GENETIC DISSECTION OF COTTON FIBER QUALITY TRAITS IN RECIPROCAL ADVANCED BACKCROSS POPULATIONS

by

JEEVAN ADHIKARI

(Under the Direction of Andrew H. Paterson)

ABSTRACT

Introgression is an important source of genetic variation for crop improvement and interspecific introgression lines are important resources for plant breeders to access novel alleles. In addition to combining alleles from diverse genotypes and creating novel allelic combinations, interspecific hybridization also provides the opportunity to study gene flow and transmission of chromatin between species. Experimental advanced-generation backcross populations contain individuals with genomic compositions resembling those resulting from natural interspecific hybridization. Individual members of such advanced-generation populations usually retain some genomic features of the donor parent while they most closely resemble their recurrent (backcross) parent. In this study, we developed a reciprocal set of advanced backcross populations using two elite cotton cultivars, Acala Maxxa (*Gossypium hirsutum*) and Pima S6 (*G. barbadense*) as parents and investigated these populations, segregating for a few chromosomal segments, for the nature and pattern of reciprocal chromatin transmission in interspecific crosses. Limited correspondence in genomic regions recalcitrant for donor chromatin between the two reciprocal populations suggested the effect of species background on the introgressed chromatin segments. Furthermore, we selected near-isogenic lines (NILs) containing only one chromosomal segment introgressed

from the donor parent in the recipient genome in such a way that the donor segments in the recipient genome would combinedly tile most of the donor parent's genome. A total of 399 and 423 NILs were selected in the Acala Maxxa and Pima S6 backgrounds, representing 78.72% and 71.48% of the Pima S6 genome and Acala Maxxa genomes respectively. These populations (both advanced backcross lines and NILs) were evaluated for six major fiber quality traits as well as for lint percentage and for five phenological traits. A total of 206 QTLs were identified for these traits, majority of which were small effect QTLs (i.e., explaining <10% of phenotypic variance) exemplifying the merit of these populations in identification of small-effect QTLs. Although some reciprocal QTLs were identified, limited reciprocity of majority of the QTLs in the two backgrounds shows strong influence of recipient genome, in addition to the combined consequences of epistasis, small phenotypic effects and imperfect coverage of donor chromatin in the recipient background.

INDEX WORDS: *Gossypium hirsutum*, *Gossypium barbadense*, Fiber quality traits, Near-isogenic lines, Interspecific Introgression, Advanced backcross population, QTL analyses, Transmission genetics

GENETIC DISSECTION OF COTTON FIBER QUALITY TRAITS IN RECIPROCAL ADVANCED BACKCROSS POPULATIONS

by

JEEVAN ADHIKARI

B.S., Tribhuvan University, Nepal, 2010

M.S., University of Georgia, 2015

A Dissertation Submitted to the Graduate Faculty of The University of Georgia in Partial Fulfillment of the Requirements for the Degree

DOCTOR OF PHILOSOPHY

ATHENS, GEORGIA

2022

© 2022

JEEVAN ADHIKARI

All Rights Reserved

GENETIC DISSECTION OF COTTON FIBER QUALITY TRAITS IN RECIPROCAL ADVANCED BACKCROSS POPULATIONS

by

JEEVAN ADHIKARI

Major Professor:	Andrew H Paterson
Committee:	Peng-Wah Chee
	Jason Wallace
	Paul Schliekelman
	Brian Schwartz

Electronic Version Approved:

Ron Walcott
Vice Provost for Graduate Education and Dean of the Graduate School
The University of Georgia
December 2022

DEDICATION

I would like to dedicate this dissertation to my parents, my lovely wife Roshani, and my two beautiful kids - Aayan and Aaditya.

ACKNOWLEDGEMENTS

I would like to send great appreciation to my major professor, Dr Andrew H Paterson, for his inspiration, guidance, and support. I would also like to thank my respectable committee members for their support and countless valuable advice on this research. Special thanks to Dr. Chee and his lab members for their help and support in cotton planting in Plains. I would also like to thank all members of Plant Genome Mapping Laboratory for their help and support.

TABLE OF CONTENTS

Page

CHAPTER 1

INTRODUCTION AND LITERATURE REVIEW

Cotton in a nutshell

Cultivated cotton is a major source of natural fiber for textile industries all around the world and one of the important sources of oil seeds. This high value commodity crop of the arid and semi-arid tropics is grown in over 75 countries globally, the major share of which is accounted for by China, India, United States and Brazil. Cotton has been used for at least 5000 years (Lubbers & Chee, 2009) for yarn spinning, weaving, and dyeing to make clothes in some of the oldest civilizations such as those in Indus valley and Nile valley. The United States cotton industry, ranking third in the world after China and India in production; but first in export of raw fiber, accounts for more than $21 billion in products and services annually, generating more than 125,000 jobs in the industry sectors from farm to textile mills (USDA, 2016). Quality of cotton fiber, which refers to a set of measurements describing the physical properties of samples extracted from a cotton bale, determine, in part, the market value of the raw cotton (Bradow & Davidonis, 2000). Modern high speed spinning yarn producing technologies have not only enhanced the production efficiency of textiles, but also raised the requirements of long and strong cotton fibers (Zeng et al., 2007). Cotton fibers with desirable quality not only help in maintaining and enhancing yarn processing efficiencies but also influence the quality of the end- product. Recent breeding activities have aimed at increasing some of the parameters defining cotton fiber quality. Knowledge on the biological processes and genetic mechanism underlying the changes and

progresses during the growth and development of cotton fiber is important to achieve higher success in breeding for different agronomic and fiber quality traits.

Taxonomy and gene pools / breeding pools

Cotton belongs to the family of Mallows (tribe *Gossypieae* and family *Malvaceae*) and genus *Gossypium* L., which represents 45 diploid and 5 allopolyploid extant species (Fryxell & Craven, 1992; Vollesen, 1987). Diploid *Gossypium* spp. have 13 pairs of chromosomes (2n=26) and are divided into eight different genome types, designated as A, B, C, D, E, F, G, or K. They are mostly distributed in African, Asian, Australian, and North American tropics and subtropics (Endrizzi et al., 1985; Wendel et al.). On the other hand, the allopolyploids evolved in the New World (Central America) and have 26 pairs of chromosomes (2n=52) with complete A and D chromosome complements in their somatic cells (Wendel & Albert, 1992). Two old world diploids (*G. arboreum* and *G. herbaceum*) and two new world allopolyploids (*G. hirsutum* and *G. barbadense*) are domesticated and cultivated for their spinnable fibers. However, Upland cotton (*G. hirsutum*) alone accounts for more than 90% of global cotton production followed by Pima (also Egyptian; *G. barbadense*) cotton which contributes around 9% of the total (Abdurakhmonov, 2012).

Gossypium species are grouped into different genepools based on their compatibility for hybridization with each other (Beasley and Brown 1942, Stephens 1949, Hutchinson 1951) and retention of introgressed chromatin. Based on classical concepts (Harlan and de Wet 1971), the primary gene pool consists of species that are easy hybridize with one another, with normal chromosome pairing, high recombination rates and only minimal distortion. Classically, the primary gene-pool of cultivated AD-genome cotton consisted of all cross compatible AD-genomes

including *G. tomentosum*, *G. darwinii* and *G. mustelinum* as well as the wild, landrace, cultigens, and feral ecotypes of G. barbadense and 12 G. hirsutum (Khadi et al. 2010). Although AD-genome species are cross-compatible, natural hybrids or introgressed mosaics among sympatric species do not appear to persist and colonize over time and space, thereby limiting natural shuffling of favorable alleles among these genomes (Lehman et al. 2014; Pereira et al. 2012; Wendel et al. 1992). The reasons may be multilocus epistatic interactions and divergent gene regulatory systems restricting chromatin transmission (Jiang et al. 2000) or gene order rearrangements and cryptic chromosomal aberrations. As such, the four tetraploid species viz. *G. barbadense*, *G. darwinii*, *G. mustelinum*, and *G. tomentosum* could be reclassified as secondary gene pool for G. hirsutum (Lubbers and Chee 2009). These species form hybrids with *Gossypium hirsutum* that are viable and fertile but with noticeable segregation distortion and hybrid breakdown to some extent (Hutchinson 1951). The tertiary genepool consists of species that are difficult to hybridize with *Gossypium hirsutum*, usually requiring embryo rescue or chromosome doubling to obtain advanced generations after hybridization, including all diploid *Gossypium* species.

Upland cotton accessions in the United States are classified into four distinct breeding pools, namely *Acala*, *Plains*, *Delta*, and *Eastern*, representing cultivars adapted to four partially overlapping production regions of the country (Niles and Feaster 1984). *Acala* types have superior fiber and spinning qualities and are better adapted to the irrigated Southwest. The second breeding pool constitutes "*Plains*" types which roughly cover one half of the US cotton belt and are mostly suited to Texas, Oklahoma, and eastern New Mexico. The third, *Delta* types are primarily grown in the rain-belt area from southern Texas to Alabama and cover approximately one third of the US cotton production area while *Eastern* types are mostly adapted to Southeastern states of Georgia

and the Carolinas. Most of the Pima cotton is grown in and around California because of its warm springs, hot summers, and dry falls.

Cotton fiber and its development

Cotton fibers are highly elongated unicellular seed trichomes that develop from the outer epidermal cells of the ovule at or around the day of anthesis (DOA). They are primarily composed of cellulose and generally grow up to 30 to 40 mm in length and about 15 micrometer in diameter upon full maturity (Basra & Malik, 1984). About 100,000 fine epidermal cells grow out non-synchronously in each ovule (Guan et al., 2014; Hee Jin Kim & Barbara A. Triplett, 2001) and go through a series of developmental stages ending up becoming fully mature cotton fibers that are post-developmentally processed in the textile industries to produce end products of human needs and benefits. Thus, improvements in cotton fiber properties for improved quality of the textiles depend on the physiological changes occurring during the growth and development of cotton fibers.

The process of transformation of these primitive seed trichomes to economically important cotton fibers, progresses through four well-defined but overlapping developmental stages: fiber initiation, cell elongation, secondary wall deposition and maturation or dehydration (Basra & Malik, 1984; Jasdanwala et al., 1977). The first three stages occur while cotton fiber is still alive and actively growing, while the final stage of development occurs after the boll opens and describes drying of fully formed cotton fiber (Seagull et al., 2000). The first stage is the initial development stage, extending from about 3 days before anthesis (DBA) to about 3-days post anthesis (DPA). This stage is characterized by several waves of initial isodiametric expansion of

the epidermal cells across the surface of the ovule, thus making it possible to observe fiber initials at the ovule surface for a period of about 5-6 DPA (James Mc, 1975).

Following the formation of initials, fibers enter the stage of rapid expansion and exponential growth, which depending upon genotypes, lasts for several weeks. It is during this stage, when a thin primary cell wall, composed of different carbohydrate polymers, gets deposited around the fiber (Meinert & Delmer, 1977). The third stage in the development of cotton fibers is marked by the gradual cessation of the elongation process and biosynthesis of secondary cell wall. Unlike many fiber- or trichome-bearing crops, the secondary cell wall in cotton fibers is primarily cellulose (Meinert & Delmer, 1977; Seagull et al., 2000). This stage usually begins around 12 to 15 DPA and persists until fiber maturation (Seagull et al., 2000), and is characterized by progressive thickening of the cell wall and gradual decrease in the volume of living protoplast. The final stage is maturation or dehydration stage where the fibers lose water gradually and growth or elongation of the fibers is ceased.

Cotton fiber properties

Although higher yields and yield stability remain top priorities for cotton improvement in the US, the paradigm has shifted to include better fiber quality in responses to the evolution of efficient spinning technologies and to the formidable challenge posed by the synthetic fibers, which created export directed demands for high quality cotton fibers. Since the quality of raw fiber determines yarn processing efficiency as well as the quality of the final produce (for example, textile), market-value of raw cotton is adjusted based on its quality. However, fiber quality is a collective term encompassing a suite of physical properties that are measured on fiber samples from a cotton bale (Bradow & Davidonis, 2000). As such, several determinants dictate the market

value of a bale of cotton. The major quality determinants used by textile industries for cotton pricing and quality control include length, strength, elongation, fineness, and uniformity of the cotton fibers, which are measured by high-volume instrument (HVI) system (www.cottoninc.com).

Cotton lint fiber is the longest unicellular extension of a plant cell (H. J. Kim & B. A. Triplett, 2001), the property which makes it a key attribute to the textile industry. HVI reports 'the upper-half-mean length' (UHM), which is basically the mean length of the longest 50% of fibers in a sample measured, in hundredths of an inch (www.cottoninc.com). UHMs less than 0.99" is short, around 1.00" is medium, between 1.11" and 1.26" is long, and above 1.26" is extra-long (www.cotton.org). Longer fibers produce better and stronger yarns as they have greater resistance to friction during the processing (Broughton et al. 1992).

Fiber strength is generally reported in grams per tex (grams/tex), which is the force required to break a bundle of sample fibers clamped 1/8 inch between the two sets of jaws. Grams/tex implies force (in grams) required to break a tex (weight in grams of 1,000 meters of fiber) of sample (www.cottoninc.com). Strength of less than 26 grams/tex is weak, between 26-29 grams/tex is medium, 30-32 grams/tex is strong, and above 33 grams/tex is very strong. Stronger fibers resist snapping during processing, which correlate with higher yarn strength (P. W. Chee & B. T. Campbell, 2009). Fiber elongation is a measure of elasticity in percentage, which is simultaneously measured when force is applied to break a bundle of sample fibers to find the fiber strength (Bradow & Davidonis, 2000). In general, less than 5.9% elongation is low, 5.9% to 6.7% elongation is medium, 6.8% - 7.6% is high, and above 7.6% is very high. Higher elongation percent means higher yarn stretchiness which helps withstand fiber and yarn processing stresses but does not directly contribute to yarn strength (May, 1999; May & Green, 1994).

Fiber fineness is indirectly measured as micronaire, which is an airflow measurement on a weighed fiber sample compressed to a specific volume in a chamber (www.cottoninc.com). Compressed fiber sample resist airflow and this resistance is proportional to the linear density of the fibers, which is expressed 11 in micrograms per inch and adjusted for the maturity of the fiber. Fiber fineness-based market value adjustment considers a range of 3.7 to 4.2 micronaire as premium quality, ranges of 3.5 to 3.6 and 4.3 to 4.9 as base quality, and range of 3.7 to 4.2 as premium quality (www.cotton.org). Thick fibers result in coarse textile with lower thread counts, while very fine (immature) fibers adversely affect the spinning efficiency and fabric properties (Basra & Malik, 1984; Grover & Hamby, 1960). In fact, fiber fineness affects yarn strength more than the fiber strength (Sattar and Hussain 1985).

QTL mapping of fiber quality traits

Most fiber quality traits are controlled by groups of genes reflected as groups of quantitative trait loci (QTL). Phenotypic variance of these traits is explained by the joint action of QTL and their interactions with environment (Paterson et al., 2003). Rapid advances in molecular marker technology have facilitated the construction of detailed genetic and QTL maps of cotton. The first QTL mapping in cotton was reported by (Shappley et al.) in 1996 and later on various studies on QTL mapping of fiber qualities were reported (Jiang et al., 1998; Lacape et al., 2005; Paterson et al., 2003; Saranga et al., 2001; Tang et al., 2015; Ulloa et al., 2002; Ulloa & Meredith, 2000; Zhang et al., 2003). Recently, high throughput sequencing technologies have widened the avenues for rapid identification of SNP and SSR markers and consequently have aided in the construction of high-density genetic maps (Liu et al., 2015; Shi et al., 2015; Zhen et al., 2016).

To date, only a handful of studies have dealt with intraspecific crosses to map various agronomic and fiber quality traits to the cotton genome. Due to higher quality fiber of *G. barbadense* and high yield potential of *G. hirsutum*, together with the relatively high level of DNA polymorphism between these species, many QTL mapping studies have been conducted using crosses between them (Lacape et al., 2005). A pioneering study mapped 14 QTLs for agronomic and fiber quality traits on a linkage map derived from an interspecific cross of *G. hirsutum* cv. 'CAMD-E' and *G. barbadense* cv. 'Sea Island Seaberry' (Jiang et al., 1998) and also found that the most QTLs influencing fiber quality and yield were located on the "D" sub-genome, derived from an ancestor that does not produce spinnable fibers. Using an interspecific map developed from RFLP, SSR and AFLP markers; seven QTLs for various fiber-related traits were identified (Mei et al., 2004). A comprehensive analysis of advanced generation backcross populations from a cross between *G. hirsutum* cv. Tamcot 2111 and *G. barbadense* cv. Pima S6 using RFLP markers detected 22 QTLs for fiber elongation (P. Chee et al., 2005), 32 and 9 QTLs for fiber fineness (X. Draye et al., 2005), 28, 9 and 8 QTLs for fiber length, uniformity and short fiber content, respectively (P. W. Chee et al., 2005).

To date more than 2000 QTLs related to fiber quality traits have been reported (J. Said et al., 2015). However, QTL meta-analyses show that many new QTL remain to be discovered, especially those of small phenotypic effect (Junkang Rong et al., 2007). Moreover, fine mapping of QTL may improve the ability to use them across species (for example in combining the yield of *G. hirsutum* with the quality of *G. barbadense*). Together, these needs warrant the development of populations which can be used in fine-scale and sensitive mapping of small effect QTLs. Among the various possible mapping populations (F_2, Recombinant Inbred Lines, Doubled-haploids,

Backcrosses, Chromosome Substitution Lines, etc.), near-isogenic lines developed by repeated backcrossing of the progenies to the maternal parent best serve the purpose.

Genotyping for QTL mapping

Establishment of genetic or physical maps of the genome of an organism is prerequisite for mapping of QTLs related to essential traits. In cotton, this has been facilitated by the establishment of plethora of genetic maps in both inter- and intra-specific crosses. From the very first genetic map constructed using AFLP makers (Reinisch et al., 1994b) to the recent high density maps made using SNP data obtained from high throughput next generation sequencing technologies, they have served as a backbone to establish the location of the causal genomic segment for the variation in observed phenotypes. Recent advances in next generation sequencing (NGS) technologies have boosted the process of identification of large number of single nucleotide polymorphism (SNP) markers and consequently the construction of very high-density genetic maps.

Genotyping by sequencing (GBS) has proven to be a cost-effective method of identifying and employing large number of SNP makers in making genetic maps and tagging QTLs to these maps (Peter Andolfatto et al., 2011). In this method, polymorphisms are scored using NGS technologies followed by a bioinformatics pipeline. The advantage of GBS is that it reduces cost through an enzyme-based genomic complexity reduction step and the use of barcoded adapters for multiplexing (Poland & Rifeb, 2012). For those species that have a reference genome available, GBS has proven to be more efficient for sequence-based genotyping. A reference genome makes ordering and imputing low coverage marker data generated through GBS and other sequence based genotyping approaches straightforward. Although GBS approaches greatly benefit from a reference genome, the rapid discovery and ordering (through genetic mapping) of sequence-based

molecular markers can assist with the development and refinement of a reference genome (Poland & Rifeb, 2012). In addition, high-density genetic maps developed through GBS can be used to anchor and order physical maps and refine or correct unordered sequence contigs.

Near-isogenic lines (NILs)

Fiber quality traits are quantitative in nature and are characterized by continuous variation of phenotypes. Over the last few decades, hundreds of QTLs related to fiber quality have been identified and mapped to different positions on the cotton genome, mostly using the F_2 generation of inter- and intra-specific hybrids. Mapping of QTLs using early generation populations (F_2 or backcross) pose two major limitations. First, early generation populations cannot be truly replicated and scored in different years or locations (unless they could be clonally propagated) and hence, it would be necessary to use different genotypes each time. This would require repeated marker genotyping which is inefficient and prone to errors. Second, the amount of recombination events in such populations is very small which might result in the presence of large linkage blocks and consequently tight unwanted linkages (Newbury, 2003). Thus, in these types of populations the confidence interval (CI) associated with a QTL could be very large. The CI would be much larger if the traits under study have low heritability and smaller QTL effect and the size of the population is small.

Even for advanced populations like the recombinant inbred lines (RILs) with about 100 to 200 lines, CIs for detectable QTLs are seldom less than 5 cM and most often greater than 20 cM (Kearsey & Luo, 2003). Having additional markers above one every 10 cM does not have any significant decrease in the CI, unless the population size is increased exponentially (Melchinger et al., 2000; Utz et al., 2000). Thus, identification of approximate location of a QTL requires such

kind of genetic materials that have defined and narrow limits on the possible region of chromosome that might possibly harbor the QTL. This ideally involves producing genotypes that are genetically identical except for a short chromosomal region and then demonstrating that they differ only for the traits of interest. Obviously, the smaller the region, the more precisely a QTL can be located.

NILs are lines containing a single (or small number of) homozygous genomic introgression from a donor parent in a different and otherwise homogeneous genomic background. The initial progenies selected from crosses are heterogeneous for the given segment, which are then selfed to obtain homozygous lines. Depending on the available resources, NILs can be constructed in various ways, the simplest being crossing of two diverse parents and repeatedly backcrossing the progenies to one of the parents to eventually retrieve the required genomic constitution. NILs serve many functions, ranging from breeding purposes to genetic analysis of complex traits (Rik Kooke et al., 2012). The ultimate objective of these lines, in large, determines the choice of parents, the crossing scheme and eventually their genomic composition. The size and number of genomic regions introgressed from a donor genome into a recipient background also depends upon the objectives of the study, but generally a single small segment is preferred.

NILs have mostly been used to verify the effects of one or a few QTLs that were previously identified using RILs or other types of mapping populations (Szalma et al., 2007a). However, NILs developed for such purposes contain only a small portion of the donor genome. If the purpose of the developed NILs is to map incorporated QTLs, then these lines should be developed in a way that each line carries only a small portion of the donor genome, but across the whole set of NILs, the overall genome (ideally) of the donor is represented. The introgression of as many chromosomal regions of the donor parent as possible (possibly with some replications) allows testing of their individual effects in the near-isogenic backgrounds.

Apart from their use in verification of QTLs introgressed into a genomic background, NILs offer advantage of being a good mapping population. NILs can be used to simultaneously map, verify and incorporate QTLs into elite genetic backgrounds (Eshed & Zamir, 1995). The genetic and genomic composition of these lines make them a good population for the mapping of complex quantitative traits using single factor analysis. In addition, they allow simultaneous introgression and discovery of QTLs for important agronomic as well as plant architectural traits. Similar strategy was suggested for maize by Stuber et al. (1999) to improve the utilization of DNA markers in mapping, verifying and incorporating QTLs. Since each line contains very small introgression(s), NILs offer an important advantage over traditional QTL mapping approaches in that they improve the ability to identify QTLs/genes with small phenotypic effects and providing for testing of the main effects of small genomic regions. In addition, the presence of small introgressed regions in NILs, following several generations of backcrossing, removes (to a large extent) the linkage drag which could otherwise be a significant issue in early generation populations. By homogenizing all genetic factors outside of the focal genomic region, the true effect of a QTL on the phenotype can be estimated relative to the line (background) into which the introgression was introduced (Landi et al., 2005).

Another major advantage of NILs is that the resolution to which a gene/QTL can be mapped can be improved by minimizing the size of the introgressed segments. This certainly comes with the drawback of having to maintain a larger population (to create a genome wide library of NILs) and in some instances, perform extra rounds of backcrossing. In addition to the simplification of genetic analyses, NILs are considered genetically 'immortal' which allows for replicated experiments across multiple environments resulting in more accurate estimates of effect size for complex traits. These genetically immortal populations share the advantage that they can

easily be maintained through seeds and allow endless study of multiple, even invasive or destructive, traits. Statistical power of such analyses is increased because replicate measurements of genetically identical individuals can be done (Fletcher et al., 2013).

Finally, NILs have proven to be an effective resource for QTL validation and a logical starting point for the creation of fine-mapping populations. NILs are also very useful for comparative physiological and biochemical studies of the function of a single gene as the introgressed segments they carry are often the size of many functional genes. However, the presence of a single introgression segment does not allow testing for genetic interactions and thereby the detection of QTL expressed in specific genetic backgrounds i.e. epistasis (Fletcher et al., 2013; Keurentjes et al., 2007). In addition, because most of the genetic background is identical for all lines, NILs show more limited developmental and growth variation, increasing the homogeneity of growth stage within experiments. This would allow identification and characterization of a specific phenotype of interest. Nevertheless, lethality and sterility might sometimes hinder the obtaining of specific single introgression lines (Keurentjes et al., 2007).

The statistical power of QTL detection is an important consideration in QTL mapping studies. Kaeppler (1997) demonstrated lesser statistical power in NIL-based QTL mapping tests than in RIL-based tests, when each NIL was paired with the recurrent parent in the experimental design. This is not a potential drawback of the NIL based mapping strategy because the experimental design can be improved by comparing multiple NILs to the recurrent parent at the same time. Furthermore, albeit the power to detect a single QTL is higher in RILs than in NILs, NILs offer more accurate QTL effect estimates than RILs if multiple QTLs are segregating in the population (Szalma et al., 2007a). In addition, the effects of multiple segregating QTLs are overestimated in typical population sizes used in RIL mapping studies (Beavis, 1997; Melchinger

et al., 1998), while in NILs, at similar population sizes, much of the noise caused by the effects of the background is reduced, thus the resulting differences in phenotypes between the recurrent parent and the NILs is primarily due to the allelic differences at the introgressed locus (Szalma et al., 2007a). Homogeneous genetic background in NILs eliminates collinearity between QTLs that typically occur in RIL populations. The same property of NILs make them useful in developing trait-linked molecular markers. If a NIL is showing different phenotype than the recurrent parent, then DNA sequence difference in or nearby the introgressed region is most possibly linked tightly to the trait of interest.

Albeit several QTL mapping studies have reported significant QTLs in the literature to date, the utilization of these QTLs in germplasm improvement is not common (Holland, 2004). This gap between the research community and the breeding community has been one of the major obstacles in germplasm development and utilization. Employing NIL-based approach of QTL discovery might help get around this obstacle (Stuber et al., 1999). First, use of NILs would reduce the number of generations of backcrossing required to incorporate favorable QTL alleles into an elite line. Second, NILs that perform superior performances can released immediately (Stuber et al., 1999). In such superior NILs, the epistatic interaction between the introgressed region and the genetic background, if any, must either be favorable or insignificantly negative (Tanksley & Nelson, 1996b), thus facilitating their direct release as cultivars.

NILs, nevertheless, are one of the complex populations to create and deal with. The major setback in working with NILs is the time required just to develop them. In general, it takes about 6 to 8 years just to construct these lines and some additional years of trial and testing for genetic analyses. As mentioned previously, another major drawback of NILs is the study of epistatic interaction of donor loci in the recipient background. Finally, power to detect QTLs in NILs is

relatively lower than that in RILs, with the relative power in NILs being highest for traits with high heritability and high precision gain due to increased replication of entries (Kaeppler, 1997).

Construction of a NIL-library

Creation of a single near-isogenic line generally starts by crossing a line carrying the targeted QTL region to one of the parental lines of the population, thus creating a backcross population. Each backcross line is then repeatedly crossed back to the recipient parent, until each line contains one (usually) small segment of the donor genome in the recipient background. Additionally, heterogeneous inbred families derived from RILs can also be used to derive NILs. The RIL populations consist of largely homogeneous genetic background with some segregating loci. RILs that are heterozygous at the given QTL/loci can be selfed to produce NILs (Tuinstra et al., 1997). NILs derived from such method may serve some purposes depending upon the number of introgressed loci and the portion of the donor genome they cover. Genome-wide genotyping of the backcross progeny is performed to identify recombination events allowing for selection of progeny which carry the target chromosomal introgression derived from the donor and recurrent parent genome elsewhere. Subsequent generations of self-pollination (selfing) are normally required to achieve homozygosity of the introgressed region and the process can take several backcrossing cycles to produce a NIL carrying an introgression of acceptable size and genomic location (Keurentjes et al., 2007). Producing NILs with smaller introgressions requires greater effort. Large populations are needed to break up small chromosomal segments, and high-density genotyping is required to discover them (Fletcher et al., 2013).

A NIL-library is a family of near-isogenic lines where each line carries one random donor parent fragment, and the family carries introgressions spanning the entire donor genome.

Development of a NIL-library can serve three purposes; a) identify a QTL, b) Fine map the identified QTL to a small genomic location, and c) incorporate a QTL into elite germplasm (Eshed & Zamir, 1994). Development of NILs using exotic germplasm as a donor also helps amplify the genetic diversity of elite germplasm. The approach here is to produce sets of lines that differ only in a small genomic region of our interest, so that all other genes/QTLs affecting the region of our interest (either through masking or through epistatic interactions) are same in all the lines, thus, "Mendelizing" the locus of our interest.

These genetic stocks not only provide for reduction in the complexity of QTL studies, but also offer a valuable resource for identification of high-likelihood candidates for QTLs for virtually any trait, by identifying and characterizing additional recombinants in a region of interest. If there is more than one introgressed segment differing between a NIL and the recipient genotype then epistatic effects of one QTL on another can be studied (Tanksley & Nelson, 1996b). NILs that have superior hybrid performance can be directly released as new breeding lines or cultivars (Stuber et al., 1999). QTL mapping greatly benefits from minimization of genetic background noises as such effects create complications in studying exact effects of QTL on a specific trait. NILs offer a pragmatic solution in studying the effect of a particular QTL with minimal influence of background noises (Tanksley & Nelson, 1996b).

Rationale of the study

Mature cotton fibers, produced from tiny seed trichomes after passing through the four stages of development, are the primary subject of interest in textile industries. The value of these fibers is mainly dependent on their quality as defined by length, strength, fineness, elongation, and uniformity. A lot of physiological changes, accompanied by associated changes in transcriptome

alternations, occur during these four stages of growth and development of cotton fiber (Hinchliffe et al., 2010). Identification and understanding of these alterations is important to dissect the stages involved in transforming primitive trichomes to the economically important fibers of the modern cotton cultivars. To better understand the stages involved in transformation of the epidermal cell into mature cotton fiber and identify associated transcriptomic alterations, we constructed two panels of reciprocal near-isogenic lines (NILs), each of which contain one and only one introgressed segment from the donor genotype, but collectively cover a major portion of the donor genome. These NILs, each consisting about 0.5% of the donor genome in a reference background, also provide a powerful tool for genetically dissecting complex traits like fiber quality traits, thereby increasing the precision with which phenotypic changes can be mapped to transcriptomic and genetic alterations.

References

Abdurakhmonov, I. Y. (2012). Genetic Diversity in Gossypium genus. In: IntechOpen.

Adhikari, J. (2015a). *Effects of exotic genotypes and genetic backgrounds on fiber quality and plant architectural traits in Upland cotton (Gossypium hirsutum L.)* University of Georgia].

Adhikari, J. (2015b). *Effects of exotic genotypes and genetic backgrounds on fiber quality and plant architectural traits in Upland cotton (Gossypium hirsutum L.)* [BibliographiesTheses, Electronic, Non-fiction].

Adhikari, J., Das, S., Wang, Z., Khanal, S., Chandnani, R., Patel, J. D., Goff, V. H., Auckland, S., Rainville, L. K., & Jones, D. (2017). Targeted identification of association between cotton fiber quality traits and microsatellite markers. *213*(3), 1-15.

Al-Ghazi, Y., Bourot, S., Arioli, T., Dennis, E. S., & Llewellyn, D. J. (2009). Transcript profiling during fiber development identifies pathways in secondary metabolism and cell wall structure that may contribute to cotton fiber quality. *Plant Cell Physiol*, *50*(7), 1364-1381.

Anderson, E. (1949). *Introgressive hybridization*. J. Wiley.

Andolfatto, P., Davison, D., Erezyilmaz, D., Hu, T. T., Mast, J., Sunayama-Morita, T., & Stern, D. L. (2011). Multiplexed shotgun genotyping for rapid and efficient genetic mapping. *Genome Research*, *21*(4), 610-617.

Andolfatto, P., Davison, D., Erezyilmaz, D., Hu, T. T., Mast, J., Sunayama-Morita, T., & Stern, D. L. (2011). Multiplexed shotgun genotyping for rapid and efficient genetic mapping. *21*(4), 610-617.

Baack, E., Melo, M. C., Rieseberg, L. H., & Ortiz-Barrientos, D. (2015). The origins of reproductive isolation in plants. *New Phytologist*, *207*(4), 968-984.

Basra, A. S., & Malik, C. P. (1984). Development of the Cotton Fiber. In J. F. D. G.H. Bourne & K. W. Jeon (Eds.), *International Review of Cytology* (Vol. Volume 89, pp. 65-113). Academic Press.

Beasley, J. O. (1942). Meiotic chromosome behaviour in species, species hybrids, haploids, and induced polyploids of Gossypium [Journal Article]. *Genetics*, *27*, 25-54.

Beavis, W. (1997). QTL Analyses. In *Molecular Dissection of Complex Traits* (pp. 146-162). CRC Press. 10.1201/9781420049381.ch10

Bradow, J. M., & Davidonis, G. H. (2000). Quantitation of fiber quality and the cotton production-processing interface: a physiologist's perspective. *Journal of Cotton Science*, *4*(1), 34-64.

Broman, K. W. (2001). Resource Review of Statistical Methods for QTL Mapping in Experimental Crosses. *30*(7), 44-52.

Broman, K. W., & Sen, S. (2009). *A guide to QTL mapping with R/qtl* [Bibliographies, Online, Non-fiction, Electronic document]. Springer.

Brown, N., Kumar, P., Khanal, S., Singh, R., Suassuna, N. D., McBlanchett, J., Lubbers, E., Jones, D., Paterson, A. H., & Chee, P. W. (2020). Registration of eight upland cotton (Gossypium hirsutum L.) germplasm lines with qFL-Chr.25, a fiber-length QTL introgressed from Gossypium barbadense. *Journal of Plant Registrations*, *14*(1), 57-63.

Brown, N., Kumar, P., Singh, R., Lubbers, E., Campbell, B. T., Myers, G. O., Wright, R. J., Subramani, J., Jones, D., Paterson, A. H., & Chee, P. W. (2019). Evaluation of a Chromosome Segment from Gossypium barbadense Harboring the Fiber Length QTL qFL-Chr.25 in Four Diverse Upland Cotton Genetic Backgrounds. *Crop Science*, *59*(6), 2621-2633.

Brubaker, C. L., Paterson, A. H., & Wendel, J. F. (1999). Comparative genetic mapping of allotetraploid cotton and its diploid progenitors. *42*(2), 184-203.

Chandnani, R., Kim, C., Guo, H., Shehzad, T., Wallace, J. G., He, D., Zhang, Z., Patel, J. D., Adhikari, J., Khanal, S., & Paterson, A. H. (2018). Genetic Analysis of Gossypium Fiber Quality Traits in Reciprocal Advanced Backcross Populations. *The Plant Genome*, *11*(1), 170057.

Chandnani, R., Wang, B., Draye, X., Rainville, L. K., Auckland, S., Zhuang, Z., & Lubbers, E. L. (2017). Segregation distortion and genome-wide digenic interactions affect transmission of introgressed chromatin from wild cotton species [Report]. *Theoretical and Applied Genetics*, *130*(10), 2219.

Chee, P., Bredhauer, R., Smith, C. W., Paterson, A. H., Delmonte, T. A., Draye, X., Jiang, C. X., & Decanini, L. (2005). Molecular dissection of interspecific variation between Gossypium hirsutum and Gossypium barbadense (cotton) by a backcross-self approach. I. Fiber elongation. *Theoretical and Applied Genetics*, *111*(4), 757-763.

Chee, P., Draye, X., Jiang, C.-X., Decanini, L., Delmonte, T. A., Bredhauer, R., Smith, C. W., & Paterson, A. H. (2005). Molecular dissection of interspecific variation between Gossypium hirsutum and Gossypium barbadense (cotton) by a backcross-self approach: I. Fiber elongation. *Theoretical and Applied Genetics*, *111*(4), 757-763.

Chee, P. W., & Campbell, B. T. (2009). Bridging Classical and Molecular Genetics of Cotton Fiber Quality and Development. In A. H. Paterson (Ed.), *Genetics and Genomics of Cotton* (pp. 283-311). Springer US.

Chee, P. W., & Campbell, B. T. (2009). Bridging Classical and Molecular Genetics of Cotton Fiber Quality and Development. *3*, 283-312.

Chee, P. W., Draye, X., Jiang, C. X., Decanini, L., Delmonte, T. A., Bredhauer, R., Smith, C. W., & Paterson, A. H. (2005). Molecular dissection of phenotypic variation between Gossypium hirsutum and Gossypium barbadense (cotton) by a backcross-self approach: III. Fiber length [Journal Article]. *TAG Theoretical and Applied Genetics*, *111*(4), 772-781.

Cheng, H.-Q., Han, L.-B., Yang, C.-L., Wu, X.-M., Zhong, N.-Q., Wu, J.-H., Wang, F.-X., Wang, H.-Y., & Xia, G.-X. (2016). The cotton MYB108 forms a positive feedback regulation loop with CML11 and participates in the defense response against Verticillium dahliae infection. *Journal of Experimental Botany*, *67*(6), 1935-1950.

Desai, A., Chee, P. W., Junkang, R., May, O. L., Paterson, A. H., & Gustafson, J. P. (2006). Chromosome structural changes in diploid and tetraploid A genomes of Gossypium [Article]. *Genome*, *49*(4), 336-345.

Dou, L., Li, Z., Wang, H., Li, H., Xiao, G., & Zhang, X. (2022). The hexokinase Gene Family in Cotton: Genome-Wide Characterization and Bioinformatics Analysis [Original Research]. *Frontiers in Plant Science*, *13*.

Draye, X., Chee, P., Jiang, C.-X., Decanini, L., Delmonte, T. A., Bredhauer, R., Smith, C. W., & Paterson, A. H. (2005). Molecular dissection of interspecific variation between Gossypium hirsutum and G. barbadense (cotton) by a backcross-self approach: II. Fiber fineness [Article]. *Theoretical & Applied Genetics*, *111*(4), 764-771.

Draye, X., Chee, P., Jiang, C. X., Decanini, L., Delmonte, T. A., Bredhauer, R., Smith, C. W., & Paterson, A. H. (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*, *111*(4), 764-771.

El-Sharkawy, I., Sherif, S., Mila, I., Bouzayen, M., & Jayasankar, S. (2009). Molecular characterization of seven genes encoding ethylene-responsive transcriptional factors during plum fruit development and ripening. *Journal of Experimental Botany, 60*(3), 907-922.

Endrizzi, J. E., Turcotte, E. L., & Kohel, R. J. (1985). Genetics, Cytology, and Evolution of Gossypium. In E. W. Caspari & J. G. Scandalios (Eds.), *Advances in Genetics* (Vol. 23, pp. 271-375). Academic Press.

Eshed, Y., & Zamir, D. (1994). A genomic library of Lycopersicon pennellii in L. esculentum: a tool for fine mapping of genes. *Euphytica*, *79*(3), 175-179.

Eshed, Y., & Zamir, D. (1995). An introgression line population of Lycopersicon pennellii in the cultivated tomato enables the identification and fine mapping of yield-associated QTL. *Genetics*, *141*.

Fang, D. (2015). Molecular Breeding. In. https://doi.org/10.2134/agronmonogr57.2013.0027

Fang, D. D., Jenkins, J. N., Deng, D. D., McCarty, J. C., Li, P., & Wu, J. (2014). Quantitative trait loci analysis of fiber quality traits using a random-mated recombinant inbred population in Upland cotton (Gossypium hirsutum L.). *BMC Genomics*, *15*(1), 397.

Fang, L., Wang, Q., Hu, Y., Jia, Y., Chen, J., Liu, B., Zhang, Z., Guan, X., Chen, S., Zhou, B., Mei, G., Sun, J., Pan, Z., He, S., Xiao, S., Shi, W., Gong, W., Liu, J., Ma, J., . . . Zhang, T. (2017). Genomic analyses in cotton identify signatures of selection and loci associated with fiber quality and yield traits. *Nature Genetics*, *49*(7), 1089-1098.

Fankhauser, C., & Chory, J. (1997). Light control of plant development. *Annual review of cell and developmental biology*, *13*, 203-229.

Feltus, F. A., Hart, G. E., Schertz, K. F., Casa, A. M., Kresovich, S., Abraham, S., Klein, P. E., Brown, P. J., & Paterson, A. H. (2006). Alignment of genetic maps and QTLs between inter- and intra-specific sorghum populations. *Theor Appl Genet, 112*(7), 1295-1305.

Feng, J. X., Ji, S. J., Shi, Y. H., Xu, Y., Wei, G., & Zhu, Y. X. (2004). Analysis of five differentially expressed gene families in fast elongating cotton fiber. *Acta Biochim Biophys Sin (Shanghai), 36*(1), 51-56.

Flagel, L. E., & Wendel, J. F. (2010). Evolutionary rate variation, genomic dominance and duplicate gene expression evolution during allotetraploid cotton speciation [research-article]. *The New Phytologist, 186*(1), 184-193.

Fletcher, R. S., Mullen, J. L., Yoder, S., Bauerle, W. L., Reuning, G., Sen, S., Meyer, E., Juenger, T. E., & McKay, J. K. (2013). Development of a next-generation NIL library in Arabidopsis thaliana for dissecting complex traits. *BMC Genomics, 14*(1), 655.

Fryxell, P. A., & Craven, L. A. (1992). A Revision of Gossypium Sect. Grandicalyx (Malvaceae), Including the Description of Six New Species [research-article]. *Systematic Botany*(1), 91.

Fu, Y., Dong, C., Wang, J., Wang, Y., & Li, C. (2019). Genome-wide association study reveals the genetic control underlying node of the first fruiting branch and its height in upland cotton (Gossypium hirsutum L.). *Euphytica, 215*(2), 35.

Ghosal, S., Zheng, B., Chapman, S. C., Potgieter, A. B., Jordan, D. R., Wang, X., Singh, A. K., Singh, A., Hirafuji, M., Ninomiya, S., Ganapathysubramanian, B., Sarkar, S., & Guo, W. (2019). A Weakly Supervised Deep Learning Framework for Sorghum Head Detection and Counting. *Plant Phenomics, 2019*, 1525874.

Glaubitz, J. C., Casstevens, T. M., Lu, F., Harriman, J., Elshire, R. J., Sun, Q., & Buckler, E. S. (2014). TASSEL-GBS: A High Capacity Genotyping by Sequencing Analysis Pipeline [Article]. *PLoS ONE*, *9*(2), 1-11.

Grant, V. (1981). *Plant speciation* (2nd ed.) [Bibliographies, Non-fiction]. Columbia University Press.

Grover, C. E., Gallagher, J. P., & Wendel, J. F. (2015). Candidate Gene Identification of Flowering Time Genes in Cotton. *The Plant Genome*, *8*(2), plantgenome2014.2012.0098.

Grover, C. E., Yoo, M.-J., Lin, M., Murphy, M. D., Harker, D. B., Byers, R. L., Lipka, A. E., Hu, G., Yuan, D., Conover, J. L., Udall, J. A., Paterson, A. H., Gore, M. A., & Wendel, J. F. (2020). Genetic Analysis of the Transition from Wild to Domesticated Cotton (Gossypium hirsutum L.). *G3 Genes|Genomes|Genetics*, *10*(2), 731-754.

Grover, E. B., & Hamby, D. S. (1960). *Handbook of textile testing and quality control* [Bibliographies Non-fiction]. Textile Book Publishers.

Guan, X., Song, Q., & Chen, Z. J. (2014). Polyploidy and small RNA regulation of cotton fiber development. *Trends in Plant Science*, *19*(8), 516-528.

Guo, Y., McCarty, J. C., Jenkins, J. N., An, C., & Saha, S. (2009). Genetic detection of node of first fruiting branch in crosses of a cultivar with two exotic accessions of upland cotton. *Euphytica*, *166*(3), 317-329.

Guo, Y. F., McCarty, J. C., Jenkins, J. N., & Saha, S. (2008). QTLs for node of first fruiting branch in a cross of an upland cotton, Gossypium hirsutum L., cultivar with primitive accession Texas 701 [Journal Article]. *Euphytica*, *163*(1), 113-122.

He, X., Wang, T., Xu, Z., Liu, N., Wang, L., Hu, Q., Luo, X., Zhang, X., & Zhu, L. (2018). The cotton HD-Zip transcription factor GhHB12 regulates flowering time and plant architecture via the GhmiR157-GhSPL pathway. *Communications Biology*, *1*(1), 229.

Heiser, C. B. (1979). Hybrid Populations of Helianthus divaricatus and H. microcephalus after 22 Years [research-article]. *Taxon*, *28*(1/3), 71-75.

Hinchliffe, D. J., Meredith, W. R., Yeater, K. M., Kim, H. J., Woodward, A. W., Chen, Z. J., & Triplett, B. A. (2010). Near-isogenic cotton germplasm lines that differ in fiber-bundle strength have temporal differences in fiber gene expression patterns as revealed by comparative high-throughput profiling. *TAG. Theoretical And Applied Genetics. Theoretische Und Angewandte Genetik*, *120*(7), 1347-1366.

Hodgkin, T., & Hajjar, R. (2007). Using crop wild relatives for crop improvement: trends and perspectives. Wallingford, UK.

Holland, J. B. (2004, September 26 - October 1, 2004). Implementation of molecular markers for quanititative traits in a breeding program - challenges and opportunities. New directions for a diverse plannet: Proceedings for the 4th International Crop Scinece Congress, Brisbane, Australia.

Hu, Y., Chen, J., Fang, L., Zhang, Z., Ma, W., Niu, Y., Ju, L., Deng, J., Zhao, T., Lian, J., Baruch, K., Fang, D., Liu, X., Ruan, Y.-l., Rahman, M.-u., Han, J., Wang, K., Wang, Q., Wu, H., . . . Zhang, T. (2019). Gossypium barbadense and Gossypium hirsutum genomes provide insights into the origin and evolution of allotetraploid cotton. *Nature Genetics*, *51*(4), 739-748.

Huijser, P., & Schmid, M. (2011). The control of developmental phase transitions in plants. *138*(19), 4117-4130.

Hutchinson, J. B. (1951). Intra-specific differentiation in Gossypium hirsutum. *Heredity*, *5*(2), 161-193.

Imran, M., Shakeel, A., Azhar, F. M., Farooq, J., Saleem, M. F., Saeed, A., Nazeer, W., Riaz, M., Naeem, M., & Javaid, A. (2012). Combining ability analysis for within-boll yield components in upland cotton (Gossypium hirsutum L.). *Genet Mol Res*, *11*(3), 2790-2800.

James Mc, D. S. (1975). Fiber Initiation on the Cotton Ovule (Gossypium Hirsutum). *American Journal of Botany*, *62*(7), 723-730.

