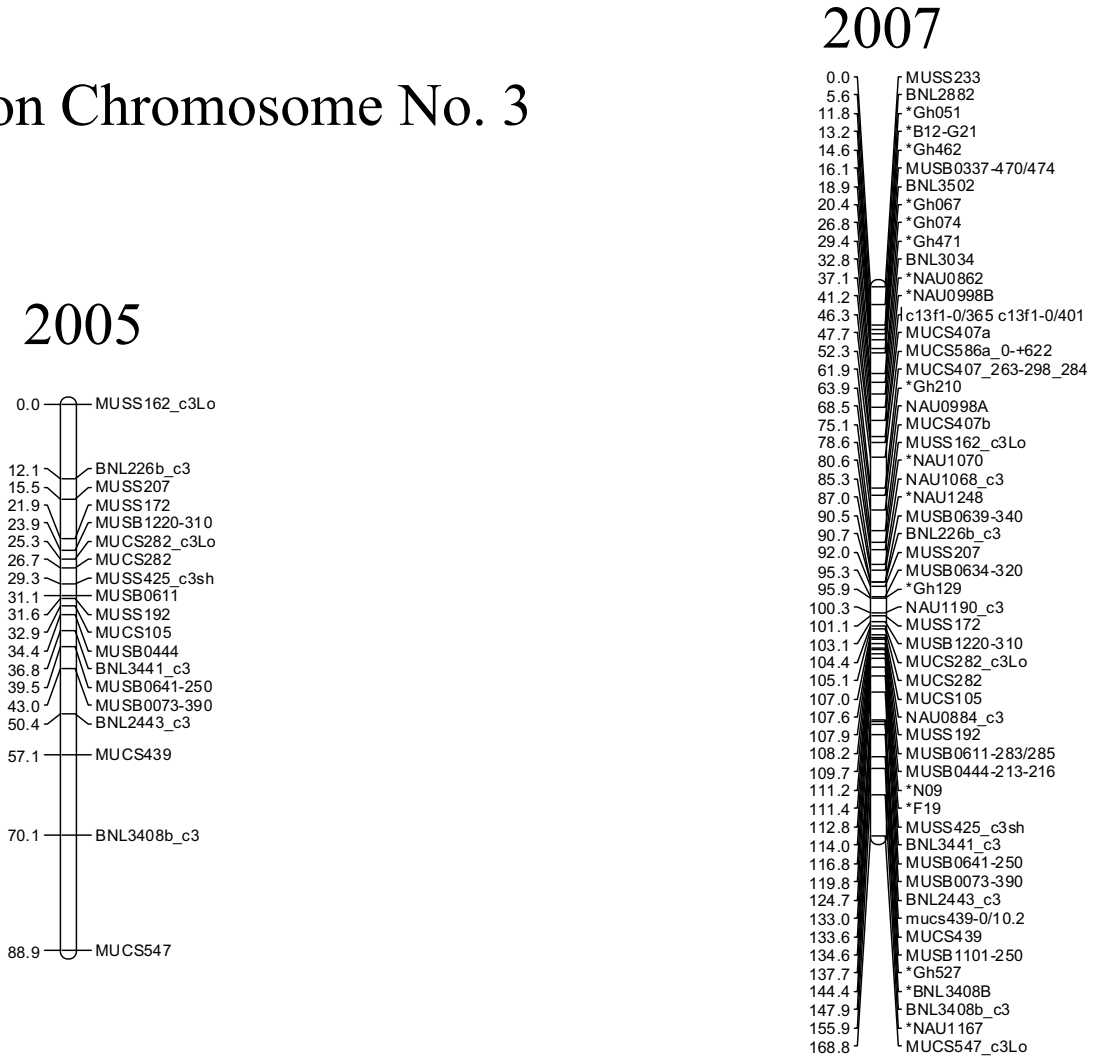

FIGURE 1.

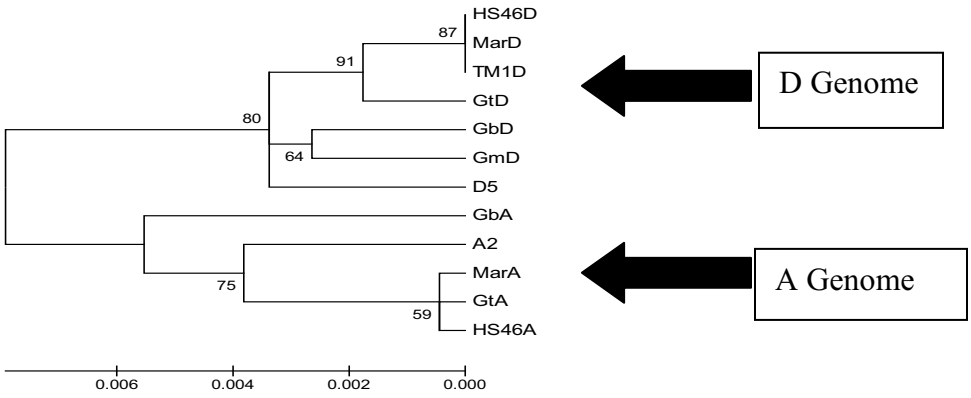
HS46D
MarD
TM1D
GtD
GbD
GmD
D5
GbA
A2
MarA
GtA
HS46A
87
91
80
64
75
59
0.006
0.004
0.002
0.000
D Genome
A Genome

FIGURE 2.