Jasdanwala, R. T., Singh, Y. D., & Chinoy, J. J. (1977). Auxin Metabolism in Developing Cotton Hairs. (106), 1111.

Jia, X., Pang, C., Wei, H., Wang, H., Ma, Q., Yang, J., Cheng, S., Su, J., Fan, S., Song, M., Wusiman, N., & Yu, S. (2016). High-density linkage map construction and QTL analysis for earliness-related traits in Gossypium hirsutum L. *BMC Genomics*, *17*(1), 909.

Jiang, C., Wright, R. J., El-Zik, K. M., & Paterson, A. H. (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), 4419-4424.

Jiang, C.-x., Chee, P. W., Draye, X., Morrell, P. L., Smith, C. W., & Paterson, A. H. (2000). MULTILOCUS INTERACTIONS RESTRICT GENE INTROGRESSION IN INTERSPECIFIC POPULATIONS OF POLYPLOID GOSSYPIUM (COTTON). *Evolution*, *54*(3), 798-814.

Jiang, Y., Li, C., & Paterson, A. H. (2016). High throughput phenotyping of cotton plant height using depth images under field conditions. *Computers and Electronics in Agriculture*, *130*, 57-68.

Jiang, Y., Li, C., Sun, S., Paterson, A. H., & Xu, R. (2017, 2017/01/01/). GPHENOVISION: A GROUND MOBILE SYSTEM WITH MULTI-MODAL IMAGING FOR FIELD-BASED HIGH THROUGHPUT PHENOTYPING OF COTTON.

Jiang, Y., Snider, J. L., Li, C., Rains, G. C., & Paterson, A. H. (2020). Ground Based Hyperspectral Imaging to Characterize Canopy-Level Photosynthetic Activities [Article]. *Remote Sensing*, *12*(2), 315-315.

Kaeppler, S. M. (1997). Quantitative trait locus mapping using sets of near-isogenic lines: relative power comparisons and technical considerations. *Theoretical and Applied Genetics*, *95*(3), 384-392.

Kearsey, M. J., & Luo, Z. W. (2003). Mapping, characterization and deployment of quantitative trait loci. In H. J. Newbury (Ed.), *Plant Molecular Breeding* (pp. 1-29). Blackwell Publishing Ltd.

Kelly, S., Huihui, L., Navarro, J. A. R., Dong, A., Maria Cinta, R., Sarah, H., Charlotte, A., Jeffrey, C. G., Sharon, M., Robert, J. E., Edward, S. B., & Peter, J. B. (2014). Novel Methods to Optimize Genotypic Imputation for Low-Coverage, Next-Generation Sequence Data in Crop Plants [article]. *The Plant Genome*, *7*(3).

Keurentjes, J. J. B., Bentsink, L., Alonso-Blanco, C., Hanhart, C. J., Blankestijn-De Vries, H., Effgen, S., Vreugdenhil, D., & Koornneef, M. (2007). Development of a Near-Isogenic Line Population of Arabidopsis thaliana and Comparison of Mapping Power With a Recombinant Inbred Line Population [10.1534/genetics.106.066423]. *Genetics*, *175*(2), 891.

Khanal, S. (2018). *Dissection of quantitative variation introgressed into bermudagrass and upland cotton* University of Georgia].

Kim, C., Guo, H., Kong, W., Chandnani, R., Shuang, L.-S., & Paterson, A. H. (2016, 2016/01/01/). Application of genotyping by sequencing technology to a variety of crop breeding programs. Netherlands.

Kim, H. J., & Triplett, B. A. (2001). Cotton Fiber Growth in Planta and in Vitro. Models for Plant Cell Elongation and Cell Wall Biogenesis. *127*(4), 1361-1366.

Kim, H. J., & Triplett, B. A. (2001). Cotton Fiber Growth in Planta and in Vitro. Models for Plant Cell Elongation and Cell Wall Biogenesis. *Plant Physiology*, *127*(4), 1361-1366.

Kohel, R. J., Richmond, T. R., & Lewis, C. F. (1974). Genetics of flowering response in cotton. VI. Flowering behavior of Gossypium hirsutum L. and G. barbadense L. hybrids [Journal Article]. *Crop Science*, *14*(5), 696-699.

Komeda, Y. (2004). GENETIC REGULATION OF TIME TO FLOWER IN ARABIDOPSIS THALIANA [Article]. *Annual Review of Plant Biology*, *55*(1), 521-535.

Kooke, R., Wijnker, E., & Keurentjes, J. J. (2012). Backcross populations and near isogenic lines. *Methods Mol Biol*, *871*, 3-16.

Kooke, R., Wijnker, E., & Keurentjes, J. J. B. (2012). Backcross Populations and Near Isogenic Lines. In S. A. Rifkin (Ed.), *Quantitative Trait Loci (QTL): Methods and Protocols* (pp. 3-16). Humana Press.

Kumar, P., Singh, R., Lubbers, E. L., Shen, X., Paterson, A. H., Campbell, B. T., Jones, D. C., & Chee, P. W. (2019). Genetic Evaluation of Exotic Chromatins from Two Obsolete Interspecific Introgression Lines of Upland Cotton for Fiber Quality Improvement. *Crop Science*, *59*(3), 1073-1084.

Kushanov, F. N., Buriev, Z. T., Shermatov, S. E., Turaev, O. S., Norov, T. M., Pepper, A. E., Saha, S., Ulloa, M., Yu, J. Z., Jenkins, J. N., Abdukarimov, A., & Abdurakhmonov, I. Y. (2017).

QTL mapping for flowering-time and photoperiod insensitivity of cotton Gossypium darwinii Watt [Journal Article]. *PLoS ONE*, *12*(10), e0186240.

Kushanov, F. N., Turaev, O. S., Ernazarova, D. K., Gapparov, B. M., Oripova, B. B., Kudratova, M. K., Rafieva, F. U., Khalikov, K. K., Erjigitov, D. S., Khidirov, M. T., Kholova, M. D., Khusenov, N. N., Amanboyeva, R. S., Saha, S., Yu, J. Z., & Abdurakhmonov, I. Y. (2021). Genetic Diversity, QTL Mapping, and Marker-Assisted Selection Technology in Cotton (Gossypium spp.) [Review]. *Frontiers in Plant Science*, *12*.

Lacape, J. M., Trung Bieu, N., Courtois, B., Belot, J. L., Giband, M., Gourlot, J. P., Gawryziak, G., Roques, S., & Hau, B. (2005). QTL analysis of cotton fiber quality using multiple Gossypium hirsutum×Gossypium barbadense backcross generations. *Crop Science*, *45*(1), 123-140.

Lai, D., Li, H., Fan, S., Song, M., Pang, C., Wei, H., Liu, J., Wu, D., Gong, W., & Yu, S. (2011). Generation of ESTs for Flowering Gene Discovery and SSR Marker Development in Upland Cotton. *PLoS ONE*, *6*(12), e28676.

Landi, P., Sanguineti, M. C., Salvi, S., Giuliani, S., Bellotti, M., Maccaferri, M., Conti, S., & Tuberosa, R. (2005). Validation and characterization of a major QTL affecting leaf ABA concentration in maize. *Molecular Breeding*, *15*(3), 291-303.

Leal Valentim, F., Mourik, S. v., Posé, D., Kim, M. C., Schmid, M., van Ham, R. C. H. J., Busscher, M., Sanchez-Perez, G. F., Molenaar, J., Angenent, G. C., Immink, R. G. H., & van Dijk, A. D. J. (2015). A Quantitative and Dynamic Model of the Arabidopsis Flowering Time Gene Regulatory Network [Article]. *PLoS ONE*, *10*(2), 1-18.

Levi, A., Ovnat, L., Paterson, A. H., & Saranga, Y. (2009). Photosynthesis of cotton near-isogenic lines introgressed with QTLs for productivity and drought related traits [Article]. *Plant Science, 177*(2), 88-96.

Lewis, C. F., & Richmond, T. R. (1960). The genetics of flowering response in cotton. II. Inheritance of flowering response in a Gossypium barbadense cross [Journal Article]. *Genetics, 45*, 79-85.

Li, C., Wang, C., Dong, N., Wang, X., Zhao, H., Converse, R., Xia, Z., Wang, R., & Wang, Q. (2012). QTL detection for node of first fruiting branch and its height in upland cotton (Gossypium hirsutum L.). *Euphytica*, *188*(3), 441-451.

Li, C., Wang, X., Dong, N., Zhao, H., Xia, Z., Wang, R., Converse, R. L., & Wang, Q. (2013). QTL analysis for early-maturing traits in cotton using two upland cotton (Gossypium hirsutum L.) crosses. *Breeding Science*, *63*(2), 154-163.

Li, C., Wang, X., Dong, N., Zhao, H., Xia, Z., Wang, R., Converse, R. L., & Wang, Q. (2013). QTL analysis for early-maturing traits in cotton using two upland cotton (Gossypium hirsutum L.) crosses. *63*(2), 154-163.

Li, C., Zhang, J., Hu, G., Fu, Y., & Wang, Q. (2016). Association mapping and favorable allele mining for node of first fruiting/sympodial branch and its height in Upland cotton (Gossypium hirsutum L.). *Euphytica*, *210*(1), 57-68.

Li, F., Fan, G., Lu, C., Xiao, G., Zou, C., Kohel, R. J., Ma, Z., Shang, H., Ma, X., Wu, J., Liang, X., Huang, G., Percy, R. G., Liu, K., Yang, W., Chen, W., Du, X., Shi, C., Yuan, Y., . . . Yu, S. (2015). Genome sequence of cultivated Upland cotton (Gossypium hirsutum TM-1) provides insights into genome evolution. *Nature biotechnology*, *33*(5), 524-530.

Li, L., Zhao, S., Su, J., Fan, S., Pang, C., Wei, H., Wang, H., Gu, L., Zhang, C., Liu, G., Yu, D., Liu, Q., Zhang, X., & Yu, S. (2017). High-density genetic linkage map construction by F2 populations and QTL analysis of early-maturity traits in upland cotton (Gossypium hirsutum L.). *PLoS ONE*, *12*(8), e0182918.

Li, Y., Jiang, J., Du, M. L., Li, L., Wang, X. L., & Li, X. B. (2013). A cotton gene encoding MYB-like transcription factor is specifically expressed in pollen and is involved in regulation of late anther/pollen development. *Plant Cell Physiol*, *54*(6), 893-906.

Li, Y. L., Sun, J., & Xia, G. X. (2005). Cloning and characterization of a gene for an LRR receptor-like protein kinase associated with cotton fiber development. *Mol Genet Genomics*, *273*(3), 217-224.

Li, Z., Wang, P., You, C., Yu, J., Zhang, X., Yan, F., Ye, Z., Shen, C., Li, B., Guo, K., Liu, N., Thyssen, G. N., Fang, D. D., Lindsey, K., Zhang, X., Wang, M., & Tu, L. (2020). Combined GWAS and eQTL analysis uncovers a genetic regulatory network orchestrating the initiation of secondary cell wall development in cotton [Report]. *New Phytologist*, *226*(6), 1738.

Lin, Y., Schertz, K. F., & Paterson, A. H. (1995). Comparative analysis of QTLs affecting plant height and maturity across the Poaceae, in reference to an interspecific sorghum population [Journal Article]. *Genetics*, *141*(1), 391-411.

Lin, Z., He, D., Zhang, X., Nie, Y., Guo, X., Feng, C., & Stewart, J. M. (2005). Linkage map construction and mapping QTL for cotton fibre quality using SRAP, SSR and RAPD [Article]. *Plant Breeding*, *124*(2), 180-187.

Liu, D., Liu, F., Shan, X., Zhang, J., Tang, S., Fang, X., Liu, X., Wang, W., Tan, Z., Teng, Z., Zhang, Z., & Liu, D. (2015). Construction of a high-density genetic map and lint

percentage and cottonseed nutrient trait QTL identification in upland cotton (Gossypium hirsutum L.) [Article]. *Molecular Genetics & Genomics*, *290*(5), 1683-1700.

Lubbers, E. L., & Chee, P. W. (2009). The Worldwide Gene Pool of G. hirsutum and its Improvement. *3*, 23-52.

Ma, Q., Wu, M., Pei, W., Li, H., Li, X., Zhang, J., Yu, J., & Yu, S. (2014). Quantitative phosphoproteomic profiling of fiber differentiation and initiation in a fiberless mutant of cotton. *BMC Genomics*, *15*(1), 466.

May, O. L. (1999). Producing quality cotton by conventional breeding, marker assisted selection, and transgenic methods [Journal Article]. *Journal of New Seeds*, *1*(2), 65-82.

May, O. L., & Green, C. C. (1994). Genetic variation for fiber properties in elite Pee Dee cotton populations [Journal Article]. *Crop Science*, *34*(3), 684-690.

Mei, M., Syed, N. H., Gao, W., Thaxton, P. M., Smith, C. W., Stelly, D. M., & Chen, Z. J. (2004). Genetic mapping and QTL analysis of fiber-related traits in cotton (Gossypium). *TAG. Theoretical And Applied Genetics. Theoretische Und Angewandte Genetik*, *108*(2), 280-291.

Meinert, M. C., & Delmer, D. P. (1977). Changes in biochemical composition of the cell wall of the cotton fiber during development. *Plant Physiology*, *59*(6), 1088-1097.

Melchinger, A. E., Utz, H. F., & Schon, C. C. (2000). From Mendel to Fisher: The power and limits of QTL mapping for quantitative traits. . *Vortr. Pflanzenzuchtg.*, *48*, 132-142.

Melchinger, A. E., Utz, H. F., & Schön, C. C. (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*(1), 383.

Meyn, O., & Emboden, W. A. (1987). Parameters and Consequences of Introgression in Salvia apiana x S. mellifera (Lamiaceae) [research-article]. *Systematic Botany*, *12*(3), 390-399.

Monforte, A. J., & Tanksley, S. D. (2000). Development of a set of near isogenic and backcross recombinant inbred lines containing most of the Lycopersicon hirsutum genome in a L. esculentum genetic background: A tool for gene mapping and gene discovery. *Genome*, *43*(5), 803-813.

Newbury, H. J. (2003). *Plant molecular breeding*. Oxford.

Ninh, T. T., Gao, W., Trusov, Y., Zhao, J.-R., Long, L., Song, C.-P., & Botella, J. R. (2021). Tomato and cotton G protein beta subunit mutants display constitutive autoimmune responses. *Plant Direct*, *5*(11), e359.

Niu, E., Shang, X., Cheng, C., Bao, J., Zeng, Y., Cai, C., Du, X., & Guo, W. (2015). Comprehensive Analysis of the COBRA-Like (COBL) Gene Family in Gossypium Identifies Two COBLs Potentially Associated with Fiber Quality. *PLoS ONE*, *10*(12), e0145725.

Niu, H., Ge, Q., Shang, H., & Yuan, Y. (2022). Inheritance, QTLs, and Candidate Genes of Lint Percentage in Upland Cotton. *Front Genet*, *13*, 855574.

Paran, I., & Zamir, D. (2003). Quantitative traits in plants: beyond the QTL. *Trends in Genetics*, *19*(6), 303-306.

Paterson, A. H., Boman, R. K., Brown, S. M., Chee, P. W., Gannaway, J. R., Gingle, A. R., May, O. L., & Smith, C. W. (2004). Reducing the Genetic Vulnerability of Cotton. *Crop Science*, *44*(6), 1900-1901.

Paterson, A. H., Boman, R. K., Brown, S. M., Chee, P. W., Gannaway, J. R., Gingle, A. R., May, O. L., & Smith, C. W. (2004). Reducing the genetic vulnerability of cotton. Madison, USA.

Paterson, A. H., Brubaker, C. L., & Wendel, J. F. (1993). A rapid method for extraction of cotton (Gossypium spp.) genomic DNA suitable for RFLP or PCR analysis. *Plant Molecular Biology Reporter, 11*(2), 122-127.

Paterson, A. H., Lin, Y. R., Li, Z., Schertz, K. F., Doebley, J. F., Pinson, S. R., Liu, S. C., Stansel, J. W., & Irvine, J. E. (1995). Convergent domestication of cereal crops by independent mutations at corresponding genetic Loci. *Science, 269*(5231), 1714-1718.

Paterson, A. H., Saranga, Y., Menz, M., Jiang, C. X., & Wright, R. J. (2003). QTL analysis of genotype × environment interactions affecting cotton fiber quality [Article]. *Theoretical & Applied Genetics, 106*(3), 384-396.

Paterson, A. H., Wendel, J. F., Gundlach, H., Guo, H., Jenkins, J., Jin, D., Llewellyn, D., Showmaker, K. C., Shu, S., Udall, J., Yoo, M.-j., Byers, R., Chen, W., Doron-Faigenboim, A., Duke, M. V., Gong, L., Grimwood, J., Grover, C., Grupp, K., . . . Schmutz, J. (2012). Repeated polyploidization of Gossypium genomes and the evolution of spinnable cotton fibres. *Nature, 492*(7429), 423-427.

Perkins Jr., H. H., Ethridge, D. E., & Bragg, C. K. (1984). Fiber. In *Cotton* (pp. 437-509).

Poland, J. A., & Rifeb, T. W. (2012). Genotyping-by-sequencing for plant breeding and genetics. *Plant Genome, 5*(3), 92-102.

Qin, H., Guo, W., Zhang, Y. M., & Zhang, T. (2008). QTL mapping of yield and fiber traits based on a four-way cross population in Gossypium hirsutum L. *Theor Appl Genet, 117*(6), 883-894.

Ray, L. L., & Richmond, T. R. (1966). Morphological measures of earliness of crop maturity in cotton [Journal Article]. *Crop Science, 6*, 527-531.

Reinisch, A. J., Dong, J. M., Brubaker, C. L., Stelly, D. M., Wendel, J. F., & Paterson, A. H. (1994a). A detailed RFLP map of cotton, Gossypium hirsutum x Gossypium barbadense: chromosome organization and evolution in a disomic polyploid genome. *Genetics*, *138*(3), 829-847.

Reinisch, A. J., Dong, J. M., Brubaker, C. L., Stelly, D. M., Wendel, J. F., & Paterson, A. H. (1994b). A detailed RFLP map of cotton, Gossypium hirsutum×Gossypium barbadense: chromosome organization and evolution in a disomic polyploid genome. *Genetics*, *138*(3), 829-847.

Rieseberg, L. H., Kim, M. J., & Seiler, G. J. (1999). Introgression between the Cultivated Sunflower and a Sympatric Wild Relative, Helianthus petiolaris (Asteraceae). *International Journal of Plant Sciences*, *160*(1), 102-108.

Rong, J., Abbey, C., Bowers, J. E., Brubaker, C. L., Chang, C., Chee, P. W., Delmonte, T. A., Ding, X., Garza, J. J., & Marler, B. S. (2004). A 3347-locus genetic recombination map of sequence-tagged sites reveals features of genome organization, transmission and evolution of cotton (Gossypium). *166*(1), 389-417.

Rong, J., Feltus, F. A., Waghmare, V. N., Pierce, G. J., Chee, P. W., Draye, X., Saranga, Y., Wright, R. J., Wilkins, T. A., May, O. L., Smith, C. W., Gannaway, J. R., Wendel, J. F., & Paterson, A. H. (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 [Article]. *Genetics*, *176*(4), 2577-2588.

Rong, J., Feltus, F. A., Waghmare, V. N., Pierce, G. J., Chee, P. W., Draye, X., Saranga, Y., Wright, R. J., Wilkins, T. A., May, O. L., Smith, C. W., Gannaway, J. R., Wendel, J. F., & Paterson, A. H. (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), 2577-2588.

Rong, J., Wang, X., Schulze, S. R., Compton, R. O., Williams-Coplin, T. D., Goff, V., Chee, P. W., & Paterson, A. H. (2012). Types, levels and patterns of low-copy DNA sequence divergence, and phylogenetic implications, for Gossypium genome types. *Heredity (Edinb)*, *108*(5), 500-506.

Said, J., Song, M., Wang, H., Lin, Z., Zhang, X., Fang, D., & Zhang, J. (2015). A comparative meta-analysis of QTL between intraspecific Gossypium hirsutum and interspecific G. hirsutum × G. barbadense populations [Article]. *Molecular Genetics & Genomics*, *290*(3), 1003-1025.

Said, J. I., Knapka, J. A., Song, M., & Zhang, J. (2015). Cotton QTLdb: a cotton QTL database for QTL analysis, visualization, and comparison between Gossypium hirsutum and G. hirsutum × G. barbadense populations. *Mol Genet Genomics*, *290*(4), 1615-1625.

Said, J. I., Lin, Z., Zhang, X., Song, M., & Zhang, J. (2013). A comprehensive meta QTL analysis for fiber quality, yield, yield related and morphological traits, drought tolerance, and disease resistance in tetraploid cotton. *BMC Genomics*, *14*(1), 776.

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

Said, J. I., Song, M., Wang, H., Lin, Z., Zhang, X., Fang, D. D., & Zhang, J. (2015). A comparative meta-analysis of QTL between intraspecific Gossypium hirsutum and interspecific G. hirsutum × G. barbadense populations. *Mol Genet Genomics*, *290*(3), 1003-1025.

Saranga, Y., Menz, M., Jiang, C. X., Wright, R. J., Yakir, D., & Paterson, A. H. (2001). Genomic dissection of genotype × environment interactions conferring adaptation of cotton to arid conditions. *Genome Research*, *11*(12), 1988-1995.

Seagull, R. W., Oliveri, V., Murphy, K., Binder, A., & Sushma, K. (2000). Cotton fiber growth and development 2. Changes in cell diameter and wall birefringence. *Journal of Cotton Science*, *4*(2), 97-104.

Shappley, Z. W., Jenkins, J. N., Watson, C. E., Jr., Kahler, A. L., & Meredith, W. R. (1996). Establishment of molecular markers and linkage groups in two F2 populations of Upland cotton. *Theoretical and Applied Genetics*, *92*(8), 915-919.

Shen, X., Cao, Z., Singh, R., Lubbers, E. L., Xu, P., Smith, C. W., Paterson, A. H., & Chee, P. W. (2011). Efficacy of qFL-chr1, a Quantitative Trait Locus for Fiber Length in Cotton (Gossypium spp.). *Crop Science*, *51*(5), 2005-2010.

Shen, X., Guo, W., Zhu, X., Yuan, Y., Yu, J. Z., Kohel, R. J., & Zhang, T. (2005). Molecular mapping of QTLs for fiber qualities in three diverse lines in Upland cotton using SSR markers [Article]. *Molecular Breeding*, *15*(2), 169-181.

Shi, Y., Li, W., Li, A., Ge, R., Zhang, B., Li, J., Liu, G., Li, J., Liu, A., Shang, H., Gong, J., Gong, W., Yang, Z., Tang, F., Liu, Z., Zhu, W., Jiang, J., Yu, X., Wang, T., & Wang, W. (2015). Constructing a high-density linkage map for Gossypium hirsutum × Gossypium barbadense and identifying QTLs for lint percentage [Article]. *Journal of Integrative Plant Biology*, *57*(5), 450-467.

Stephens, S. G. (1949). The cytogenetics of speciation in Gossypium I. Selective elimination of the donor parent genotype in interspecific backcrosses [Journal Article]. *Genetics 1949.*, *34*, 627-637.

Stuber, C. W., Polacco, M., & Lynn, M. (1999). Synergy of empirical breeding, marker-assisted selection, and genomics to increase crop yield potential. *Crop Science*, *39*(6), 1571.

Su, J., Pang, C., Wei, H., Li, L., Liang, B., Wang, C., Song, M., Wang, H., Zhao, S., Jia, X., Mao, G., Huang, L., Geng, D., Wang, C., Fan, S., & Yu, S. (2016). Identification of favorable SNP alleles and candidate genes for traits related to early maturity via GWAS in upland cotton. *BMC Genomics*, *17*(1), 687.

Sun, S., Li, C., Chee, P. W., Paterson, A. H., Jiang, Y., Xu, R., Robertson, J. S., Adhikari, J., & Shehzad, T. (2020). Three-dimensional photogrammetric mapping of cotton bolls in situ based on point cloud segmentation and clustering. *ISPRS Journal of Photogrammetry and Remote Sensing*, *160*, 195-207.

Szalma, S. J., Hostert, B. M., LeDeaux, J. R., Stuber, C. W., & Holland, J. B. (2007a). QTL mapping with near-isogenic lines in maize. *Theoretical and Applied Genetics*, *114*(7), 1211-1228.

Szalma, S. J., Hostert, B. M., LeDeaux, J. R., Stuber, C. W., & Holland, J. B. (2007b). QTL mapping with near-isogenic lines in maize [Article]. *Theoretical & Applied Genetics*, *114*(7), 1211-1228.

Taliercio, E., & Kloth, R. (2004). Expression and Characterization of a UDP-Glucose Pyrophosphorylase Gene in Cotton. *Journal of Cotton Science*, *8*(2), 91-98.

Tan, C., Li, C., He, D., & Song, H. (2021). *Anchor-free deep convolutional neural network for plant and plant organ detection and counting* 2021 ASABE Annual International Virtual Meeting, St. Joseph, MI.

Tang, S., Teng, Z., Zhai, T., Fang, X., Liu, F., Liu, D., Zhang, J., Liu, D., Wang, S., Zhang, K., Shao, Q., Tan, Z., Paterson, A., & Zhang, Z. (2015). Construction of genetic map and QTL

analysis of fiber quality traits for Upland cotton (Gossypium hirsutum L.) [Article]. *Euphytica*, *201*(2), 195-213.

Tanksley, S. D., & Nelson, J. C. (1996a). Advanced backcross QTL analysis: a method for the simultaneous discovery and transfer of valuable QTLs from unadapted germplasm into elite breeding lines. *92*(2), 191-203.

Tanksley, S. D., & Nelson, J. C. (1996b). Advanced backcross QTL analysis: a method for the simultaneous discovery and transfer of valuable QTLs from unadapted germplasm into elite breeding lines. *Theoretical and Applied Genetics*, *92*(2), 191-203.

Tuinstra, M. R., Ejeta, G., & Goldsbrough, P. B. (1997). Heterogeneous inbred family (HIF) analysis: a method for developing near-isogenic lines that differ at quantitative trait loci. *Theoretical and Applied Genetics*, *95*(5), 1005-1011.

Ulloa, M., Meredith Jr, W. R., Shappley, Z. W., & Kahler, A. L. (2002). RFLP genetic linkage maps from four F2.3 populations and a joinmap of Gossypium hirsutum L [Article]. *Theoretical & Applied Genetics*, *104*(2/3), 200-208.

Ulloa, M., & Meredith, W. R., Jr. (2000). Genetic linkage map and QTL analysis of agronomic and fiber quality traits in an intraspecific population. *Journal of Cotton Science*, *4*(3), 161-170.

USDA. (2016, May 26, 2016). *Cotton and Wool*. USDA. Retrieved September 12 from

Utz, H. F., Melchinger, A. E., & Schön, C. C. (2000). Bias and Sampling Error of the Estimated Proportion of Genotypic Variance Explained by Quantitative Trait Loci Determined From Experimental Data in Maize Using Cross Validation and Validation With Independent Samples. *Genetics*, *154*(4), 1839-1849.

Vollesen, K. (1987). The Native Species of Gossypium (Malvaceae) in Africa, Arabia and Pakistan [research-article]. *Kew Bulletin*, *42*(2), 337-349.

Waddle, B. M., Lewis, C. F., & Richmond, T. R. (1961). The genetics of flowering response in cotton. III. Fruiting behaviour of Gossypium hirsutum race lati-folium in a cross with a variety of cultivated American upland cotton [Journal Article]. *Genetics*, *46*, 427-437.

Waghmare, V. N., Rong, J., Rogers, C. J., Bowers, J. E., Chee, P. W., Gannaway, J. R., Katageri, I., & Paterson, A. H. (2016). Comparative transmission genetics of introgressed chromatin in Gossypium (cotton) polyploids [research-article]. *American Journal of Botany*, *103*(4), 719-729.

Waghmare, V. N., Rong, J., Rogers, C. J., Pierce, G. J., Wendel, J. F., & Paterson, A. H. (2005). Genetic mapping of a cross between Gossypium hirsutum (cotton) and the Hawaiian endemic, Gossypium tomentosum [Article]. *Theoretical & Applied Genetics*, *111*(4), 665-676.

Wang, B., Guo, W., Zhu, X., Wu, Y., Huang, N., & Zhang, T. (2007). QTL mapping of yield and yield components for elite hybrid derived-RILs in upland cotton. *J Genet Genomics*, *34*(1), 35-45.

Wang, C. C. N., Pei-Chun, C., Ka-Lok, N., Chun-Ming, C., Sheu, P. C. Y., & Tsai, J. J. P. (2014). A model comparison study of the flowering time regulatory network in Arabidopsis [Article]. *BMC Systems Biology*, *8*(1), 1-25.

Wang, F., Gong, Y., Zhang, C., Liu, G., Wang, L., Xu, Z., & Zhang, J. (2011). Genetic effects of introgression genomic components from Sea Island cotton (Gossypium barbadense L.) on fiber related traits in upland cotton (G. hirsutum L.). *Euphytica*, *181*, 41-53.

Wang, G. L., Dong, J. M., & Paterson, A. H. (1995). The distribution of Gossypium hirsutum chromatin in G. barbadense germ plasm: molecular analysis of introgressive plant breeding. *Theoretical and Applied Genetics*, *91*(6), 1153-1161.

Wang, M., Tu, L., Lin, M., Lin, Z., Wang, P., Yang, Q., Ye, Z., Shen, C., Li, J., Zhang, L., Zhou, X., Nie, X., Li, Z., Guo, K., Ma, Y., Huang, C., Jin, S., Zhu, L., Yang, X., . . . Zhang, X. (2017). Asymmetric subgenome selection and cis-regulatory divergence during cotton domestication. *Nature Genetics*, *49*(4), 579-587.

Wang, Y., Secco, D., & Poirier, Y. (2008). Characterization of the PHO1 gene family and the responses to phosphate deficiency of Physcomitrella patens. *Plant Physiol*, *146*(2), 646-656.

Wang, Z., Zhang, D., Wang, X., Tan, X., Guo, H., & Paterson, A. H. (2013). A Whole-Genome DNA Marker Map for Cotton Based on the D-Genome Sequence of Gossypium raimondii L. *G3 Genes|Genomes|Genetics*, *3*(10), 1759-1767.

Wendel, J. F. (1989). New World Tetraploid Cottons Contain Old World Cytoplasm [research-article]. *Proceedings of the National Academy of Sciences of the United States of America*, *86*(11), 4132-4136.

Wendel, J. F., & Albert, V. A. (1992). Phylogenetics of the Cotton Genus (Gossypium): Character-State Weighted Parsimony Analysis of Chloroplast-DNA Restriction Site Data and Its Systematic and Biogeographic Implications [research-article]. *Systematic Botany*(1), 115.

Wendel, J. F., Schnabel, A., & Seelanan, T. (1995). Bidirectional Interlocus Concerted Evolution Following Allopolyploid Speciation in Cotton (Gossypium) [research-article]. *Proceedings of the National Academy of Sciences of the United States of America*, *92*(1), 280-284.

Wendel, J. F., Small, R. L., Cronn, R. C., & Brubaker, C. L. Genes, jeans, and genomes: reconstructing the history of cotton.

Xu, R., & Li, C. (2022). A modular agricultural robotic system (MARS) for precision farming: Concept and implementation [Report]. *Journal of Field Robotics*, *39*(4), 387.

Xu, Z., He, J., Tehseen Azhar, M., Zhang, Z., Fan, S., Jiang, X., Jia, T., Shang, H., & Yuan, Y. (2022). UDP-glucose pyrophosphorylase: genome-wide identification, expression and functional analyses in Gossypium hirsutum. *PeerJ*, *10*, e13460.

Yano, M., Harushima, Y., Nagamura, Y., Kurata, N., Minobe, Y., & Sasaki, T. (1997). Identification of quantitative trait loci controlling heading date in rice using a high-density linkage map. *95*(7), 1025-1032.

Yu, J., Yu, S., Lu, C., Wang, W., Fan, S., Song, M., Lin, Z., Zhang, X., & Zhang, J. (2007). High-density Linkage Map of Cultivated Allotetraploid Cotton Based on SSR, TRAP, SRAP and AFLP Markers. *Journal of Integrative Plant Biology*, *49*(5), 716-724.

Yu, J., Zhang, K., Li, S., Yu, S., Zhai, H., Wu, M., Li, X., Fan, S., Song, M., Yang, D., Li, Y., & Zhang, J. (2013). Mapping quantitative trait loci for lint yield and fiber quality across environments in a Gossypium hirsutum × Gossypium barbadense backcross inbred line population. *Theor Appl Genet*, *126*(1), 275-287.

Yu, Y., Yuan, D., Liang, S., Li, X., Wang, X., Lin, Z., & Zhang, X. (2011). Genome structure of cotton revealed by a genome-wide SSR genetic map constructed from a BC1 population between gossypium hirsutum and G. barbadense. *BMC Genomics*, *12*(1), 15.

Yufang, G., McCarty, J. C., Saha, S., Jenkins, J. N., & Jixiang, W. (2007/01//). QTL ANALYSES OF FLOWERING TIME IN A PRIMITIVE COTTON ACCESSION (GOSSYPIUM HIRSUTUM L.).

Zeng, L. H., Meredith, W. R., Boykin, D. L., & Taliercio, E. (2007). Evaluation of an exotic germplasm population derived from multiple crosses among Gossypium tetraploid species. *Journal of Cotton Science*, *11*(3), 118-127.

Zeng, Q.-w., Qin, S., Song, S.-q., Zhang, M., Xiao, Y.-h., Luo, M., Hou, L., & Pei, Y. (2012). Molecular Cloning and Characterization of a Cytokinin Dehydrogenase Gene from Upland Cotton (Gossypium hirsutum L.). *Plant Molecular Biology Reporter*, *30*(1), 1-9.

Zhang, J., Jia, X., Guo, X., Wei, H., Zhang, M., Wu, A., Cheng, S., Cheng, X., Yu, S., & Wang, H. (2021). QTL and candidate gene identification of the node of the first fruiting branch (NFFB) by QTL-seq in upland cotton (Gossypium hirsutum L.). *BMC Genomics*, *22*(1), 882.

Zhang, J., Zhang, P., Huo, X., Gao, Y., Chen, Y., Song, Z., Wang, F., & Zhang, J. (2021). Comparative Phenotypic and Transcriptomic Analysis Reveals Key Responses of Upland Cotton to Salinity Stress During Postgermination [Original Research]. *Frontiers in Plant Science*, *12*.

Zhang, J. F., Percy, R. G., & McCarty, J. C. (2014). Introgression genetics and breeding between Upland and Pima cotton: a review. *Euphytica*, *198*, 1-12.

Zhang, Q., Zhang, H., Sun, L., Fan, G., Ye, M., Jiang, L., Liu, X., Ma, K., Shi, C., Bao, F., Guan, R., Han, Y., Fu, Y., Pan, H., Chen, Z., Li, L., Wang, J., Lv, M., Zheng, T., . . . Cheng, T. (2018). The genetic architecture of floral traits in the woody plant Prunus mume. *Nature Communications*, *9*(1), 1702.

Zhang, T., Hu, Y., Jiang, W., Fang, L., Guan, X., Chen, J., Zhang, J., Saski, C. A., Scheffler, B. E., Stelly, D. M., Hulse-Kemp, A. M., Wan, Q., Liu, B., Liu, C., Wang, S., Pan, M., Wang, Y., Wang, D., Ye, W., . . . Chen, Z. J. (2015). Sequencing of allotetraploid cotton

(Gossypium hirsutum L. acc. TM-1) provides a resource for fiber improvement. *Nature biotechnology*, *33*(5), 531-537.

Zhang, T., Yuan, Y., Yu, J., Guo, W., & Kohel, R. J. (2003). Molecular tagging of a major QTL for fiber strength in Upland cotton and its marker-assisted selection [Article]. *Theoretical & Applied Genetics*, *106*(2), 262-268.

Zhen, Z., Haihong, S., Yuzhen, S., Long, H., Junwen, L., Qun, G., Juwu, G., Aiying, L., Tingting, C., Dan, W., Yanling, W., Palanga, K. K., Muhammad, J., Weijie, L., Quanwei, L., Xiaoying, D., Yunna, T., Weiwu, S., Juan, C., & Pengtao, L. (2016). Construction of a high-density genetic map by specific locus amplified fragment sequencing (SLAF-seq) and its application to Quantitative Trait Loci (QTL) analysis for boll weight in upland cotton (Gossypium hirsutum.) [Article]. *BMC Plant Biology*, *16*, 1-18.

Zheng, L., Wu, H., Qanmber, G., Ali, F., Wang, L., Liu, Z., Yu, D., Wang, Q., Xu, A., & Yang, Z. (2020). Genome-Wide Study of the GATL Gene Family in Gossypium hirsutum L. Reveals that GhGATL Genes Act on Pectin Synthesis to Regulate Plant Growth and Fiber Elongation. *Genes (Basel)*, *11*(1).

Zhu, G., Hou, S., Song, X., Wang, X., Wang, W., Chen, Q., & Guo, W. (2021). Genome-wide association analysis reveals quantitative trait loci and candidate genes involved in yield components under multiple field environments in cotton (Gossypium hirsutum). *BMC Plant Biology*, *21*(1), 250.

Zhu, L. L., & Kuraparthy, V. (2014). Molecular genetic mapping of the major effect photoperiod response locus in Pima cotton (Gossypium barbadense L.) [Journal Article]. *Crop Science*, *54*(6), 2492-2498.

CHAPTER 2

COMPARATIVE TRANSMISSION GENETICS OF INTROGRESSED CHROMATIN IN RECIPROCAL ADVANCED BACKCROSS POPULATIONS IN *GOSSYPIUM* (COTTON) POLYPLOIDS

Jeevan Adhikari, Rahul Chandnani, Deepak Vitrakoti, Sameer Khanal, Wiriyanat Ployaram and Andrew H. Paterson. Submitted to *Heredity*.

Abstract

Introgression is a potential source of valuable genetic variation and interspecific introgression lines are important resources for plant breeders to access novel alleles. Experimental advanced-generation backcross populations contain individuals with genomic compositions similar to those resulting from natural interspecific hybridization and provide opportunities to study the nature and transmission pattern of donor chromatin in recipient genomes. Here, we analyze transmission of donor chromatin in reciprocal backcrosses between *G. hirsutum* and *G. barbadense*. Across the genome, recurrent backcrossing in both backgrounds yielded donor chromatin at slightly higher frequencies than the Mendelian expectation. The average retention of donor alleles across BC_5F_1 plants was higher than expected while the average frequency of donor alleles in BC_5F_2 segregating families was less than expected. Although the recipient genome tolerated donor chromatin in general, 21 regions recalcitrant to donor alleles were identified. Only limited correspondence is observed between the recalcitrant regions in the two backgrounds, suggesting the effect of species background on introgression of donor segments. Skewed chromatin transmission in the reciprocal crosses further suggests that genetic background may profoundly affect introgression of donor chromosomal regions. In the two subgenomes of polyploid cotton, the rate of donor chromatin introgression was similar. Investigation into reciprocal transmission suggested that certain genomic regions favored heterozygote advantage while some regions offered selective advantage to one of the two parents.

Introduction

Interspecific gene transfer is a potential source of valuable genetic variation, and interspecific hybridization has been an attractive natural means for introducing novel and selectable variation for important traits into crop improvement (Anderson, 1949; Levi et al., 2009; Tanksley & Nelson, 1996a). Gene flow via interspecific hybridization can provide raw material for natural selection and evolutionary change. Introgression of chromosomal segments is one of the consequences of interspecific hybridization, which may result from backcrossing following the initial hybridization (Grant, 1981). In addition to introducing genes for adaptive traits (Heiser, 1979; Waghmare et al., 2016), introgression can reduce reproductive isolation barriers (Meyn & Emboden, 1987) and broaden the genetic base of a crop species by incorporating novel alleles / allele combinations (Adhikari et al., 2017; A. H. Paterson et al., 2004). While gene flow via hybridization and introgression can be a significant substrate for evolution (Anderson, 1949), genomic regions acting as barriers to gene flow are important for species integrity. As such, identification and investigation of such regions might shed light on factors responsible for reproductive isolation (Baack et al., 2015).

Experimental advanced-generation backcross populations contain individuals with genomic compositions resembling those resulting from natural interspecific hybridization. Individual members of such advanced-generation populations usually retain some genomic features of the donor parent while they most closely resemble their recurrent (backcross) parent. Experimental introgression populations are important genetic resources not only for crop improvement but also to study gene flow between species (Jiang et al., 2000). Although introgression has been widely acknowledged as a potential source of valuable genetic variation to enrich crop gene pools, it has had varying and often limited effect in practice (Hodgkin & Hajjar, 2007) due to limited

availability of genetic markers and genetic resources in the past. Availability of genomic resources such as reference genomes and abundant generation of genetic markers at modest cost have facilitated the study of gene flow among populations (Kim et al., 2016; Paterson et al., 2012).

Cotton belongs to the genus *Gossypium*, comprised of more than 50 species, of which about 45 are diploid (2n=26) and 7 are allotetraploid (2n=4x=52). In addition to being an important economic crop and leading textile fiber, cotton is well suited for studies of introgressive hybridization and in particular the influences of polyploidy on levels and patterns of introgression. Tetraploids contain two distinct subgenomes -- the At subgenome resembles the extant A genome of *G. herbaceum* L and the Dt subgenome resembles the D genome of *G. raimondii* Ulbrich or *G. gossypoides* Ulbrich (Wendel et al., 1995). The A- and D-genome species are estimated to have diverged from a common ancestor 6-11 million years ago (mya) and hybridized (followed by polyploidization) about 1-2 mya (Wendel, 1989). After polyploidization, several chromosomal rearrangements occurred distinguishing the tetraploid (AD) genomes from their diploid progenitors (Brubaker et al., 1999; Desai et al., 2006; Rong et al., 2004). Although normal meiotic chromosome pairing has suggested little structural rearrangement since the divergence of *G. hirsutum* and *G. barbadense* (Beasley, 1942), comprehensive linkage and genetic maps (Rong et al., 2004; Waghmare et al., 2005; Yu et al., 2007) have suggested some possible small rearrangements among the chromosomes of tetraploid Gossypium species. These comprehensive linkage maps and high-contiguity genome sequences (Paterson et al., 2012; Zhang et al., 2015) provided for detailed study of *Gossypium* transmission genetics.

Introgression and retention patterns of *G. barbadense*, *G. tomentosum* and *G. mustelinum* chromosome segments in *G. hirsutum* background have been studied previously (Jiang et al., 2000; Waghmare et al., 2016), with multilocus interactions suggested to play major roles in

determination of the genomic composition of populations. Large and widespread deficiencies of donor (*G. barbadense*) chromatin were found, including seven independent chromosomal regions showing no introgression into *G. hirsutum*. The At and Dt subgenomes of allotetraploid cotton have been suggested to play different roles in evolution and consequently differ in retention of donor chromatin.

Although *G. barbadense* introgression into *G. hirsutum* has been studied (Jiang et al., 2000), *G. hirsutum* introgression into *G. barbadense* background remains to be explored extensively. A survey of elite genotypes revealed five genomic regions of prominent historical introgression of *G. hirsutum* chromatin into *G. barbadense* (Wang et al, 1995), but provides no information about early generations or whether these introgressions were related to natural differences between the taxa or selection for a trait(s) by plant breeders. A detailed genetic recombination map of cotton provided further insights into the transmission genetics of *G. hirsutum* into *G. barbadense* (Rong et al., 2004) in addition to features of genome organization and evolution of cotton.

In this study we examine the transmission genetics of advanced-generation backcross progenies and resulting near isogenic lines developed from a cross between *Gossypium hirsutum* L. and *G. barbadense* L. In this paper we address the levels and patterns of introgression and retention of donor chromatin in the recurrent genome after several generations of backcrossing. We show that these cultivated species have differential introgression permeability and donor genome retention. Segregation patterns across genomes provide insights into reproductive barriers that affect both natural populations and crop gene pools. We also investigate segregation pattern of introgressed alleles and their deviation from expected Mendelian ratios. The segregation distorted regions (SDRs) identified based on these segregation patterns and the availability of reference genome enabled us to study gene family enrichment in genomic regions that are

significantly resistant to introgression and might be important in species isolation. This study contributes to understanding gene flow between cultivated species of cotton and provides a platform for hypotheses about possible roles of specific genomic regions or genes that influence genome composition of these species.

Materials and Methods

Plant materials and population development

Plant materials used in this study were developed from a set of reciprocal crosses between *Gossypium hirsutum* acc. Acala Maxxa and *G. barbadense* acc. Pima S6 (both inbred lines). These genotypes have been extensively used to produce molecular tools and resources including BAC libraries and Illumina genome sequences. Reciprocal advanced backcross populations were developed by first crossing the parents reciprocally (Acala Maxxa (♀) × Pima S6 (♂) – hereafter referred to as *G. hirsutum* background; and Pima S6 (♀) × Acala Maxxa (♂) – hereafter referred to as *G. barbadense* background), then independently backcrossing F_1 plants to the respective maternal parent to create 300 to 400 BC_1 progenies for each cross. The backcrossing scheme included planting only one seed from each preceding backcross to generate the next generation (Figure 2.1). After five generation of backcrossing, 179 BC_5F_1 plants from the *G. hirsutum* background and 190 BC_5F_1 plants from *G. barbadense* were self-pollinated and a total of 8364 BC_5F_2 plants (2-32 individuals in each BC_5F_2 family) were grown at Iron Horse Farm, Watkinsville, Georgia in 2019 under cultural conditions consistent with commercial irrigated cotton production.

Genotyping

The genomic composition of the BC_5F_1 plants was inferred based on genotyping by sequencing (GBS). DNA was extracted from the parents and 369 BC_5F_1 plants using a scaled-down version of a published CTAB protocol (Paterson et al., 1993). A total of four multiplexed GBS libraries were constructed according to P. Andolfatto et al. (2011) wherein the DNA were double digested with HinP1I-HaeIII enzymes. The libraries were sequenced on Illumina MiSeq (in-house) with 75 bp single end reads (SE75). The TASSEL5 GBSV2 pipeline was used for sequence data processing and genotype calling (Glaubitz et al., 2014). Reads were aligned to *G. hirsutum* acc. *TM-1* (Zhang et al., 2015) using Burrow-Wheeler Alignment (bwa) and exported to variant call format (VCF). To minimize sequencing errors, only the first 64 base pairs were used to map reads to the reference genome. Filtering of the VCF was done for bi-allelic SNPs using Fisher's exact test with a threshold *P*-value < 0.001, considering that true variants should represent biallelic homozygous state for inbred accessions. Genotypes for lines in *G. hirsutum* background were called together and those for lines in *G. barbadense* background cross were also called together. The SNPs were filtered for MAF > 0.01, missing < 30% and heterozygous < 10% at the population level. The retained SNPs were imputed using the Fast Inbred Line Library Imputation (FILLIN) pipeline available in TASSEL5 GBSv2 (Kelly et al., 2014).

The genomic composition of BC_5F_2 plants were inferred based on targeted microsatellite (SSR) genotyping of the introgressed chromosomal segments identified in their respective BC_5F_1 parents. At least two (and at most four) SSR markers were used to verify most of the introgressed regions while for small introgressions only one SSR marker was deployed. A total of 852 polymorphic SSR markers spanning the introgressed regions were derived from several published genetic maps of crosses between *G. barbadense* and *G. hirsutum* stored in the CottonGen SSR

database (https://www.cottongen.org/data/download/marker). A total of 47 candidate SSRs were monomorphic in our lines and discarded, as were 23 with ambiguous bands. Among the 8364 BC_5F_2 individuals planted in 2019, the remaining 782 SSR markers were used to genotype 5315 plants (from BC_5F_1 parents carrying 2 to 5 introgressions) for the presence (or absence) and nature (homozygous vs heterozygous) of the respective introgression/s.

Data analysis

All statistical data analysis was performed in R programming software. BC_5F_1 and BC_5F_2 families were tested for deviation from expected Mendelian ratio using chi-square tests. Allele and genotype frequencies were obtained from SNP data for the BC_5F_1 families while for the BC_5F_2 families, these frequencies were obtained from SSR genotyping of the subset of donor segments segregating in respective families. Genomic regions spanning at least 5 Mb and containing 3 or more consecutive SNPs with significant distortion ($p<0.001$) were defined as segregation distortion regions (SDRs). Genomic regions that completely lacked donor alleles for 3 or more consecutive markers spanning at least 5 Mb were defined as Introgression Devoid Regions (IDRs). These definitions have been modified from the ones described in Jiang et al. (2000) and Waghmare et al. (2016) to represent these regions in terms of physical lengths as the expected segregation ratio in the BC_5F_1 generation precluded the construction of a genetic map. The reference genome sequence, genomic sequences spanning the SDRs and IDRs and the list of cotton genes in these regions were extracted from CottonGen (https://www.cottongen.org/data/genome). Gene ontology (GO) enrichment analysis was carried out on all SDRs and IDRs by using the Enrichment Analysis feature on Cotton Functional Genomics Database (https://www.cottonfgd.org).

Results

Marker distribution and genome coverage

Raw sequence data processing, SNP filtering and post-processing was done separately for the two populations, thus resulting in the retention of different number of total SNP markers. A total of 2542 SNP markers ranging from 19 to 174 per chromosome and averaging one marker per 716 kb (Table 2.1) was used to characterize the *G. barbadense* population while a total of 3345 SNP markers ranging from 65 to 218 per chromosome and averaging one marker per 536 kb was used to characterize the *G. hirsutum* population. In total, the reported physical length of the tetraploid cotton genome is ~2.5 Gb, out of which 1.9 Gb has been anchored to the 26 chromosomes in Jbrowse CottonGen (Zhang et al., 2015). The 2542 SNPs in the *G. barbadense* populations cover 94.15% (1.82 Gb) of the anchored genome ranging from 82 to 99 % for individual chromosomes while for the *G. hirsutum* background, the 3345 SNPs cover 92.64% (1.79 Gb) of the anchored genome ranging from 68 to 99% for individual chromosomes (Table 2.1).

Genomic and sub genomic distribution of *G. hirsutum* introgression into *G. barbadense*

In all, 2471 (97.21%) of the 2542 loci showed *G. hirsutum* introgression in one or more BC_5F_1 plants. One or more introgressed loci were detected on all 26 chromosomes (Figure 2.2). For the 190 BC_5F_1 plants genotyped, there were a total of 617 introgressed chromosomal segments (averaging 3.25 segments per BC_5F_1 plant) ranging in size from 1.64 Mb to 83.55 Mb averaging 23.31 Mb (Table 2.2). A few chromosomes showed introgression over virtually their entire lengths (Figure 2.3). However, some chromosomes contained one or more regions that appeared "resistant" to introgression as shown by absence of *G. hirsutum* alleles on three or more

consecutive SNP markers spanning at least 5 Mb. At least 16 such regions localized to 12 chromosomes were devoid of *G. hirsutum* alleles (Figures 2.2, Table 2.3). These chromosomal regions lacking *G. hirsutum* alleles spanned lengths of 0.09 to 14.71 Mb with an average span of 3.11 Mb (Table 2.3).

The At subgenome retained *G. hirsutum* alleles at a significantly higher (p-value = 0.012) rate (4.92%) than the Dt subgenome (2.90%) (Figure 2.2). Among the 1832 informative At subgenome loci, 1795 (97.98%) showed introgression. Among the 710 informative Dt subgenome loci, 676 (95.21%) showed introgression. Among the 617 chromosomal segments introgressed in the 190 BC_5F_1 families, 426 (69.04%) were introgressed in the At subgenome and 191 (30.96%) were introgressed in the Dt subgenomes (Table 2.2).

Genomic and sub genomic distribution of *G. barbadense* introgression into *G. hirsutum*

In all, 3292 (98.41%) of the 3345 loci showed *G. barbadense* introgression in one or more BC_5F_1 plants. One or more introgressed loci were detected on all 26 chromosomes. For the 179 BC_5F_1 plants genotyped, there were a total of 722 introgressed chromosomal segments (averaging 4.03 segments per BC_5F_1 plant) ranging in size from 0.12 Mb to 101.05 Mb and averaging 20.48 Mb (Table 2.2). While a few chromosomes showed introgression over virtually their entire lengths (Figure 2.4), some chromosomes contained regions that resisted introgression (Figure 2.2, Table 2.3). At least 5 such regions localized to 4 chromosomes were devoid of *G. barbadense* alleles. These chromosomal regions spanned lengths of 3.92 to 56.75 Mb with an average span of 14.42 Mb (Table 2.3).

Introgression of *G. hirsutum* chromatin into *G. barbadense* occurred at similar rates (p-value = 0.91) in the At and the Dt subgenomes (Figure 2.2). Unlike the reciprocal population, the At

subgenome retained *G. hirsutum* alleles at a slightly lower rate (4.46%) than the Dt subgenome (5.52%). Among the 2182 informative At subgenome loci, 2131 (97.66%) showed introgression. Among the 1163 informative Dt subgenome loci, 1161 (99.82%) showed introgression. Among the 722 chromosomal segments introgressed in the 190 BC_5F_1 families, 355 (49.17%) were introgressed in the At subgenome and 367 (50.83%) were introgressed in the Dt subgenomes (Table 2.2).

Segregation distortion and segregation distorted regions (SDRs)

Ideally a BC_5F_1 population is expected to segregate in a 31:1 ratio. A total of 793 markers in the *G. hirsutum* background and 488 in the *G. barbadense* background significantly deviated (χ^2 test, $P < 0.01$) from the expected segregation ratio (Figure 2.5). Twelve (1.51%, so marginally above the false positive rate) of the distorted loci showed retention towards the recipient parent in the *G. hirsutum* background while 781 (98.49%) of the distorted loci retained donor alleles more than expected. In the *G. barbadense* background, all the distorted loci retained donor alleles more than expected. In the *G. hirsutum* background, 462 (58.26%) distorted markers originated from the At subgenome and 331 (41.74%) from the Dt subgenome while in the *G. barbadense* background, 242 distorted markers originated from the At subgenome and 246 originated from the Dt subgenome.

Regions with prominent introgression

Among the 34 SDRs identified in both backgrounds (at $p<0.01$), five regions in *G. barbadense* background and eight in *G. hirsutum* were significant even at a very stringent

statistical measure ($p<0.0001$). Under further scrutiny, these regions were found to harbor donor alleles in two- to five times the number of individuals than would be expected. The highly introgressed regions, referred to as "regions of prominent introgression" by Wang et al., (1995), were all in the At subgenome in *G. barbadense* background while in the *G. hirsutum* background, 6 (of 8) of these regions were found in the Dt subgenome (Table 2.4). Among the five *G. hirsutum* allele rich regions in the *G. barbadense* background, SDRGh1.1 was also identified in the study by Wang et al., (1995), where the authors identified *G. hirsutum* chromatin in a collection of 54 Sea Island, Egyptian and Pima cottons (*G. barbadense*). The sequence of the RFLP marker A1097 delineating the *G. hirsutum* allele rich region showed DNA sequence correspondence to the same cotton reference genome sequence used in this study. The location of this marker (72708857 – 72708647 bp, Chr 1, *G. hirsutum* acc TM-1 NAU-NBI genome assembly) was found within the boundaries of SDRGh1.1 (33.5Mb to 76.8 Mb, Table 2.4).

Introgression Devoid Regions (IDRs) and Gene Ontology enrichment analysis in IDRs

Deviation from expected donor allele frequencies is one of the important features studied in transmission genetics. While significant number of markers spanning several genomic regions (may) deviate from expected frequencies, some genomic regions are totally devoid of donor alleles. While occasional genotyping (sequencing) errors can account for occasional anomalous DNA marker loci, 'runs' of consecutive markers in the genome that are all devoid of introgression cannot realistically be attributed to chance. Regions where 3 or more consecutive markers lack donor alleles were defined as introgression devoid regions (IDRs). A total of 16 IDRs distributed over 12 chromosomes were identified in the *G. barbadense* background while 5 IDRs distributed over four chromosomes were identified in *G. hirsutum* (Table 2.3).

As genomic regions devoid of donor alleles might harbor genes that are biologically significant for the recipient genome, we looked for genes enriched in these IDRs. A total of 1593 genes were identified in the *G. hirsutum* background, of which 298 belonged to GO terms. Two significantly enriched GO terms were identified, one in chromosome 6 and the other in chromosome 11 (Table 2.5). Both GO terms were involved in molecular functions related to oxidoreductase and fatty-acid binding activity. In the *G. barbadense* background, a total of 3656 genes were identified, of which 721 belonged to GO terms. Ten significantly enriched GO terms were identified (Table 2.5), of which three were involved in biological functions (cellulose biosynthesis, recognition of pollen), six in molecular functions and one in cellular functions.

Segregation of Donor Chromatin in BC_5F_2 Families

A total of 190 BC_5F_2 families comprising 2973 individuals (ranging from 2 to 32 and averaging 15.64 individuals per family) in the *G. barbadense* background and 179 BC_5F_2 families comprising 2342 individuals (ranging from 2 to 32 and averaging 13.15 individuals per family) in *G. hirsutum* background were subjected to study of segregation ratios. At the subset of loci retaining the donor allele in BC_5F_1 plants, segregation ratios observed in the BC_5F_2 progeny showed bias against donor chromatin in both backgrounds. Across all *G. barbadense* chromosomal segments introgressed into *G. hirsutum* BC_5F_1 plants, the average frequency of *G. barbadense* allele retention was 35.42%, much less than the expected 50% (p-value < 0.0001). At codominant marker loci, heterozygotes occurred at an average frequency of 32.54 % (versus 50% expected), whereas *G. barbadense* homozygotes occurred at 18.58% (versus 25% expected). At dominant marker loci, 30.61% of individuals had at least one copy of the *G. barbadense* allele (versus 75% expected). In the reciprocal cross, across all *G. hirsutum* chromosomal segments introgressed into

G. barbadense BC_5F_1 plants, the average frequency of *G. hirsutum* allele retention was 25.85%, much less than the expected 50% ($\chi^2_{2\ df} = 2959.14$, p-value < 0.0001). At codominant marker loci, heterozygotes occurred at an average frequency of 28.49%, whereas *G. hirsutum* homozygotes occurred at 12.32%. At dominant marker loci, 35.02% of individuals had at least one copy of the *G. hirsutum* allele.

Significant deviation from expected genotypic and allelic frequencies was observed for several loci tested in the BC_5F_2 families. In the *G. hirsutum* background, 63.59% of DNA markers distorted significantly from the expected 1:2:1 ratio across individual families ($p<0.01$) while only 22.82% of the markers were significantly distorted in at least one family for allelic segregation (1:1). Significant distortion from expected genotypic frequency was observed in all chromosomes except 14 and 24 (p value < 0.01) while significant deviation from expected allelic frequencies were observed in all but three chromosomes (14, 18 and 24) (Table 2.6). Similarly, in the *G. barbadense* background, 76.79% of the markers distorted significantly from the expected 1:2:1 ratio across individual families ($p<0.01$) while only 39.32% of the markers were significantly distorted in at least one family for allelic segregation (1:1). Significant distortion from expected genotypic frequency was observed in all chromosomes except 17 and 24, while significant deviation from expected allelic frequencies were observed in all chromosomes except 14, 17 and 24 (Table 2.6).

Individual loci showed significant differences in segregation patterns in different BC_5F_2 families. A total of six loci segregating in three or more families in *G. hirsutum* background are shown in Table 2.7. Locus DPL0085 is exemplary, showing donor allele retention of 80.95% in family 9037 and 73.08% in family 9103 but only 13.51% in family 9127. Locus CIR0185 shows similar pattern (with 88% donor allele retention). However, for most of the other alleles shown

here (and not shown because of retention in less than three families), donor allele retention is significantly lower than expected. In the *G. barbadense* background, a total of eight loci segregating in three or more families in *G. barbadense* background are shown in Table 2.8. For all eight loci, donor allele retention was significantly lower than expected segregation at the genotypic or allelic level.

Selection against donor alleles in BC_5F_2 families

A total of 13 loci showed no *G. hirsutum* (HH) homozygotes in 181 cases across all the segregating families in the *G. barbadense* background, and a total of 7 loci showed no *G. barbadense* (BB) homozygotes in 205 cases across all the segregating families in *G. hirsutum* background. This suggests a mild level of negative selection against donor alleles at or near these loci. Selection against *G. barbadense* homozygotes was nominally stronger ($\chi^2 = 3.57$, p value = 0.058) at At (85.71%) than Dt subgenomic loci (14.29%) in *G. hirsutum* background. Similarly, in the *G. barbadense* background, selection against *G. hirsutum* homozygotes was nominally stronger ($\chi^2 = 3.53$, p value = 0.061) at At (76.92%) than Dt subgenomic loci (23.08%). Segregation distortion as reflected by genotypic versus allelic frequency ratios indicates the type of selection. In both backgrounds, genotypic distortion was higher than allelic distortion: in *G. hirsutum* background, 63.59% of markers deviated significantly from genotypic expectations while only 22.82% deviated from allelic expectations; and in *G. barbadense* background, 76.79% markers deviated significantly from genotypic expectations and only 39.32% deviated from allelic expectations.

Discussion

A handful of studies has been carried out on transmission genetics of tetraploid cotton species, most of them focusing on the nature and patterns of introgression of donor alleles to the *G. hirsutum* background from *G. barbadense* (Jiang et al., 2000; Stephens, 1949), *G. tomentosum* (Waghmare et al., 2016) and *G. mustelinum* (Chandnani et al., 2017). The present study extends our knowledge of the patterns of introgression from *G. barbadense* to *G. hirsutum* while also providing novel insights on the nature and patterns of introgression of *G. hirsutum* chromatin into *G. barbadense* in a reciprocal experimental population. Our experimental data provide a glimpse into the consequences of natural exchange of chromatin between these two species. More generally, we provide rich empirical data useful to investigate many issues related to levels and patterns of introgression among species.

Genomic composition of the BC_5F_1 plants

Across the entire genome, the average retention of *G. hirsutum* alleles at the 2542 assayed loci among the 190 BC_5F_1 families was 4.35%, slightly but not significantly higher than the expected 3.125% (z = 0.97 and p-value = 0.165). Individual loci retained from 0.53% - 10.53% of *G. hirsutum* alleles (Figure 2.2) while individual BC_5F_1 families retained *G. hirsutum* alleles at 0.12% - 27.65%. A total of 48.9% of individuals retained *G. hirsutum* alleles at a rate lower than expected while 51.1% of individuals retained alleles at higher rate.

Similarly, the average retention of *G. barbadense* alleles at 3345 assayed loci among the 179 BC_5F_1 families was 4.79%, slightly but not significantly higher than the expected 3.125% (z = 1.277 and p-value = 0.101). Individual loci retained from 0.55 to 11% of *G. barbadense* alleles

while individual BC_5F_1 families retained *G. barbadense* alleles at 0.08 – 33.45% (Figure 2.2). A total of 46.6% of individuals retained *G. barbadense* alleles at a rate lower than expected while 53.4% of individuals retained alleles at higher rate. Average retention of donor alleles was not significantly different for the two reciprocal advanced populations both at the whole-genome level and for individual chromosomes (Table 2.1).

Persistence of donor chromatin in recipient genome

This study reveals consequences of reciprocal introgression between elite cultivars Acala Maxxa (*G. hirsutum*) and Pima S6 (*G. barbadense*). Higher than expected average levels of introgression of donor chromatin in both backgrounds (*G. hirsutum* background = 4.79 %, *G. barbadense* background = 4.35%) suggest favorability of donor alleles in general. Higher fitness of heterozygotes over homozygous genotypes might be a major cause of these results. Most previous studies showed unintentional selection against donor chromatin in interspecific crosses (Chandnani et al., 2017; Jiang et al., 2000; Waghmare et al., 2016). Small population sizes and lack of genome-wide genetic markers might have caused previous studies to underestimate the level of introgression. A previous study with population size similar to this experiment also reported twice the frequency of heterozygotes than homozygotes (Yu et al., 2011). Although we have used a high density of markers to scan the genome, still 0.2 Gb of genetically anchored genome was lacking polymorphic or segregating markers. One reason for the lack of polymorphism in this proportion of the genome might be the history of introgression from *G. barbadense* to *G. hirsutum* background for Acala cultivar development (Wang et al., 1995). Lack of genetic markers also might reflect uneven genome sampling in GBS libraries.

In the self-pollinated progeny of the BC_5F_1 plants, there was a conspicuous deficiency of donor alleles in both backgrounds. Patterns of segregation in the BC_5F_2 families were similar to those found in previous studies of *G. barbadense*, *G. tomentosum* and *G. mustelinum*, with frequencies of donor alleles that are lower than expected across most segregating families (Chandnani et al., 2017; Jiang et al., 2000; Waghmare et al., 2016). These results suggest that segregation in BC_5F_2 families favor the recipient haplotype with a higher average frequency of the recipient homozygotes than the donor homozygotes, which adds to prior evidence of non-random maintenance of integrity of the recipient genome and further supports the notion that higher fitness of heterozygotes than homozygotes may contribute to persistence of donor chromatin in recipient (recurrent) parent genomes.

Introgression devoid regions

Some regions in the recipient genome are not as tolerant of donor chromatin as other genomic regions. We found a total of 5 regions in the *G. hirsutum* background completely lacking *G. barbadense* introgression, accounting for 1.16% (22.10 Mb) of the total physical length of the anchored cotton genome (Table 2.3). Curiously, these do not correspond to 7 regions completely lacking *G. barbadense* introgression in a prior study (Jiang et al., 2000), suggesting that even among different combinations of *G. barbadense* and *G. hirsutum*, different chromosomal regions may be devoid of introgression. In the *G. barbadense* background, there were 16 regions completely devoid of *G. hirsutum* introgression, accounting for 2.62% (49.8 Mb) of the anchored genome. This indicates that although each BC_5F_1 individual was introgressed with donor alleles at a slightly higher than expected fraction of loci, introgression was possible in certain regions only. Similar results were reported in studies of introgression of chromatin of wild cotton relatives into

G. hirsutum (Chandnani et al., 2017; Waghmare et al., 2016) and in the study of introgression of *G. hirsutum* chromatin into *G. barbadense* (Wang et al., 1995).

Segregation distortion was generally evident from multiple linked markers, thus was clearly not attributable to sequencing errors but was a result of biological factors. If the elimination of donor chromatin were to occur randomly after a backcross, then the probability of any one unlinked region lacking introgression in the BC_5F_1 line would be (1 – 0.03125) = 0.96875. With the simplifying assumption that each of the 5 unlinked regions (in the *G. hirsutum* background) of segregation distortion behaves as a single unit of inheritance and all segregate independently, the probability of all 4 unlinked regions lacking introgressions in all 179 BC_5F_1 plants would be $[(0.96875)^5]^{179} = 4.56 \times 10^{-13}$. Thus, it is unlikely that all these regions lack introgression in all 179 BC_5F_1 plants by chance. In the *G. barbadense* background where there were 16 regions completely devoid of *G. hirsutum* introgression, this probability is $[(0.96875)^{16}]^{190} = 1.21 \times 10^{-42}$, providing even stronger evidence that this lack of introgression is not by chance but due to some biological factors.

To investigate possible biological factors involved in some genomic regions being recalcitrant to introgression, we studied all 21 genomic regions that were devoid of donor chromatin for gene ontology (GO) enrichment. A total of 2 GO enriched terms were identified for the *G. hirsutum* background, both being identified as having molecular functions (Table 2.5). A total of 10 GO enriched terms were identified in *G. barbadense* background, three related to biological processes (pollen recognition, cellulose biosynthesis, and coenzyme A metabolism) and one to cellular function (chloroplast). All the identified GO terms are basic biological processes, and the identification of these functional candidates offer testable hypotheses why certain genomic regions exclude alien chromatin over others.

Regions of prominent donor chromatin introgression

While some genomic regions are recalcitrant to donor chromatin, others were more tolerant. In general, both backgrounds allowed a slightly higher frequency of donor alleles than expected. All SDRs identified retained higher frequencies of donor alleles than expected. However, certain genomic regions showed greater richness of donor alleles than others. In the *G. hirsutum* background, eight of 19 SDRs were significantly richer in *G. barbadense* alleles than nominally significant SDRs. To investigate whether the transmission of these chromatin segments from *G. barbadense* to *G. hirsutum* occurs randomly, we compared our findings to those of a study to identify *G. barbadense* chromatin in *G. hirsutum* "Sealand"cultivars developed by the Pee Dee breeding program. Among a total of 22 putative *G. barbadense* chromosome segments in Sealand 542 and Sealand 883 backgrounds (Kumar et al., 2019) and 19 SDRs identified in our study, five regions clearly overlapped and two more were in close proximity (Table 2.9). These results hint at the possibility that the transfer of certain chromatin regions in interspecific crosses might potentially be related to cellular, molecular, or biological functions and might be informative for crop improvement. Indeed, a total of 13 quantitative trait loci (QTLs) related to six fiber quality traits were identified on the *G. barbadense* introgressed chromosomal segments (Kumar et al., 2019).

Similarly, in *G. barbadense* background, five (of 13) SDRs had enriched *G. hirsutum* chromatin (Table S2). Several *G. barbadense* cultivar groups (Pima, Egyptian and Sea Island) are known to harbor prominent *G. hirsutum* enrichment in five regions, one each in chromosomes 1, 5, 14 and 25, and one in unlinked linkage group U01 Wang et al. (1995). Pima S6, the *G. barbadense* cultivar used in our study, contained *G. hirsutum* chromatin in one G. hirsutum rich

region identified in our study, SDRGb01.1 in chromosome 1 (as verified by RFLP marker A1097 within the bounds of SDRGb01.1). Therefore, this 'apparent' SDR may be an artifact of a lack of *G. barbadense* alleles in Pima S6 in the region. However, the other four prominent regions as well as the remaining nominal SDRs identified in our study did not find correspond to *G. hirsutum* rich regions in Pima S6 (Wang et al., 1995).

Retention of donor alleles at numbers higher than the expected numbers of loci may reflect fitness consequences in recipient backgrounds; while reduced introgression might be linked with factors such as structural rearrangement, multilocus interaction, species integrity and reproductive isolation. GO and GO enrichment analysis was carried out on all SDRs as retention of donor alleles at higher-than-expected frequencies might reflect fitness consequences in recipient backgrounds. We were especially interested in GO terms related to fitness and adaptation in these regions. A total of 2503 GO terms and 16 GO enriched terms were identified (Tables 2.4, 2.10). Among the 16 GO enriched terms identified in these SDRs, 11 were related to biological processes. GO terms related to fitness and adaptation (response to freezing, response to biotic stimulus, defense response, photosynthesis and light reaction and different metabolic processes) were identified in regions rich in donor chromatin (Tables 2.4, 2.10) providing a starting point to investigate the hypothesis that genomic regions rich in donor chromatin introgression reflect fitness and adaptation behavior in recipient genome.

Subgenomic differentiation in introgression of donor chromatin

Our data about selection against donor alleles further support the notion that different subgenomes have different evolutionary fates. Selection against At subgenomic loci was slightly stronger than Dt subgenomic loci in the *G. hirsutum* background. Perhaps, this may be related to

the observations that D genome has higher expression than A genome and most fiber quality QTLs have been mapped on Dt subgenomic loci in allotetraploid cotton (Flagel & Wendel, 2010; Rong et al., 2004). Although the respective progenitor genomes for both subgenomes mostly contain common repertoires of genes, they differ largely in DNA quantities and transposable (repeat) element content; the A subgenome having significantly higher amount of these repeat elements than the D subgenome. In addition, the A subgenome is almost as twice as large as D subgenome in terms of DNA content. Despite these facts, studies have shown more genes with expression bias towards the D subgenome and asymmetrically higher gene loss in A subgenome than in D subgenome (Zhang et al., 2015). Li et al. (2015) showed significantly higher mutation frequency and rate of formation of SNPs within intergenic collinear regions of the D subgenome than in the A subgenome, which is consistent with the observation that disproportionately higher frequency of mutation were observed in Cot-filtered non-coding (CFNC) DNA of the D subgenome than the A subgenome (Rong et al., 2012). Albeit the A subgenome is evolving more rapidly than the D subgenome, more domestication pressure towards selecting higher yield and relaxed selection pressure in the A subgenome (Zhang et al., 2015) might have resulted in more fiber related QTLs being mapped into the D subgenome and in more D subgenome homeologs showing higher expression than their A genome counterparts.

Despite some genomic regions (or loci) showing complete selection against donor homozygotes, others showed different levels of permeability. Such differences in permeability of donor alleles by various regions of the recipient genome may indicate differential levels of fitness for the donor alleles (Rieseberg et al., 1999). Complete absence of recipient homozygotes at a few loci (3 cases each in *G. barbadense* and *G. hirsutum* backgrounds) and/or fixation of donor homozygotes at some loci suggest that a single introgression event can be sufficient to fix the

donor allele in a population. At the same time, complete absence of donor homozygotes at other loci suggests that some donor alleles dramatically reduce fitness in the recipient genome. The nature of selection has also been an important aspect of segregation studies (Li et al., 2011). In both backgrounds, more loci deviated from genotypic expectations than from allelic expectation in both BC_5F_1 and BC_5F_2 families. This suggests that zygotic selection may be more important than gametic selection in these populations.

Genetic backgrounds and their effects in transmission of donor alleles

Genetic backgrounds can profoundly affect the introgression of a particular chromosomal regions. Albeit introgression was observed across all chromosomes and the rate of overall as well as chromosome wise introgression was not significantly different in the two populations, transmission of certain genomic locations reveals a contrast in how these two backgrounds appeal each other. A total of 16 introgression devoid regions (IDRs) were identified in *G. barbadense* background while only five were identified in *G. hirsutum* background (Table 2.3). This clearly suggests that *G. barbadense* offers more resistance to *G. hirsutum* chromatin than the reciprocal; and is supported by the fact that a higher number of SDRs enriched in donor alleles were identified in *G. hirsutum* background than in *G. barbadense* background (Table 2.4). Occasional crosses between improved forms of *G. hirsutum* and *G. barbadense* have led to a degree of genetic exchange that may have mitigated the resistance of Pima S6 chromatin to *G. hirsutum*. Indeed, most improved genotypes of *G. barbadense* are comprised of 5-10% of *G. hirsutum* chromatin, with about two-thirds of those being clustered at five specific locations (Wang et al., 1995). Efforts to introduce *G. barbadense* traits into *G. hirsutum* cultivars (Kumar et al., 2019) have had much

less impact on the elite gene pool, perhaps leaving the inherent isolation mechanisms of *G. hirsutum* more intact.

Other differences between the behavior of introgressed chromatin between these two species are not readily explained by experimental design or breeding history. For example, among those loci that were heterozygous in BC_5F_1 and were segregating in the BC_5F_2 families, both backgrounds showed considerable tolerance of homozygosity from the donor (18.58% homozygosity tolerance by *G. hirsutum* and 12.32% homozygosity tolerance by *G. barbadense*). Similar level of tolerance (18.6%) of *G. barbadense* by *G. hirsutum* was reported by (Waghmare et al., 2016). These levels of tolerance were much higher than those reported on a wide cross involving *G. tomentosum* (1.27%) in *G. hirsutum* background (Waghmare et al., 2016). These results are incongruous with the closer geographic proximity of wild *G. hirsutum* (Central America) to *G. barbadense* (Peru) that would seem to confer a greater selective advantage to reproductive-isolation mechanisms acting between these species than with *G. tomentosum* (Hawaii). The greater evolutionary distance between *G. hirsutum* and *G. barbadense* (representing different polyploid clades) should also have provided greater opportunity for such mechanisms to evolve.

Reciprocal transmission of donor chromatin

The reciprocal populations described here offer a broader scope of understanding the genetics of transmission of donor chromatin than can be achieved by more conventional, unidirectional, studies. The overall rate of introgression of donor chromatin in the reciprocal populations was similar (4.22 in *G. barbadense* background vs 4.79% in *G. hirsutum* background), but the nature of this retention was very different when we look closely at specific genomic regions

(Figure 2.2, Table 2.1). For example, in the *G. barbadense* background, the retention of donor alleles along the length of chromosome 1 (5.99% for this chromosome) is almost always greater than the expected rate of 3.125% suggesting favorability of *G. hirsutum* chromatin along the length of chromosome 1. Interestingly, in the reciprocal (*G. hirsutum*) background, the donor (*G. barbadense*) alleles are retained at rates lower than expected along the length of chromosome 1, suggesting selection favoring *G. hirsutum* alleles. This hypothesis is further supported by the fact that no *G. barbadense* introgressions were identified in chromosome 1 in two crosses involving upland cotton Suyuan 7235 as female parent and Sealand 542 and Sealand 883 as male parents respectively in just two generations after initial crossing (Kumar et al., 2019). In addition, prominent regions of *G. hirsutum* chromatin introgression was observed in a large and wide collection of *G. barbadense* cultivars (Wang et al., 1995) with some Pima and Sea Island accessions harboring *G. hirsutum* chromatin along the entire length of chromosome 1. Recent deep sequencing study conducted by Hu et al. (2019) in nine *G. barbadense* and ten *G. hirsutum* accessions also revealed an *G. hirsutum* chromatin introgressed region (43.10 Mb to 92.00 Mb) on chromosome 1 in all nine *G. barbadense* accessions collected from Egyptian, American Pima and Central Asian ELS cottons.

Selection favoring *G. barbadense* alleles is exemplified by chromosome 23. The *G. hirsutum* background retained *G. barbadense* alleles at 6.4% for chromosome 23 while the reciprocal background harbored *G. hirsutum* alleles at significantly lower rates (p value < 0.0001) than expected. Indeed, chromosomes 3, 6, 9, 12, 13, 16, 17, 18, 19, 20 and 26 contain short regions favoring chromatin from one species over the other. Certain regions, however, have shown heterozygote advantage over the recipient alleles, retaining donor alleles in higher frequencies than expected in both backgrounds (Figures 2.2 and 2.3). For example, almost the entire length of

chromosome 2 shows favorability for donor alleles in both backgrounds (6.33% retention in *G. hirsutum* background and 5.4% retention in *G. barbadense* background). Similar patterns of heterozygote advantage have been observed in regions of chromosomes 4, 5, 10, 15 and 25. Other genomic regions completely alienated donor chromatin in both backgrounds. Although almost the entire length of chromosome 1 favored *G. hirsutum* alleles, the distal end of both this chromosome and chr. 11 lacked donor alleles in both reciprocal crosses (Figures 2.2 and 2.3, Table 2.3). Avoidance of donor chromatin may be related to species integrity via preservation of important cellular, molecular, and/or biological functions.

Conclusion

In summary, reciprocal transmission genetic study between *G. hirsutum* and *G. barbadense* shows that the extent of introgression and the fate of introgressed chromatin depends on several factors including genetic background, fitness of substituted alleles and allelic combinations, and location of transmitted chromatin. An important motivation for the analysis of advanced-generation interspecific population involving these two species is that they have been frequently analyzed for discovery of novel variation that might enhance agricultural productivity. Commercially, *G. barbadense* fiber has qualities superior to those of most if not all *G. hirsutum;* and commands a premium price (currently near 3x!) though is much lower yielding. Valuable phenotypic attributes associated with *G. barbadense* introgression have been reported (Jiang et al., 2000). The same population also revealed a rich set of QTLs with potentially desirable attributes (P. W. Chee et al., 2005; Xavier Draye et al., 2005), a subset of which have been studied in detail for their value in elite germplasm (Shen et al., 2011). The transfer of desirable attributes to/from *G. hirsutum* to/from *G. barbadense*, long a goal of many cotton breeders, has generally failed.

Genetic analysis now provides insight into the biological complexity of this undertaking, which is complicated by interactions between unlinked loci, pronounced differences among genetic backgrounds, uncertain predictive value across generations, and difficulties associated with obtaining fixed (homozygous) genotypes for many introgressed segments.

References

Adhikari, J., Das, S., Wang, Z., Khanal, S., Chandnani, R., Patel, J. D., Goff, V. H., Auckland, S., Rainville, L. K., & Jones, D. (2017). Targeted identification of association between cotton fiber quality traits and microsatellite markers. *213*(3), 1-15.

Anderson, E. (1949). *Introgressive hybridization* [Bibliographies Non-fiction]. J. Wiley.

Andolfatto, P., Davison, D., Erezyilmaz, D., Hu, T. T., Mast, J., Sunayama-Morita, T., & Stern, D. L. (2011). Multiplexed shotgun genotyping for rapid and efficient genetic mapping. *21*(4), 610-617.

Baack, E., Melo, M. C., Rieseberg, L. H., & Ortiz-Barrientos, D. (2015). The origins of reproductive isolation in plants. *New Phytologist*, *207*(4), 968-984.

Beasley, J. O. (1942). Meiotic chromosome behaviour in species, species hybrids, haploids, and induced polyploids of Gossypium [Journal Article]. *Genetics*, *27*, 25-54.

Brubaker, C. L., Paterson, A. H., & Wendel, J. F. (1999). Comparative genetic mapping of allotetraploid cotton and its diploid progenitors. *42*(2), 184-203.

Chandnani, R., Wang, B., Draye, X., Rainville, L. K., Auckland, S., Zhuang, Z., & Lubbers, E. L. (2017). Segregation distortion and genome-wide digenic interactions affect transmission of introgressed chromatin from wild cotton species [Report]. *Theoretical and Applied Genetics*, *130*(10), 2219.

Chee, P. W., Draye, X., Jiang, C. X., Decanini, L., Delmonte, T. A., Bredhauer, R., Smith, C. W., & Paterson, A. H. (2005). Molecular dissection of phenotypic variation between Gossypium hirsutum and Gossypium barbadense (cotton) by a backcross-self approach: III. Fiber length [Journal Article]. *TAG Theoretical and Applied Genetics*, *111*(4), 772-781.

Desai, A., Chee, P. W., Junkang, R., May, O. L., Paterson, A. H., & Gustafson, J. P. (2006). Chromosome structural changes in diploid and tetraploid A genomes of Gossypium [Article]. *Genome*, *49*(4), 336-345.

Draye, X., Chee, P., Jiang, C.-X., Decanini, L., Delmonte, T. A., Bredhauer, R., Smith, C. W., & Paterson, A. H. (2005). Molecular dissection of interspecific variation between Gossypium hirsutum and G. barbadense (cotton) by a backcross-self approach: II. Fiber fineness [Article]. *Theoretical & Applied Genetics*, *111*(4), 764-771.

Flagel, L. E., & Wendel, J. F. (2010). Evolutionary rate variation, genomic dominance and duplicate gene expression evolution during allotetraploid cotton speciation [research-article]. *The New Phytologist*, *186*(1), 184-193.

Glaubitz, J. C., Casstevens, T. M., Lu, F., Harriman, J., Elshire, R. J., Sun, Q., & Buckler, E. S. (2014). TASSEL-GBS: A High Capacity Genotyping by Sequencing Analysis Pipeline [Article]. *PLoS ONE*, *9*(2), 1-11.

Grant, V. (1981). *Plant speciation* (2nd ed.) [BibliographiesNon-fiction]. Columbia University Press.

Heiser, C. B. (1979). Hybrid Populations of Helianthus divaricatus and H. microcephalus after 22 Years [research-article]. *Taxon*, *28*(1/3), 71-75.

Hodgkin, T., & Hajjar, R. (2007). Using crop wild relatives for crop improvement: trends and perspectives. Wallingford, UK.

Hu, Y., Chen, J., Fang, L., Zhang, Z., Ma, W., Niu, Y., Ju, L., Deng, J., Zhao, T., Lian, J., Baruch, K., Fang, D., Liu, X., Ruan, Y.-l., Rahman, M.-u., Han, J., Wang, K., Wang, Q., Wu, H., . . . Zhang, T. (2019). Gossypium barbadense and Gossypium hirsutum genomes provide

insights into the origin and evolution of allotetraploid cotton. *Nature Genetics, 51*(4), 739-748.

Jiang, C.-x., Chee, P. W., Draye, X., Morrell, P. L., Smith, C. W., & Paterson, A. H. (2000). MULTILOCUS INTERACTIONS RESTRICT GENE INTROGRESSION IN INTERSPECIFIC POPULATIONS OF POLYPLOID GOSSYPIUM (COTTON). *Evolution, 54*(3), 798-814.

Kelly, S., Huihui, L., Navarro, J. A. R., Dong, A., Maria Cinta, R., Sarah, H., Charlotte, A., Jeffrey, C. G., Sharon, M., Robert, J. E., Edward, S. B., & Peter, J. B. (2014). Novel Methods to Optimize Genotypic Imputation for Low-Coverage, Next-Generation Sequence Data in Crop Plants [article]. *The Plant Genome, 7*(3).

Kim, C., Guo, H., Kong, W., Chandnani, R., Shuang, L.-S., & Paterson, A. H. (2016, 2016/01/01/). Application of genotyping by sequencing technology to a variety of crop breeding programs. Netherlands.

Kumar, P., Singh, R., Lubbers, E. L., Shen, X., Paterson, A. H., Campbell, B. T., Jones, D. C., & Chee, P. W. (2019). Genetic Evaluation of Exotic Chromatins from Two Obsolete Interspecific Introgression Lines of Upland Cotton for Fiber Quality Improvement. *Crop Science, 59*(3), 1073-1084.

Levi, A., Ovnat, L., Paterson, A. H., & Saranga, Y. (2009). Photosynthesis of cotton near-isogenic lines introgressed with QTLs for productivity and drought related traits [Article]. *Plant Science, 177*(2), 88-96.

Meyn, O., & Emboden, W. A. (1987). Parameters and Consequences of Introgression in Salvia apiana x S. mellifera (Lamiaceae) [research-article]. *Systematic Botany, 12*(3), 390-399.

Paterson, A. H., Boman, R. K., Brown, S. M., Chee, P. W., Gannaway, J. R., Gingle, A. R., May, O. L., & Smith, C. W. (2004). Reducing the genetic vulnerability of cotton. Madison, USA.

Paterson, A. H., Brubaker, C. L., & Wendel, J. F. (1993). A rapid method for extraction of cotton (Gossypium spp.) genomic DNA suitable for RFLP or PCR analysis. *Plant Molecular Biology Reporter*, *11*(2), 122-127.

Paterson, A. H., Wendel, J. F., Gundlach, H., Guo, H., Jenkins, J., Jin, D., Llewellyn, D., Showmaker, K. C., Shu, S., Udall, J., Yoo, M.-j., Byers, R., Chen, W., Doron-Faigenboim, A., Duke, M. V., Gong, L., Grimwood, J., Grover, C., Grupp, K., . . . Schmutz, J. (2012). Repeated polyploidization of Gossypium genomes and the evolution of spinnable cotton fibres. *Nature*, *492*(7429), 423-427.

Rieseberg, L. H., Kim, M. J., & Seiler, G. J. (1999). Introgression between the Cultivated Sunflower and a Sympatric Wild Relative, Helianthus petiolaris (Asteraceae). *International Journal of Plant Sciences*, *160*(1), 102-108.

Rong, J., Abbey, C., Bowers, J. E., Brubaker, C. L., Chang, C., Chee, P. W., Delmonte, T. A., Ding, X., Garza, J. J., & Marler, B. S. (2004). A 3347-locus genetic recombination map of sequence-tagged sites reveals features of genome organization, transmission and evolution of cotton (Gossypium). *166*(1), 389-417.

Shen, X., Cao, Z., Singh, R., Lubbers, E. L., Xu, P., Smith, C. W., Paterson, A. H., & Chee, P. W. (2011). Efficacy of qFL-chr1, a Quantitative Trait Locus for Fiber Length in Cotton (Gossypium spp.). *Crop Science*, *51*(5), 2005-2010.

Stephens, S. G. (1949). The cytogenetics of speciation in Gossypium I. Selective elimination of the donor parent genotype in interspecific backcrosses [Journal Article]. *Genetics 1949.*, *34*, 627-637.

Tanksley, S. D., & Nelson, J. C. (1996). Advanced backcross QTL analysis: a method for the simultaneous discovery and transfer of valuable QTLs from unadapted germplasm into elite breeding lines. *92*(2), 191-203.

Waghmare, V. N., Rong, J., Rogers, C. J., Bowers, J. E., Chee, P. W., Gannaway, J. R., Katageri, I., & Paterson, A. H. (2016). Comparative transmission genetics of introgressed chromatin in Gossypium (cotton) polyploids [research-article]. *American Journal of Botany*, *103*(4), 719-729.

Waghmare, V. N., Rong, J., Rogers, C. J., Pierce, G. J., Wendel, J. F., & Paterson, A. H. (2005). Genetic mapping of a cross between Gossypium hirsutum (cotton) and the Hawaiian endemic, Gossypium tomentosum [Article]. *Theoretical & Applied Genetics*, *111*(4), 665-676.

Wang, G. L., Dong, J. M., & Paterson, A. H. (1995). The distribution of Gossypium hirsutum chromatin in G. barbadense germ plasm: molecular analysis of introgressive plant breeding. *Theoretical and Applied Genetics*, *91*(6), 1153-1161.

Wendel, J. F. (1989). New World Tetraploid Cottons Contain Old World Cytoplasm [research-article]. *Proceedings of the National Academy of Sciences of the United States of America*, *86*(11), 4132-4136.

Wendel, J. F., Schnabel, A., & Seelanan, T. (1995). Bidirectional Interlocus Concerted Evolution Following Allopolyploid Speciation in Cotton (Gossypium) [research-article]. *Proceedings of the National Academy of Sciences of the United States of America*, *92*(1), 280-284.

Yu, J., Yu, S., Lu, C., Wang, W., Fan, S., Song, M., Lin, Z., Zhang, X., & Zhang, J. (2007). High-density Linkage Map of Cultivated Allotetraploid Cotton Based on SSR, TRAP, SRAP and AFLP Markers. *Journal of Integrative Plant Biology*, *49*(5), 716-724.

Yu, Y., Yuan, D., Liang, S., Li, X., Wang, X., Lin, Z., & Zhang, X. (2011). Genome structure of cotton revealed by a genome-wide SSR genetic map constructed from a BC1 population between gossypium hirsutum and G. barbadense. *BMC Genomics*, *12*(1), 15.

Zhang, T., Hu, Y., Jiang, W., Fang, L., Guan, X., Chen, J., Zhang, J., Saski, C. A., Scheffler, B. E., Stelly, D. M., Hulse-Kemp, A. M., Wan, Q., Liu, B., Liu, C., Wang, S., Pan, M., Wang, Y., Wang, D., Ye, W., . . . Chen, Z. J. (2015). Sequencing of allotetraploid cotton (Gossypium hirsutum L. acc. TM-1) provides a resource for fiber improvement. *Nature biotechnology*, *33*(5), 531-537.

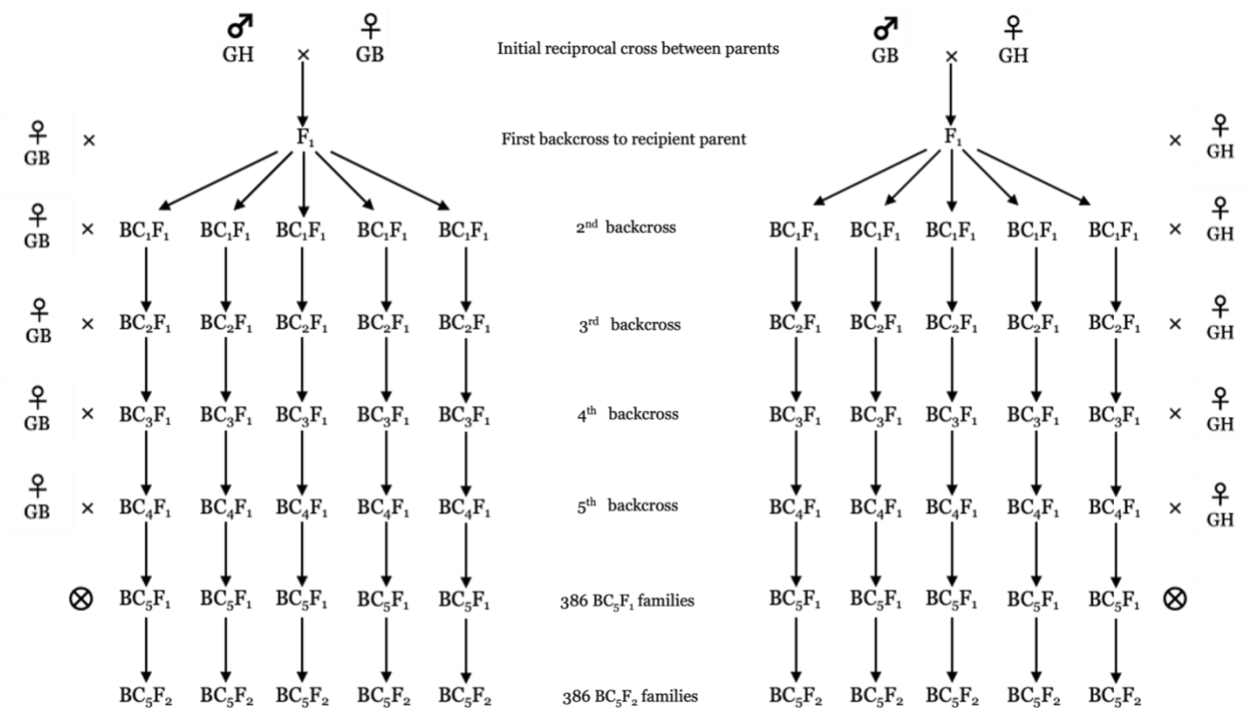

Figure 2.1. Development of reciprocal set of advanced-backcross populations. GH and GB denote *G. hirsutum* and *G. barbadense* respectively. Each backcross lineage was advanced by single-seed descent.

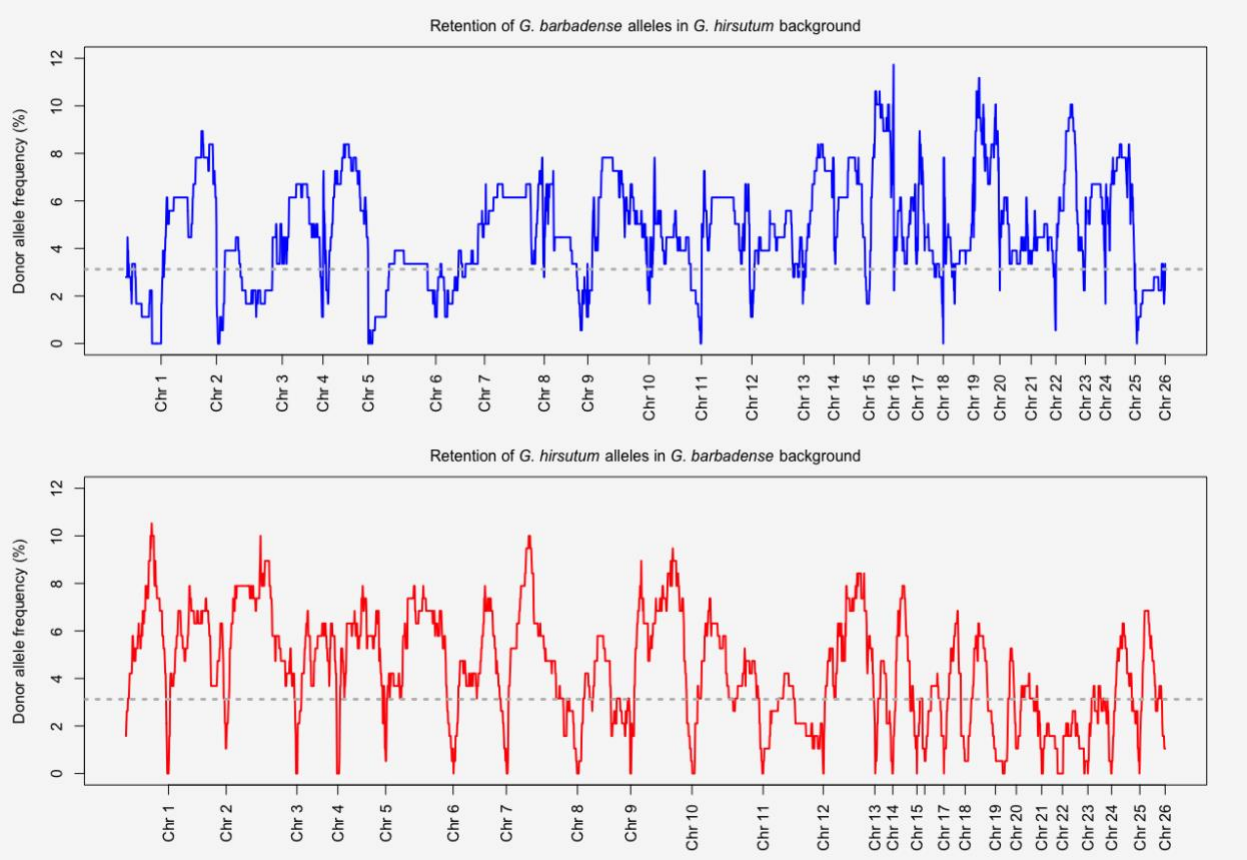

Figure 2.2. Retention of donor alleles in reciprocal interspecific populations. X-axis shows markers across the genome separated by chromosomes and y-axis shows the frequency of donor allele. The dotted lines show expected donor allele frequency (3.125%) for BC_5F_1 generation.

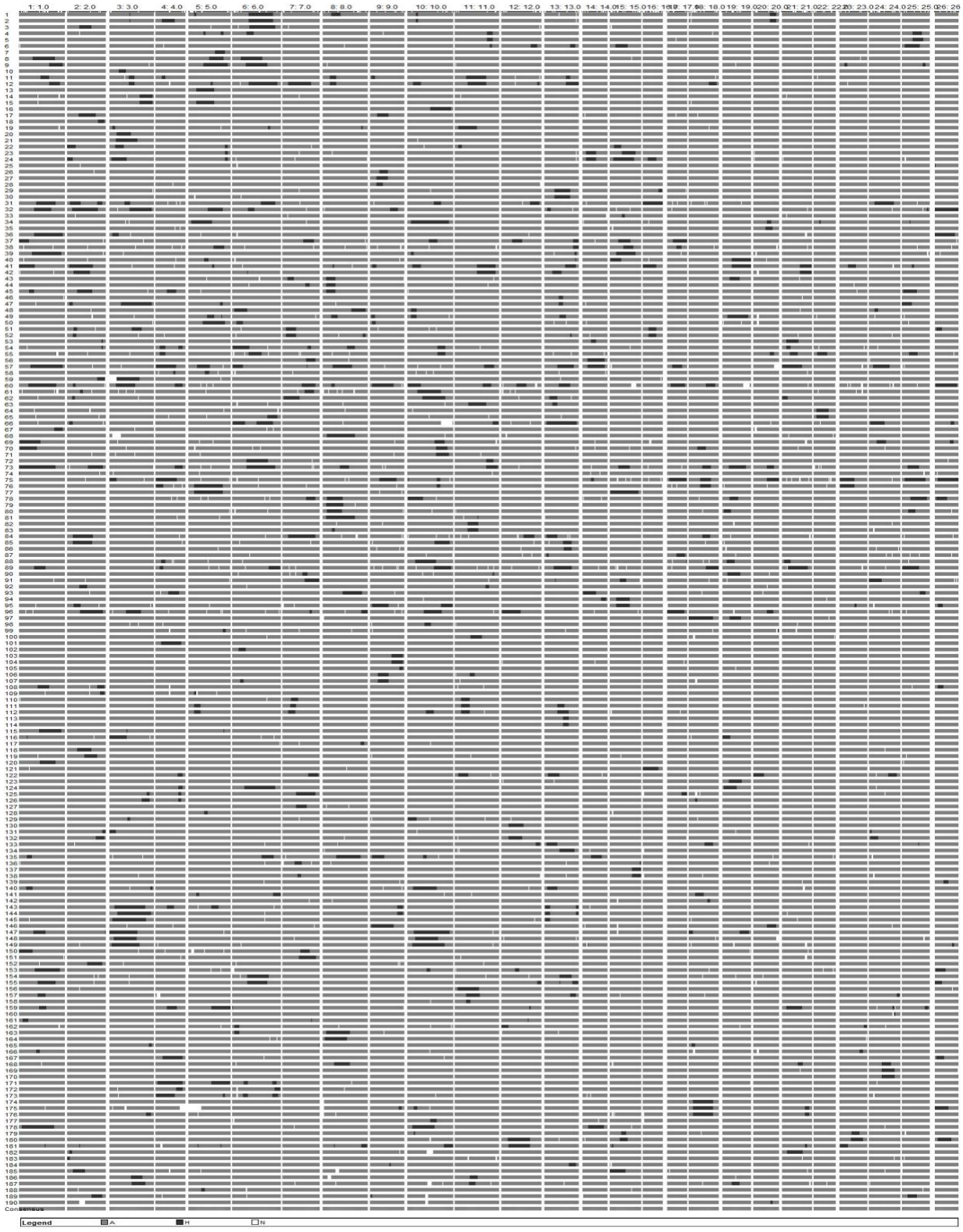

Figure 2.3. Genomic composition of 190 BC_5F_1 lines in *G. barbadense* background. Gray areas represent *G. barbadense* homozygotes (BB), black areas represent heterozygotes (BH) and white areas represent missing genotypes (Chromosomes on x-axis, individuals on y-axis).

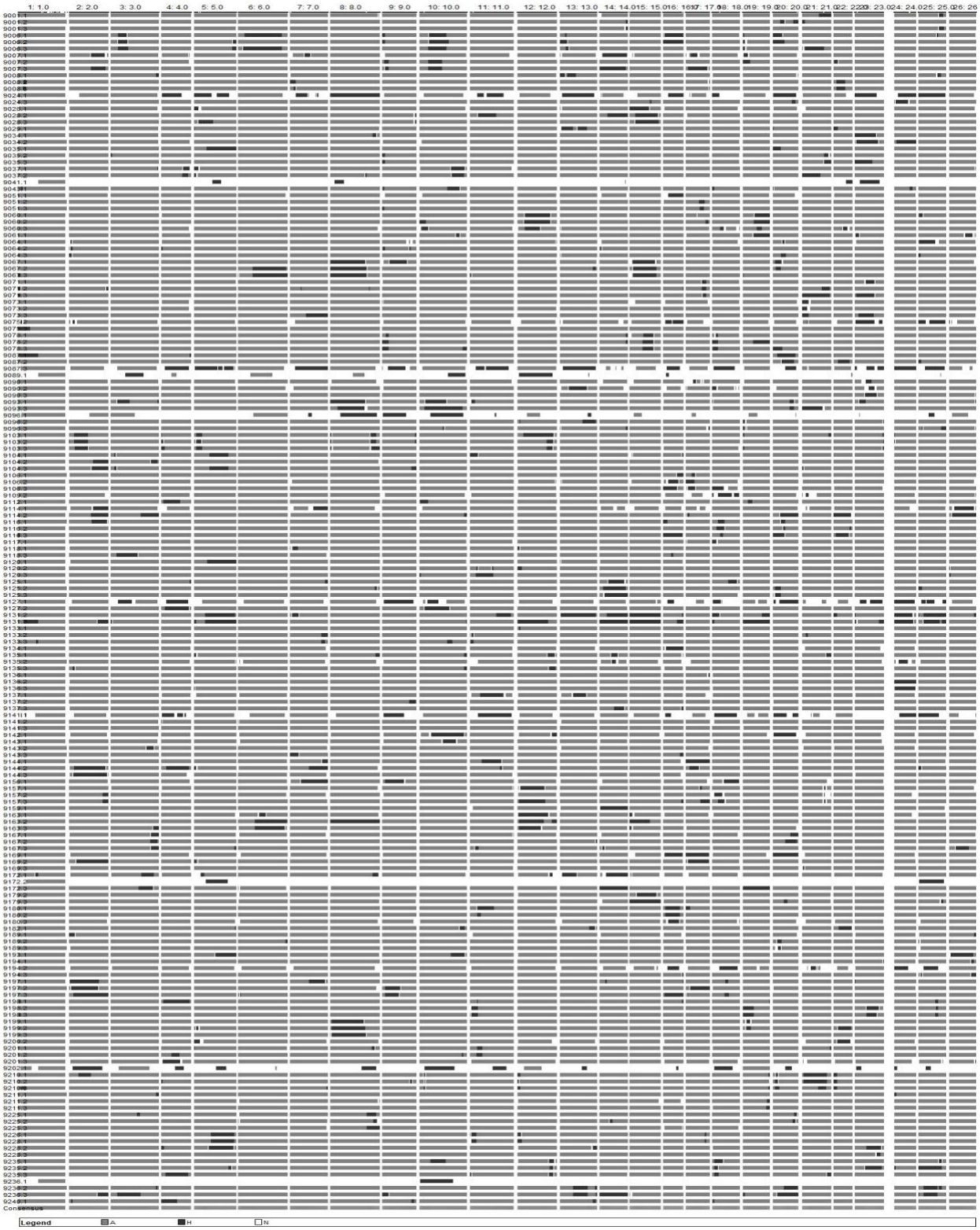

Figure 2.4. Genomic composition of 179 BC_5F_1 lines in *G. hirsutum* background. Gray areas represent *G. hirsutum* homozygotes (HH), black areas represent heterozygotes (HB) and white areas represent missing genotypes (Chromosomes on x-axis, individuals on y-axis).

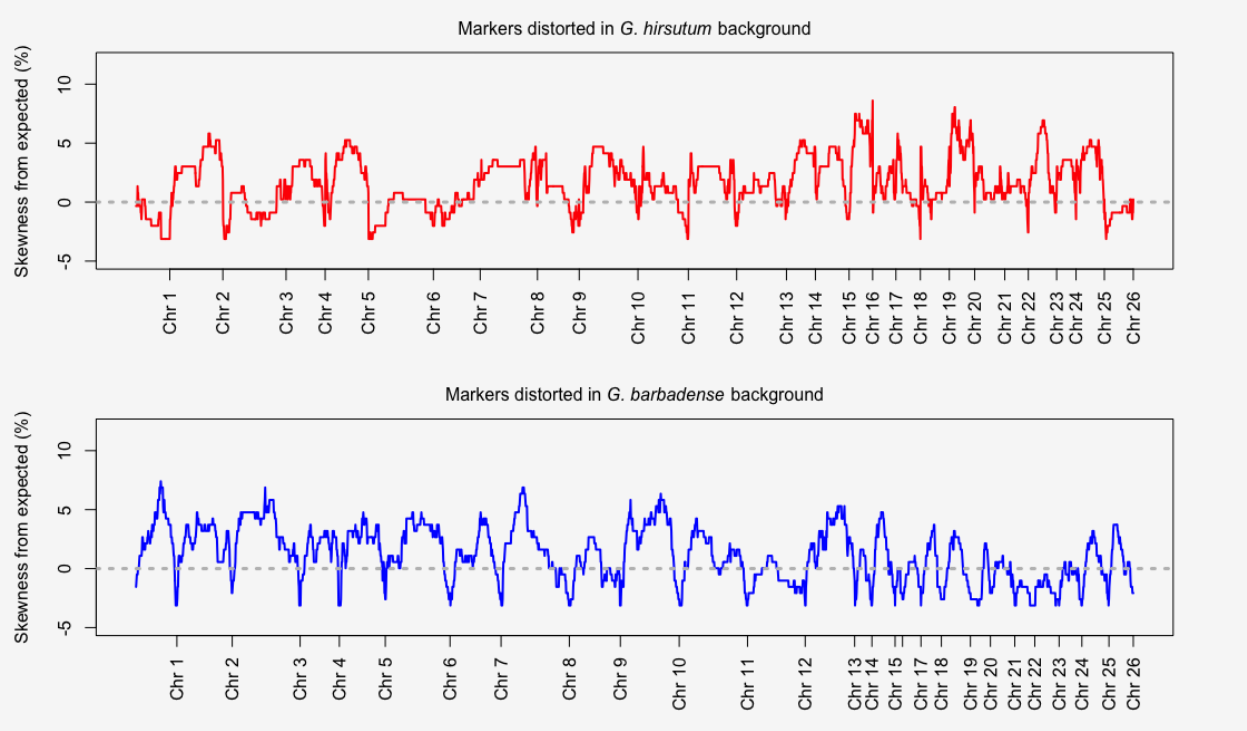

Figure 2.5. Segregation distortion in BC_5F_1.

Table 2.1. Distribution of markers, anchored genome coverage and average retention of donor alleles in *G. hirsutum* and *G. barbadense* backgrounds.

Chr	*G. hirsutum* background			*G. barbadense* background		
	# markers	% genome covered	% donor alleles	# markers	% genome covered	% donor alleles
1	114	99.01	1.53	105	95.74	5.99
2	178	95.53	6.33	141	96.78	5.40
3	212	97.76	2.67	173	92.14	6.33
4	131	94.61	5.28	100	97.63	4.78
5	145	92.67	6.28	117	96.01	5.33
6	218	97.61	2.55	165	98.02	5.32
7	157	96.06	3.05	130	97.86	4.25
8	192	97.57	5.89	174	90.78	4.95
9	140	91.76	4.03	130	94.78	3.33
10	197	94.89	5.73	150	94.78	6.12
11	169	95.28	3.79	174	99.16	4.13
12	162	91.16	5.35	147	95.47	2.29
13	167	94.58	4.12	126	90.34	5.77
14	97	92.19	6.63	44	85.64	2.67
15	113	93.58	5.83	59	95.81	4.63
16	79	89.38	8.58	19	90.15	2.02
17	78	93.18	4.73	47	82.67	2.68
18	82	89.78	4.33	52	98.82	3.67
19	97	84.08	4.07	74	92.49	3.53
20	85	93.21	8.22	51	95.50	1.82
21	101	90.89	4.37	62	94.55	2.76
22	79	76.74	4.14	51	90.40	1.04
23	95	91.56	6.40	62	91.18	1.47
24	65	68.29	5.94	58	98.17	2.41
25	95	94.16	6.82	69	99.19	3.67
26	97	93.69	2.13	62	84.90	4.13

Table 2.2. Genomic distribution of introgressed chromosomal segments.

	G. barbadense background	*G. hirsutum* background
Total individuals	190	179
Total introgressions	617	722
Average introgressions per plant	3.25	4.03
Min size of int (Mb)	1.67	0.12
Max size of int (Mb)	83.55	101.05
Average size (Mb)	23.31	20.48
At subgenome introgressions	426	355
Dt subgenome introgressions	191	367

Table 2.3. Distribution of introgression devoid regions (IDRs) in *G. hirsutum* and *G. barbadense* backgrounds.

Background	Chr	DDR id	Location (Mb)	# Genes	# GO terms	GO Enriched
G. hirsutum	1	IDRGh01.1	93.02 - 99.77	873	0	0
	3	IDRGh03.1	5.03 - 8.29	116	0	0
	6	IDRGh06.1	0.30 - 4.48	269	178	1
	6	IDRGh06.2	5.86 - 9.86	134	0	0
	11	IDRGh11.1	87.81 - 91.73	201	120	1
G. barbadense	1	IDRGb01.1	91.87 - 96.53	198	0	0
	3	IDRGb03.1	93.61 - 96.23	131	87	2
	4	IDRGb04.1	56.67 - 61.88	285	191	1
	5	IDRGb05.1	2.82 - 4.61	169	0	0
	8	IDRGb08.1	3.12 - 6.96	228	164	2
	8	IDRGb08.2	97.10 - 97.19	3	0	0
	9	IDRGb09.1	1.68 - 2.53	36	0	0
	9	IDRGb09.2	71.51 - 72.76	150	101	3
	11	IDRGb11.1	0.22 - 4.59	467	0	0
	11	IDRGb11.2	91.63 - 92.76	54	0	0
	14	IDRGb14.1	55.75 - 56.62	23	0	0
	20	IDRGb20.1	24.48 - 26.51	1072	0	0
	22	IDRGb22.1	0.28 - 4.75	299	178	2
	22	IDRGb22.2	31.66 - 46.38	449	0	0
	23	IDRGb23.1	53.83 - 55.01	0	0	0
	24	IDRGb24.1	64.86 - 65.59	92	0	0

Table 2.4. Segregation distortion regions (SDRs) with number of genes, GO terms and GO enriched terms. SDRs in bold represented genomic regions with prominent introgressions.

Background	Chr	SDR ID	Location (Mb)	# Genes	GO terms	GO Enriched
	1	**SDRGb01.1**	33.59 - 76.82	231	0	0
	2	SDRGb02.1	15.24 - 51.11	284	0	0
	3	**SDRGb03.1**	17.33 - 61.20	333	0	0
	5	SDRGb05.1	51.74 - 56.45	113	0	0
	6	SDRGb06.1	34.69 - 83.85	687	421	1
G. barbadense	7	SDRGb07.1	49.47 - 59.07	61	0	0
	8	**SDRGb08.1**	20.67 - 30.32	92	0	0
	10	SDRGb10.1	16.07 - 25.25	200	131	3
	10	**SDRGb10.1**	43.04 - 80.28	421	0	0
	11	SDRGb11.1	29.51 - 31.74	20	13	1
	13	**SDRGb13.1**	34.25 - 59.72	272	0	0
	15	SDRGb15.1	22.68 - 35.25	186	0	0
	26	SDRGb26.1	13.48 - 21.73	119	0	0
	2	**SDRGh02.1**	47.56 - 80.39	395	238	4
	4	SDRGh04.1	22.20 - 37.43	88	0	0
	5	**SDRGh05.1**	36.57 - 72.80	381	218	1
	8	SDRGh08.1	23.17 - 29.83	32	0	0
	8	SDRGh08.2	64.54 - 93.60	597	0	0
	9	SDRGh09.1	2.93 - 13.98	200	0	0
	10	SDRGh10.1	25.05 - 55.54	117	0	0
	10	SDRGh10.2	67.73 - 69.42	9	0	0
	12	SDRGh12.1	69.54 - 83.55	751	0	0
G. hirsutum	14	**SDRGh14.1**	19.36 - 58.54	798	0	0
	15	SDRGh15.1	27.28 - 48.82	381	0	0
	16	**SDRGh16.1**	6.81 - 43.48	1744	0	0
	18	**SDRGh18.1**	1.91 - 22.23	769	0	0
	19	SDRGh19.1	3.22 - 4.97	231	0	0
	20	**SDRGh20.1**	7.33 - 50.59	1123	692	2
	23	**SDRGh23.1**	22.33 - 41.64	947	606	1
	24	SDRGh24.1	29.03 - 43.78	296	184	3
	25	SDRGh25.1	1.81 - 5.16	195	0	0
	25	**SDRGh25.2**	12.63 - 48.66	744	0	0

Table 2.5. GO enriched terms identified in IDRs in both backgrounds.

IDR ID	Accession	Name	GO Type	Q-value
IDRGh06.1	GO:0016705	oxidoreductase activity	molecular	9.00E-05
IDRGh11.1	GO:0000062	fatty-acyl-CoA binding	molecular	6.80E-06
IDRGb03.1	GO:0015936	coenzyme A metabolic process	biological	4.70E-07
	GO:0004420	NADPH activity	molecular	4.70E-07
IDRGb04.1	GO:0004506	squalene monooxygenase activity	molecular	1.30E-06
IDRGb08.1	GO:0016853	isomerase activity	molecular	1.40E-05
	GO:0009507	chloroplast	cellular	4.00E-05
IDRGb09.2	GO:0004869	endopeptidase inhibitor activity	molecular	1.00E-05
	GO:0016760	cellulose synthase activity	molecular	4.80E-05
	GO:0030244	cellulose biosynthetic process	biological	8.80E-05
IDRGb22.1	GO:0048544	recognition of pollen	biological	9.00E-07
	GO:0004674	protein serine/threonine kinase activity	molecular	1.00E-05

Table 2.6. Chromosome wise segregation distortion in BC_5F_2 families.

	G. hirsutum background				*G. barbadense* background			
	Allelic		Genotypic		Allelic		Genotypic	
Chr	χ^2	p value	χ^2	p value	χ^2	p value	χ^2	p value
1	9.05	0.003	86.32	0	153.37	0	297.75	0
2	111.39	0	231.27	0	128.03	0	211.00	0
3	74.20	0	211.39	0	67.78	0	218.51	0
4	16.82	0	231.53	0	322.38	0	609.45	0
5	313.49	0	606.62	0	259.50	0	570.53	0
6	30.09	0	121.09	0	72.66	0	282.75	0
7	121.16	0	349.12	0	238.42	0	392.36	0
8	40.70	0	83.17	0	254.43	0	388.81	0
9	36.39	0	166.11	0	214.09	0	396.30	0
10	128.03	0	211.65	0	188.16	0	474.89	0
11	80.98	0	183.64	0	115.33	0	261.22	0
12	15.75	0	188.10	0	100.52	0	150.78	0
13	9.50	0.002	186.50	0	136.03	0	603.24	0
14	2.20	0.138	3.09	0.213	0.11	0.738	72.11	0
15	74.98	0	110.65	0	196.27	0	390.65	0
16	67.37	0	112.69	0	19.51	0	56.51	0
17	6.90	0.008	55.90	0	2.82	0.093	4.52	0.104
18	0.91	0.339	13.51	0.001	176.75	0	407.51	0
19	17.63	0	156.63	0	165.31	0	263.31	0
20	15.25	0	110.47	0	22.53	0	37.53	0
21	6.57	0.010	44.69	0	68.00	0	166.94	0
22	73.80	0	180.84	0	20.57	0	34.57	0
23	56.64	0	96.85	0	48.13	0	86.53	0
24	1.09	0.290	1.43	0.488	0.50	0.479	9.50	0.008
25	18.98	0	115.06	0	19.60	0	39.60	0
26	11.11	0	29.11	0	239.41	0	389.41	0

Table 2.7. Loci showing significant variation in segregation patterns as tested by χ^2 test (at $p<0.01$) in BC_5F_2 families (*G. hirsutum* background).

BC_5F_2	BNL3441			BNL3580			CIR0185		
Family	HH	HB	BB	HH	HB	BB	HH	HB	BB
9006	15	25	17	-	-	-	3	20	12
9035	-	-	-	-	-	-	12	2	1
9037	-	-	-	-	-	-	-	-	-
9075	-	-	-	4	4	0	-	-	-
9087	-	-	-	23	4	1	-	-	-
9093	12	5	3	-	-	-	-	-	-
9103	-	-	-	-	-	-	-	-	-
9118	16	1	0	-	-	-	-	-	-
9120	-	-	-	-	-	-	9	1	9
9127	13	5	7	-	-	-	-	-	-
9131	-	-	-	-	-	-	20	18	6
9157	-	-	-	-	-	-	-	-	-
9167	-	-	-	-	-	-	20	3	1
9198	-	-	-	11	11	9	-	-	-
9201	-	-	-	-	-	-	-	-	-
BC_5F_2	DPL0085			DPL0176			MUSB1307		
Family	HH	HB	BB	HH	HB	BB	HH	HB	BB
9006	-	-	-	-	-	-	-	-	-
9035	-	-	-	-	-	-	-	-	-
9037	6	3	24	-	-	-	-	-	-
9075	-	-	-	-	-	-	-	-	-
9087	-	-	-	-	-	-	-	-	-
9093	-	-	-	-	-	-	-	-	-
9103	7	2	18	10	24	3	8	3	8
9118	-	-	-	-	-	-	-	-	-
9120	-	-	-	-	-	-	-	-	-
9127	32	6	5	-	-	-	-	-	-
9131	-	-	-	-	-	-	13	5	9
9157	-	-	-	-	-	-	3	1	4
9167	-	-	-	-	-	-	-	-	-
9198	-	-	-	-	-	-	-	-	-
9201	-	-	-	1	1	2	-	-	-

Table 2.8. Loci showing significant variation in segregation patterns as tested by χ^2 test (at $p<0.01$) in BC_5F_2 families (*G. barbadense* background).

BC_5F_2	BNL3267			BNL3790			BNL3903			BNL4029		
Family	HH	HB	BB	HH	HB	BB	HH	HB	BB	HH	HB	BB
10001	-	-	-	-	-	-	29	2	6	-	-	-
10002	-	-	-	-	-	-	-	-	-	43	4	19
10004	-	-	-	-	-	-	-	-	-	-	-	-
10019	30	1	32	-	-	-	-	-	-	-	-	-
10104	-	-	-	30	1	4	-	-	-	-	-	-
10109	-	-	-	-	-	-	-	-	-	-	-	-
10110	-	-	-	-	-	-	-	-	-	20	1	10
10117	-	-	-	-	-	-	-	-	-	-	-	-
10126	-	-	-	-	-	-	-	-	-	-	-	-
10131	-	-	-	-	-	-	-	-	-	35	1	8
10141	-	-	-	-	-	-	-	-	-	-	-	-
10145	-	-	-	-	-	-	-	-	-	-	-	-
10146	-	-	-	38	2	7	-	-	-	-	-	-
10148	-	-	-	-	-	-	27	1	14	-	-	-
10150	-	-	-	-	-	-	-	-	-	-	-	-
10152	-	-	-	54	11	5	-	-	-	-	-	-
10156	-	-	-	-	-	-	-	-	-	-	-	-
10158	14	12	8	-	-	-	-	-	-	-	-	-
10166	-	-	-	-	-	-	-	-	-	-	-	-
10178	16	1	7	-	-	-	-	-	-	-	-	-
10217	-	-	-	-	-	-	7	2	1	-	-	-
10244	-	-	-	-	-	-	-	-	-	-	-	-
BC_5F_2	DPL0637			DPL0652			NAU3207			NAU5180		
Family	HH	HB	BB	HH	HB	BB	HH	HB	BB	HH	HB	BB
10001	-	-	-	-	-	-	-	-	-	-	-	-
10002	25	5	9	-	-	-	-	-	-	-	-	-
10004	12	17	8	-	-	-	-	-	-	-	-	-
10019	-	-	-	-	-	-	-	-	-	-	-	-
10104	-	-	-	-	-	-	-	-	-	-	-	-
10109	-	-	-	36	8	2	-	-	-	-	-	-
10110	-	-	-	-	-	-	-	-	-	-	-	-
10117	7	17	3	-	-	-	-	-	-	-	-	-
10126	-	-	-	34	6	6	-	-	-	-	-	-
10131	-	-	-	-	-	-	-	-	-	-	-	-
10141	-	-	-	-	-	-	-	-	-	37	13	6

10145	-	-	-	-	-	-	14	2	4	-	-	-
10146	-	-	-	-	-	-	-	-	-	-	-	-
10148	-	-	-	-	-	-	-	-	-	39	2	1
10150	-	-	-	-	-	-	33	6	4	-	-	-
10152	-	-	-	-	-	-	-	-	-	-	-	-
10156	-	-	-	37	4	6	-	-	-	-	-	-
10158	-	-	-	-	-	-	-	-	-	-	-	-
10166	-	-	-	-	-	-	-	-	-	36	20	19
10178	-	-	-	-	-	-	-	-	-	-	-	-
10217	-	-	-	-	-	-	-	-	-	-	-	-
10244	-	-	-	-	-	-	18	11	14	-	-	-

Table 2.9. Comparison of SDRs with previously identified *G. barbadense* introgressions in *G. hirsutum* background. * Indicates the introgression and their locations identified by (Kumar et al., 2019). Physical locations for the SSR markers used in the study were obtained by BLASTing sequences of the markers to the reference genome.

Chr	SDR ID	Location (Mb)	Introgression ID*	Location (Mb)*
5	SDRGh05.1	36.57 - 72.80	5.4	27.92 - 32.93
			5.5	80.78 - 86.88
9	SDRGh09.1	2.93 - 13.98	9.1	2.84 - 16.49
15	SDRGh15.1	27.28 - 48.82	15.1	35.57 - 44.81
16	SDRGh16.1	6.81 - 43.48	16.1	10.83 - 55.33
18	SDRGh18.1	1.91 - 22.23	18.1	2.91 - 51.88
23	SDRGh23.1	22.33 - 41.64	23.1	28.92 - 46.87
25	SDRGh25.1	1.81 - 5.16	25.1	0.58 - 1.97

Table 2.10. GO enriched terms in SDRs

SDR ID	Accession	Name	GO Type	Q-value
SDRGh02.1	GO:0051015	actin filament binding	molecular	1.90E-07
	GO:0007015	actin filament organization	biological	3.30E-07
	GO:0003840	gamma-glutamyltransferase activity	molecular	1.80E-06
	GO:0006749	glutathione metabolic process	biological	1.40E-05
SDRGh05.1	GO:0050826	response to freezing	biological	2.50E-07
SDRGh20.1	GO:0009607	response to biotic stimulus	biological	1.90E-09
	GO:0006952	defense response	biological	8.90E-07
SDRGh23.1	GO:0080019	fatty-acyl-CoA reductase activity	molecular	1.90E-05
SDRGh24.1	GO:0045156	electron transporter	molecular	2.20E-06
	GO:0009772	electron transport in photosystem II	biological	5.70E-05
	GO:0019684	photosynthesis, light reaction	biological	8.40E-05
SDRGb06.1	GO:0050826	response to freezing	biological	1.10E-05
SDRGb10.1	GO:0072488	ammonium transmembrane transport	biological	1.10E-05
	GO:0015696	ammonium transport	biological	1.10E-05
	GO:0008519	ammonium transmembrane transport	molecular	1.10E-05
SDRGb11.1	GO:0006629	lipid metabolic process	biological	2.70E-06

CHAPTER 3

GENETIC ANALYSIS OF FLOWERING HABIT IN A RECIPROCAL SET OF INTERSPECIFIC NEAR-ISOGENIC LINES OF CULTIVATED COTTON (GOSSYPIUM SPP)

Jeevan Adhikari, Daniel Petti, Changying Li, and Andrew H Paterson. To be submitted to *G3: Genes | Genomes | Genetics*.

Abstract

Flowering response in cotton is an important phenological characteristic which not only determines the adjustment of a cotton species to a particular geographical region, but also determines the transition of cotton from its vegetative growth phase to reproductive maturity. The quality and quantity of the final product, the cotton fiber, is heavily dependent on the flowering behavior of cotton varieties. Due to the complex nature of flowering habit and its importance on the economic product, investigation into the underlying molecular mechanisms and identification of genetic signatures of flowering habit requires suitable genetic mapping population. Here we investigate the flowering response in cotton in a reciprocal set of near isogenic lines (NILs) and characterize the genomic locations controlling five flowering related traits (flowering start time, flowering end time, flowering duration, peak flowering time and rate of change of flower count during peak flower production period). Flower counts for each genotype were collected using state-of-art high-throughput unmanned aerial vehicle (UAV) phenotyping system and efficient image processing machine learning algorithm. The distribution of these five flowering related traits mostly aligns with the recurrent parent consistent with the genomic structure of the NIL populations. Genetic analysis of these lines revealed a total of 29 flowering related QTLs in the reciprocal backgrounds. We identified two flowering related QTL hotspots, one each on chromosome 2 and chromosome 3. *In silico* analysis of tightly linked regions around the most significantly associated marker revealed potential candidate genes related to flowering response in cotton, some of which have previously been verified to have expression in floral organs during flowering in Arabidopsis. Further investigation of these QTL hotspots and functional analysis of candidate genes identified in this study would help decipher the genetic architecture of flowering response in cotton.

Introduction

Understanding the molecular mechanisms underlying the adaptability of woody plants to local environment has been a long-standing goal plant biology. The switch from vegetative to reproductive growth in plants is called floral transition and the timing of floral transition determines the reproductive success and ultimate fitness for the adaptation of plants to their local environment (Huijser & Schmid, 2011; Hutchinson, 1951). Genetic control of floral transition and flowering patterns in trees (perennial woody plants) is complex and has yet to be adequately investigated due to their long vegetative growth period and difficulties in genetic modification. The molecular genetic networks controlling floral transition in model annual and herbaceous crops like Arabidopsis and rice has been extensively investigated (Fankhauser & Chory, 1997; Leal Valentim et al., 2015; Wang et al., 2014). However, in perennial woody trees, the regulation of flowering time has just begun to be elucidated (Grover et al., 2015), but still face a long-life cycle of these plants and the extravagant waiting period for the first floral initiation as the major challenges in investigating the underlying molecular mechanisms. Cotton (*Gossypium spp.*) can be used a model plant to decipher the molecular mechanism underlying these biological phenomena since, albeit being a woody perennial by nature, cotton is grown as an annual crop in most parts of the world.

Cotton is one of the most important fiber crops in the world. As in other sexually reproducing angiosperms, flowering time control in cotton is critically important in both ecological and agronomic settings (Grover et al., 2015). The primary product of value, the cotton fiber, is a result of several molecular and physiological processes occurring in the cotton plant from initiation of flowering to maturation and opening of the cotton bolls. The architecture of the cotton plant,

including fruit branch formation and flowering pattern, is the most important characteristic that directly influences light exploitation, crop productivity and management. In cotton, the formation of floral bud indicates the beginning of reproductive growth based on prior vegetative growth and most cultivated cotton varieties produce lateral branch from the leaf axils, then the lateral branch develops and differentiates into vegetative branches and fruit branches. Initiation of flowering and fruit branching are major events in the development of cotton architecture and are among the most important productivity-related traits considered in breeding programs and for cultivating farmers.

Flowering habit in cotton, defined by the initiation of flowering, its duration, the peak flowering time, and the cessation of flowering, differs greatly by species. Although, most cultivated cotton cultivars are day-neutral, species-specific differences in flowering habit are still observable and these differences make a particular species acclimated to specific environmental regimes or in some cases to special ecological niches around the world. In their native habitat, most primitive cottons flower and set fruit during the winter months and they accumulate vegetative masses during summer months which coincides with dry winter and wet summer. Photoperiod, however, is not the sole determinant of flowering response in cotton (Kohel et al., 1974). Other factors like temperature, water balance and maturity are important factors in induction of flowering in cotton (Hutchinson, 1951).

Compared with the number of studies evaluating fiber yield and quality in cotton, very few systematic studies have been conducted to identify genetic signatures for flowering habit in cotton. Albeit initial conventional genetic analysis suggested multigenic inheritance of flowering response in *G. hirsutum*, the mode of inheritance and the number of genes involved depended on the materials used (Kohel et al., 1974; Lewis & Richmond, 1960; Waddle et al., 1961). Other comprehensive studies investigating flower initiation in plants revealed complex nature of the trait

(Komeda, 2004; Lin et al., 1995; Yano et al., 1997). Considering the complex nature of flowering response in cotton and limited knowledge on the flowering mechanism of cotton at the molecular level, identification of genomic locations / QTLs governing these traits would provide an insight to this vital biological phenomenon in cotton.

The two most widely cultivated species of cotton, *G. hirsutum* and *G. barbadense*, differ greatly not only in their habitats and growth habits, but also in flowering response and fiber yield and quality. Among the two mostly cultivated species, *G. hirsutum*, in general, is an early flowering type while *G. barbadense* is a late flowering and late maturing type. They have observable differences in flowering habits and because of these differences, breeding practices often requires planting / seed sowing of these species at different times to perform interspecific crosses and other cultivation practices. On the other hand, these differences make these two species valuable resources to study molecular mechanism underlying flowering, in addition to several other traits, that these species differ in.

Utilizing diverse marker technologies, a few reports have been previously published on QTL and/or gene mapping of flowering time response in cotton. More than 30 candidate genes that are involved in various flowering processes in Upland cotton have been identified (Lai et al., 2011). Zhu and Kuraparthy (2014) were able to localize the photoperiod-sensitive locus Gb Ppd1 and associated SSR markers on the cotton chromosome 25. Guo et al. (2008) presented the mapping of flowering-time QTL, assessed by node of first fruiting branches in cotton. C. Li et al. (2013) identified 60 loci associated with early maturing traits in cotton. Kushanov et al. (2017) identified 6 QTLs that were directly associated with flowering time and photoperiodic flowering in an F_2 population and 7 QTLs in the F_3 generation using 212 SSR and 3 cleaved amplified polymorphic sequence (CAPS) markers, Li et al. (2020) reported a cotton genome variation map

generated by re-sequencing of 436 cotton accessions. Whole-genome scans for sweep regions identified 357 putative selection sweeps covering 112 Mb of the upland cotton genome, containing 5,184 genes. These genes were functionally associated with flowering-time control, hormone catabolism, aging, and defense response adaptations to climate changes (Kushanov et al., 2021).

Flowering response in cotton is extremely important as it determines the timing of floral initiation and consequently affects boll formation, boll growth and maturation leading to the overall quantity and quality of fiber produced. Thus, investigation and identification of genomic regions governing these traits is necessitated by their direct association with fiber yield and quality as well as to the adaptation of cotton to various ecological habitats. Despite the economic significance of genomic regions governing these traits, they have not been a frequent subject of molecular investigation. One reason behind the paucity of investigations into these traits might be the complex nature of these traits and the need of specialized populations with suitable genomic composition needed to crack the genetic code controlling these traits. Most early generation populations with large linkage blocks do not serve the purpose of fine tuning the genomic location governing flowering response in cotton.

Flowering response in cotton is complex, not only in terms of molecular mechanisms underlying it, but also in terms of quantifying its related traits and their extraction and data mining. Flowering in cotton is spread over several months and particularly in crosses involving diverse parents, the non-synchronous behavior of flowering pattern makes collecting flowering related traits in cotton extensively labor demanding and prone to human errors. One of the major problems faced by breeders and cotton genomic scientists in this regard is the non-correspondence in defining the traits and their collection/extraction procedure. For example, some published reports have used days to 50% flowering as the measure of flowering start time (Adhikari, 2015b; Grover

et al., 2015; Kushanov et al., 2021) while others have used nodes of first fruiting branch to define same set of traits (Guo et al., 2008; C. Li et al., 2013; Jingjing Zhang et al., 2021). Others have used days to 50% boll opening to study earliness while others have used days from flowering to boll opening as the measure of flowering duration. All these ambiguity in defining and extracting flowering related traits come from, in addition to several other factors, the ease with which researchers want to collect such data.

With the advent of high throughput phenotyping systems, such ambiguity and irregularities in extracting flowering related data can be minimized. Using high throughput phenotyping systems, we can easily record flowering data over the entire growing season of the crop easily and more accurately than manual recording. Ground based systems have been used to collect flowering related data from cotton (Jiang et al., 2016; Jiang et al., 2017; Jiang et al., 2020; Sun et al., 2020; Xu & Li, 2022), but cotton canopies growing into each other at later periods in the growing season makes this system less efficient. UAV systems offer greater advantage over ground-based systems in that they can be used at any stage of crop growth, and they also cover more ground within a smaller period of time than ground-based systems. Albeit UAV systems have a slight disadvantage of not capturing flowers embedded deep inside the crop canopy thus curtailing the extraction of exact flower counts, the data collected from this system is sufficient enough to get the pattern/trend on flowering behavior in cotton and to investigate how genotypes differ in terms of flowering imitation, duration, cessation, and peak flowering times.

In this study, we utilize an advanced set of reciprocal NILs constructed from a two-way interspecific cross between *G. hirsutum* acc Acala Maxxa and *G. barbadense* acc Pima S6 to dissect the genetic control of flowering behavior in cotton. These reciprocal sets of NILs containing only one introgressed region from the donor parent in the genome of the recurrent

parent serve as important genetic resources in deciphering flowering behavior in cotton due to their genomic structure and presence of small donor chromosomal segments. Here, we utilized advanced high throughput UAV systems to collect flower data over the entire growing season. Using state-of-art machine learning algorithms, we extracted and studied the flowering patterns in the two reciprocal populations in terms of earliness and lateness in flowering, duration of flowering, peak flowering time and rate of change of flower production during the inflection period, defined as the period from initiation of flowering to the peak flowering time.

Materials And Methods

Population development

Plant materials used in this study were developed from a set of reciprocal crosses between *Gossypium hirsutum* acc. Acala Maxxa and *G. barbadense* acc. Pima S6 (both inbred lines). These genotypes have been extensively used to produce several molecular tools and resources including BAC libraries and Illumina genome sequences. Reciprocal advanced backcross populations were developed by first crossing the parents in a two-way cross (Acala Maxxa (♀)× Pima S6 (♂) – hereafter referred to as *G. hirsutum* background; and Pima S6 (♀) × Acala Maxxa (♂) – hereafter referred to as *G. barbadense* background), then independently backcrossing F_1 plants to the respective female parent to create 300 to 400 BC_1 progenies for each cross. The backcrossing scheme included planting only one seed from each preceding backcross to generate the next generation (Figure 3.1).

After five generation of backcrossing, 179 BC_5F_1 plants from the *G. hirsutum* background and 190 BC_5F_1 plants from the *G. barbadense* were self-pollinated and a total of 8364 BC_5F_2 plants (2-32 individuals in each BC_5F_2 family) were grown at Iron Horse Farm,

Watkinsville, Georgia in 2019 under cultural conditions consistent with commercial irrigated cotton production. Individual BC_5F_2 plants that contained only one introgressed segment from the donor parent were deemed as NILs. We identified a total of 397 NILs in the Acala Maxxa background and a total of 423 NILs in the Pima S6 background. These NILs were grown at Iron Horse Farm, Watkinsville, Georgia in 2021 under cultural conditions consistent with commercial irrigated cotton production.

Genotyping

The genomic composition of the BC_5F_1 plants were inferred based on genotyped by sequencing (GBS). DNA was extracted from the two parents and 369 BC_5F_1 plants using a scaled-down version of the CTAB protocol described by (Paterson et al., 1993). A total of four multiplexed GBS libraries were constructed according to P. Andolfatto et al. (2011) wherein the DNA were double digested with HinP1I-HaeIII enzymes. The libraries were sequenced on Illumina MiSeq (in-house) with 75 bp single end reads (SE75). The TASSEL5 GBSV2 pipeline was used for sequence data processing and genotype calling (Glaubitz et al., 2014). Reads were aligned to *G. hirsutum acc TM-1* (Zhang et al., 2015) using Burrow-Wheeler Alignment (bwa) and exported to variant call format (VCF). To minimize sequencing errors, only the first 64 base pairs were used to map reads to the reference genome. Filtering of the VCF was done for bi-allelic SNPs using the Fisher exact test with a threshold p-value < 0.001 considering that true variants should represent biallelic homozygous state for inbred accessions. Genotypes for lines in *G. hirsutum* background were called together and those for lines in *G. barbadense* background cross were also called together. The SNPs were filtered for MAF > 0.01, missing < 30% and heterozygous < 10%

at the population level. The retained SNPs were imputed using Fast Inbred Line Library Imputation (FILLIN) pipeline available in TASSEL5 GBSv2 (Kelly et al., 2014).

The genomic composition of BC_5F_2 plants were inferred based on targeted microsatellite (SSR) genotyping of the introgressed chromosomal segments identified in their respective BC_5F_1 parents. At least two (and at most four) SSR markers were used to verify most of the introgressed regions while for small introgressions only one available SSR marker was deployed. A total of 852 polymorphic SSR markers spanning the introgressed regions were derived from several published genetic maps in crosses between *G. barbadense* and *G. hirsutum* stored in CottonGen SSR database (https://www.cottongen.org/data/download/marker). 47 of them were found to be monomorphic in pilot studies and were discarded. Another 23 with ambiguous bands were also discarded and a total of 782 SSR markers were used to genotype the BC_5F_2 progenies. Among the 8364 BC_5F_2 individuals planted in 2019, a total of 5315 (with corresponding BC_5F_1 parents carrying 2 to 5 introgressions) plants were SSR genotyped for the presence (or absence) and nature (homozygous vs heterozygous) of the respective introgression/s that their BC_5F_1 parent carried. Individual BC_5F_2 plants that were verified by SSR markers to carry only one introgression (homozygous or heterozygous) from the donor parent were deemed as NILs.

Phenotypic Evaluation

A total of 397 NILs in the *G. hirsutum* (Acala Maxxa) background and a total of 423 NILs in the *G. barbadense* (Pima S6) background were planted at the Iron Horse Farm, Watkinsville, Georgia in 2021 under cultural conditions consistent with commercial irrigated cotton production. A total of two replications of each genotype was planted in completely randomized design. Flower counts for each plot was recorded using UAV system (described below) during the reproductive

growth stage of the cotton plants. A total of 23 evenly spaced data collection sessions was conducted throughout the flowering period and the data from these 23 sessions was utilized to extract five flowering behavior specific phenotypes as follows:

A. Flowering start date: Defined as the start of flowering and estimated by the identification of one or more first flowers by the UAV system, measured as days after planting (DAP).
B. Flowering end time: Defined as the cessation of flowering and estimated by no or reduced flower counts at the end of the season, measured as DAP.
C. Flowering peak time: Defined as the floral flush time when each genotype produced the highest number of flowers and estimated by the session with most flower counts, measured as DAP.
D. Flowering duration: Defined as the difference between the flowering end date and the flowering start data, measured as days.
E. Inflection period slope: Defined as the rate of change of flowers produced each day by the genotypes during the inflection period, measured as flowers/day.

Data Collection and Image Processing

The field was photographed repeatedly using a small UAV for the duration of the flowering period. Individual plot images were then extracted, and a coevolutionary neural network (CNN) was applied in order to generate per-plot flower counts. Data were collected at a bi-weekly cadence from 2021-08-09 through 2021-11-05. Some sessions were skipped due to inclement weather, resulting in a total of 23 sessions. By the end of the data collection process, many of the plants had open bolls present. Data collection was halted after the first overnight freeze. A total of six ground

control points were distributed throughout the field, with their exact positions measured using an RTK GPS.

Images of the field were collected using a Matrice 100 drone (DJI, Shenzhen, China) fitted with a custom mount and equipped with a Lumix G7 camera (Panasonic Corporation of North America, Newark, N.J., USA) and a 17 mm lens. The drone was flown at a height of 15 meters, resulting in a GSD of 0.23 cm/pixel. In a few cases, technical issues with the Matrice 100 data required the substitution of equivalent data from a DJI Phantom 4 Pro v2 drone.

Orthomosaic images were generated from each flight using Agisoft Metashape (Agisoft LLC, St. Petersburg, Russia). Extraction of individual plot images was performed manually using QGIS (QGIS Development Team, 2009). For each session, 2,150 individual plot images were extracted, resulting in a total of 41,696 images for the entire season. The analysis process was automated using Kedro and required several hours to produce flower count estimates for every plot when running on an Nvidia 1080 Ti GPU.

Flower counting was performed using a YOLOv5-small detector trained on a custom dataset of flower images. (Ghosal et al., 2019) proposed a pseudo-active-learning approach to training detector models, which was adopted here in order to expand the dataset. The model was first bootstrapped on the "UGA2020V" dataset from (Tan et al., 2021), which consists of images of flowering cotton plants collected using a tractor-mounted camera. Afterwards, ten of the original, uncropped images from the aforementioned drone data were selected at random. The model was applied to these images in order to produce initial flower detections, which were then corrected by humans using the CVAT tool. Finally, the model was re-trained with the new annotations, and the process was repeated with ten new random images, until the model

performance saturated. In this way, a large, annotated dataset was constructed with minimal human effort.

Data analysis

Phenotypes for the two reciprocal populations were analyzed in R programming language. Single marker analyses were done in R/qtl (Broman & Sen, 2009). The significance threshold was set to LOD of 3, to mitigate the multiple-comparison problem. Filtration of significant markers adopts the method proposed by Szalma et al. (2007b). If several markers on the same introgressed segment show significant association with phenotype, the most significant one was reported. For the co-segregation of multiple introgressions, the QTL location is examined as follows. First, if multiple families show significance for the trait and carry overlapping introgression, the introgression is considered to carry QTL. Second, if the co-segregation of introgressions is in single families, the most significant introgression is considered to carry QTL.

Phenotypic variance explained by each locus was reported by taking the most significant marker as independent variable and phenotypic value as dependent variable in R (R Core Team 2016). Additive effects were estimated by half the difference of phenotypic values between the lines carrying the homozygous introgression and lines not carrying the introgression. Dominance effects were estimated by the difference of phenotypic values between the lines carrying the heterozygous introgression and the remaining lines that do not carry the introgression. If multiple or overlapping introgressions were present at both homozygous and heterozygous state, the estimation of additive effects utilized the lines carrying the introgression at homozygous state only and, the estimation of dominance utilized the lines carrying the introgression at heterozygous state only.

Gene actions for the QTLs were determined by calculating the degree of dominance (absolute values) for each QTL. The degree of dominance is the ratio of dominance effect to additive effect (d/a) of the QTL and based on this ratio, gene action of the QTLs can be categorized as (i) additive (0 <d/a < 0.2) (ii) partially dominant (0.2 < d/a <0.8) (iii) dominant (0.8 < d/a < 1.2 and (iv) over-dominant (d/a > 1.2). QTLs with dominant and over-dominant effects are considered to have heterotic effect or heterozygous advantage. To evaluate non-random correspondence of QTLs detected in our study with those reported in prior studies, we used a method of aligning genetic maps and QTLs that uses the hypergeometric probability function (Feltus et al., 2006). QTLs were assumed to be orthologous (co-occurring) if they explained a significant proportion of variation for a directly comparable trait measured and the 1-LOD confidence intervals overlapped. The hypergeometric probability function equation is as follows:

$$p = \frac{\binom{l}{m}\binom{n-1}{s-m}}{\binom{n}{s}}$$

where, *p* is the probability of non-random correspondence of QTLs being compared for a given trait; *n* is the number of intervals which can be compared (each interval is usually defined as approximating a QTL likelihood interval i.e., 30 cM) along the entire genome; *m* is the number of matches declared between QTLs; *l* is the total number of QTLs identified in the larger sample and *s* is the number of QTLs identified in the smaller sample.

Candidate gene identification

In silico annotation was performed on the identified QTLs to look for candidate genes related to flowering habit in cotton. For each QTL identified in the study, the genomic region spanning 50 kb on each side of the most significantly associated marker was used for in silico analysis. The DNA sequence from this tightly linked region was used to look for *G. hirsutum* genes

in the CottonGen database and these genes were then analyzed for biological functions, with particular focus on flowering related functions and expression on floral organs.

Results

Genomic structure of the NILs

The NIL population in the Acala Maxxa (*G. hirsutum*) background consisted of 397 individuals consisting of only one chromosomal segment introgressed from the donor parent (*G. barbadense* 'Pima S6'). In total, these lines covered 78.72% of the donor genome (Table 3.1, Figure 3.2). Single introgressed segments were identified for all chromosomes except chromosome 24. Apart from chromosome 24, which lacked any introgressions and chromosome 6, which contained only 1.3% of donor chromatin, the chromosome-wise donor coverage ranged from 39.17% to 97.14% (Table 3.1). Individual lines contained an average of 0.904% of the donor genome, ranging from 0.07% (1.37 Mb) to 2.92% (56.65 Mb).

In the Pima S6 (*G. barbadense*) background, a total of 423 NILs were identified. In total these lines covered 71.48% of the donor (*G. hirsutum* 'Acala Maxxa') genome (Table 3.1, Figure 3.3). Single introgressed segment was identified for all chromosomes except chromosome 25. Apart from chromosome 24, which lacked any introgressions, the chromosome-wise donor coverage ranged from 16.71 to 94.58% (Table 3.1). Individual lines contained an average of 0.97% of the donor genome, ranging from 0.15% to 3.07%.

Sub-genomic distribution of the introgressed segments

In the *G. hirsutum* background, introgression contained in the A (hereafter At) -subgenome were identified in 269 lines while those contained in the D (hereafter Dt) -subgenome were

identified in 128 lines. These introgressions covered 89.54% of the At subgenome while they only covered 74.67% of the Dt subgenome (Table 3.1). In the *G. barbadense* background, introgressions contained in the At subgenome were identified in 332 lines while those contained in the Dt subgenome were identified in 91 lines. These introgressions covered 83.25% of the At subgenome while they only covered 59.72% of the Dt subgenome (Table 3.1).

Phenotypic performance of the NIL populations

The cumulative flower counts over the 23 sessions of data collection for all the NILs are presented in Figure 3.4(a). Individual lines and population mean for both backgrounds follow the expected sigmoidal pattern for cumulative flower counts. Curves for overall population mean as well for individual lines within the respective population show that individual Acala Maxxa NILs produce flowers earlier than that in the Pima S6 NILs. The results are concurrent with the flowering habit of the parents. Over the course of the growing season, as plants reach their peak growth and cessation of flowering, the cumulative counts of Pima S6 NILs (and the population average) exceeds that of Acala Maxxa NILs, indicating slow start but extended production of flowers by Pima S6 NILs. The results are concurrent with the flowering habit of Pima S6, which flowers late and keeps flowering late in the season when Acala Maxxa has already reached its maturity and has stopped flowering.

An interesting pattern seen in the figure is that the spread of the total flower count, as seen towards of the end of the season, is larger in Pima S6 population than in the Acala Maxxa population. This might be due to different number of plants in each plot with plots having more plants showing higher total counts and those with fewer plants showing lower total counts. In general, the curves indicate early flowering followed by rapid increase in the number of flowers

and early cessation of flowering in Acala Maxxa NILs while a slow and gradual but extended period of flowering in Pima S6 lines resulting in higher total counts in Pima S6 lines. In all, the NILs in each population and the population average shows the trends that would be expected from the respective parents.

To further investigate the flowering response in these populations, we utilized the flower counts taken over the growing season and extracted phenotypes that quantified components of their flowering habits. These phenotypes include flowering start time, peak flowering time, flowering end time, duration of flowering and change in flowers per day (slope) during the inflection period defined as the duration from initiation of flowering to the peak flowering time of the lines and indicated by the exponential growth phase of the sigmoid curve. Figure 3.4(b) shows the distribution of flowering start time in the two reciprocal populations. In accord with the pattern seen in Figure 3.4(a), Acala Maxxa NILs show significantly earlier flowering than Pima S6 NILs (Figure 3.5).

Figure 3.4(c) shows the distribution of flowering end date in the two reciprocal populations. Flowering end dates were significantly different between the two populations (Figure 3.5). Two factors might have played a role in this behavior. First, freezing temperatures on the second and third weeks of November might have stopped flowering. Second, owing to the reduced flowers produced by the plants following the low temperature weeks, we stopped scanning the field after December 12, 2021. Nonetheless, the data we collected so far provided valuable information in observing the spread of flowering end times in these populations in the 2-3 weeks span before we stopped collecting data. Figure 5d shows the distribution of peak flowering time in the reciprocal populations. As would be expected from the flowering trend graph and the flowering habit of the

parents, Acala Maxxa NILs had significantly earlier peak flowering than Pima S6 NILs (Figure 3.5).

The average flowering duration in Acala Maxxa NILs was significantly shorter than that of Pima S6 NILs (Figure 3.4(e)) and the rate of increase in flowers per day was higher in Pima S6 NILs (Figure 3.4(f)). The population averages for both these traits were significantly different between the two reciprocal populations (Figure 3.5). The results presented in Figure 3.4(b-f) sums up the flowering pattern shown in Figure 3.4(a). With shorter flowering duration and with fewer flowers produced per day, the total flower counts in Acala Maxxa population lag Pima S6 population. On the other hand, longer growing season and a greater number of flowers production per day boosts the total number of flowers in Pima S6 population albeit they start flowering late in the season.

Marker-trait association and overview of identified QTLs

Marker-trait associations were carried out for all five traits related to flowering habit in the reciprocal set of NILs and QTL were identified following the procedure described in the Materials and Methods section. A total of 29 QTLs were identified. QTLs for all five traits were identified in both backgrounds. Most QTLs were identified for flowering end date in both backgrounds while the least number of QTLs (one) were identified for peak flowering date in the Acala Maxxa background and for flowering duration in the Pima S6 background. The phenotypic variance explained by these QTLs ranged from 1.37 to 18.22% (Table 3.2 and Table 3.3). Only 4 (13.33%) of the 30 QTLs identified were ‘major’ QTLs, i.e., explaining more than 10% of the total phenotypic variance. Among the 29 QTLs identified for flowering related traits in the reciprocal set of NILs, 21 were located in the At subgenome and 8 in the Dt subgenome (Table 3.4).

Chromosomes 2 and 3 harbored three QTLs each while chromosomes 6, 8, 9, 11, 12 and 17 harbored two QTLs each.

QTLs in the Acala Maxxa background

In the Acala Maxxa background, a total of 16 QTLs were identified for the five flowering related traits (Table 3.2). The phenotypic variation explained by these QTLs ranged from 1.37% to 11.66%. Two QTLs were identified for flowering start date, one each on chromosome 8 (qSDGh.08.01) and chromosome 18 (qSDGh.18.01). The marker most significantly associated with these QTLs is shown in Table 3.2. These QTLs explained 4.17 and 4.68% of the total phenotypic variation respectively. Both these QTLs have a total effect of increasing the flowering start time. Of these two QTLs, qSDGh.08.01was over-dominant while qSDGh.18.01was additive.

A total of 6 small-effect QTLs were identified for flowering end date in the Acala Maxxa background explaining 2.72 to 5.65% of the total phenotypic variation (Table 3.2). One QTL were identified each on chromosomes 9, 12, 17, 21 and 23. Except for two QTLs (qEDGh.02.01 and qEDGh.23.01) all other QTLs have a total effect of increasing the flowering end date. Among the six QTLs, one was over-dominant, one was dominant, three were partially dominant and one was additive. One QTL (qPDGh.10.01) was identified for peak flowering date in the Acala Maxxa background on chromosome 10 (Table 3.2). This QTL explained 3.39% of the total phenotypic variation and had a total additive effect of increasing the peak flowering date by about four days.

Two QTLs (qDUGh.02.01 and qDUGh.17.01) were identified for flowering duration in the Acala Maxxa background on chromosomes 2 and 17. These QTLs explained 6.35 and 3.48% of the total phenotypic variation (Table 3.2). Both these QTLs increased the total duration of flowering in the Acala Maxxa background. One of these QTLs was additive while the other was

over-dominant. A total of 5 QTLs were identified for slope (rate of change of flower count per day during the inflection period) on chromosomes 1, 2, 11, 12 and 22. These QTLs explained 1.37 to 11.66% of the total phenotypic variation (Table 3.2). One QTL (qSLGh.12.01) was a major QTL explaining 11.66% of the total phenotypic variation and it had a total effect of increasing the number of flowers. The remaining four small-effect QTLs also increased the number of flowers during the inflection period. Four of these five QTLs had an additive effect while one was over-dominant.

QTLs in the Pima S6 background

A total of 13 QTLs were identified in the Pima S6 background. The phenotypic variation explained by these QTLs ranged from 3.61 to 18.22% (Table 3.3). Two QTLs were identified for flowering start time in the Pima S6 background on chromosome 3 (qSDGb.03.01) and on chromosome 9 (qSDGb.09.01). These QTLs explained 6.34 and 8.57% of the phenotypic variation respectively. Both QTLs had additive gene action and both QTLs shortened the flowering start time. Four QTLs were identified for flowering end time in the Pima S6 background explaining 4.87 to 18.22% of the phenotypic variation. Three of these QTLs had additive effect while the fourth QTL was partially dominant with an overall effect of shortening the end date of flowering in Acala Maxxa background by about 11 days. All four QTLs shortened the flower end time in the Pima S6 background and the QTL qEDGb.19.01, in particular, had greater influence on shortening the flowering end time by about 19 days.

A total of four QTLs were identified for peak flowering date in the Pima S6 background explaining 3.93 to 13.55% of the phenotypic variation (Table 3.3). Among these four QTLs, one was additive, two were partially dominant and one was over-dominant. One QTL (qEDGb.04.01)

had a total effect of increasing the peak flowering data while the remaining three QTLs, including the major effect QTL (qEDGb.11.01), decreased the peak flowering time in the Pima S6 background. One partially dominant QTL (qDUGb.06.601) explaining 3.61% of the phenotypic variation was identified for flowering duration and this QTL had a total effect of decreasing the flowering duration in the Pima S6 background. Two QTLs, both with additive gene action, were identified for slope, one on chromosome 3 and chromosome 5 each, explaining 5.79% and 3.99% of the total phenotypic variation. These QTLs had a total effect of increasing the number of flowers produced per day.

Co-occurrence of QTLs in multiple NILs and NIL sub-families

The estimated effect of the identified QTL is more stable and statistically powerful if the introgressed region carrying the QTL of interest is present in multiple NILs and/or NIL sub-families. Among the 29 QTLs identified in the study, 27 were identified in multiple NILs and 22 were identified in multiple NIL sub-families (Table 3.5). A total of 11 QTLs were present in 10 or more NILs and nine QTLs were present in introgressions carried by 3 or more families, providing some validation of the effects of these QTLs.

Discussion

Most cotton in their natural habitat respond to photoperiod and flower during short days. However, almost all cultivated species of cotton in the US have been bred for day-neutrality. Although this has had significant effects on their flowering morphology including their adaptation to various ecological niches, they still show their distinctive natural variation in flowering habit. The two mostly grown species of cotton, *G. hirsutum* and *G. barbadense*, vary not only in terms

of their growing habits and fiber quantity and quality, but also in their flowering response. The upland cotton, *G. hirsutum* is usually early flowering with short and rapid flowering phase followed by early boll formation and early maturity, while *G. barbadense* is late flowering with extended flower production periods. These differences, although, sometimes troubling for breeders to sync interspecific crosses and get high percentage of successes in those crosses, provide valuable genetic variation to study flowering behavior in cotton.

Here we present a comprehensive study of genomic regions controlling flowering related traits it's a reciprocal set of NILs constructed using Acala Maxxa and Pima S6, two diverse lines representing the two most cultivated species of cotton, and showing significantly different flowering behavior, as parents. Despite cotton being an important economic crop and flowering response in cotton being an important factor determining the final yield and quality of cotton fiber, investigation into the molecular mechanisms driving these traits has been scanty. Flowering control in cotton, like other angiosperms, is a complex quantitative trait and investigation into such traits is facilitated by well-structured mapping populations. The reciprocal set of NILs we developed not only provide a well-suited set of individuals for efficient mapping of these traits but also help investigate the reciprocal effects of the introgressed segments into the recipient genome.

Flowering response in the reciprocal NILs

The two reciprocal NIL populations showed flowering behavior like their recurrent parents (Figure 4). These two populations not only performed like their recurrent parents, but their performance was significantly different for all five flowering related traits investigated in our study (Figure S1). Albeit the average performance of these two populations behaved like their recurrent parents, which is expected given the exceptionally large proportion of their genome coming from

the recurrent parent, a discretely large amount of variation was observable for all five traits (Figure 4). This variation is accounted for by the chromosomal segments introgressed from the donor parents as shown by the phenotypic performance (QTL effect) of the lines carrying the respective introgression.

Effect of species background

Identification of similar numbers of total flowering related QTLs in both species' backgrounds suggest that similar numbers of genes might control flowering traits irrespective of background species. On the other hand, many findings indicate the effect of species background on numbers and phenotypic effects of QTL for individual traits, contributions of subgenomes, and patterns of QTL and clustering.

A perplexing observation is the complete absence of QTLs identified with opposite phenotypic effects at corresponding locations in the reciprocal genetic backgrounds. In principle, one would expect the majority of QTLs to show such reciprocity if alternative alleles at a QTL show additive or dominant-recessive effects. Limited correspondence of identified QTLs in the two backgrounds could be a result of several factors. One factor may be the small phenotypic effects of most of the identified QTL, increasing the likelihood that one or both members of a reciprocal pair elude detection (Broman, 2001). For markers that showed association with traits but did not reach the LOD threshold of 3, there were 5 additional cases of loci within 3Mb of their homeologous location with opposite phenotypic effects in the reciprocal background.

Another important reason for lack of correspondence of QTLs in reciprocal backgrounds could be the unavailability of reciprocally introgressed segments and/or presence of reciprocally introgressed segments in very few individuals in one background leading to little power in

precisely estimating QTL effect or in some cases missing phenotypic data for those handful of lines carrying the respective introgression. For example, a QTL for peak flowering date was identified in chromosome 24 in the Pima S6 background, which lacked introgression in the Acala Maxxa background. Similarly, for a few chromosomal segments that were introgressed reciprocally, missing phenotypes resulted in the loss of reciprocal estimates.

While most QTL effects were consistent with the phenotypic differences between the parental lines, some exceptional results show the potentiality of improving each parental lines by introgression from the other. For example, the QTL qDUGh.01.01 which carried introgression from Pima S6 to the Acala Maxxa on chromosome 1 reduced the duration of flowering. All the QTLs related to slope (change in flowers/day) had positive effects irrespective of the parental line contributing the introgression, suggesting that both parents could potentially deliver chromosomal segments that could reciprocally improve the performance of both parents. Results like these further support the existence of favorable effect alleles in both species whose effects might have been masked or neutralized by unfavorable alleles or allele combinations.

QTL clustering

Colocalization (clustering) of QTL with common allele effects makes marker-assisted selection much more efficient. Clustered QTL for different traits were prevalent and quite variable between the two backgrounds (Table 3.2 and Table 3.3). On chromosome 3, the same marker was most significantly associated with flowering start date, flowering end date and slope. The QTLs associated with this marker had effects that were favorable to a breeder or farmer, reducing the start and end day of flowering and increasing the number of flowers per day. The effects shown by this QTL could potentially improve synchronous flowering between these two diverse species

of cotton. Another QTL cluster was identified on chromosome 2 in the Acala Maxxa background. Three QTLs, each related to flowering end date, flowering duration, and slope, were identified in the same introgressed region. The same marker was most significantly associated with all three traits and had total effect of decreasing the flowering end date and duration while increasing the number of flowers per day. This is an interesting observation because this introgression carrying these three QTLs is the only introgressed segment contributed by Pima S6 to decrease days to cessation of flowering and to decrease flowering duration. The results entail further investigation and verification of this introgression.

Subgenomic distribution of flowering related QTLs

With regard to the distribution of QTL in A and D subgenomes, many prior studies have concluded that more QTL for growing period, yield and fiber quality were distributed in Dt subgenome than in At subgenome (Chandnani et al., 2018; J. Rong et al., 2007; Joseph I. Said et al., 2013; Wang et al., 2013). However, Lin et al. (2005) and Shen et al. (2005) reported that more QTL for fiber quality were located in A subgenome than in D subgenome. In a study of early-maturing traits by Chengqi Li et al. (2013), more QTL for early-maturing traits were stably distributed in Dt subgenome (16) than in At subgenome (12) in Pop I, whereas more QTL were stably distributed in At subgenome (9) than in Dt subgenome (5) in Pop II. In our study, we identified 21 QTLs in the At subgenome and 8 QTLs in the Dt subgenome (Table 3.5). Various factors might affect the distribution of QTLs in At and Dt subgenomes including, but not limited to, species used in developing experimental populations, the number and type of genetic markers used, and the type of trait under investigation. Therefore, whether the genes and genomic regions

controlling flowering response in cotton are more frequently distributed in At or Dt subgenome should be established in subsequent studies.

Similarity with QTLs previously reported

Flowering response in cotton is a very complex trait and it has not been as comprehensively studied as fiber quality or yield traits. An additional level of complexity comes from the definition of flowering related traits that are being extracted and utilized in genetic mapping. Some mapping studies have defined flowering related traits as time of flowering initiation, flowering duration and flowering cessation while many others have early maturity traits as synonymous to early flowering traits (Fu et al., 2019; Guo et al., 2009; Jia et al., 2016; Li et al., 2012; Li et al., 2017; Yufang et al.). A few studies have also used time to 50% flowering and days to 50% boll opening as well as well as percentage of open flowers and open bolls to characterize early maturity in cotton (Adhikari, 2015b; Jia et al., 2016).

Early maturity is one of the desirable traits in cotton and is an important target trait in cotton breeding, especially with modern mechanization demanding cotton with short uniform stature. Early-maturing cotton has a short growth period and generally shows a dwarf and compact growth architecture (Li et al., 2017). Early maturity is a complex agronomic trait involving budding date, flowering initiation, flowering duration, nodes of first fruiting branches (NFFB), height of NFFB and many more (Fu et al., 2019). NFFB, referring to the number of nodes that generate the first fruiting branch, is a trait associated with flowering time and a lower NFFB of a cotton cultivar indicates early maturity (Ray & Richmond, 1966). Genomic regions controlling this trait thus control flowering time and flowering duration in cotton. Some previous studies have identified QTLs related to NFFB, flowering time (FT), the number of days from flowering to boll

opening (FBP), whole growth period (WGP) and height of NFFB (HNFFB) to characterize flowering response in cotton (Fu et al., 2019; Guo et al., 2009; Jia et al., 2016; Li et al., 2012; Li et al., 2017; Yufang et al.). While NFFB, FT and HNFFB are closely related to flowering start time and flowering end time in our study, FGB and WGP are more closely related to flowering duration.

Using an interspecific recombinant inbred line (RIL) population developed from the cross between an early maturing variety CCRI36 and a late maturing accession G2005, Jia et al. (2016) identified a total of 247 flowering related QTLs, 19 of which (10 for days to 50% flowering and 9 for whole growth period) were colocalized in the same chromosomes in which we identified QTLs for flowering start date and flowering duration. In a study using 169 upland cotton backbone cultivars and breeding lines (Fu et al., 2019), genome wide association studies identified two significant marker trait associations for NFFB in chromosome 3, three in chromosome 8, one in chromosome 9 and 1 in chromosome 18. The four QTLs for flowering start date identified in our study are each located in these same chromosomes. C. Li et al. (2013) also identified a suggestive NFFB QTL on chromosome 9 in the study of early maturing traits in upland cotton.

Identification of QTL correspondence for flowering related QTLs in cotton is a daunting task given the variation in trait names and methods used to collect and report these traits. Utilizing the hypergeometric probability method described by Feltus et al. (2006), we looked for correspondence to the flowering related QTLs identified in our study with those previously reported. A total of 499 previously reported QTLs (17 publications from 1998 to 2021, Table 3.6) were used to identify correspondence to the QTLs identified in our study (Table 3.7). To account for the non-uniformity in the type of traits measured to quantify flowering behavior, traits of similar nature were used for correspondence. For example, QTLs for FT, NFFB/FFBN/NFB and

HFFBN, which were used to quantify the initiation of flowering in cotton, were used to correspond "Start Time" QTLs in our study. Likewise, bud period (BP), flower and bud period (FBP) and growth period (GP) QTLs were used to find correspondence to flowering duration QTLs. YPBF (yield percent before frost) QTLs were used as correspondence to flower end time QTLs. No QTLs of corresponding nature were identified for peak flowering time and slope.

A total of five QTLs (related to three traits described in current study) corresponded non-randomly ($p < 0.05$) with at least one of the previously reported QTLs (Table S5). For example, flowering start date QTL reported here showed non-random overlap ($p=0.0003$) with NFFB QTL reported by Fu et al. (2019). Similarly, flowering duration QTL identified in our study corresponded with FBP QTL reported by Su et al. (2016) in a natural population of cotton. Likewise, three non-randomly overlapping QTLs were identified for flowering end date (Fu et al., 2019; Li et al., 2012; Li et al., 2016). These few, but statistically significant, non-randomly correspondent QTLs provide for validation of the QTLs identified in our study as well as for the need of further exploration of these genomic regions to dissect these important agroeconomic traits in cotton.

The four major effect flowering related QTLs identified in this study are on chromosomes 3, 11, 12 and 13. Numerous published studies show that these chromosomes also host QTLs for fiber quality traits (J. Rong et al., 2007; Joseph I. Said et al., 2013; Wang et al., 2013). Guo et al. (2008) also identified QTLs for NFB clustered with fiber quality QTLs and suggested that most NFB QTLs were linked with some QTLs related to fiber quality. In addition, several other fiber quality QTL hotspots previously identified might harbor the small-effect QTLs identified in our study. If there is a positive linkage between flowering related and fiber quality QTLs co-occurring in these clusters, selection of genotypes with favorable flowering behavior would also select fiber

and yield genes/ gene combinations from the donor parent. Simultaneous selection of genotypes that flower early and have short flowering duration with desired fiber quality traits by marker assisted selection may accelerate cotton improvement. This demands further exploration of these immortal lines for relationship between flowering related traits and fiber quality related traits. Investigation into these relationships could be one of our next adventures in further deciphering the genomic organization of these QTL hotspots in the cotton genome.

In silico annotation of potential candidate genes

Availability of a tetraploid reference genome enabled us to scrutinize physical regions surrounding the identified QTLs for genes / gene families known or suspected to affect flowering response in cotton. This investigation was limited to tightly linked regions i.e., 50 kb on both sides of the SNP marker that is most significantly associated with the QTL (Table 3.8). On chromosome 17, the nearest gene of interest was *CCR2* (Gh D03G0508). TAIR (The Arabidopsis Information Resource, https://www.arabidopsis.org/) reports that this gene is expressed in flower, flower pedicel, inflorescence meristem, petal, petiole, sepal and pollen during flowering stage, petal differentiation and expansion stage and plant embryo bilateral stage. This gene has also been identified as one of the important genes involved in flowering time in cotton (Grover et al., 2020). Another gene in the same close vicinity, AT5G16730 (Gh D030511) is also presumed to have similar functions as *CCR2*. On chromosome 2, the nearest gene of interest was *ERF1B* (Gh A02G0377) which is expressed during four leaves visible stage and flowering stage and is responsible for regulation of transcription and ethylene-activated signaling pathway (mostly defense related) during flowering stages and during fruit development and ripening in woody plants (El-Sharkawy et al., 2009; Ninh et al., 2021; Jingxia Zhang et al., 2021). Another gene *PRP*

(Gh A02G0378 and Gh A02G0379) in close proximity to *ERF1B* has functions and expression patterns similar to *ERF1B.* This gene (*PRP)* also has been found to have important roles in restructuring cell wall during cotton fiber elongation (Feng et al., 2004). Similarly, GATA15 (Gh A121163) was identified as the nearest gene of interest on chromosome 12. This gene is responsible for developmental growth and response to temperature stimulus and is shown to be expressed during flowering and petal differentiation stages in flower, flower pedicel and inflorescence meristem.

On chromosome 11, *At5g39250* was the nearest gene to peak flowering time QTL qPDGb.11.01 in the Pima S6 background. This gene belongs to F-box protein family and is involved in biological processes during flowering stage and petal differentiation stages in flower and flower pedicel, inflorescence meristem and petals. *GhHB12*, a HD ZIP I transcription factor, is located very close (Chr 11: 10190958-10192498) to this QTL and has been found to regulate flowering time and plant architecture via the GhmiR157-GhSPL pathway (He et al., 2018). An important flowering related gene, *MYB108*, was identified close to peak flowering duration QTL, qPDGb.24.01, on chromosome 24. This gene encodes a *MYB* transcription factor involved in anther dehiscence as well as regulating cell death in response to fungi (Cheng et al., 2016). This gene has been found to be expressed in flower, flower pedicel, petal and pollens during flowering and petal differentiation stages in woody plants (Zhang et al., 2018) and in cotton, it has been found to be specifically expressed in pollen and involved in late anther/pollen development (Y. Li et al., 2013). The genes identified in tightly linked regions flanking the marker most significantly associated with the identified QTLs could potentially affect flowering response in cotton. Some of these genes have been extensively studied in Arabidopsis with their flowering related functions experimented and verified, while the functions of others are still unknown. While all these genes

are potential candidates for flowering response genes in cotton, further investigation is necessary to verify their role in floral initiation and flowering habit in cotton.

Conclusion

This study investigates into the flowering behavior of cotton in a reciprocal set of NILs using state-of-art high throughput phenotyping system. Flowering habit in cotton is complex, both in terms of its genetic architecture, and in terms of the process of collection of phenotypic data related to flowering. In one hand, the genomic composition of the reciprocal NILs provide a suitable platform to decipher the genetic mechanism underlying complex traits like flowering response in cotton. On the other hand, our utilization of automated aerial-based phenotyping system and machine learning algorithms has not only simplified the data acquisition process but also allowed us to capture the flowering pattern in cotton over the entire growth season of the crop. The genetic materials and the phenotyping platform utilized in this study has not only provided a new dimension to the scientific study of complex flowering traits but has also created benchmark to upgrade our conventional phenotyping systems to account for the season-long variation of such important traits.

Flowering in cotton has deviated significantly since its domestication into an annual crop from its innate perennial nature and this has had profound effects on its overall adaptation, and morphological and agronomic proliferation. Deciphering the molecular mechanism underling flowering response in cotton helps us understand the changes that have occurred in this important crop in regard to its flowering behavior and how these changes relate to flowering pattern in trees versus annual plants. Our study provides an initial exploration of genomic regions that control this economically and ecologically important characteristic of cotton and further investigation into

functional aspects of these genomic regions will help identify the causal variants in the cotton genome. Subsequent works include deeper investigation into such causal variants and their utilization in developing early maturing cotton cultivars with better fiber quality.

References

Adhikari, J. (2015). *Effects of exotic genotypes and genetic backgrounds on fiber quality and plant architectural traits in Upland cotton (Gossypium hirsutum L.)* [BibliographiesTheses, Electronic, Non-fiction].

Andolfatto, P., Davison, D., Erezyilmaz, D., Hu, T. T., Mast, J., Sunayama-Morita, T., & Stern, D. L. (2011). Multiplexed shotgun genotyping for rapid and efficient genetic mapping. *21*(4), 610-617.

Broman, K. W. (2001). Resource Review of Statistical Methods for QTL Mapping in Experimental Crosses. *30*(7), 44-52.

Broman, K. W., & Sen, S. (2009). *A guide to QTL mapping with R/qtl* [Bibliographies, Online, Non-fiction, Electronic document]. Springer.

Chandnani, R., Kim, C., Guo, H., Shehzad, T., Wallace, J. G., He, D., Zhang, Z., Patel, J. D., Adhikari, J., Khanal, S., & Paterson, A. H. (2018). Genetic Analysis of Gossypium Fiber Quality Traits in Reciprocal Advanced Backcross Populations. *The Plant Genome*, *11*(1), 170057.

Cheng, H.-Q., Han, L.-B., Yang, C.-L., Wu, X.-M., Zhong, N.-Q., Wu, J.-H., Wang, F.-X., Wang, H.-Y., & Xia, G.-X. (2016). The cotton MYB108 forms a positive feedback regulation loop with CML11 and participates in the defense response against Verticillium dahliae infection. *Journal of Experimental Botany*, *67*(6), 1935-1950.

El-Sharkawy, I., Sherif, S., Mila, I., Bouzayen, M., & Jayasankar, S. (2009). Molecular characterization of seven genes encoding ethylene-responsive transcriptional factors during plum fruit development and ripening. *Journal of Experimental Botany*, *60*(3), 907-922.

Fankhauser, C., & Chory, J. (1997). Light control of plant development. *Annual review of cell and developmental biology, 13*, 203-229.

Feltus, F. A., Hart, G. E., Schertz, K. F., Casa, A. M., Kresovich, S., Abraham, S., Klein, P. E., Brown, P. J., & Paterson, A. H. (2006). Alignment of genetic maps and QTLs between inter- and intra-specific sorghum populations. *Theor Appl Genet, 112*(7), 1295-1305.

Feng, J. X., Ji, S. J., Shi, Y. H., Xu, Y., Wei, G., & Zhu, Y. X. (2004). Analysis of five differentially expressed gene families in fast elongating cotton fiber. *Acta Biochim Biophys Sin (Shanghai), 36*(1), 51-56.

Fu, Y., Dong, C., Wang, J., Wang, Y., & Li, C. (2019). Genome-wide association study reveals the genetic control underlying node of the first fruiting branch and its height in upland cotton (Gossypium hirsutum L.). *Euphytica*, *215*(2), 35.

Glaubitz, J. C., Casstevens, T. M., Lu, F., Harriman, J., Elshire, R. J., Sun, Q., & Buckler, E. S. (2014). TASSEL-GBS: A High Capacity Genotyping by Sequencing Analysis Pipeline [Article]. *PLoS ONE, 9*(2), 1-11.

Grover, C. E., Gallagher, J. P., & Wendel, J. F. (2015). Candidate Gene Identification of Flowering Time Genes in Cotton. *The Plant Genome*, *8*(2), plantgenome2014.2012.0098.

Grover, C. E., Yoo, M.-J., Lin, M., Murphy, M. D., Harker, D. B., Byers, R. L., Lipka, A. E., Hu, G., Yuan, D., Conover, J. L., Udall, J. A., Paterson, A. H., Gore, M. A., & Wendel, J. F. (2020). Genetic Analysis of the Transition from Wild to Domesticated Cotton (Gossypium hirsutum L.). *G3 Genes|Genomes|Genetics*, *10*(2), 731-754.

Guo, Y., McCarty, J. C., Jenkins, J. N., An, C., & Saha, S. (2009). Genetic detection of node of first fruiting branch in crosses of a cultivar with two exotic accessions of upland cotton. *Euphytica*, *166*(3), 317-329.

Guo, Y. F., McCarty, J. C., Jenkins, J. N., & Saha, S. (2008). QTLs for node of first fruiting branch in a cross of an upland cotton, Gossypium hirsutum L., cultivar with primitive accession Texas 701 [Journal Article]. *Euphytica*, *163*(1), 113-122.

He, X., Wang, T., Xu, Z., Liu, N., Wang, L., Hu, Q., Luo, X., Zhang, X., & Zhu, L. (2018). The cotton HD-Zip transcription factor GhHB12 regulates flowering time and plant architecture via the GhmiR157-GhSPL pathway. *Communications Biology*, *1*(1), 229.

Huijser, P., & Schmid, M. (2011). The control of developmental phase transitions in plants. *138*(19), 4117-4130.

Hutchinson, J. B. (1951). Intra-specific differentiation in Gossypium hirsutum. *Heredity*, *5*(2), 161-193.

Jia, X., Pang, C., Wei, H., Wang, H., Ma, Q., Yang, J., Cheng, S., Su, J., Fan, S., Song, M., Wusiman, N., & Yu, S. (2016). High-density linkage map construction and QTL analysis for earliness-related traits in Gossypium hirsutum L. *BMC Genomics*, *17*(1), 909.

Jiang, Y., Li, C., & Paterson, A. H. (2016). High throughput phenotyping of cotton plant height using depth images under field conditions. *Computers and Electronics in Agriculture*, *130*, 57-68.

Jiang, Y., Li, C., Sun, S., Paterson, A. H., & Xu, R. (2017, 2017/01/01/). GPHENOVISION: A GROUND MOBILE SYSTEM WITH MULTI-MODAL IMAGING FOR FIELD-BASED HIGH THROUGHPUT PHENOTYPING OF COTTON.

Jiang, Y., Snider, J. L., Li, C., Rains, G. C., & Paterson, A. H. (2020). Ground Based Hyperspectral Imaging to Characterize Canopy-Level Photosynthetic Activities [Article]. *Remote Sensing*, *12*(2), 315-315.

Kelly, S., Huihui, L., Navarro, J. A. R., Dong, A., Maria Cinta, R., Sarah, H., Charlotte, A., Jeffrey, C. G., Sharon, M., Robert, J. E., Edward, S. B., & Peter, J. B. (2014). Novel Methods to Optimize Genotypic Imputation for Low-Coverage, Next-Generation Sequence Data in Crop Plants [article]. *The Plant Genome*, *7*(3).

Kohel, R. J., Richmond, T. R., & Lewis, C. F. (1974). Genetics of flowering response in cotton. VI. Flowering behavior of Gossypium hirsutum L. and G. barbadense L. hybrids [Journal Article]. *Crop Science*, *14*(5), 696-699.

Komeda, Y. (2004). GENETIC REGULATION OF TIME TO FLOWER IN ARABIDOPSIS THALIANA [Article]. *Annual Review of Plant Biology*, *55*(1), 521-535.

Kushanov, F. N., Buriev, Z. T., Shermatov, S. E., Turaev, O. S., Norov, T. M., Pepper, A. E., Saha, S., Ulloa, M., Yu, J. Z., Jenkins, J. N., Abdukarimov, A., & Abdurakhmonov, I. Y. (2017). QTL mapping for flowering-time and photoperiod insensitivity of cotton Gossypium darwinii Watt [Journal Article]. *PLoS ONE*, *12*(10), e0186240.

Kushanov, F. N., Turaev, O. S., Ernazarova, D. K., Gapparov, B. M., Oripova, B. B., Kudratova, M. K., Rafieva, F. U., Khalikov, K. K., Erjigitov, D. S., Khidirov, M. T., Kholova, M. D., Khusenov, N. N., Amanboyeva, R. S., Saha, S., Yu, J. Z., & Abdurakhmonov, I. Y. (2021). Genetic Diversity, QTL Mapping, and Marker-Assisted Selection Technology in Cotton (Gossypium spp.) [Review]. *Frontiers in Plant Science*, *12*.

Lai, D., Li, H., Fan, S., Song, M., Pang, C., Wei, H., Liu, J., Wu, D., Gong, W., & Yu, S. (2011). Generation of ESTs for Flowering Gene Discovery and SSR Marker Development in Upland Cotton. *PLoS ONE*, *6*(12), e28676.

Leal Valentim, F., Mourik, S. v., Posé, D., Kim, M. C., Schmid, M., van Ham, R. C. H. J., Busscher, M., Sanchez-Perez, G. F., Molenaar, J., Angenent, G. C., Immink, R. G. H., &

van Dijk, A. D. J. (2015). A Quantitative and Dynamic Model of the Arabidopsis Flowering Time Gene Regulatory Network [Article]. *PLoS ONE*, *10*(2), 1-18.

Lewis, C. F., & Richmond, T. R. (1960). The genetics of flowering response in cotton. II. Inheritance of flowering response in a Gossypium barbadense cross [Journal Article]. *Genetics*, *45*, 79-85.

Li, C., Wang, C., Dong, N., Wang, X., Zhao, H., Converse, R., Xia, Z., Wang, R., & Wang, Q. (2012). QTL detection for node of first fruiting branch and its height in upland cotton (Gossypium hirsutum L.). *Euphytica*, *188*(3), 441-451.

Li, C., Wang, X., Dong, N., Zhao, H., Xia, Z., Wang, R., Converse, R. L., & Wang, Q. (2013). QTL analysis for early-maturing traits in cotton using two upland cotton (Gossypium hirsutum L.) crosses. *Breeding Science*, *63*(2), 154-163.

Li, C., Zhang, J., Hu, G., Fu, Y., & Wang, Q. (2016). Association mapping and favorable allele mining for node of first fruiting/sympodial branch and its height in Upland cotton (Gossypium hirsutum L.). *Euphytica*, *210*(1), 57-68.

Li, L., Zhao, S., Su, J., Fan, S., Pang, C., Wei, H., Wang, H., Gu, L., Zhang, C., Liu, G., Yu, D., Liu, Q., Zhang, X., & Yu, S. (2017). High-density genetic linkage map construction by F2 populations and QTL analysis of early-maturity traits in upland cotton (Gossypium hirsutum L.). *PLoS ONE*, *12*(8), e0182918.

Li, Y., Jiang, J., Du, M. L., Li, L., Wang, X. L., & Li, X. B. (2013). A cotton gene encoding MYB-like transcription factor is specifically expressed in pollen and is involved in regulation of late anther/pollen development. *Plant Cell Physiol*, *54*(6), 893-906.

Li, Z., Wang, P., You, C., Yu, J., Zhang, X., Yan, F., Ye, Z., Shen, C., Li, B., Guo, K., Liu, N., Thyssen, G. N., Fang, D. D., Lindsey, K., Zhang, X., Wang, M., & Tu, L. (2020).

Combined GWAS and eQTL analysis uncovers a genetic regulatory network orchestrating the initiation of secondary cell wall development in cotton [Report]. *New Phytologist*, *226*(6), 1738.

Lin, Y., Schertz, K. F., & Paterson, A. H. (1995). Comparative analysis of QTLs affecting plant height and maturity across the Poaceae, in reference to an interspecific sorghum population [Journal Article]. *Genetics*, *141*(1), 391-411.

Lin, Z., He, D., Zhang, X., Nie, Y., Guo, X., Feng, C., & Stewart, J. M. (2005). Linkage map construction and mapping QTL for cotton fibre quality using SRAP, SSR and RAPD [Article]. *Plant Breeding*, *124*(2), 180-187.

Ninh, T. T., Gao, W., Trusov, Y., Zhao, J.-R., Long, L., Song, C.-P., & Botella, J. R. (2021). Tomato and cotton G protein beta subunit mutants display constitutive autoimmune responses. *Plant Direct*, *5*(11), e359.

Paterson, A. H., Brubaker, C. L., & Wendel, J. F. (1993). A rapid method for extraction of cotton (Gossypium spp.) genomic DNA suitable for RFLP or PCR analysis. *Plant Molecular Biology Reporter*, *11*(2), 122-127.

Ray, L. L., & Richmond, T. R. (1966). Morphological measures of earliness of crop maturity in cotton [Journal Article]. *Crop Science*, *6*, 527-531.

Rong, J., Feltus, F. A., Waghmare, V. N., Pierce, G. J., Chee, P. W., Draye, X., Saranga, Y., Wright, R. J., Wilkins, T. A., May, O. L., Smith, C. W., Gannaway, J. R., Wendel, J. F., & Paterson, A. H. (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), 2577-2588.

Said, J. I., Lin, Z., Zhang, X., Song, M., & Zhang, J. (2013). A comprehensive meta QTL analysis for fiber quality, yield, yield related and morphological traits, drought tolerance, and disease resistance in tetraploid cotton. *BMC Genomics*, *14*(1), 776.

Shen, X., Guo, W., Zhu, X., Yuan, Y., Yu, J. Z., Kohel, R. J., & Zhang, T. (2005). Molecular mapping of QTLs for fiber qualities in three diverse lines in Upland cotton using SSR markers [Article]. *Molecular Breeding*, *15*(2), 169-181.

Su, J., Pang, C., Wei, H., Li, L., Liang, B., Wang, C., Song, M., Wang, H., Zhao, S., Jia, X., Mao, G., Huang, L., Geng, D., Wang, C., Fan, S., & Yu, S. (2016). Identification of favorable SNP alleles and candidate genes for traits related to early maturity via GWAS in upland cotton. *BMC Genomics*, *17*(1), 687.

Sun, S., Li, C., Chee, P. W., Paterson, A. H., Jiang, Y., Xu, R., Robertson, J. S., Adhikari, J., & Shehzad, T. (2020). Three-dimensional photogrammetric mapping of cotton bolls in situ based on point cloud segmentation and clustering. *ISPRS Journal of Photogrammetry and Remote Sensing*, *160*, 195-207.

Szalma, S. J., Hostert, B. M., LeDeaux, J. R., Stuber, C. W., & Holland, J. B. (2007). QTL mapping with near-isogenic lines in maize [Article]. *Theoretical & Applied Genetics*, *114*(7), 1211-1228.

Waddle, B. M., Lewis, C. F., & Richmond, T. R. (1961). The genetics of flowering response in cotton. III. Fruiting behaviour of Gossypium hirsutum race lati-folium in a cross with a variety of cultivated American upland cotton [Journal Article]. *Genetics*, *46*, 427-437.

Wang, C. C. N., Pei-Chun, C., Ka-Lok, N., Chun-Ming, C., Sheu, P. C. Y., & Tsai, J. J. P. (2014). A model comparison study of the flowering time regulatory network in Arabidopsis [Article]. *BMC Systems Biology*, *8*(1), 1-25.

Wang, Z., Zhang, D., Wang, X., Tan, X., Guo, H., & Paterson, A. H. (2013). A Whole-Genome DNA Marker Map for Cotton Based on the D-Genome Sequence of Gossypium raimondii L. *G3 Genes|Genomes|Genetics*, *3*(10), 1759-1767.

Xu, R., & Li, C. (2022). A modular agricultural robotic system (MARS) for precision farming: Concept and implementation [Report]. *Journal of Field Robotics*, *39*(4), 387.

Yano, M., Harushima, Y., Nagamura, Y., Kurata, N., Minobe, Y., & Sasaki, T. (1997). Identification of quantitative trait loci controlling heading date in rice using a high-density linkage map. *95*(7), 1025-1032.

Yufang, G., McCarty, J. C., Saha, S., Jenkins, J. N., & Jixiang, W. (2007/01//). QTL ANALYSES OF FLOWERING TIME IN A PRIMITIVE COTTON ACCESSION (GOSSYPIUM HIRSUTUM L.).

Zhang, J., Jia, X., Guo, X., Wei, H., Zhang, M., Wu, A., Cheng, S., Cheng, X., Yu, S., & Wang, H. (2021). QTL and candidate gene identification of the node of the first fruiting branch (NFFB) by QTL-seq in upland cotton (Gossypium hirsutum L.). *BMC Genomics*, *22*(1), 882.

Zhang, J., Zhang, P., Huo, X., Gao, Y., Chen, Y., Song, Z., Wang, F., & Zhang, J. (2021). Comparative Phenotypic and Transcriptomic Analysis Reveals Key Responses of Upland Cotton to Salinity Stress During Postgermination [Original Research]. *Frontiers in Plant Science*, *12*.

Zhang, Q., Zhang, H., Sun, L., Fan, G., Ye, M., Jiang, L., Liu, X., Ma, K., Shi, C., Bao, F., Guan, R., Han, Y., Fu, Y., Pan, H., Chen, Z., Li, L., Wang, J., Lv, M., Zheng, T., . . . Cheng, T. (2018). The genetic architecture of floral traits in the woody plant Prunus mume. *Nature Communications*, *9*(1), 1702.

Zhang, T., Hu, Y., Jiang, W., Fang, L., Guan, X., Chen, J., Zhang, J., Saski, C. A., Scheffler, B. E., Stelly, D. M., Hulse-Kemp, A. M., Wan, Q., Liu, B., Liu, C., Wang, S., Pan, M., Wang, Y., Wang, D., Ye, W., . . . Chen, Z. J. (2015). Sequencing of allotetraploid cotton (Gossypium hirsutum L. acc. TM-1) provides a resource for fiber improvement. *Nature biotechnology*, *33*(5), 531-537.

Zhu, L. L., & Kuraparthy, V. (2014). Molecular genetic mapping of the major effect photoperiod response locus in Pima cotton (Gossypium barbadense L.) [Journal Article]. *Crop Science*, *54*(6), 2492-2498.

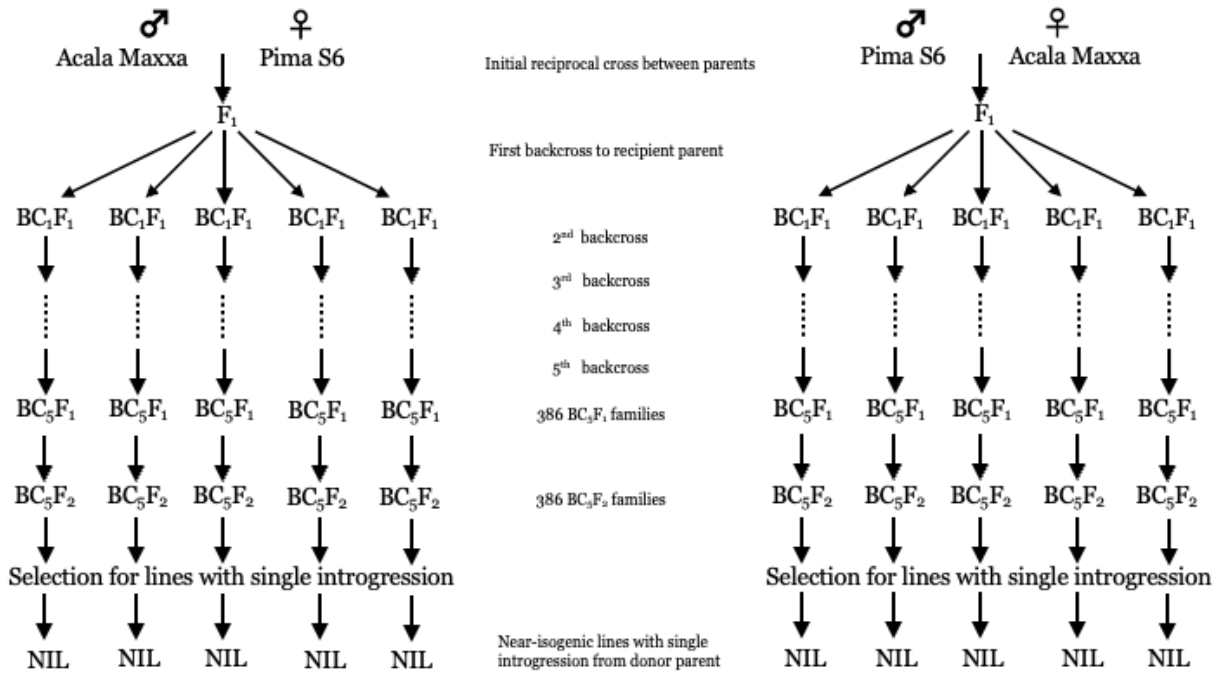

Figure 3.1. Development of reciprocal set of Near-isogenic lines. The panel on the left shows the development of NILs in the Acala Maxxa background and the panel on the right shows the development of NILs in the Pima S6 background.

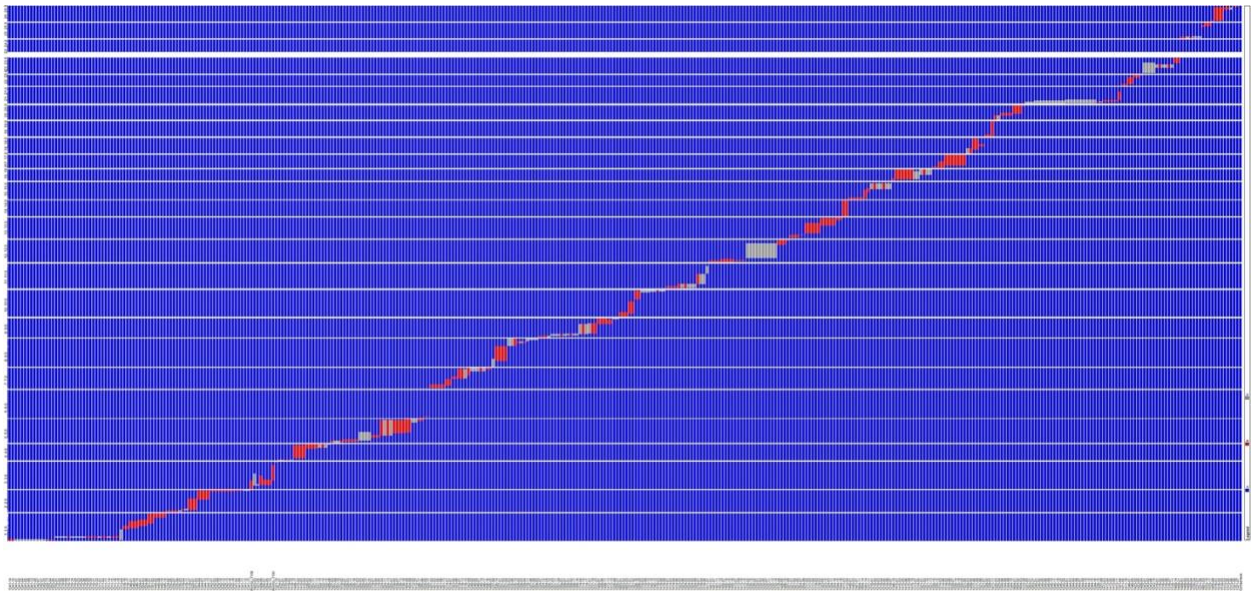

Figure 3.2. Graphical genotypes of NILs in Acala Maxxa background. Figure represents the genomic constitution of the genome wide NILs in Acala Maxxa background. The 397 NILs (left to right) are shown in the x-axis and the 26 chromosomes (down to up in ascending order) are shown in the y-axis. Blue color represents Acala Maxxa loci in homozygous state (HH), grey color represents heterozygous loci (HB) and red color represents Pima S6 loci in homozygous state (BB)

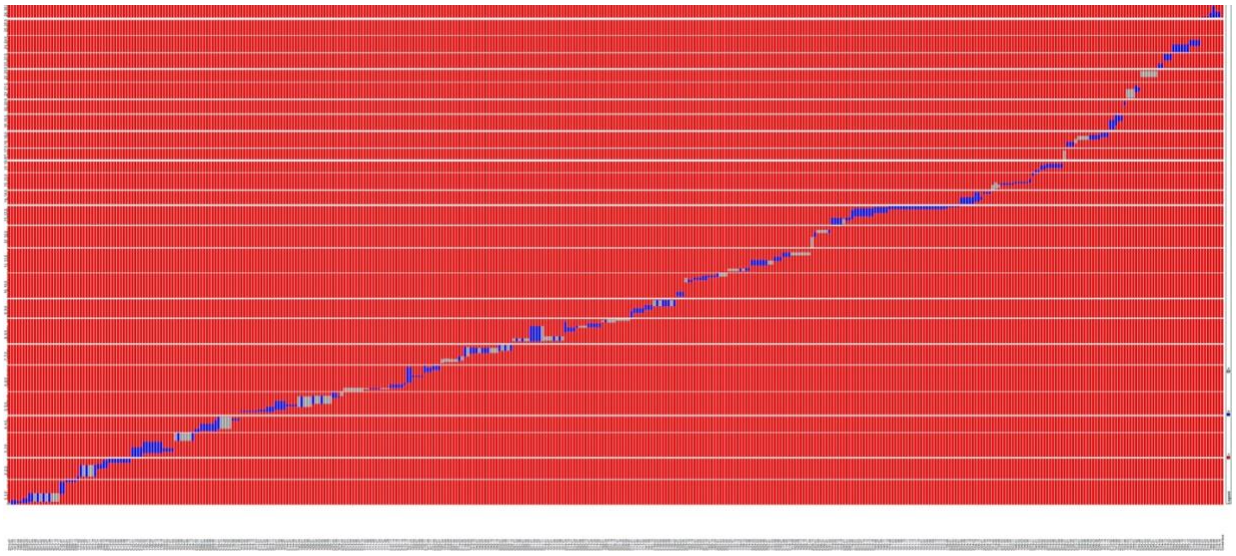

Figure 3.3. Graphical genotypes of NILs in Pima S6 background. Figure represents the genomic constitution of the genome wide NILs in Pima S6 background. The 423 NILs (left to right) are shown in the x-axis and the 26 chromosomes (down to up in ascending order) are shown in the y-axis. Blue color represents Acala Maxxa loci in homozygous state (HH), grey color represents heterozygous loci (HB) and red color represents Pima S6 loci in homozygous state (BB)

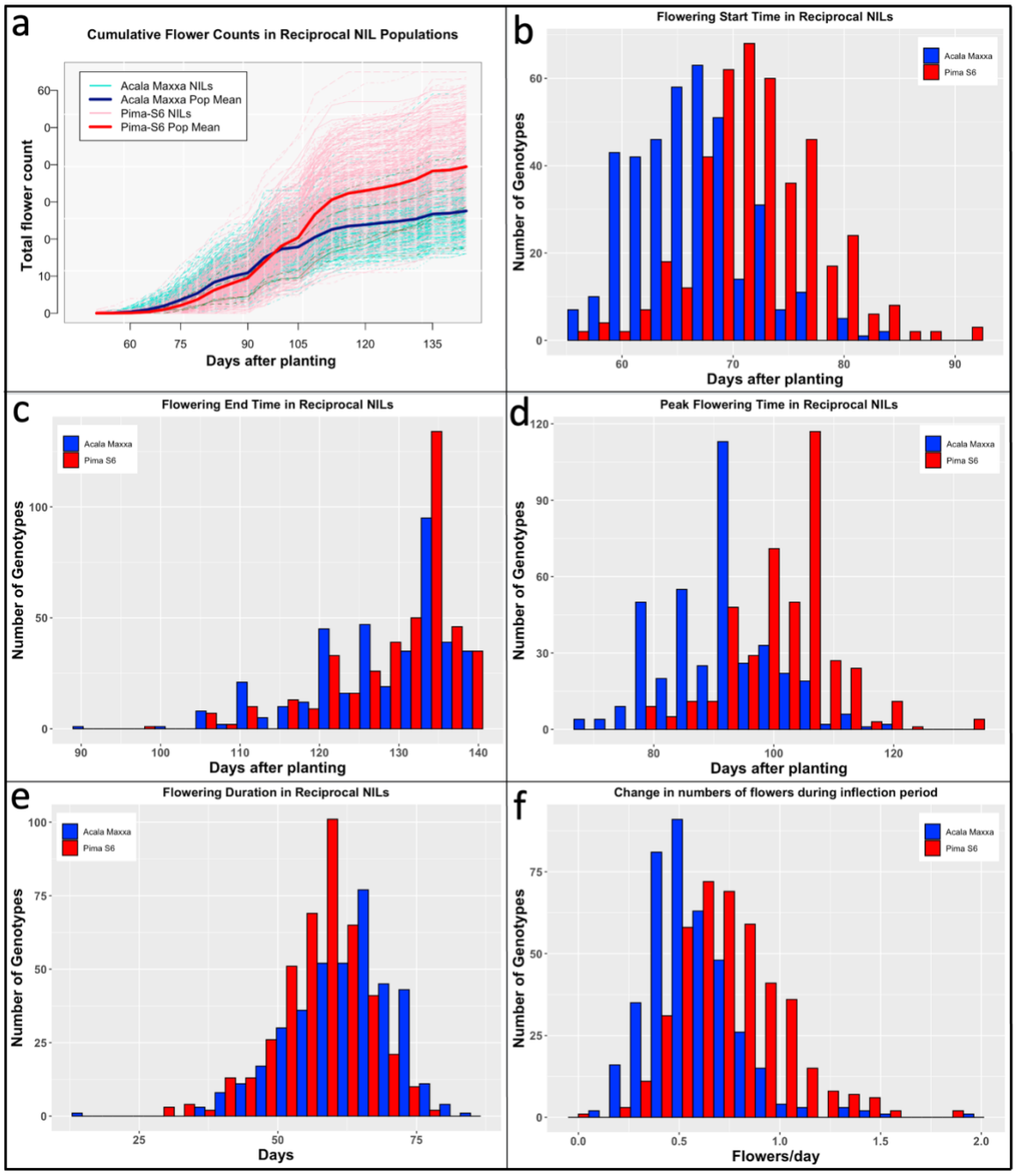

Figure 3.4. Distribution of flowering related traits in the reciprocal set of Near-isogenic lines.

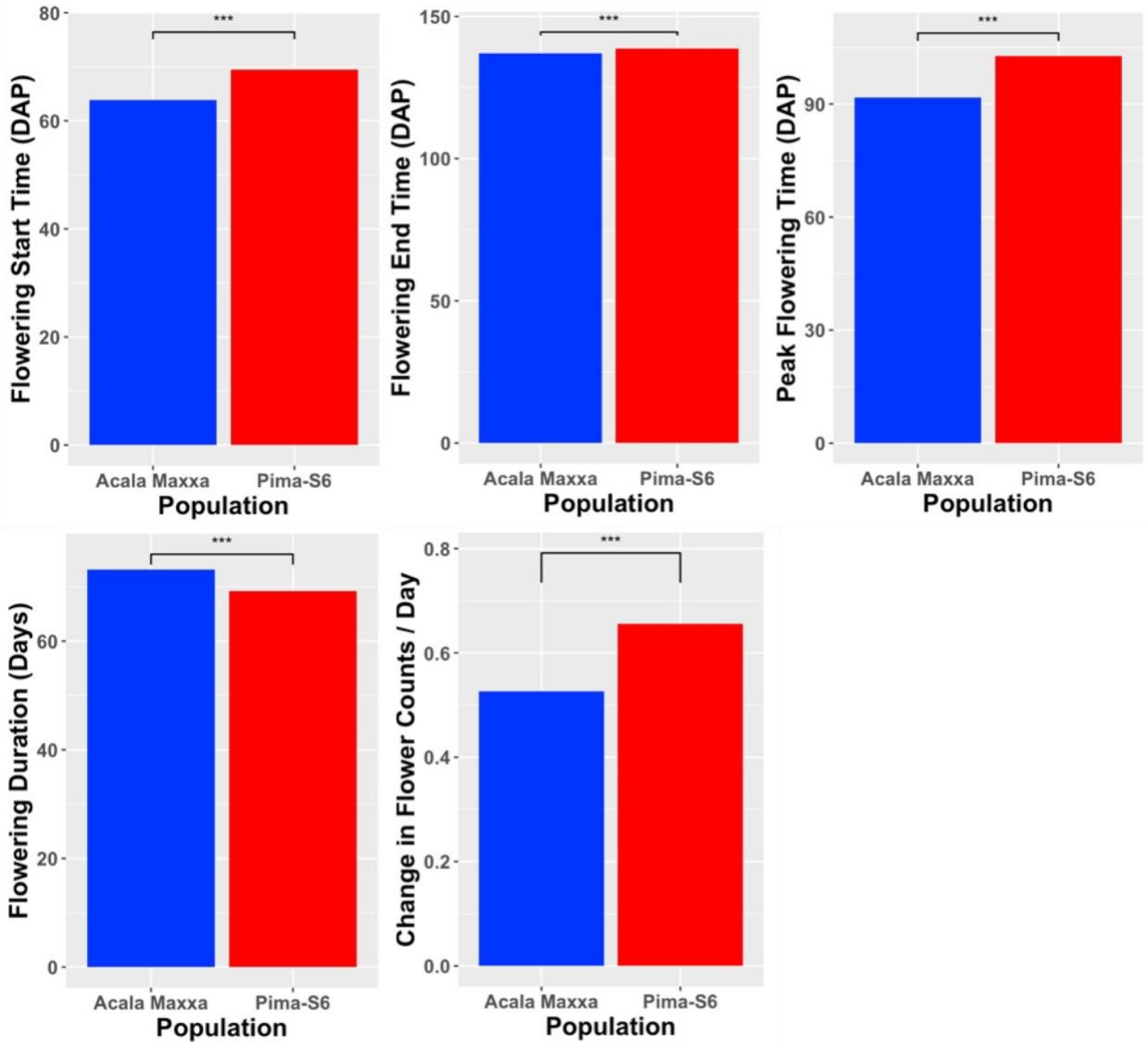

Figure 3.5. Population averages for flowering related traits in the reciprocal set of NILs
*** Significantly different at $\alpha = 0.05$

Table 3.1. Coverage of donor genome in recipient background in the reciprocal NIL populations.

	Acala Maxxa Background		Pima S6 Background	
Chromosome	No. of NILs	Coverage (%)	No. of NILs	Coverage (%)
Chr 01	51	97.14	24	91.09
Chr 02	27	95.53	19	94.08
Chr 03	9	88.58	22	87.66
Chr 04	17	85.93	16	92.44
Chr 05	30	92.67	36	82.66
Chr 06	2	1.33	34	94.59
Chr 07	20	90.01	25	87.79
Chr 08	15	93.61	41	82.72
Chr 09	26	91.76	16	89.52
Chr 10	15	85.33	18	51.15
Chr 11	14	82.49	29	63.76
Chr 12	26	85.74	7	74.46
Chr 13	17	85.68	45	90.34
Chr 14	2	92.19	11	73.13
Chr 15	14	86.35	15	74.88
Chr 16	13	89.38	10	57.51
Chr 17	11	87.79	1	70.20
Chr 18	6	83.70	15	80.20
Chr 19	3	84.08	5	86.26
Chr 20	10	91.75	1	16.71
Chr 21	31	64.55	5	72.92
Chr 22	7	76.74	6	42.33
Chr 23	12	83.20	5	64.41
Chr 24	0	0.00	10	66.88
Chr 25	10	39.18	0	0.00
Chr 26	9	91.85	7	71.01
Total	397	78.71	423	71.49

Table 3.2. QTLs identified in the Acala Maxxa (*G. hirsutum*) background. a is the additive effect of the QTL, d is the dominance effect of the QTL and d/a is gene action defined as the ratio of dominance to additive effect.

Trait	QTL name	Chr	Marker	LOD	PVE	a	d	d/a
Start Date	qSDGh.08.01	8	S8_25984638	3.19	4.17	-2.71	24.79	9.14
	qSDGh.18.01	18	S18_28039634	3.48	4.68	20.18		
End date	qEDGh.02.01	2	S2_4683354	5.32	5.65	3.91	-17.08	4.35
	qEDGh.09.01	9	S9_36767391	3.49	4.68	3.52	-3.13	0.88
	qEDGh.12.01	12	S12_65273956	3.78	3.49	24.04	2.15	0.08
	qEDGh.17.01	17	S17_8895832	16.94	2.72	15.61		
	qEDGh.21.01	21	S21_16292082	3.89	3.78		9.10	
	qEDGh.23.01	23	S23_11637539	5.78	3.99		-2.36	
Peak Date	qPDGh.10.01	10	S10_31157912	3.24	3.39	3.51		
Duration	qDUGh.02.01	2	S2_4683354	5.52	6.35	6.13	-15.21	2.48
	qDUGh.17.01	17	S17_8895832	3.38	3.48	14.95		
Slope	qSLGh.01.01	1	S1_55935911	10.51	1.37	0.19		
	qSLGh.02.01	2	S2_4683354	12.89	5.16	0.50	0.91	1.82
	qSLGh.11.01	11	S11_61184623	7.61	5.88		0.55	
	qSLGh.12.01	12	S12_50388561	6.53	11.66		0.15	
	qSLGh.22.01	22	S22_37477542	5.19	6.13	0.39		

Table 3.3 QTLs identified in the Pima S6 (*G. barbadense*) background. a is the additive effect of the QTL, d is the dominance effect of the QTL and d/a is gene action defined as the ratio of dominance to additive effect.

Trait	QTL name	Chr	Marker	LOD	PVE	a	d	d/a
Start Date	qSDGb.03.01	3	S3_40357376	3.76	6.34	-5.83		
	qSDGb.09.01	9	S9_26439375	3.09	8.57	-14.74		
End Date	qEDGb.03.01	3	S3_20315421	5.93	18.22	-6.23		
	qEDGb.06.01	6	S6_12962936	4.13	5.67	-7.97	-3.61	0.45
	qEDGb.13.01	13	S13_48094017	5.53	11.73	-4.21		
	qEDGb.19.01	19	S19_64167229	7.87	4.87	-18.54		
Peak Date	qPDGb.04.01	4	S4_53065395	4.43	5.11	19.33	7.19	0.37
	qPDGb.08.01	8	S8_27114194	3.19	3.93	4.17	-12.76	3.05
	qPDGb.11.01	11	S11_6426512	6.98	13.55	-12.33	3.66	0.29
	qPDGb.24.01	24	S24_43802683	3.84	4.47	-13.31		
Duration	qDUGb.06.01	6	S6_12962936	3.27	3.61	-13.25	-7.42	0.56
Slope	qSLGb.03.01	3	S3_20315421	4.64	5.79	0.44		
	qSlGb.05.01	5	S5_23021069	3.78	3.99	0.21		

Table 3.4. Sub-genomic distribution of flower related QTLs identified in reciprocal NIL populations.

Traits	**Acala Maxxa**		**Pima S6**	
	At	**Dt**	**At**	**Dt**
Start Date	1	1	2	0
End Date	3	3	3	1
Peak Date	1	0	3	1
Duration	1	1	1	0
Slope	4	1	2	0

Table 3.5. Co-occurrence of QTLs in multiple families and lines. BB‡ denotes the number of lines containing the marker in homozygous state, BH† denotes the number of lines containing the marker in heterozygous state, * denotes the number of NILs carrying the most significant marker and # indicates the number of NIL sub-families carrying the significant marker.

Background	Trait	QTL name	Marker	BB‡	BH†	NILs*	Families#
Pima S6	Start Date	qSDGb.03.01	S3_20315421	11	0	11	3
		qSDGb.09.01	S9_26439375	4	0	4	2
	End Date	qEDGb.03.01	S3_20315421	11	0	11	3
		qEDGb.06.01	S6_12962936	4	12	16	2
		qEDGb.13.01	S13_48094017	13	0	13	3
		qEDGb.19.01	S19_64167229	1	0	1	1
	Peak Date	qPDGb.04.01	S4_53065395	4	4	8	3
		qPDGb.08.01	S8_27114194	6	6	12	3
		qPDGb.11.01	S11_6426512	1	5	6	2
		qPDGb.24.01	S24_43802683	4	0	4	2
	Duration	qDUGb.06.01	S6_12962936	4	12	16	2
	Slope	qSLGb.03.01	S3_20315421	11	0	11	3
		qSlGb.05.01	S5_23021069	10	0	10	4
Acala Maxxa	Start Date	qSDGh.08.01	S8_25984638	4	1	5	3
		qSDGh.18.01	S18_28039634	4	0	4	2
	End date	qEDGh.02.01	S2_4683354	4	1	5	2
		qEDGh.09.01	S9_36767391	3	3	6	2
		qEDGh.12.01	S12_65273956	1	10	11	3
		qEDGh.17.01	S17_8895832	2	0	2	1
		qEDGh.21.01	S21_16292082	0	10	10	1
		qEDGh.23.01	S23_11637539	0	4	4	1
	Peak Date	qPDGh.10.01	S10_31157912	2	0	2	1
	Duration	qDUGh.01.01	S2_4683354	4	1	5	2
		qDUGh.17.01	S17_8895832	2	0	2	1
	Slope	qSLGh.01.01	S1_55935911	6	0	0	2
		qSLGh.02.01	S2_4683354	4	1	5	2
		qSLGh.11.01	S11_61184623	0	1	1	1
		qSLGh.12.01	S12_50388561	0	10	10	2
		qSLGh.22.01	S22_37477542	3	0	3	2

Table 3.6. Previous publications reporting flowering related QTLs in cotton. Trait*: BP (Bud period), FBP (Flower and bud period), FDR (Flowering duration), FFBN/NFFB/NFB (First fruiting branching node), HFFBN (Height of first fruiting branching node), FT (Flowering time), GP (Growth period), YPBF (Yield percent before frost)

Trait*	QTLs identified	Chromosomes	Reference	Population
BP	7	1, 10, 12, 17, LG2	Li et al., 2013	F2:3
FBP	8	3, 7, 8, 9, 16, 17, 22, LG2	Li et al., 2013	F2:3
FBP	40	1, 2, 3, 4, 5, 8, 9, 10, 11, 12, 14, 16, 17, 18, 20, 23, 24, 25, 26	Jia et al., 2016	RIL
FBP	10	8, 10, 17, 18, 20, 24, 25, 26	Li et al., 2017	F2, F2:3
FBP	1	17	Su et al., 2016	Natural Population
FBP	11	8, 14, 23	Li et al., 2018	Natural Population
FDR	6	11, 19, 24, 26	Kushanov et al., 2017	F2, F3
FFBN	8	1, 5, 6, 9, 11, 12, 13, 17	Li et al., 2012	F2:3
FFBN	1	14	Zhang et al., 2009	RIL
FFBN	43	1, 3, 4, 7, 9, 13, 14, 15, 17, 18, 20, 21, 22, 23, 24, 26	Jia et al., 2016	RIL
FFBN	9	5, 7, 12, 14, 17, 18, 19, 20, 25	Li et al., 2017	F2, F2:3
FFBN	18	2, 6, 9, 11, 12, 13, 17, 19, 21, 23, 25	Li et al., 2016	Natural Population
FFBN	1	17	Su et al., 2016	Natural Population
FT	2	15, 26	Kushanov et al., 2017	F2, F3
FT	3	15, 21	Zhang et al., 2016	F2
FT	1	16	Ren et al., 2002	Substitution Lines
FT	1	14	Zhang et al., 2008	RIL
FT	1	14	Zhang et al., 2009	RIL
FT	11	2, 6, 7, 8, 14, 15	Lacape et al., 2013	RIL
FT	39	2, 3, 5, 7, 8, 9, 10, 11, 12, 14, 15, 17, 19, 22, 24, 26	Jia et al., 2016	RIL
FT	10	2, 5, 7, 8, 12, 17, 25	Li et al., 2017	F2, F2:3

FT	5	4, 12, 17	Su et al., 2016	Natural Population
GP	5	13, 17, 17, 21, 17	Li et al., 2013	F2:3
GP	1	24	Zhang et al., 2009	RIL
GP	47	2, 4, 6, 8, 9, 10, 12, 14, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26	Jia et al., 2016	RIL
GP	9	8, 12, 15, 17, 24, 20	Li et al., 2017	F2, F2:3
GP	5	3, 17	Su et al., 2016	Natural Population
GP	5	23	Li et al., 2018	Natural Population
HFFBN	6	9, 17, 25, 9, 17, LG3	Li et al., 2012	F2:3
HFFBN	45	2, 3, 4, 5, 7, 9, 11, 13, 15, 14, 17, 24, 23, 21, 18	Jia et al., 2016	RIL
HFFBN	4	17,25,23,26	Li et al., 2017	F2, F2:3
HFFBN	19	5, 7, 9, 11, 12, 14, 15, 18, 21, 22, 23, 25	Li et al., 2016	Natural Population
HNFFB	22	3, 5, 7, 8, 11, 21, 24, 25	Fu et al., 2019	Natural Population
NFB	5	15, 16, 21, 25	Guo et al., 2008	F2
NFB	4	4, 5, 26	Kushanov et al., 2017	F2, F3
NFFB	55	1, 2, 3, 6, 7, 8, 10, 12, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26	Zhang et al., 2021	BC4F2
NFFB	12	3, 4, 6, 8, 9, 13, 18, 23, 26	Fu et al., 2019	Natural Population
YPBF	14	9, 10, 11, 12, 17, 20, 24,12, LG5	Li et al., 2013	F2:3
YPBF	1	22	Jiang et al., 1998	F2
YPBF	2	14, 25	Zhang et al., 2008	RIL
YPBF	1	25	Zhang et al., 2009	RIL
YPBF	1	17	Su et al., 2016	Natural Population

Table 3.7. QTL correspondence probability

Present study (NILs)		Previous study				No of matching	
Trait	# QTLs	Trait	Pop	# QTLs	Reference	QTLs	Prob.
Start Date	4	FT	RIL	11	Lacape et al., 2013	1	0.3644
		FT	RIL	39	Jia et al., 2016	3	0.2436
		FT	F2, F2:3	10	Li et al., 2017	1	0.3461
		FFBN	F2:3	8	Li et al., 2012	1	0.3016
		FFBN	RIL	43	Jia et al., 2016	3	0.2887
		FFBN	F2, F2:3	9	Li et al., 2017	1	0.3252
		FFBN	Natural	18	Li et al., 2016	1	0.4304
		NFFB	BC4F2	55	Zhang et al., 2021	3	0.4146
		NFFB	Natural	12	Fu et al., 2019	4	**0.0003**
		HFFBN	F2:3	6	Li et al., 2012	1	0.2451
		HFFBN	RIL	45	Jia et al., 2016	3	0.3140
		HFFBN	Natural	19	Li et al., 2016	2	0.1978
		HNFFB	Natural	22	Fu et al., 2019	2	0.2414
Duration	3	FBP	F2:3	8	Li et al., 2013	1	0.2488
		FBP	RIL	40	Jia et al., 2016	2	0.3797
		FBP	F2, F2:3	10	Li et al., 2017	1	0.2939
		FBP	Natural	1	Su et al., 2016	1	**0.0375**
		GP	F2:3	5	Li et al., 2013	1	0.1688
		GP	RIL	47	Jia et al., 2016	3	0.1973
		GP	F2, F2:3	9	Li et al., 2017	1	0.2722
		GP	Natural	5	Su et al., 2016	1	0.1688
		BP	F2:3	7	Li et al., 2013	1	0.2239
End Date	10	YPBF	F2:3	14	Li et al., 2013	3	0.1721
		YPBF	Natural	1	Su et al., 2016	1	0.1250
		FT	F2	3	Zhang et al., 2016	1	0.2939

FT	RIL	11	Lacape et al., 2013	2	0.2793
FT	RIL	39	Jia et al., 2016	6	0.2006
FT	F2, F2:3	10	Li et al., 2017	3	0.0873
FT	Natural	5	Su et al., 2016	2	0.1024
FFBN	F2:3	8	Li et al., 2012	4	**0.0066**
FFBN	RIL	43	Jia et al., 2016	6	0.2445
FFBN	F2, F2:3	9	Li et al., 2017	3	0.0678
FFBN	Natural	18	Li et al., 2016	8	**<0.0001**
FFBN	Natural	1	Su et al., 2016	1	0.1250
HFFBN	F2:3	6	Li et al., 2012	2	0.1373
HFFBN	RIL	45	Jia et al., 2016	7	0.1803
HFFBN	F2, F2:3	4	Li et al., 2017	2	0.0687
HFFBN	Natural	19	Li et al., 2016	4	0.1307
HNFFB	Natural	22	Fu et al., 2019	2	0.2689
NFB	F2	5	Guo et al., 2008	1	0.3814
NFFB	BC4F2	55	Zhang et al., 2021	8	0.2218
NFFB	Natural	12	Fu et al., 2019	4	**0.0329**

Table 3.8. Genes present in regions tightly linked to the most significant SNP marker. Gray highlighted genes are verified to have roles in flowering in Arabidopsis.

QTL	Gene ID	Gene	Description
qDUGh.17.01	Gh_D03G0506	psbK	Photosystem II reaction center protein K
	Gh_D03G0507	ycf2-A	Protein Ycf2
	Gh_D03G0508	CCR2	Cinnamoyl-CoA reductase 2
	Gh_D03G0509	NA	NA
	Gh_D03G0510	NA	NA
	Gh_D03G0511	At5g16730	WEB family protein, chloroplastic
qEDGh.02.01	Gh_A02G0377	ERF1B	Ethylene-responsive transcription factor 1B
	Gh_A02G0378	PRP	Repetitive proline-rich cell wall protein
	Gh_A02G0379	PRP	Repetitive proline-rich cell wall protein
qEDGh.12.01	Gh_A12G1162	NA	NA
	Gh_A12G1163	GATA15	GATA transcription factor 15
	Gh_A12G1164	NA	NA
	Gh_A12G1165	EIF-5A2	Eukaryotic translation initiation factor 5A-2
	Gh_A12G1166	NA	NA
	Gh_A12G1167	NA	NA
	Gh_A12G1168	GATL7	Probable galacturonosyltransferase-like 7
	Gh_A12G1169	CSLD5	Cellulose synthase-like protein D5
	Gh_A12G1170	Dnajb6	DnaJ homolog subfamily B member 6
qPDGb.08.01	Gh_A08G0763	VDAC4	Mitochondrial outer membrane protein porin4
qPDGb.11.01	Gh_A11G0652	At5g39250	F-box protein At5g39250
	Gh_A11G0653	UGT83A1	UDP-glycosyltransferase 83A1
	Gh_A11G0654	NA	NA
	Gh_A11G0655	NA	NA
	Gh_A11G0656	NA	NA
	Gh_A11G0657	RIK	Protein RIK
	Gh_A11G0658	AMY3	Alpha-amylase 3, chloroplastic
	Gh_A11G0659	RPK2	LRR receptor-like serine/threonine-protein
qPDGb.24.01	Gh_D08G1339	NA	NA
	Gh_D08G1340	MYB108	Transcription factor MYB108
qSDGh.08.01	Gh_A08G0755	ssx2ip-a	Afadin- and alpha-actinin-binding protein A
qSDGh.18.01	Gh_D13G1032	HSP26-A	Probable glutathione S-transferase
qSLGh.22.01	Gh_D04G1148	NA	NA
	Gh_D04G1149	IP5P8	Type I inositol polyphosphate 5-phosphatase8

CHAPTER 4

DISSECTING QUANTITATIVE VARIATION IN FIBER LENGTH PARAMETERS USING ADVANCED RECIPROCAL BACKCROSS POPULATIONS IN COTTON

Jeevan Adhikari, Deepak Vitrakoti, Wiriyanat Ployaram, Sameer Khanal, Peng W Chee, and Andrew H Paterson. To be submitted to *Theoretical and Applied Genetics*.

Abstract

Estimation of precise locations and accurate effects of 'quantitative trait loci' (QTLs) affecting fiber quality traits is an essential prerequisite for using these QTLs in breeding programs via marker assisted selection. Advanced backcross lines (ABLs) and/or near-isogenic lines (NILs) carrying one or few chromosomal segments introgressed from a donor parent into the recipient background serve an important purpose of delineating the boundaries of a QTL to a small and rather precise genomic location, and also adds to the accuracy of estimating the effects of the QTL by reducing background noise caused by multiple segregating donor chromosome segments. In a set of reciprocal interspecific advanced backcross populations for two elite cotton cultivars, Acala Maxxa (*G. hirsutum*) and Pima S6 (*G. barbadense*) representing the two major domesticated species of cotton, we identified genomic locations underpinning three important fiber length parameters – upper half mean length, fiber uniformity index and short fiber content. Phenotypic evaluation of these lines in three environments revealed a total of 58 QTLs, with one QTL for short fiber index identified in all three environments and five (four for upper half mean length and one for short fiber index) identified in at least two environments. Most of the QTLs identified in the study were small-effect (explaining <10% variation) QTLs justifying the genomic composition of the lines used for analysis. Limited reciprocity of QTLs in the two backgrounds shows the strong influence of recipient genome, in addition to the combined consequences of epistasis, small phenotypic effects and imperfect coverage of donor chromatin in the recipient background.

Introduction

Cotton is the most important natural fiber and the leading textile crop of global economic importance. Most of the world's textile fiber comes from two domesticated species of cotton -

Upland cotton (*Gossypium hirsutum* L.) and Pima Cotton (*G. barbadense* L.) which collectively account for more than 95% of global cotton fiber production. Despite stiff competition from synthetic fibers, consumer preference towards natural fiber and elevated market demands for high quality textile fibers have generated renewed interest in breeding for better fiber quality (Peng W. Chee & B. Todd Campbell, 2009). However, there is a low level of genetic differentiation in the cultivated cotton gene pool (Fang et al., 2017; Wang et al., 2017) resulting from the depletion of favorable variation due to evolutionary and breeding (domestication) bottlenecks (Adhikari, 2015a; Andrew H. Paterson et al., 2004). In addition, negative correlations between yield components and fiber quality parameters have restricted simultaneous breeding gains.

Improving fiber quality parameters simultaneously with increasing lint yield has been an important goal for cotton improvement. Genetic dissection of complex fiber quality traits and yield components into underlying quantitative trait loci (QTL) creates opportunities to manipulate quantitative traits based on Mendelian principles, using techniques such as marker-assisted selection (MAS), QTL pyramiding, fine mapping, and map-based cloning. These techniques have the potential to mitigate some of the challenges hindering improvement of fiber quality (Peng W. Chee & B. Todd Campbell, 2009; Fang, 2015).

Introgressive breeding has been one of the successful techniques employed to transfer or combine desirable traits from two or more related species. Experimental populations created by/for introgressive breeding not only serve the purpose of combining favorable alleles, but also serve as suitable platform to identify genomic locations underpinning complex traits.

Among the two widely cultivated cotton species, Upland cotton is known for its high yield potential while Pima cotton has superior fiber quality. Transferring desirable traits from Pima cotton to Upland cotton has been a long-standing goal for cotton breeders and geneticists.

Numerous introgressive breeding programs have been conducted in cotton in the past six or seven decades to simultaneously improve fiber yield and quality by interspecific hybridization (Zhang et al., 2014). In the past two decades alone, geneticists have tagged hundreds of QTLs underlying cotton fiber yield and quality parameters (J. Rong et al., 2007; J. I. Said, J. A. Knapka, et al., 2015; J. I. Said et al., 2013; J. I. Said, M. Song, et al., 2015). With progress in genetic mapping technologies, more QTLs are being tagged and published, but the use of these QTLs in marker-assisted breeding for cotton improvement has been limited (Khanal, 2018; J. I. Said, J. A. Knapka, et al., 2015). Stability (consistency) of QTL expression across environments and among generations and/or genetic backgrounds, variable effects and low precision in mapped locations are among the limiting factors to marker-assisted cotton improvement (Fang, 2015). Elevated noise from multiple chromosomal segments segregating independently, some of which span entire chromosomes, often result in over- or underestimation of the effects of mapped QTLs and thus in lack of reproducible effects across environments and/or generations. Experimental populations with reduced background noise and smaller size of introgressed chromatin segments facilitate estimation of the location and effects of these genomic benchmarks more precisely.

NILs, with only one chromosomal segment from the donor parent, and / or advanced backcross lines (ABLs), with only a few donor segments segregating in the recipient genome, reduce the background noise present in early generation populations. These populations also serve as a good resource for both genetic mapping and breeding (R. Kooke et al., 2012). Since there are few (ideally one) introgressed segment(s) in each NIL, phenotypes due to QTLs on the segment(s) are rendered much more discrete than in F_2 or backcross populations, often behaving as simple Mendelian factors (Paran & Zamir, 2003). QTL mapping based on NILs can thus increase accuracy of QTL position and detect small effect QTLs that might otherwise be obscured by larger-effect

genes in more complex populations. In addition, because of the fixed genotype of NILs, they can be replicated in different environments to test interaction between genetic and environmental factors (Monforte & Tanksley, 2000). By crossing NILs to the recurrent parent, fine mapping of specific QTLs toward their cloning is facilitated.

Fiber length is one of the most important aspects of fiber quality – defined as the set of physical properties that relate to its spinnability into yarn, contribute to textile performance and enhance end-product quality. The three important parameters defining fiber length are average length of fiber bundle or Upper Half Mean (UHM) length, length uniformity of a population of fibers or fiber Uniformity Index (UI), and content of short fibers in a population of cotton fibers or short fiber content / index (SFC or SFI). Longer fibers can be processed at greater efficiencies and produce finer yarns while shorter fibers require increased twisting during spinning, causing low-strength, poor-quality yarns (Perkins Jr. et al., 1984). In addition, lower uniformity in fiber length and higher proportion of short fibers increase manufacturing waste and decrease spinning efficiency during yarn processing. Advances and sophistication in modern spinning technology often incentivize cotton fiber with improved length and uniformity, in addition to other fiber quality properties.

To better understand the genetic architecture of three important fiber quality parameters in the two most cultivated species of cotton, we constructed a set of reciprocal interspecific advanced backcross populations using Acala Maxxa (*G. hirsutum*) and Pima S6 (*G. barbadense*) as parents, tiling 71.48% of the Acala Maxxa genome in Pima S6 (hereafter GB) background and 78.72% of the Pima S6 genome in the Acala Maxxa (hereafter GH) background. With advanced backcross populations and NILs, we found 58 QTLs for these three important fiber length parameters.

Materials and Methods

Population development

Plant materials used in this study were developed from a set of reciprocal crosses between *Gossypium hirsutum* acc. Acala Maxxa and *G. barbadense* acc. Pima S6 (both inbred lines). These genotypes have been extensively used to produce several molecular tools and resources including BAC libraries and Illumina genome sequences. Reciprocal advanced backcross populations were developed by first crossing the parents in a two-way cross (Acala Maxxa (♀)× Pima S6 (♂) – GH background; and Pima S6 (♀) × Acala Maxxa (♂) – GB background), then independently backcrossing F_1 plants to the respective female parent to create 300 to 400 BC_1 progenies for each cross. The backcrossing scheme included planting only one seed from each preceding backcross to generate the next generation (Figure 3.1).

After five generation of backcrossing, 179 BC_5F_1 plants from the GH background and 190 BC_5F_1 plants from the GB background were self-pollinated and a total of 369 BC_5F_2 families were grown at Iron Horse Farm (IHF), Watkinsville, Georgia in 2019 and 2021 and at Southwest Georgia Research Station (Plains), Plains, Georgia in 2021 under cultural conditions consistent with commercial irrigated cotton production. Individual BC_5F_2 plants that contained only one introgressed segment from the donor parent were deemed as NILs. We identified a total of 397 NILs in the GH background and a total of 423 NILs in the GB background. Selfed seeds of these NILs (BC_5F_3 seeds) were grown at IHF and Plains in 2021 under cultural conditions consistent with commercial irrigated cotton production.

Phenotypic evaluation and data analysis

Two replications of each BC_5F_2 families and NILs were planted in a randomized complete block design (RCBD) in three environments (IHF-2019, IHF-2021 and Plains-2021). Six replications each of the two parents were included in all three environments. Fiber samples were collected by harvesting 25 bolls from each plot, ginned in a laboratory gin, evaluated by HVI (Cotton Incorporated Textile Service Laboratory, Cary, NC). Phenotypic data was collected and analyzed for three important fiber length parameters: UHM, UI and SFI.

All statistical analyses were conducted in R programming language. Single marker analyses were done in R/qtl (Broman & Sen, 2009). The significance threshold was set to LOD of 3, to mitigate the multiple-comparison problem. Filtration of significant markers adopts the method proposed by Szalma et al. (2007b). If several markers on the same introgressed segment show significant association with phenotype, the most significant one was reported. For the co-segregation of multiple introgressions, the QTL location is examined as follows. First, if multiple families show significance for the trait and carry overlapping introgression, the introgression is considered to carry QTL. Second, if the co-segregation of introgressions is in single families, the most significant introgression is considered to carry QTL.

Phenotypic variance explained by each locus was reported by taking the most significant marker as independent variable and phenotypic value as dependent variable in R (R Core Team 2021). Additive effects were estimated by half the difference of phenotypic values between the lines carrying the homozygous introgression and lines not carrying the introgression. Dominance effects were estimated by the difference of phenotypic values between the lines carrying the heterozygous introgression and the remaining lines that do not carry the introgression. If multiple or overlapping introgressions were present at both homozygous and heterozygous state, the

estimation of additive effects utilized the lines carrying the introgression at homozygous state only and, the estimation of dominance utilized the lines carrying the introgression at heterozygous state only. Dunnett's test was performed to compare means of individual lines with the recurrent parent and with five cotton cultivars recommended in the state of Georgia.

Gene actions for QTLs were determined by calculating the degree of dominance (absolute values) for every QTL that has both additive and dominance effects. The degree of dominance is the ratio of dominance effect to additive effect (d/a) of the QTL and based on this ratio, gene action of the QTLs can be categorized as (i) additive ($0 < d/a < 0.2$) (ii) partially dominant ($0.2 < d/a < 0.8$) (iii) dominant ($0.8 < d/a < 1.2$ and (iv) over-dominant ($d/a > 1.2$). QTLs with dominant and over-dominant effects are considered to have heterotic effect or heterozygous advantage.

Identification of common QTLs

Common QTL is defined as either the same marker is detected in the two reciprocal populations, or two different markers are detecting exactly the same introgression(s) in each population. Correspondence between QTLs for a trait across the entire genome is inferred using the hypergeometric probability function. The model was adopted from (Feltus et al., 2006): *p* is the probability of non-random correspondence of QTLs being compared for a given trait, *n* is the number of comparable intervals which is calculated by dividing the total genome size by average introgression size in both populations; *m* is the number of common QTLs; *l* is the number of QTLs in the GH background; *s* is the number of QTLs in the GB background. The same model was also adopted to detect correspondence between QTLs reported in this study with those previously published. In this case, *l* is the total number of QTLs identified in the larger sample (study reporting higher number of QTLs) and *s* is the number of QTLs identified in the smaller sample.

$$p = \frac{\binom{l}{m}\binom{n-1}{s-m}}{\binom{n}{s}}$$

Candidate gene identification

In silico annotation was performed on the identified QTLs to look for candidate genes related to flowering habit in cotton. For each QTL identified in the study, the genomic region spanning 50 kb on each side of the most significantly associated marker was used for in silico analysis. The DNA sequence from this tightly linked region was used to look for *G. hirsutum* genes in the CottonGen database and these genes were then analyzed for biological functions, with particular focus on fiber growth and development.

Results

Genomic composition of NILs and ABLs

The distribution of SNP markers in the two reciprocal ABLs is presented in Table 2.1. Genomic distribution of the introgressed chromosomal segments in ABLs is shown in Table 2.2 and Figure 2.3 and 2.4, while coverage of donor genome by the NILs is shown in Table 3.1 and Figures 3.2 and 3.3. Relevant information about the genomic composition of NILs and ABLs are presented in the Results section of chapters 2 and 3.

Phenotypic performance of parents and experimental populations

The phenotypic performance of the two parents, reciprocal backcross populations, and reciprocal NIL populations is shown in Figure 4.1. The distribution of traits was approximately normal (Shapiro and Wilk test; $p > 0.05$) and typical of quantitative inheritance. Pima S6

outperformed Acala Maxxa in all three environments (Table 4.1, Table 4.2, Figure 4.1). Both advanced backcross and NIL populations in the GB background performed better than in the reciprocal background for fiber length while they did not differ much for fiber uniformity ratio. Both population types had improved short fiber content in the GB background relative to the GH background. Transgressive segregation is seen for all three traits for all populations across all environments tested. Transgressive segregants outperforming both parents were identified in both population types and both backgrounds (Figure 4.1).

To identify the effect of genotypes and environment in the overall performance of the advanced backcross populations and the NILs, we conducted analysis of variance (ANOVA) keeping all variables as fixed factors. Results showed significant effects of both genotype (GEN), and genotype-by-environment (GXE) (Table 4.3). GEN captured the most variation for all traits in both population types. GXE also captured significant amount of variation in the phenotypes and thus precluded the use of combined phenotypic values in identification of QTLs for these traits. Thus, marker trait association and identification of fiber quality QTLs for the three traits under study was performed separately for each environment tested.

Marker trait association and overview of QTLs

A total of 58 marker trait associations were identified (32 in the GH background (Table 4.4) and 26 in the GB background (Table 4.5)). Phenotypic variances explained by these QTLs ranged from 1.19% to 25.85%. Among the 58 QTLs identified, 17 were of large effect, explaining > 10% of total phenotypic variation while the remaining 41 were of small effect (explaining <10% of total phenotypic variation). The highest number of QTLs (22) was identified for SFI, followed by UHM (21) and UI having the lowest number of identified QTLs (15).

QTLs conferring fiber length (UHM)

A total of 13 QTLs were identified for fiber length in the GH background, explaining 2.51 % to 12.94% of total phenotypic variation (Table 4.4). Of the 13 QTLs, six were in the At subgenome and seven were in the Dt subgenome. Eight of the 13 QTLs were small effect QTLs while the remaining five were major effect QTLs. Seven of the 13 QTLs decreased fiber length while only six increased fiber length. One QTL on chromosome 1 was identified at two environments, IHF-2021 and Plains-2021. Two other QTLs, one on chromosome 2 and the other on chromosome 21, were also identified in two environments (IHF-2019 and IHF-2021).

In the GB background, a total of eight QTLs were identified for fiber length, explaining 3.21% to 10.45% of total phenotypic variation (Table 4.5). Among the eight QTLs, three each were identified at IHF-2019 and Plains-2021 and two were identified at IHF-2021. Seven of the eight QTLs were identified in the At subgenome and only one was identified in the Dt subgenome. Five of the eight QTLs decreased fiber length while the remaining three QTLs increased fiber length. One QTL on chromosome 11 was identified in two environments (IHF-2021 and Plains-2021).

QTLs conferring fiber uniformity index (UI)

A total of 10 QTLs were identified for fiber length in the GH background, explaining 1.19 % to 10.77% of total phenotypic variation (Table 4.4). Of the 10 QTLs, five were in the At subgenome and the remaining five were in the Dt subgenome. Eight of the 13 QTLs were small effect QTLs while only two were major effect QTLs. Half of QTLs identified in the GH background increased fiber uniformity ratio while the other half decreased fiber uniformity ratio.

In the GB background, a total of five QTLs were identified for fiber uniformity ratio, explaining 1.37% to 12.44% of total phenotypic variation (Table 4.5). Among the five QTLs, two each were identified at IHF-2019 and IHF-2021 and one was identified at Plains-2021. Three of the eight QTLs were identified in the At subgenome and two were identified in the Dt subgenome. All five QTLs identified in the Pima S6 background decreased fiber uniformity ratio.

QTLs conferring short fiber index (SFI)

In the GH background, a total of nine QTLs were identified for short fiber index, explaining 2.96% to 12.69% of total phenotypic variation (Table 4.4). Among the nine QTLs, one was identified at IHF-2019, six were identified at IHF-2021 and two were identified at Plains-2021. Three of the nine QTLs were identified in the At subgenome and the remaining six were identified in the Dt subgenome. All QTLs except the one on chromosome 22 increased short fiber content.

A total of 13 QTLs were identified for fiber length in the GB background, explaining 2.97% to 25.85% of total phenotypic variation (Table 4.5). Of the 13 QTLs, eight were in the At subgenome and five were in the Dt subgenome. Six of the 13 QTLs were small effect QTLs while the remaining seven were major effect QTLs. All but one QTLs increased short fiber content. One QTL on chromosome 5 was identified in all three environments. Another QTL on chromosome 14 was identified at two environments (IHF-201 and Plains-2021).

Subgenomic distribution of QTLs

The QTLs for the three fiber quality parameters were almost evenly distributed across the two subgenomes in the GH background (14 in the At subgenome vs 18 in the Dt subgenome) while comparatively more QTLs were located in the At subgenome (18) than in the Dt subgenome (8)

in the GB background (Table 4.6). QTLs were identified in all chromosomes except chromosomes 3, 13 and 17 (Table 4.7). In the GH background, QTLs were identified in 19 of 26 chromosomes while in the GB background, they were identified in 15 chromosomes. In total, chromosomes 1 and 5 carried the most QTLs (6 each) followed by chromosome 21 (5 QTLS) and chromosomes 2 and 11 (4 QTLs each) (Table S4). In the GH background, chromosomes 1 and 21 carried the most QTLs (4 each) while in the GB background, chromosome 5 carried the most QTLs (5) followed by chromosomes 11 and 14 (3 QTLs each).

Clustering of fiber quality QTLs

Certain genomic regions in the cotton genome are known to harbor QTLs for two or more traits. Such regions carrying multi-trait QTLs within a 20 Mb span were considered QTL clusters or QTL hotspots. In agreement with previously published reports, we observed similar nonrandom distribution of fiber quality QTLs throughout the cotton genome. In total, we observed three such clusters in the GH background and three more clusters in the Pima S6 background (Table 4.8). One cluster in the GH background on chromosome 21 (cQTL.Gh.21.1) contained three QTLs while the other two clusters contained two QTLs each. In the GB background, cluster cQTL.Gb.05.1 contained four QTLs (three for SFI and one for UM). Another cluster on chromosome 14 contained three QTLs (two SFI QTLs and one UI QTL) and the cluster on chromosome 2 contained two QTLs (one for UHM and one for UI).

Discussion

Introgressive breeding approaches not only introduce a preponderance of novel allelic variation into cultivated gene pools, but interspecific populations developed using these

approaches can be widely used for molecular dissection of complex fiber yield and quality parameters. Past studies have focused on investigating the effects of GB chromatin segments introgressed into GH genome, however, the reverse has not been routinely studied. In the current study, we developed a reciprocal set of advanced backcross lines and NILs selected from among the selfed progenies of these advanced backcross lines and tested these populations to assess the effects of reciprocal chromatin transfer on three important fiber length parameters – UHM, UI and SFI. We identified a total of 58 marker-trait associations for these three traits with variable genetic effects. This study adds to the resources and observations available to study the quantitative nature of fiber quality traits reciprocally in two elite cotton backgrounds.

Performance of ABL and NIL populations

NILs and advanced backcross populations in both backgrounds showed average phenotypes consistent with their recurrent parent (Figure 4.1). Albeit the average performance of these two populations behaved like their recurrent parents, which is expected given the exceptionally large proportion of their genome coming from the recurrent parent, a large amount of variation was observable for all three parameters. Presence of transgressive segregants in both directions for all three fiber length parameters suggests that the chromatin segments introgressed from the donor parents have effects that could significantly alter the performance of individual lines. In fact, these alterations and their effects is shown by the phenotypic performance (QTL effects) of the lines carrying respective introgression from the donor parent.

Effect of species background

The total number of QTLs identified for the three fiber length parameters did not differ by much (32 in GH background vs 26 in the GB background) in the two backgrounds. While identification of similar total numbers of total QTLs for comparable traits in the reciprocal backgrounds might suggest the involvement of similar number of genes in controlling these traits, many other findings indicate the effect of species background on effects, locations, and patterns of these QTLs.

A perplexing observation is the complete absence of QTLs identified with opposite phenotypic effects at corresponding locations in the reciprocal genetic backgrounds. In principle, one would expect the majority of QTLs to show such reciprocity if alternative alleles at a QTL show additive or dominant-recessive effects. Limited correspondence of identified QTLs in the two backgrounds could be a result of several factors. One factor may be the small phenotypic effects of most of the identified QTL, increasing the likelihood that one or both members of a reciprocal pair elude detection (Broman, 2001). Another intriguing factor that could account for some failures to identify correspondence of QTLs in the reciprocal backgrounds, especially in advanced backcross lines with multiple introgressed chromosomal segments, is epistasis. Interaction between introgressed loci might result in underestimation of their effects which might have resulted in some QTLs failing to reach the biometric thresholds required to declare them as QTLs per se. The widespread observation that fiber quality parameters generally have high heritability (Fang et al., 2014) suggests a limited role of epistasis, but it could contribute to failures to identify reciprocal QTLs with relatively small effects (Chandnani et al., 2018). This might be the case here as most of the QTLs detected in this study show low genetic contribution to the total phenotypic variation explained by the phenotype.

Another interesting observation for most of the QTLs detected in our study is the direction of the effect in reciprocal backgrounds. In general, one would expect chromatin segments introgressed from GB to GH background to show significant positive effects for fiber length and uniformity as Pima S6 is well known for its superior quality for these fiber quality parameters. Surprisingly, 7 of 13 (53.84%) of the QTLs introgressed from Pima S6 decreased fiber length and only 6 QTLs showed positive effects for fiber length in the GH background. This suggests that only certain genomic regions in the GB genome might be useful for improvement of quality parameters in the GH background. Similarly, in the reciprocal background, contrary to what most would expect, not all QTLs introgressed from Acala Maxxa reduced fiber length. A total of 3 (of 8) QTLs for UHM increased fiber length while the remaining 5 decreased fiber length consistent with the parental phenotypes. These results suggest that novel genetic / genomic combinations of chromatin segments might result in favorable phenotypic effects even if the introgressed chromatin is from an inferior genotype, further strengthening the notion that introgressive breeding is a valuable source of novel genetic variation.

QTL clustering

Colocalization (clustering) of QTL with common allele effects makes marker-assisted selection much more efficient. QTL clusters for the three fiber quality traits studied were prevalent and were quite variable between the two backgrounds, with clusters identified in different sets of chromosomes in the two backgrounds (Table 4.8). In the GH background, cluster cQTL.Gh.19.1 on chromosome 19 contained two QTLs, one each for UI and SFI, both of which had unfavorable alleles introgressed from Pima S6 into Acala Maxxa. Similarly, the second cluster cQTL.Gh.21.1 identified in the GH background had three QTLs all of which rendered negative effects in the

recipient background. While the first two of the three QTLs clusters carried QTLs all of which carried unfavorable alleles to the GH background, the third cluster cQTL.Gh.23.1 carried two QTLs, one for UHM and one for SFI, both of which conferred favorable alleles to GH background. Negative selection on chromosomes 19 and 21 may purge unfavorable alleles whereas positive selection on chromosome 23 may enrich favorable alleles from GB to GH.

In the GB background, cluster cQTL.Gb.02.1 on chromosome 2 carried two QTLs, one of which (UHM) conferred positive effect on Pima S6 while the other (UI) carried unfavorable allele. The largest cluster identified in this study cQTL.Gb.05.1 on chromosome 5 carried four QTLs, one of which had favorable effects towards Pima S6 and three contributed unfavorable alleles. While clusters of QTLs carrying both favorable and unfavorable alleles are not uncommon, the presence of both makes fiber quality improvement by conventional approaches in GB background more challenging than in GH background because of greater linkage drag for unfavorable alleles in GB (Chandnani et al., 2018). NILs with one or a very few QTLs confined to a small introgressed genomic region would serve as a good starting material to get rid of such linkage drag in the GB background.

Subgenomic distribution of fiber quality QTLs

With regard to the distribution of QTL in A and D subgenomes, many prior studies have concluded that more QTL for growing period, yield and fiber quality were distributed in Dt subgenome than in At subgenome (Adhikari et al., 2017; Chandnani et al., 2018; J. Rong et al., 2007; Joseph I. Said et al., 2013; Wang et al., 2013). However, Shen et al. (2005) and Lin et al. (2005) reported that more QTL for fiber quality were located in At subgenome than in Dt subgenome. In the current study, we identified more QTLs in the At subgenome than in the Dt

subgenome. For NILs and advanced backcross populations in the GH background, the subgenomic affinities of QTLs were not significant ($p>0.05$) but nominally agreed with most previous findings, with Dt subgenome harboring 22% more QTLs than the At subgenome. In the reciprocal background, however, significantly higher ($p<0.05$) number of QTLs were identified in the At subgenome. QTLs for individual traits also followed the overall trend with almost similar or a greater number of QTLs in the Dt subgenome in the GH background while significantly higher number of QTLs in the At subgenome in the GB background. These results also show the effect of genetic background on subgenomic distribution of QTLs.

Stability of fiber quality QTLs

While different loci associated with the same trait under various environments might suggest interaction between genotype and environment, QTLs being detected across environments might indicate environmental stability. Although most QTLs identified in this study were single-environment QTLs, some QTLs stood out to be stable and were identified in at least two of the three environments tested (Results section and Tables 4.4 and 4.5). In addition to one genomic region on chromosome 5 consistently identified for SFI on all three environments, we also found four other genomic regions on chromosomes 1, 2, 11 and 21 consistently associated with UHM in two different environments.

Similarity with QTLs previously reported

The genetic composition of our experimental populations provides a platform to identify novel small-effect QTLs in addition to major QTLs. Since NILs serve as a resource to not only identify marker trait associations but also an important tool to verify QTLs previously identified,

mostly those using early generation populations, the correspondence identified here could be used as a means of validation of previously published QTLs. Given that there are hundreds of studies reporting fiber quality QTLs and owing to the genetic structure of our experimental populations, we mostly compared our results with populations of similar genomic composition (Brown et al., 2019; P. W. Chee et al., 2005; Xavier Draye et al., 2005; Yu et al., 2013). We also performed elaborate statistical comparisons with other previous reports on the correspondence of QTLs identified in our study.

First, we studied the correspondence of QTLs identified in our study with those reported previously in populations of similar genomic composition. Using an interspecific backcross-self approach, (P. W. Chee et al., 2005) identified a QTL (introgressed from GB to GH) for UHM on chromosome 2 (FL02.1), which is located within the bounds of the same GB chromatin segment introgressed in GH background in our study (Table 4.4). The introgressed segment in chromosome 2 (spanning from 47155918 to 78271231 bp) harbors the closest marker (pGH399a) tagged to this QTL. This QTL (FL02.1) had an additive effect of reducing fiber length and the two QTLs (qUHM.gh.19-IH.01 and qUHM.Gh.21-IH.02) identified in our study residing in the same introgressed chromatin segment also have a total effect of reducing fiber length (Table 4.4). Another fiber length QTL identified in chromosome 7 in the same study was also identified in our study. QTL qUHM.Gh.19-PL.02 identified at Plains-2021 was identified in the same introgressed segment (5.62 Mb to 17.41 Mb) as QTL FL07.1 whose closest marker G1185a is located just 8 Mb upstream of the most significant marker identified in our study. This QTL identified in our study has an additive effect of increasing fiber length by 0.08 mm (Table 4.4). One more fiber length QTL identified in the same study was also identified in the current study. A QTL on chromosome 15 (FL15.1) was identified in the same chromosomal segment introgressed from GB

to GH. The marker (pAR077a) closely linked to this QTL was only 10 Mb upstream from the marker closely linked to the QTL (qUHM.Gh.19-IH.02) identified in our study (Table 4.4). These results provide a stout two-way validation of the QTLs identified in these genomic locations; firstly, the additional evidence identified in this study bolstering effects of QTLs from P. W. Chee et al. (2005) and secondly, previously reported QTLs forming a basis of verification for QTLs identified in our study.

In another study (Brown et al., 2019), the authors used bulked sister lines (BSLs) to study the effect of qFL-Chr.25, a fiber length QTL introgressed from GB to GH and originally identified in 'Sealand 883', in four different commercial genotypic backgrounds. This is an important QTL as it significantly increased the staple length by 1.4 mm in DP50 and by 1 mm in GA089 background (Brown et al., 2020; Brown et al., 2019). We also identified a fiber length QTL in the same region in our advanced backcross population in the GB background, with the most significant marker 11 Mb downstream of the most significant marker in the previous study. This is an interesting observation because in our study we could not find any reciprocity for this QTL, however, we were able to observe reciprocity of our result with the one previously published, both with similar effects. While the QTL previously identified had a significant effect in increasing fiber length, the one identified in the current study had positive yet smaller effect in increasing fiber length. One possible reason for the lower effect observed in our study might be the distance of the most significant marker from the actual location of the QTL. Nevertheless, these results suggest that these four QTLs identified in both these studies in similar populations but is quite different environment conditions are stable QTLs with promise in improving fiber length in cultivated upland cotton.

While correspondence of individual marker-trait associations may reflect, for example, a gene for which the recurrent parent, Acala Maxxa or Pima S6, has a rare allele, non-random patterns of association across an entire genome can reflect other properties such as convergent domestication (Paterson et al., 1995). To investigate such genome wide non-random patterns of association, we also conducted an elaborate study on the correspondence of QTLs identified in our study with other previous reports, irrespective of the population type. Information on a total of 2050 QTLs related to fiber length, fiber uniformity index and short fiber content reported in 67 previous reports were downloaded from CottonGen (https://www.cottongen.org) and the hypergeometric probability distribution was used to conduct a thorough analysis of QTL correspondence. The hypergeometric function provides a means to infer statistically whether QTLs for a trait are randomly distributed between two populations or environments. Correspondence was identified in 25 of the 67 previously reported studies for the QTLs reported in GH background as well as in GB background (Table 4.9). Common QTLs were identified in most of the 67 reports listed, however, only 25 significant P values based on the hypergeometric distribution suggest that across the genome, this correspondence is not sufficient to infer a non-random distribution of QTLs between published reports.

In silico annotation of potential candidate genes

Availability of cotton reference genomes (Paterson et al., 2012; Zhang et al., 2015) enabled us to scrutinize physical regions surrounding the identified QTLs for genes / gene families known or suspected to affect fiber quality parameters in cotton. This investigation was limited to tightly linked regions i.e., 1 Mb on both sides of the SNP marker that is most significantly associated with the QTLs (Table 4.10). Specifically, we targeted the four QTLs (qUHM.Gh.19-IH.01,

qUHM.Gh.19-PL.02, qUHM.Gh.19-IH.02, qUHM.Gb.19-IH.03) that were identified in the same introgressed regions as the ones reported in (P. W. Chee et al., 2005) and (Brown et al., 2019), as these previously reported QTLs have been thoroughly characterized, tested, and registered (Brown et al., 2020).

On chromosome 2, the nearest gene of interest (Gh_A02G1317) to the fiber length QTL qUHM.Gh.19-IH.01 was a member of the *CESA* zinc finger ring protein family (*CES_Znf_RING*), identified to be involved in cellulose biosynthesis because of the presence of UDP-forming *CELLULOSE SYNTHASE A CATALYTIC SUBNIT 2* in the members of this family. In addition to cellulose biosynthesis, this gene family members have been found to be active in protein and zinc ion binding in cell wall. In cotton, these genes have been associated with primary and secondary cell wall biosynthesis in developing cotton fibers (Hee Jin Kim & Barbara A. Triplett, 2001). Another gene (Gh_A02G1330) near the same QTL on chromosome 2 is identified to be a *PEPCK-type* gene (phosphoenolpyruvate carboxykinase) known to be involved in C4 photosynthetic carbon assimilation cycle. In cotton, this had been identified to be an important protein in fiber initiation (Ma et al., 2014)

On chromosome 7, one of the genes of interest near the QTL qUHM.Gh.19-PL.02 is *psaL* gene (Gh_A07G0818), which is involved in Photosystem I reaction center and has been found to be involved in photosynthesis and carbon assimilation, and thus could be a potential gene acting in primary and secondary cell wall biosynthesis. The most important gene (Gh_A07G0842) identified near the most significant marker for the QTL identified in chromosome 7 is the *GAUT* family genes (Galacturonosyltransferases), which are involved in pectin biosynthesis. Pectin is an essential component for secondary cell wall development. In a study of *GAUT-like* gene families (*GATL*), *GhGATL* genes were found to act on pectin synthesis to regulate fiber development

(Zheng et al., 2020). On both chromosomes 15 and 25, the nearest genes of interest to the most significant markers associated with the respective QTLs are the *LRR* receptor like protein kinases (Table 4.10). *LRR* receptor-like kinase genes are known to be expressed at specific fiber development stages with the highest expression during cellulose synthesis for secondary cell wall formation (Li et al., 2005) and have shown evidence of positive selection during cotton fiber improvement for quality and yield (Zhang et al., 2015). While the annotated functions of these genes closely align with fiber growth and development, further exploration and validation is necessary to confirm their roles.

Conclusion

The two major species of cotton have been routinely used in cotton breeding and improvement programs to transfer alleles from one species to another, GB being the donor in most cases due to its superiority in fiber length parameters and partly due to high lint yield of the recipient GH. The present study demonstrates the value of GH as a source of favorable alleles for fiber quality traits in the GB background while also reiterating the reciprocal transfer of favorable alleles from GB. Using advanced backcross lines segregating for a few donor chromosome segments as well as a set of reciprocal near-isogenic lines, we showed a strong effect of genetic background on chromatin transfer as well as the effect of these introgressed chromatin on three important fiber length components. The near-isogenic genomic composition of our population provided opportunities to estimate the effects of genomic regions more precisely, but at the cost of epistatic QTL interactions. With one of the major purposes of NILs being the ability to verify the location and effects of QTLs, in addition to their identification, we were able to demonstrate the stability of a few important fiber quality QTLs identified in previous studies. Since the parents

used in creating these populations are both elite lines representing the two major domesticated species of cotton, with their own specialties and differences, the populations we developed are not just a platform to identify genomic locations underpinning fiber quality traits, but also a good starting point from which to study divergence, domestication and evolutionary history of cotton as well as to select superior individual lines from a pool of these immortal lines for breeding purposes as well as for commercialization.

References

Adhikari, J. (2015). *Effects of exotic genotypes and genetic backgrounds on fiber quality and plant architectural traits in Upland cotton (Gossypium hirsutum L.)* University of Georgia].

Adhikari, J., Das, S., Wang, Z., Khanal, S., Chandnani, R., Patel, J. D., Goff, V. H., Auckland, S., Rainville, L. K., & Jones, D. (2017). Targeted identification of association between cotton fiber quality traits and microsatellite markers. *213*(3), 1-15.

Broman, K. W. (2001). Resource Review of Statistical Methods for QTL Mapping in Experimental Crosses. *30*(7), 44-52.

Broman, K. W., & Sen, S. (2009). *A guide to QTL mapping with R/qtl* [Bibliographies Online Non-fiction Electronic document]. Springer.

Brown, N., Kumar, P., Khanal, S., Singh, R., Suassuna, N. D., McBlanchett, J., Lubbers, E., Jones, D., Paterson, A. H., & Chee, P. W. (2020). Registration of eight upland cotton (Gossypium hirsutum L.) germplasm lines with qFL-Chr.25, a fiber-length QTL introgressed from Gossypium barbadense. *Journal of Plant Registrations*, *14*(1), 57-63.

Brown, N., Kumar, P., Singh, R., Lubbers, E., Campbell, B. T., Myers, G. O., Wright, R. J., Subramani, J., Jones, D., Paterson, A. H., & Chee, P. W. (2019). Evaluation of a Chromosome Segment from Gossypium barbadense Harboring the Fiber Length QTL qFL-Chr.25 in Four Diverse Upland Cotton Genetic Backgrounds. *Crop Science*, *59*(6), 2621-2633.

Chandnani, R., Kim, C., Guo, H., Shehzad, T., Wallace, J. G., He, D., Zhang, Z., Patel, J. D., Adhikari, J., Khanal, S., & Paterson, A. H. (2018). Genetic Analysis of Gossypium Fiber Quality Traits in Reciprocal Advanced Backcross Populations. *The Plant Genome*, *11*(1), 170057.

Chee, P. W., & Campbell, B. T. (2009). Bridging Classical and Molecular Genetics of Cotton Fiber Quality and Development. In A. H. Paterson (Ed.), *Genetics and Genomics of Cotton* (pp. 283-311). Springer US.

Chee, P. W., Draye, X., Jiang, C. X., Decanini, L., Delmonte, T. A., Bredhauer, R., Smith, C. W., & Paterson, A. H. (2005). Molecular dissection of phenotypic variation between Gossypium hirsutum and Gossypium barbadense (cotton) by a backcross-self approach: III. Fiber length [Journal Article]. *TAG Theoretical and Applied Genetics*, *111*(4), 772-781.

Draye, X., Chee, P., Jiang, C.-X., Decanini, L., Delmonte, T. A., Bredhauer, R., Smith, C. W., & Paterson, A. H. (2005). Molecular dissection of interspecific variation between Gossypium hirsutum and G. barbadense (cotton) by a backcross-self approach: II. Fiber fineness [Article]. *Theoretical & Applied Genetics*, *111*(4), 764-771.

Fang, D. (2015). Molecular Breeding. In. https://doi.org/10.2134/agronmonogr57.2013.0027

Fang, D. D., Jenkins, J. N., Deng, D. D., McCarty, J. C., Li, P., & Wu, J. (2014). Quantitative trait loci analysis of fiber quality traits using a random-mated recombinant inbred population in Upland cotton (Gossypium hirsutum L.). *BMC Genomics*, *15*(1), 397.

Fang, L., Wang, Q., Hu, Y., Jia, Y., Chen, J., Liu, B., Zhang, Z., Guan, X., Chen, S., Zhou, B., Mei, G., Sun, J., Pan, Z., He, S., Xiao, S., Shi, W., Gong, W., Liu, J., Ma, J., . . . Zhang, T. (2017). Genomic analyses in cotton identify signatures of selection and loci associated with fiber quality and yield traits. *Nature Genetics*, *49*(7), 1089-1098.

Feltus, F. A., Hart, G. E., Schertz, K. F., Casa, A. M., Kresovich, S., Abraham, S., Klein, P. E., Brown, P. J., & Paterson, A. H. (2006). Alignment of genetic maps and QTLs between inter- and intra-specific sorghum populations. *Theor Appl Genet*, *112*(7), 1295-1305.

Khanal, S. (2018). *Dissection of quantitative variation introgressed into bermudagrass and upland cotton* University of Georgia].

Kim, H. J., & Triplett, B. A. (2001). Cotton Fiber Growth in Planta and in Vitro. Models for Plant Cell Elongation and Cell Wall Biogenesis. *Plant Physiology*, *127*(4), 1361-1366.

Kooke, R., Wijnker, E., & Keurentjes, J. J. (2012). Backcross populations and near isogenic lines. *Methods Mol Biol*, *871*, 3-16.

Li, Y. L., Sun, J., & Xia, G. X. (2005). Cloning and characterization of a gene for an LRR receptor-like protein kinase associated with cotton fiber development. *Mol Genet Genomics*, *273*(3), 217-224.

Lin, Z., He, D., Zhang, X., Nie, Y., Guo, X., Feng, C., & Stewart, J. M. (2005). Linkage map construction and mapping QTL for cotton fibre quality using SRAP, SSR and RAPD [Article]. *Plant Breeding*, *124*(2), 180-187.

Ma, Q., Wu, M., Pei, W., Li, H., Li, X., Zhang, J., Yu, J., & Yu, S. (2014). Quantitative phosphoproteomic profiling of fiber differentiation and initiation in a fiberless mutant of cotton. *BMC Genomics*, *15*(1), 466.

Monforte, A. J., & Tanksley, S. D. (2000). Development of a set of near isogenic and backcross recombinant inbred lines containing most of the Lycopersicon hirsutum genome in a L. esculentum genetic background: A tool for gene mapping and gene discovery. *Genome*, *43*(5), 803-813.

Paran, I., & Zamir, D. (2003). Quantitative traits in plants: beyond the QTL. *Trends in Genetics*, *19*(6), 303-306.

Paterson, A. H., Boman, R. K., Brown, S. M., Chee, P. W., Gannaway, J. R., Gingle, A. R., May, O. L., & Smith, C. W. (2004). Reducing the Genetic Vulnerability of Cotton. *Crop Science*, *44*(6), 1900-1901.

Paterson, A. H., Lin, Y. R., Li, Z., Schertz, K. F., Doebley, J. F., Pinson, S. R., Liu, S. C., Stansel, J. W., & Irvine, J. E. (1995). Convergent domestication of cereal crops by independent mutations at corresponding genetic Loci. *Science*, *269*(5231), 1714-1718.

Paterson, A. H., Wendel, J. F., Gundlach, H., Guo, H., Jenkins, J., Jin, D., Llewellyn, D., Showmaker, K. C., Shu, S., Udall, J., Yoo, M.-j., Byers, R., Chen, W., Doron-Faigenboim, A., Duke, M. V., Gong, L., Grimwood, J., Grover, C., Grupp, K., . . . Schmutz, J. (2012). Repeated polyploidization of Gossypium genomes and the evolution of spinnable cotton fibres. *Nature*, *492*(7429), 423-427.

Perkins Jr., H. H., Ethridge, D. E., & Bragg, C. K. (1984). Fiber. In *Cotton* (pp. 437-509).

Rong, J., Feltus, F. A., Waghmare, V. N., Pierce, G. J., Chee, P. W., Draye, X., Saranga, Y., Wright, R. J., Wilkins, T. A., May, O. L., Smith, C. W., Gannaway, J. R., Wendel, J. F., & Paterson, A. H. (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), 2577-2588.

Said, J. I., Knapka, J. A., Song, M., & Zhang, J. (2015). Cotton QTLdb: a cotton QTL database for QTL analysis, visualization, and comparison between Gossypium hirsutum and G. hirsutum × G. barbadense populations. *Mol Genet Genomics*, *290*(4), 1615-1625.

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

Said, J. I., Lin, Z., Zhang, X., Song, M., & Zhang, J. (2013). A comprehensive meta QTL analysis for fiber quality, yield, yield related and morphological traits, drought tolerance, and disease resistance in tetraploid cotton. *BMC Genomics*, *14*(1), 776.

Said, J. I., Song, M., Wang, H., Lin, Z., Zhang, X., Fang, D. D., & Zhang, J. (2015). A comparative meta-analysis of QTL between intraspecific Gossypium hirsutum and interspecific G. hirsutum × G. barbadense populations. *Mol Genet Genomics*, *290*(3), 1003-1025.

Shen, X., Guo, W., Zhu, X., Yuan, Y., Yu, J. Z., Kohel, R. J., & Zhang, T. (2005). Molecular mapping of QTLs for fiber qualities in three diverse lines in Upland cotton using SSR markers [Article]. *Molecular Breeding*, *15*(2), 169-181.

Szalma, S. J., Hostert, B. M., LeDeaux, J. R., Stuber, C. W., & Holland, J. B. (2007). QTL mapping with near-isogenic lines in maize [Article]. *Theoretical & Applied Genetics*, *114*(7), 1211-1228.

Wang, M., Tu, L., Lin, M., Lin, Z., Wang, P., Yang, Q., Ye, Z., Shen, C., Li, J., Zhang, L., Zhou, X., Nie, X., Li, Z., Guo, K., Ma, Y., Huang, C., Jin, S., Zhu, L., Yang, X., . . . Zhang, X. (2017). Asymmetric subgenome selection and cis-regulatory divergence during cotton domestication. *Nature Genetics*, *49*(4), 579-587.

Wang, Z., Zhang, D., Wang, X., Tan, X., Guo, H., & Paterson, A. H. (2013). A Whole-Genome DNA Marker Map for Cotton Based on the D-Genome Sequence of Gossypium raimondii L. *G3 Genes|Genomes|Genetics*, *3*(10), 1759-1767.

Yu, J., Zhang, K., Li, S., Yu, S., Zhai, H., Wu, M., Li, X., Fan, S., Song, M., Yang, D., Li, Y., & Zhang, J. (2013). Mapping quantitative trait loci for lint yield and fiber quality across environments in a Gossypium hirsutum × Gossypium barbadense backcross inbred line population. *Theor Appl Genet*, *126*(1), 275-287.

Zhang, J. F., Percy, R. G., & McCarty, J. C. (2014). Introgression genetics and breeding between Upland and Pima cotton: a review. *Euphytica*, *198*, 1-12.

Zhang, T., Hu, Y., Jiang, W., Fang, L., Guan, X., Chen, J., Zhang, J., Saski, C. A., Scheffler, B. E., Stelly, D. M., Hulse-Kemp, A. M., Wan, Q., Liu, B., Liu, C., Wang, S., Pan, M., Wang, Y., Wang, D., Ye, W., . . . Chen, Z. J. (2015). Sequencing of allotetraploid cotton (Gossypium hirsutum L. acc. TM-1) provides a resource for fiber improvement. *Nature biotechnology*, *33*(5), 531-537.

Zheng, L., Wu, H., Qanmber, G., Ali, F., Wang, L., Liu, Z., Yu, D., Wang, Q., Xu, A., & Yang, Z. (2020). Genome-Wide Study of the GATL Gene Family in Gossypium hirsutum L. Reveals that GhGATL Genes Act on Pectin Synthesis to Regulate Plant Growth and Fiber Elongation. *Genes (Basel)*, *11*(1).

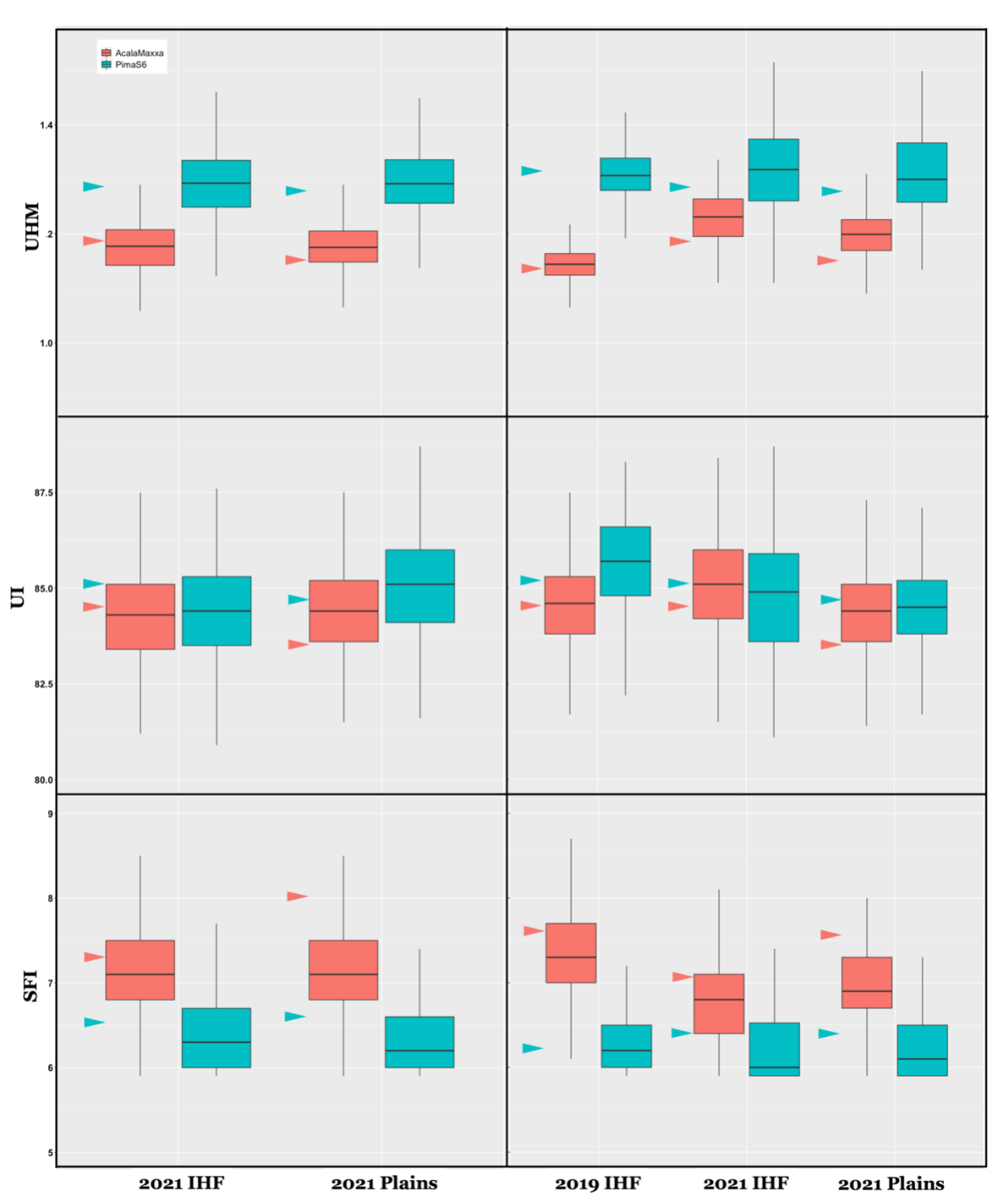

Figure 4.1: Distribution of fiber length for the parents, reciprocal advanced backcrosses and NILs. Panels on the left show distribution of NILs for the two environments tested and panels on the right show distribution of the traits for advanced backcross populations. Green arrows represent Pima S6 parent and red arrows represent Acala Maxxa.

Table 4.1. Summary statistics of fiber quality traits in advanced reciprocal backcross populations (BC_5F_2) and parental lines in three environments. Env denotes the environment in which the populations were evaluated, Sd denotes standard deviation.

Trait	Env	Parental means		Acala Maxxa Background		Pima S6 Background	
		Acala Maxxa	Pima S6	Mean (Sd)	Range	Mean (Sd)	Range
UHM	IHF 2019	01.13	01.32	01.14 (0.03)	01.04 - 01.24	01.31 (0.05)	01.12 - 01.45
	IHF 2021	01.19	01.29	01.23 (0.05)	01.04 - 01.40	01.32 (0.08)	01.11 - 01.52
	Plains 2021	01.14	01.27	01.20 (0.05)	01.06 - 01.41	01.31 (0.07)	01.09 - 01.50
UI	IHF 2019	84.30	85.82	84.54 (1.15)	81.30 - 87.80	85.66 (1.25)	81.60 - 88.30
	IHF 2021	84.26	85.32	85.03 (1.33)	81.20 - 88.40	84.81 (1.55)	81.10 - 88.70
	Plains 2021	83.21	84.54	84.35 (1.22)	80.20 - 87.90	84.50 (1.07)	81.20 - 87.10
SFI	IHF 2019	07.60	06.12	07.32 (0.52)	05.80 - 08.70	06.31 (0.41)	05.90 - 08.20
	IHF 2021	07.10	06.40	06.81 (0.56)	05.90 - 08.50	06.34 (0.59)	05.90 - 08.70
	Plains 2021	07.70	06.42	06.96 (0.49)	05.90 - 08.80	06.30 (0.49)	05.90 - 08.80

Table 4.2. A summary statistics of fiber quality traits in reciprocal NIL populations and parental lines in two environments. Sd denotes sample standard deviation. Env denotes the environment in which the populations were evaluated, Sd denotes standard deviation.

Trait	Env	Parental means		Acala Maxxa Background		Pima S6 Background	
		Acala Maxxa	Pima S6	Mean (Sd)	Range	Mean (Sd)	Range
UHM	IHF 2021	01.19	01.29	1.17 (0.05)	0.96 - 1.37	1.29 (0.07)	1.07 - 1.46
	Plains 2021	01.14	01.27	1.18 (0.05)	1.03 - 1.34	1.30 (0.06)	1.11 - 1.48
UI	IHF 2021	84.26	85.32	84.12 (1.40)	76.30 - 87.50	84.25 (1.56)	76.50 - 88.40
	Plains 2021	83.21	84.54	84.34 (1.23)	80.60 - 87.90	85.06 (1.33)	79.60 - 88.70
SFI	IHF 2021	07.10	06.40	7.23 (0.56)	5.90 - 9.00	6.44 (0.58)	5.94 - 8.60
	Plains 2021	07.70	06.42	7.18 (0.53)	5.93 - 9.40	6.34 (0.45)	5.90 - 8.63

Table 4.3. Analysis of variance of fiber length parameters. DF denote degrees of freedom, SS denotes sum of squares, MS denotes mean sum of squares, PVE denotes percent variance explained by the factor, ABL denotes advanced backcross lines, NIL denotes near-isogenic lines.

			Acala Maxxa background					Pima S6 background				
Trait	Pop	Source	DF	SS	MS	F-value	PVE	DF	SS	MS	F-value	PVE
UHM	ABL	ENV	2	0.99	0.49	262.88***	34.5	2	0.02	0.01	3.68*	0.39
		REP(ENV)	3	0	0	0.14	0.00	3	0.03	0.01	2.92*	0.59
		GEN	172	0.68	0	2.12***	23.69	230	1.92	0.01	2.69***	37.57
		GEN*ENV	287	0.51	0	0.96	17.77	410	1.68	0	1.32*	32.88
		Error	369	0.69	0		24.04	472	1.46	0		28.57
	NIL	ENV	1	0	0	0.0	0.00	1	0	0	1.21	0.00
		REP(ENV)	2	0.01	0	2.8	0.29	2	0	0	0.25	0.00
		GEN	397	1.71	0	2.68***	48.72	407	3.45	0.01	4.39***	65.59
		GEN*ENV	390	0.72	0	1.15	20.51	369	0.8	0	1.12	15.21
		Error	665	1.07	0		30.48	524	1.01	0		19.20
UI	ABL	ENV	2	63.26	31.63	25.81***	4.89	2	269.81	134.91	93.58***	12.65
		REP(ENV)	3	65.83	21.94	17.91***	5.09	3	166.49	55.5	38.5***	7.80
		GEN	172	267.56	1.56	1.27*	20.68	230	427.89	1.86	1.29*	20.05
		GEN*ENV	284	450.96	1.59	1.3*	34.86	406	602.04	1.48	1.03	28.22
		Error	364	446.06	1.23		34.48	463	667.44	1.44		31.28
	NIL	ENV	1	16.29	16.29	10.59**	0.64	1	177.69	177.69	113.42***	6.04
		REP(ENV)	2	9.72	4.86	3.16*	0.38	2	55.15	27.58	17.6***	1.88
		GEN	397	850.04	2.14	1.39***	33.33	407	1190.45	2.92	1.87***	40.48
		GEN*ENV	390	650.56	1.67	1.08	25.52	369	696.67	1.89	1.21*	23.68
		Error	665	1023.07	1.54		40.13	524	820.96	1.57		27.92
SFI	ABL	ENV	2	37.00	81.50	73.24***	13.93	2	0.54	0.27	1.38***	0.20
		REP(ENV)	3	14.46	4.82	19.08***	5.45	3	1.53	0.51	2.61***	0.56

	GEN	172	56.19	0.33	1.29*	21.16	230	84.77	0.37	1.89*	31.08
	GEN*ENV	284	67.70	0.24	0.94	25.50	410	94.33	0.23	1.18	34.59
	Error	357	90.17	0.25		33.96	469	91.55	0.20		33.57
NIL	ENV	1	0.85	0.85	10.59**	0.19	1	2.94	2.94	17.03***	0.82
	REP(ENV)	2	2.20	1.11	3.16*	0.51	2	0.02	0.01	0.05	0.02
	GEN	397	152.47	0.38	1.39***	35.07	406	184.20	0.45	2.63***	52.41
	GEN*ENV	388	106.15	0.27	1.08	24.42	369	81.59	0.22	1.28**	22.77
	Error	657	173.08	0.26		39.81	519	89.50	0.17		24.98

Table 4.4. QTLs for fiber quality traits (UHM, UI and SFI) identified in the Acala Maxxa background. Env denotes the environment in which the populations were evaluated, PVE denotes percent of phenotypic variation explained by the QTL, a denotes additive effect of the QTL, d denotes the dominance effect of the QTL and d/a denotes the ratio of dominance to additive effect.

Trait	ENV	QTL	Marker	LOD	PVE	a	d	d/a
UHM	IHF 2019	qUHM.Gh.19-IH.01	S2_76343685	3.01	10.01		-0.05	
		qUHM.Gh.19-IH.02	S15_61635493	3.14	2.51		-0.03	
		qUHM.Gh.19-IH.03	S20_54729154	3.84	5.81		0.05	
		qUHM.Gh.19-IH.04	S21_47060721	3.19	10.28		-0.06	
		qUHM.Gh.19-IH.05	S24_61134368	3.24	2.51		0.02	
	IHF 2021	qUHM.Gh.21-IH.01	S1_99771502	3.92	4.95	-0.11		
		qUHM.Gh.21-IH.02	S2_77913184	3.88	3.69	-0.05	-0.03	0.60
		qUHM.Gh.21-IH.03	S21_52944095	3.07	10.85		-0.09	
		qUHM.Gh.21-IH.04	S25_43210819	3.93	3.79	0.21		
	Plains 2021	qUHM.Gh.19-PL.01	S1_99771502	4.37	5.32	-0.09		
		qUHM.Gh.19-PL.02	S7_14432661	4.67	3.31	0.08		
		qUHM.Gh.19-PL.03	S9_56212591	4.80	11.18		0.19	
		qUHM.Gh.19-PL.04	S23_42243447	3.08	12.04		0.07	
UI	IHF 2019	qUI.Gh.19-IH.01	S11_13684212	3.74	10.77		0.38	
		qUI.Gh.19-IH.02	S18_24384539	3.71	6.05		1.22	
	IHF 2021	qUI.Gh.21-IH.01	S9_1444072	3.71	4.67	1.22	-5.22	-4.28
		qUI.Gh.21-IH.02	S10_94098404	1.89	8.24		-2.21	
		qUI.Gh.21-IH.03	S12_83553677	4.49	4.91	1.56		
		qUI.Gh.21-IH.04	S19_57285752	5.08	1.19	-2.34		
		qUI.Gh.21-IH.05	S21_49497692	2.63	10.68		-2.61	
	Plains 2021	qUI.Gh.21-PL.01	S7_64773304	2.52	3.11	-2.11	-0.71	0.34
		qUI.Gh.21-PL.02	S18_54710934	1.58	7.61		2.59	
		qUI.Gh.21-PL.03	S25_3738157	2.06	9.67		-1.75	

SFI	IHF 2019	qSFI.Gh.19-IH.01	S26_53828293	3.39	7.03		0.91	
	IHF 2021	qSFI.Gh.21-IH.01	S1_16078074	3.03	3.87	0.28	0.75	2.68
		qSFI.Gh.21-IH.02	S5_25288702	3.24	8.83		0.03	
		qSFI.Gh.21-IH.03	S16_1752648	10.07	3.09	2.14		
		qSFI.Gh.21-IH.04	S19_57285752	4.62	4.53	1.93		
		qSFI.Gh.21-IH.05	S22_36659220	3.02	7.91		-1.76	
		qSFI.Gh.21-IH.06	S23_57997331	10.67	2.96	-1.75		
	Plains 2021	qSFI.Gh.21-PL.01	S1_44737527	3.04	3.13	1.06		
		qSFI.Gh.21-PL.02	S21_4175526	4.97	12.69		0.87	

Table 4.5. QTLs for fiber quality traits (UHM, UI and SFI) identified in the Pima S6 background. Env denotes the environment in which the populations were evaluated. PVE denotes percent of phenotypic variation explained by the QTL, a denotes additive effect of the QTL, d denotes the dominance effect of the QTL and d/a denotes the ratio of dominance to additive effect.

Trait	ENV	QTL	Marker	LOD	PVE	a	d	d/a
UHM	IHF 2019	qUHM.Gb.19-IH.01	S2_4688373	3.91	10.45		0.12	
		qUHM.Gb.19-IH.02	S5_81464487	3.06	8.29		-0.08	
		qUHM.Gb.19-IH.03	S25_21824441	4.30	6.31		0.05	
	IHF 2021	qUHM.Gb.21-IH.01	S11_9508165	3.69	5.34	-0.07	-0.12	1.71
		qUHM.Gb.21-IH.02	S12_2108876	3.31	8.41		-0.15	
	Plains2021	qUHM.Gb.21-PL.01	S7_46369886	2.93	3.39		-0.03	
		qUHM.Gb.21-PL.02	S10_71211391	3.67	3.63	0.14	-0.04	-0.29
		qUHM.Gb.21-PL.03	S11_9508165	4.35	3.21	-0.06	-0.11	1.83
UI	IHF 2019	qUI.Gb.19-IH.01	S24_51061769	3.23	6.10		-1.53	
		qUI.Gb.19-IH.02	S2_4688373	3.58	3.78	-1.28		
	IHF 2021	qUI.Gb.21-IH.01	S5_22120682	3.01	1.37	-1.37		
		qUI.Gb.21-IH.02	S6_19002926	3.69	12.44		-3.13	
	Plains 2021	qUI.Gb.21-PL.01	S14_13478137	2.38	7.54		-2.10	
SFI	IHF 2019	qSFI.Gb.19-IH.01	S5_81464487	3.07	8.32		0.73	
		qSFI.Gb.19-IH.02	S14_14854428	4.98	8.06		0.65	
	IHF 2021	qSFI.Gb.21-IH.01	S1_30967195	4.39	14.61		1.29	
		qSFI.Gb.21-IH.02	S5_74778838	3.72	5.37	1.29	-3.09	-2.40
		qSFI.Gb.21-IH.03	S8_84088322	8.03	10.98	-0.55	3.04	-5.53
		qSFI.Gb.21-IH.04	S11_79404560	3.05	10.40		2.01	
		qSFI.Gb.21-IH.05	S18_51007574	3.76	2.97	1.31		
		qSFI.Gb.21-IH.06	S21_38911389	3.55	12.00		1.77	
		qSFI.Gb.21-IH.07	S24_8799673	8.31	25.85		4.46	

Plains 2021	qSFI.Gb.21-PL.01	S1_80713792	3.12	10.25		0.69	
	qSFI.Gb.21-PL.02	S4_46049389	2.51	3.43	0.67	-0.08	-0.12
	qSFI.Gb.21-PL.03	S5_76716168	4.27	6.25	-0.11	1.99	-18.09
	qSFI.Gb.21-PL.04	S14_11729890	3.77	11.67		1.90	

Table 4.6. Subgenomic distribution of fiber quality QTLs in reciprocal NIL populations.

	Acala Maxxa background		Pima S6 background		Overall	
Trait	At	Dt	At	Dt	At	Dt
UHM	6	7	7	1	13	8
UI	5	5	3	2	8	7
SFC	3	6	8	5	11	11
Total	14	18	18	8	32	26

Table 4.7. Chromosomal distribution of fiber quality QTLs in reciprocal NIL populations. PVE denotes percent variation explained the QTLs and * shows QTLs with PVE less than 10%.

	Acala Maxxa Background			Pima S6 Background		
Chr	No of QTLs	PVE* < 10%	Traits	No of QTLs	PVE* < 10%	Traits
Chr 1	4	4	SFI, UHM	2	0	SFI
Chr 2	2	1	UHM	2	1	UI, UHM
Chr 4				1	1	SFI
Chr 5	1	1	SFI	5	5	UI, UHM, SFI
Chr 6				1	0	UI
Chr 7	2	2	UI, UHM	1	1	UHM
Chr 8				1	0	SFI
Chr 9	2	1	UI, UHM			
Chr 10	1	1	UI	1	1	UHM
Chr 11	1	0	UI	3	2	UHM, SFI
Chr 12	1	1	UI	1	1	UHM
Chr 14				3	2	UI, SFI
Chr 15	1	1	UHM			
Chr 16	1	1	SFI			
Chr 18	2	2	UI	1	1	SFI
Chr 19	2	2	SFI, UI			
Chr 20	1	1	UHM			
Chr 21	4	0	UHM, UI, SFI	1	0	SFI
Chr 22	1	0	SFI			
Chr 23	2	1	UHM, SFI			
Chr 24	1	1	UHM	2	1	UI, SFI
Chr 25	2	2	UHM, UI	1	1	UHM
Chr 26	1	1	SFI			
Total	32	19 (59.37%)		26	15 (57.69%)	

Table 4.8. Clustering of fiber quality QTLs in cotton genome

Cluster name	Chr	QTLs in cluster	Physical coordinates (Mb)
		Acala Maxxa background	
cQTL.Gh.19.1	19	qUI.Gh.21-IH.04, qSFI.Gh.21-IH.04	53.12 - 59.45
cQTL.Gh.21.1	21	qUHM.Gh.19-IH.04, qUI.Gh.21-IH.05, qUHM.Gh.21-IH.03	47.06 - 52.94
cQTL.Gh.23.1	23	qUHM.Gh.19-PL.04, qSFI.Gh.21-IH.06	42.24 - 58.00
		Pima S6 background	
cQTL.Gb.02.1	2	qUHM.Gb.19-IH.01, qUI.Gb.19-IH.02	02.83 - 07.34
cQTL.Gb.05.1	5	qSFI.Gb.21-IH.02, qSFI.Gb.21-PL.03, qUHM.Gb.19-IH.02, qSFI.Gb.19-IH.01	74.78 - 81.46
cQTL.Gb.14.1	14	qSFI.Gb.21-PL.04, qUI.Gb.21-PL.01, qSFI.Gb.19-IH.02	11.73 - 14.85

Table 4.9. Correspondence of QTLs identified in the present study with those previously reported. P-value is based on hypergeometric function. Values in "bold" indicate correspondence of QTLs in the respective background (GB/GH) with those from previous studies.

	Previous Study			Common QTLs		P value	
Trait	Pop	# QTLs	Reference	GH	GB	GH	GB
UHM	RIL	18	Liu et al., 2017	4	5	**0.017**	0.000
	BC1	2	Chen et al., 2015	0	0	0.874	0.921
	F2	49	Deng et al., 2019	7	4	**0.014**	0.080
	F2:3	64	Diouf et al., 2018	6	4	0.124	0.157
	F2:3	2	Guo et al., 2014	0	1	0.874	0.077
	F2:3	10	He et al., 2007	2	0	0.111	0.658
	F2	7	Huang et al., 2015	3	2	**0.006**	**0.025**
	MAGIC	34	Huang et al., 2018	6	5	**0.009**	**0.004**
	RIL	167	Jamshed et al., 2016	5	2	**0.000**	**0.000**
	RIL	47	Jia et al, 2016	7	4	**0.011**	0.071
	F2	6	Kumar et al., 2019	0	1	0.665	0.200
	BC1	13	Lacape et al., 2005	3	2	**0.037**	0.078
	RIL	30	Li et al., 2016	2	3	0.304	0.082
	RIL	134	Li et al., 2018	6	5	0.063	0.277
	F2	47	Liang et al., 2013	0	4	**0.027**	0.071
	RIL	83	Liu et al., 2018	7	4	0.147	0.247
	RIL	34	Ma et al., 2016	2	4	0.299	**0.026**
	F2	1	Mei et al., 2004	0	0	0.935	0.960
	F2	6	Rong et al., 2007	1	2	0.285	**0.019**
	F2	3	Rong et al., 2007	2	2	**0.011**	**0.004**
	RIL	22	Shang et al., 2015	1	2	0.360	0.170
	RIL	135	Shang et al., 2016	6	4	0.059	0.163
	F2:3	15	Shen et al., 2005	5	3	**0.001**	**0.014**
	RIL	8	Shen et al., 2007	2	2	0.078	**0.033**
	F2	6	Saranga et al., 2004	1	1	0.285	0.200
	RIL	21	Tan et al., 2015	4	3	**0.029**	**0.035**
	RIL	82	Tan et al., 2018	8	4	0.071	0.244
	RIL	23	Tang et al., 2015	1	3	0.351	**0.044**
	RIL	9	Frelichowski et al., 2006	3	2	**0.013**	**0.041**
	RIL	10	Wang et al., 2007	0	1	0.502	0.288
	RIL	36	Wang et al., 2015	2	2	0.293	0.282
	BC3F2	41	Wang et al., 2016	6	4	**0.023**	**0.047**
	BIL	8	Yu et al., 2013	1	3	0.334	**0.002**

	RIL	13	Yu et al., 2014	2	3	0.160	**0.009**
	RIL	12	Zheng et al., 2009	1	2	0.386	0.068
	F2:3	3	Zhang et al., 2005	1	0	0.172	0.884
	RIL	26	Zhang et al., 2015	1	1	0.321	0.400
	CSIL	56	Zhang et al., 2016	7	5	**0.029**	**0.034**
UI	F2	26	Deng et al., 2019	2	1	0.256	0.378
	F2:3	46	Diouf et al., 2018	2	4	0.301	**0.010**
	F2	8	Huang et al., 2015	2	1	**0.049**	0.173
	MAGIC	22	Huang et al., 2018	4	3	**0.013**	**0.010**
	RIL	110	Jamshed et al., 2016	6	4	0.244	0.205
	RIL	58	Jia et al, 2016	8	3	**0.001**	0.122
	BC1	10	Lacape et al., 2005	2	2	0.073	**0.020**
	RIL	22	Li et al., 2016	2	1	0.219	0.351
	RIL	128	Li et al., 2018	6	5	0.249	0.104
	F2	14	Liang et al., 2013	1	3	0.376	**0.002**
	RIL	20	Ma et al., 2016	3	1	0.055	0.334
	F2	7	Rong et al., 2007	3	1	**0.003**	0.155
	RIL	24	Shang et al., 2016	1	2	0.388	0.097
	F2	9	Saranga et al., 2004	2	1	0.061	0.191
	RIL	20	Tan et al., 2015	4	2	**0.009**	0.072
	RIL	58	Tan et al., 2018	2	3	0.247	0.122
	RIL	15	Tang et al., 2015	5	4	**0.000**	**0.000**
	RIL	20	Wang et al., 2015	3	2	0.055	0.072
	F2	27	Wang et al., 2016	5	4	**0.004**	**0.001**
	BIL	12	Yu et al., 2013	1	1	0.356	0.239
	RIL	11	Zheng et al., 2009	2	1	0.085	0.223
	F2:3	5	Zhang et al., 2005	1	2	0.207	**0.005**
	CSIL	8	Zhang et al., 2016	2	2	**0.049**	**0.013**
SFI	F2	19	Huang et al., 2015	3	3	**0.037**	0.089
	MAGIC	23	Huang et al., 2018	4	4	**0.010**	**0.038**
	F2	2	Rong et al., 2007	1	1	0.086	0.122
	F2	30	Wang et al., 2016	4	3	**0.026**	0.195
	RIL	8	Yu et al., 2014	0	1	0.687	0.334
	RIL	5	Wang et al., 2007	1	0	0.191	0.712

Table 4.10. Genes present in regions tightly linked to most significant SNP marker. Highlighted genes are known to have roles during fiber development.

QTL	Chr	Gene ID	Gene	Description
qUHM.Gh.19-IH.01	2	**Gh_A02G1330**	*PEPCK*	C assimilation
		Gh_A02G1317	*CESA*	cellulose synthase
		Gh_A02G1312	*AP2/ERF*	transcription regulation
		Gh_A02G1327	*EIF3A*	translation initiation
		Gh_A02G1334	*HAD-SF*	hydrolase activity
qUHM.Gh.19-PL.02	7	**Gh_A07G0818**	*psaL*	C accumulation
		Gh_A07G0819	*FH2_Formin*	actin nucleation
		Gh_A07G0825	*Cu-OXIDASE*	copper ion binding
		Gh_A07G0842	*GAUT*	transferring glycosyl
		Gh_A07G0853	*2SGLOBULIN*	metabolic process
		Gh_A07G0856	*EF_HAND_1*	calcium ion binding
qUHM.Gh.19-IH.02	15	Gh_D01G2224	*FE2OG_OXY*	iron ion binding
		Gh_D01G2236	*LRR_KINASE*	Protein binding,
		Gh_D01G2238	*CCME*	cytochrome complex
		Gh_D01G2242	*HTH_MYB*	chromatin binding
		Gh_D01G2258	*F-BOX/LRR*	cellulose synthesis
		Gh_D01G2260	*GDHRDH*	redox activity
qUHM.Gb.19-IH.03	25	Gh_D06G1031	*CARP*	peptidase activity
		Gh_D06G1027	*LRR-Kinase*	transferase activity
		Gh_D06G1020	*CaM-Serine*	kinase activity
		Gh_D06G1026	*Pkinase*	kinase activity
		Gh_D06G1016	*ABC-tran*	ATP binding activity
		Gh_D06G1034	*GTPase*	protein transfer

CHAPTER 5

MOLECULAR DISSECTION OF FIBER QUALITY IN RECIPROCAL INTERSPECIFIC ADVANCED BACKCROSS POPULATIONS IN COTTON

Jeevan Adhikari, Deepak Vitrakoti, Wiriyanat Ployaram, Sameer Khanal, Peng Chee and Andrew H Paterson. To be submitted to *Theoretical and Applied Genetics*.

Abstract

Estimation of precise locations and accurate effects of quantitative trait loci (QTLs) affecting fiber quality traits is an essential prerequisite for using these QTLs in breeding programs via marker assisted selection. Advanced backcross lines (ABLs) and/or near-isogenic lines (NILs) carrying one or few chromosomal segments introgressed from the donor genotype into the recipient background serve an important purpose of delineating the boundaries of a QTL to a small and rather precise genomic location, and also add to the accuracy of the effects of the QTL by reducing background noise caused by multiple segregating donor chromatin. In reciprocal interspecific advanced backcross populations by crossing two elite cotton cultivars, Acala Maxxa (*G. hirsutum*) and Pima S6 (*G. barbadense*) representing the two major domesticated species of cotton, we identified genomic locations underpinning three important fiber quality traits – fiber strength, fiber fineness and fiber elongation. Phenotypic evaluation of these lines in three environments revealed a total of 87 QTLs, with one stable QTL for fiber elongation identified in all three environments and five other QTLs (two for fiber elongation and three for fiber strength) identified in at least two environments. Most of the QTLs identified in the study were small-effect (explaining <10% variation) QTLs justifying the genomic composition of the lines used for analysis. Limited reciprocity of QTLs in the two backgrounds shows the strong influence of recipient genome, in addition to the combined consequences of epistasis, small phenotypic effects and imperfect coverage of donor chromatin in the recipient background.

Introduction

Cotton is the most important natural fiber and the leading textile crop of global economic importance. Two domesticated species of cotton - Upland cotton (*Gossypium hirsutum* L.) and

Pima Cotton (*G. barbadense* L.), dominate global cotton production, collectively accounting for more than 95% of world's fiber yield. The quality of lint fibers produced by Upland cotton is a major focus of the cotton industry because it overwhelmingly contributes to the total global fiber production. Despite stiff competition from synthetic fibers, consumer preference towards natural fiber and elevated market demands for high quality textile fibers have generated renewed interest in breeding for better fiber quality (Peng W. Chee & B. Todd Campbell, 2009). However, there is a low level of genetic differentiation in the cultivated cotton gene pool (Fang et al., 2017; Wang et al., 2017) resulting from the depletion of favorable variation due to evolutionary and breeding (domestication) bottlenecks (Adhikari, 2015a; Andrew H. Paterson et al., 2004). In addition, negative correlations between yield components and fiber quality parameters have restricted simultaneous breeding gains.

Improving fiber quality parameters simultaneously with increasing lint yield has been an important goal for cotton improvement. Genetic dissection of complex fiber quality traits and yield components into underlying quantitative trait loci (QTL) creates opportunities to manipulate quantitative traits based on Mendelian principles, using techniques such as marker-assisted selection (MAS), QTL pyramiding, fine mapping, and map-based cloning. These techniques have the potential to mitigate some of the challenges hindering improvement of fiber quality (Peng W. Chee & B. Todd Campbell, 2009; Fang, 2015). Introgressive breeding has been one of the successful techniques employed to transfer or combine desirable traits from two or more related species. Experimental populations created by/for introgressive breeding not only serve the purpose of combining favorable alleles, but also serve as suitable platform to identify genomic locations underpinning complex traits.

Among the two widely cultivated cotton species, Upland cotton is known for its high yield potential while Pima cotton has superior fiber quality. Transferring desirable traits from Pima cotton to Upland cotton has been a long-standing goal for cotton breeders and geneticists. Numerous introgressive breeding programs have been conducted in cotton in the past six or seven decades to simultaneously improve fiber yield and quality by interspecific hybridization (Zhang et al., 2014). In the past two decades alone, geneticists have tagged hundreds of QTLs underlying cotton fiber yield and quality parameters (J. Rong et al., 2007; J. I. Said, J. A. Knapka, et al., 2015; J. I. Said et al., 2013; J. I. Said, M. Song, et al., 2015). With progress in genetic mapping technologies, more QTLs are being tagged and published, but the use of these QTLs in marker-assisted breeding for cotton improvement has been limited (Khanal, 2018; J. I. Said, J. A. Knapka, et al., 2015). Stability (consistency) of QTL expression across environments and among generations and/or genetic backgrounds, variable effects and low precision in mapped locations are among the limiting factors to marker-assisted cotton improvement (Fang, 2015). Elevated noise from several multiple chromosomal segments segregating independently, some of which span entire chromosomes, often result in over- or underestimation of the effects of mapped QTLs and thus in lack of reproducible effects across environments and/or generations. Experimental populations with reduced background noise and smaller size of introgressed chromatin segments facilitate estimation of the location and effects of these genomic benchmarks more precisely. Near-isogenic lines (NILs), with only one chromatin segment from the donor parent, or advanced backcross lines with only a few donor segments segregating in the recipient genome, can serve the purpose of eliminating much of the background noise present in early generation populations.

Near isogenic lines (NILs), with only one chromosomal segment from a donor parent in an homogeneous genetic background of a recurrent parent serve as an important resource for both

genetic mapping and breeding (R. Kooke et al., 2012). Since there are few (ideally one) introgressed segment(s) in each NIL, phenotypes due to QTLs on the segment(s) are rendered much more discrete than in F_2 or backcross populations, often behaving as simple Mendelian factors (Paran & Zamir, 2003). QTL mapping based on NILs can thus increase accuracy of QTL position and detect small effect QTLs that might otherwise be obscured by larger-effect genes in more complex populations. In addition, because of the fixed genotype of NILs, they can be replicated in different environments to test interaction between genetic and environmental factors (Monforte & Tanksley, 2000). By crossing NILs to the recurrent parent, fine mapping of specific QTLs toward their cloning is facilitated.

While long and fine cotton fibers were sought by the textile industry in the early years of mechanization, mostly due to their direct impact on the yarn, spinnability and end products (Perkins Jr. et al., 1984), other fiber properties eventually were found to contribute to overall textile performance, both in the textile mills as well as to the consumers. Modern textile mills, with their advanced and sophisticated spinning technology, require not only longer lint fibers, but also ones that are strong and finer for adopting to their high-speed and automated spinning technology (Bradow & Davidonis, 2000). Fiber quality properties such as fiber elongation, which measures the degree of extensibility or elasticity of the fibers and fiber strength, which measures the force needed to break a bundle of lint, are important quality parameters. In addition, strong and fine cotton fibers with elastic properties are second to none when it comes to consumer preference.

To better understand the genetic architecture of three important fiber quality parameters – fiber elongation (ELO), fiber strength (STR), and fiber fineness or micronaire (MIC) – in the two mostly cultivated species of cotton, we constructed a set of reciprocal interspecific advanced backcross populations using Acala Maxxa (*G. hirsutum*) and Pima S6 (*G. barbadense*) as parents,

tiling 71.48% of the Acala Maxxa genome in Pima S6 background (hereafter GB background) and 78.72% of the Pima S6 genome in the Acala Maxxa background (hereafter GH background). With advanced backcross populations and NILs, we found 87 QTLs for these three important fiber quality parameters.

Materials and Methods

Population development

Plant materials used in this study were developed from a set of reciprocal crosses between *Gossypium hirsutum* acc. Acala Maxxa and *G. barbadense* acc. Pima S6 (both inbred lines). These genotypes have been extensively used to produce several molecular tools and resources including BAC libraries and Illumina genome sequences. Reciprocal advanced backcross populations were developed by first crossing the parents in a two-way cross (Acala Maxxa (♀)× Pima S6 (♂) – GH background; and Pima S6 (♀) × Acala Maxxa (♂) – GB background), then independently backcrossing F_1 plants to the respective female parent to create 300 to 400 BC_1 progenies for each cross. The backcrossing scheme included planting only one seed from each preceding backcross to generate the next generation (Figure 3.1).

After five generation of backcrossing, 173 BC_5F_1 plants from the GH background and 231 BC_5F_1 plants from the GB background were self-pollinated and a total of 404 BC_5F_2 families were grown at Iron Horse Farm (IHF), Watkinsville, Georgia in 2019 and 2021 and at Southwest Georgia Research Station, Plains, Georgia in 2021 under cultural conditions consistent with commercial irrigated cotton production. Individual BC_5F_2 plants that contained only one introgressed segment from the donor parent were deemed as NILs. We identified a total of 397 NILs in the GH background and a total of 423 NILs in the GB background. Selfed seeds of these

NILs (BC_5F_3 seeds) were grown at IHF and Plains in 2021 under cultural conditions consistent with commercial irrigated cotton production.

Phenotypic evaluation and data analysis

Two replications of each BC_5F_2 families and NILs were planted in a randomized complete block design (RCBD) in three environments (IHF-2019, IHF-2021 and Plains-2021). Six replications each of the two parents were included in all three environments. Fiber samples were collected by harvesting 25 bolls from each plot, ginned in a laboratory gin, evaluated by HVI (Cotton Incorporated Textile Service Laboratory, Cary, NC). Phenotypic data was collected for three important fiber quality parameters – ELO, STR and MIC.

All statistical analyses were conducted in R programming language. Single marker analyses were done in R/qtl (Broman & Sen, 2009). The significance threshold was set to LOD of 3, to mitigate the multiple-comparison problem. Filtration of significant markers adopts the method proposed by Szalma et al. (2007b). If several markers on the same introgressed segment show significant association with phenotype, the most significant one was reported. For the co-segregation of multiple introgressions, the QTL location is examined as follows. First, if multiple families show significance for the trait and carry overlapping introgression, the introgression is considered to carry QTL. Second, if the co-segregation of introgressions is in single families, the most significant introgression is considered to carry QTL.

Phenotypic variance explained by each locus was reported by taking the most significant marker as independent variable and phenotypic value as dependent variable in R (R Core Team 2021). Additive effects were estimated by half the difference of phenotypic values between the lines carrying the homozygous introgression and lines not carrying the introgression. Dominance

effects were estimated by the difference of phenotypic values between the lines carrying the heterozygous introgression and the remaining lines that do not carry the introgression. If multiple or overlapping introgressions were present at both homozygous and heterozygous state, the estimation of additive effects utilized the lines carrying the introgression at homozygous state only and, the estimation of dominance utilized the lines carrying the introgression at heterozygous state only. Dunnett's test was performed to compare means of individual lines with the recurrent parent and with five cotton cultivars recommended in the state of Georgia.

Gene actions for QTLs were determined by calculating the degree of dominance (absolute values) for every QTL that has both additive and dominance effects. The degree of dominance is the ratio of dominance effect to additive effect (d/a) of the QTL and based on this ratio, gene action of the QTLs can be categorized as (i) additive ($0 < d/a < 0.2$) (ii) partially dominant ($0.2 < d/a < 0.8$) (iii) dominant ($0.8 < d/a < 1.2$ and (iv) over-dominant ($d/a > 1.2$). QTLs with dominant and over-dominant effects are considered to have heterotic effect or heterozygous advantage.

Identification of common QTLs

Common QTL is defined as either the same marker is detected in the two reciprocal populations, or two different markers are detecting exactly the same introgression(s) in each population. Correspondence between QTLs for a trait across the entire genome is inferred using the hypergeometric probability function. The model was adopted from (Feltus et al., 2006): *p* is the probability of non-random correspondence of QTLs being compared for a given trait, *n* is the number of comparable intervals which is calculated by dividing the total genome size by average introgression size in both populations; *m* is the number of common QTLs; *l* is the number of QTLs in the GH background; *s* is the number of QTLs in the GB background. The same model was also

adopted to detect correspondence between QTLs reported in this study with those previously published. In this case, *l* is the total number of QTLs identified in the larger sample (study reporting higher number of QTLs) and *s* is the number of QTLs identified in the smaller sample.

$$p = \frac{\binom{l}{m}\binom{n-1}{s-m}}{\binom{n}{s}}$$

Candidate gene identification

In silico annotation was performed on the identified QTLs to look for candidate genes related to flowering habit in cotton. For each QTL identified in the study, the genomic region spanning 1 Mb on each side of the most significantly associated marker was used for in silico analysis. The DNA sequence from this tightly linked region was used to look for *G. hirsutum* genes in the CottonGen database and these genes were then analyzed for biological functions, with particular focus on fiber growth and development.

Results

Genomic composition of NILs and ABLs

The distribution of markers in the reciprocal populations is presented in Table 2.1. Genomic distribution of the introgressed chromosomal segments in ABLs is shown in Table 2.2 and Figure 2.3 and 2.4, while coverage of donor genome by the NILs is shown in Table 3.1 and Figures 3.2 and 3.3. Relevant information about the genomic composition of NILs and ABLs are presented in the Results section of chapters 2 and 3.

Phenotypic performance of parents and experimental populations

The phenotypic performance of the two parents, reciprocal backcross populations, and reciprocal NIL populations is shown in Figure 5.1. The distribution of traits was approximately normal (Shapiro and Wilk test; $p > 0.05$) and typical of quantitative inheritance. Pima S6 outperformed Acala Maxxa in all three environments (Tables 5.1 and 5.2, Figure 5.1). Both advanced backcross and NIL populations in the GB background performed better than in the reciprocal background for fiber elongation and fiber strength while they did not differ much for fiber fineness or micronaire. Transgressive segregation is seen for all three traits for all populations across all environments tested. Transgressive segregants outperforming both parents were identified in both population types and both backgrounds (Figure 5.1).

To identify the effect of genotypes and environment in the overall performance of the advanced backcross populations and the NILs, we conducted analysis of variance (ANOVA) keeping all variables as fixed factors. Results showed significant effects of both genotype (GEN), and genotype-by-environment (GXE) (Table 5.3). GEN captured the most variation for all traits in both population types. GXE also captured significant amount of variation in the phenotypes and thus precluded the use of combined phenotypic values in identification of QTLs for these traits. Thus, marker trait association and identification of fiber quality QTLs for the three traits under study was performed separately for each environment tested.

Marker trait association and overview of QTLs

A total of 87 marker trait associations were identified (50 in the GH background (Table 5.4) and 37 in the GB background (Table 5.5)). Phenotypic variances explained by these QTLs ranged from 2.55% to 26.97% (Tables 5.4 and 5.5). Among the 87 QTLs identified, 26 were of

large effect, explaining > 10% of the total phenotypic variation while the remaining 61 were small effect QTLs (explaining <10% of total phenotypic variation). The highest number of QTLs (37) was identified for ELO, followed by STR (27) and MIC having the lowest number of identified QTLs (23).

QTLs conferring fiber elongation (ELO)

In the GH background, a total of 22 QTLs were identified for ELO in the three environments tested, explaining 3.01% to 18.41% of total phenotypic variation (Table 5.4). These QTLs were distributed over 15 chromosomes, with chromosome 12 carrying the most QTLs. A total of 11 of these QTLs were small effect and the remaining 11 were large effect. Among the 22 QTLs identified, eight were identified at IHF in 2019, nine at IHF in 2021 and five at Plains in 2021. All 22 QTLs identified in the GH background increased fiber elongation consistent with the parental phenotypes. The ELO QTL on chromosome 12 was identified in all three environments tested and is a potential candidate for fine mapping and QTL-seq studies. Another QTL on chromosome 15 was identified at IHF in 2019 and at Plains in 2021, making it a priority for further study of its effect on fiber elongation.

In the GB background, a total of 15 QTLs were identified for ELO, explaining 3.95% to 12.03% of the total phenotypic variation (Table 5.5). Among the 15 QTLs identified for ELO, two were identified at IHF in 2019, seven at IHF in 2021 and six at Plains in 2021. Of the 15 QTLs, only on chromosome 3 was a major effect QTL and the remaining 14 were small effect QTLs. Five of these QTLs decreased fiber elongation while 10 increased fiber elongation. One QTL on chromosome 6 and one on chromosome 9 were each identified in two environments (IHF 2021 and Plains 2021).

QTLs conferring fiber fineness (MIC)

A total of 10 QTLs were identified for fiber fineness (MIC) in the GH background, explaining 2.55% to 17.43% of total phenotypic variation (Table 5.4). Five QTLs each were identified at IHF in 2019 and at IHF in 2021. Among the 10 QTLs identified, two were major effect QTLs and the remaining eight were small effect QTLs. Most QTLs reduced fiber fineness (consistent with the parental phenotypes), with exceptions on chromosome 5 and chromosome 22.

In the GB background, a total of 13 QTLs were identified for fiber fineness, explaining 3.27% to 9.97% of total phenotypic variation (Table 5.5). These QTLs spanned 10 chromosomes in the cotton genome and only three of the 13 QTLs were identified in the Dt subgenome. Among the 13 QTLs, four were identified at IHF in 2019, four at IHF in 2021 and five at Plains in 2021. Only three of the 13 QTLs increased fiber fineness (the predicted change based on the parental phenotypes) while the remaining 10 decreased fiber fineness.

QTLs conferring fiber strength (STR)

A total of 18 QTLs were identified for fiber strength in the GH background, explaining 3.66% to 26.97% of the total phenotypic variation (Table 5.4). Of the 18 QTLs, seven were in the At subgenome and 11 were in the Dt subgenome. Unlike for the other two traits, most of the QTLs identified were major QTLs with only seven being small effect QTLs. A large majority, 15 QTLs, increased fiber strength as would be predicted based on parental phenotypes, with only three reducing fiber strength. One QTL on chromosome 13 was identified at IHF in both 2019 and 2021. Another QTL on chromosome 25 was also identified in two environments (IHF 2021 and Plains 2021).

In the GB background, a total of nine QTLs were identified for fiber strength, explaining 4.65% to 11.12% of total phenotypic variation (Table 5.5). Among the nine QTLs, three were identified at IHF in 2019, four at IHF in 2021 and two at Plains in 2021. Five of the nine QTLs were identified in the At subgenome and the remaining four were identified in the Dt subgenome. Seven of the nine QTLs decreased fiber strength as predicted based on the parental phenotypes, while only two QTLs increased fiber strength. One QTL on chromosome 24 was identified in two environments (IHF 2019 and IHF 2021).

Subgenomic distribution of QTLs

In total, we identified 87 QTLs for three fiber quality traits (50 in the GH background and 37 in the GB background). QTLs were almost evenly distributed across the two subgenomes in the GH background (26 in the At subgenome vs 24 in the Dt subgenome) while comparatively more QTLs were located in the At subgenome (25) than in the Dt subgenome (12) in the GB background (Table 5.6). QTLs were identified in all chromosomes except chromosome 17 (Table 5.7). In the GH background, QTLs were identified in 24 of 26 chromosomes while in the GB background, they were identified in 20 chromosomes. In total, chromosome 3 carried the most QTLs (7) followed by chromosome 10 (with 6 QTLS) and chromosomes 2, 9, 12 and 21 (with 5 QTLs each) (Table 5.6). In the GH background, chromosomes 12 and 15 carried the most QTLs (4 each) followed by chromosomes 1, 3, 9, 10, 21, 22 and 25 (3 each) while chromosomes 3 and 8 carried the most QTLs (4 each) in the GB background.

Clustering of fiber quality QTLs

Certain regions in the cotton genome are known to harbor QTLs for two or more traits. Such regions carrying multi-trait QTLs within a 20 Mb span are considered QTL clusters or QTL hotspots. In agreement with previously published reports, we observed similar nonrandom distribution of fiber quality QTLs throughout the cotton genome. In total, we observed nine such clusters in the GH background and five such clusters in the GB background (Table 5.8). Two clusters in the GH background (cQTL.Gh.03.1 and cQTL.Gh.15.1) contained three QTLs each. The QTL cluster on chromosome 3 (cQTL.Gh.03.1) contained one QTL each for ELO, MIC and STR while the cluster on chromosome 15 (cQTL.Gh.15.1) contained two QTLs for ELO and one for MIC. Two clusters, one in chromosome 2 and the other in chromosome 3, in the GB background also contained three QTLs each. Both clusters harbored one QTL for STR and two QTLs for MIC.

Discussion

Introgressive breeding approaches not only introduce a preponderance of novel allelic variation into cultivated gene pools, but interspecific populations developed using these approaches can be widely used for molecular dissection of complex fiber yield and quality parameters. Much past studies have focused on investigating the effects of GB chromatin segments introgressed into GH, however, the reverse has not been routinely studied. In the current study, we developed a reciprocal set of advanced backcross lines and NILs selected from among the selfed progenies of these advanced backcross lines and tested these populations to assess the effects of reciprocal chromatin transfer on three important fiber quality traits – ELO, MIC, and STR. We identified a total of 87 marker-trait associations for these three traits with variable genetic effects.

This study adds to the resources and observations available to study the quantitative nature of fiber quality traits reciprocally in two elite cotton backgrounds.

Performance of NILs and advanced backcross populations

NILs and advanced backcross populations in both backgrounds showed average phenotypes consistent with their recurrent parent (Figure 5.1). Albeit the average performance of these two populations behaved like their recurrent parents, which is expected given the exceptionally large proportion of their genome coming from the recurrent parent, a discretely large amount of variation was observable for all three traits. Presence of transgressive segregants on both directions for all three traits suggests that the chromatin segments introgressed from the donor parents has effects that could significantly alter the performance of individual lines. In fact, these alterations and their effects is shown by the phenotypic performance (QTL effects) of the lines carrying respective introgression from the donor parent.

Effect of species background

The total number of QTLs identified for the three fiber quality traits did not differ by much (50 in GH background vs 37 in the GB background) in the two backgrounds. While identification of similar total numbers of total QTLs for comparable traits in the reciprocal backgrounds might suggest the involvement of similar number of genes in controlling these traits, many other findings indicate the effect of species background on effects, locations, and patterns of these QTLs.

Among the 87 QTLs identified in the two backgrounds, only four QTLs were identified to be reciprocally located. Reciprocity was identified for MIC on chromosome 3 and 10 (Tables 5.4 and 5.5). While the reciprocal QTLs on chromosome 3 showed effects that were consistent with

the donor parent phenotype i.e., QTL from GH in the GB background decreasing MIC and the QTL from GB in the GH background increasing MIC, the reciprocal QTLs on chromosome 10 decreased MIC in their respective backgrounds. Two QTLs for STR were reciprocally identified, one on chromosome 11 and the other on chromosome 24. On both chromosomes, the reciprocal QTLs showed antagonistic effects towards the trait consistent with the phenotype of the donor parent, i.e., each QTL from GB on GH background increased STR while those from GH on GB background decreased STR.

A perplexing observation, however, is the relatively low frequency at which QTLs identified with opposite phenotypic effects at corresponding locations in the reciprocal genetic backgrounds. In principle, one would expect the majority of QTLs to show such reciprocity if alternative alleles at a QTL show additive or dominant-recessive effects. Limited correspondence of identified QTLs in the two backgrounds could be a result of several factors. One factor may be the small phenotypic effects of most of the identified QTL, increasing the likelihood that one or both members of a reciprocal pair elude detection (Broman, 2001). Another intriguing factor that could account for some failures to identify correspondence of QTLs in the reciprocal backgrounds, especially in advanced backcross lines with multiple introgressed chromosomal segments, is epistasis. Interaction between introgressed loci might result in underestimation of their effects which might have resulted in some QTLs failing to reach the biometric thresholds required to declare them as QTLs per se. The widespread observation that fiber quality parameters generally have high heritability (Fang et al., 2014) suggests a limited role of epistasis, but it could contribute to failures to identify reciprocal QTLs with relatively small effects (Chandnani et al., 2018). This might be the case here as most of the QTLs detected in this study show low genetic contribution to the total phenotypic variation explained by the phenotype.

QTL clustering

Colocalization (clustering) of QTL with common allele effects makes marker-assisted selection much more efficient if favorable QTLs cluster together while it makes selection even more challenging if QTLs of antagonistic effects cluster together. QTL clusters for the three fiber quality traits studied were prevalent and were quite variable between the two backgrounds, with clusters mostly identified in different sets of chromosomes in the two backgrounds (Table 5.8), the only chromosome harboring clusters in both backgrounds being chromosome 3. However, these two chromosome 3 clusters were located on different locations and formal testing using the hypergeometric probability function (see methods section) saw them as non-corresponding (p-value > 0.05). Significantly higher number of clusters was identified in GH background (9 clusters vs 5 in GB background).

In the GH background, all QTL clusters except the ones on chromosomes 9 and 21, coupled favorable alleles for STR and ELO (Tables 5.4, 5.5 and 5.8). Clusters on chromosomes 9 and 21 coupled favorable alleles for ELO with unfavorable alleles for STR. All QTL clusters containing MIC QTLs introduced alleles from GB that would decrease MIC. Given the recipient parent, Acala Maxxa, had MIC values below premium range for two of the three environments tested, the effects of these introduced QTLs would further reduce the recipient parent's MIC values. While the results suggest that selection chromosomes carrying QTL clusters with MIC QTLs would be a daunting task to separate the negative effects of MIC QTLs from the positive effects of STR and ELO QTLs, selection on other chromosomes that carry clusters containing ELO and STR QTLs only (for example cluster on chromosome 23) would enrich favorable alleles from GB to GH.

In the GB background, no clusters carried QTLs for all three traits although all clusters carried at least one MIC QTL (Table 5.8). All the MIC QTLs reduced MIC values in the Pima S6 background (Table 5.5), thus rendering favorable alleles to Pima S6 as MIC values for Pima S6 were around the upper bound of the "premium" range and the effects of these QTLs would further bring them down within the limits of the "premium" range. In that matter, cluster cQTL.Gb.03.1 on chromosome 3, would be a suitable candidate for positive selection as the STR QTL in this cluster adds favorable alleles for fiber strength to the GB background. Nevertheless, these results show cautious optimism on the performance of these lines over their recurrent parent and warrant additional multi-locational and multi-environment trails to confirm their superior performance over Pima S6. Another chromosome suitable for positive selection would be chromosome 16 as both QTLs contained in the cluster located in this chromosome brought favorable alleles to the GB background. All other clusters identified in the GB background contained QTLs, one of which rendered positive effects while the other contributed unfavorable allele. While clusters of QTLs carrying both favorable and unfavorable alleles are not uncommon, the presence of both makes fiber quality improvement by conventional approaches in GB background more challenging than in GH background because of greater linkage drag for unfavorable alleles in GB (Chandnani et al., 2018). NILs with one or a very few QTLs confined to a small introgressed genomic region would serve as a good starting material to get rid of such linkage drag in the GB background.

Subgenomic distribution of fiber quality QTLs

With regard to the distribution of QTL in A and D subgenomes, many prior studies have concluded that more QTL for growing period, yield and fiber quality were distributed in Dt subgenome than in At subgenome (Adhikari et al., 2017; Chandnani et al., 2018; J. Rong et al.,

2007; Joseph I. Said et al., 2013; Wang et al., 2013). However, Shen et al. (2005) and Lin et al. (2005) reported that more QTL for fiber quality were located in At subgenome than in Dt subgenome. In the current study, we identified more QTLs in the At subgenome than in the Dt subgenome (Table 5.6). For NILs and advanced backcross populations in the GH background, the subgenomic affinities of QTLs were not significant ($p>0.05$) but nominally agreed with most previous findings, with At subgenome harboring 8.03% more QTLs than the Dt subgenome. In the reciprocal background, however, significantly higher ($p<0.05$) number of QTLs were identified in the At subgenome (22 vs 12). QTLs for individual traits also followed the overall trend with almost similar or a greater number of QTLs in the At subgenome in both backgrounds, STR being the only exception in the GH background with more QTLs in the Dt subgenome.

Stability of fiber quality QTLs

While different loci associated with the same trait under various environments might suggest interaction between genotype and environment, QTLs being detected across environments might indicate environmental stability. Although most QTLs identified in this study were single-environment QTLs, some QTLs were identified in at least two of the three environments tested (Results section and Tables 5.4 and 5.5). In addition to one genomic region on chromosome 12 consistently identified for ELO on all three environments, we also found six other genomic regions on chromosomes 6, 9, 13, 15 and 24 consistently associated with ELO and STR in two different environments.

Similarity with QTLs previously reported

The genetic composition of our experimental populations provides a platform to identify novel small-effect QTLs in addition to major QTLs. Since NILs serve as a resource to not only identify marker trait associations but also an important tool to verify QTLs previously identified, mostly those using early generation populations, the correspondence identified here could be used as a means of validation of previously published QTLs. Given that there are hundreds of studies reporting fiber quality QTLs and owing to the genetic structure of our experimental populations, we mostly compared our results with populations of similar genomic composition (Brown et al., 2019; P. W. Chee et al., 2005; Xavier Draye et al., 2005; Yu et al., 2013). We also performed elaborate statistical comparisons with other previous reports on the correspondence of QTLs identified in our study.

First, we started studying the correspondence of QTLs identified in our study with those reported previously in populations of similar genomic composition. In a comprehensive study of biometric parameters of QTLs affecting fiber fineness or micronaire using a backcross-self approach, Xavier Draye et al. (2005) identified a fiber fineness QTL (FF01.1) on chromosome 1 with the nearest locus being an RFLP marker A1204. This marker is located in the same introgressed segment where we identified the most significant marker for our MIC QTL (qMIC.Gh.21-IH.01) on chromosome 1 in the GH background, verifying the location and effect (reducing MIC in both cases) of this QTL. To add to the stability of this QTL, it was also identified in the reciprocal population in our study with similar (reducing fineness) effect. One would expect antagonistic effect of QTLs introgressed from reciprocal backgrounds, but in this case, the reciprocal QTL identified in a corresponding genomic location has shown unidirectional effects towards the trait. As mentioned earlier, these so-called unidirectional effects could actually be

antagonistic in the reciprocal backgrounds when it comes to favorability towards the trait because of the nature of the trait. For example, negative values of the QTL effects in one background might be reducing fineness from coarse (above premium) range to the premium range while in the reciprocal background, the negative values of the QTL effects might be reducing fineness from the premium range to below premium range, thus, showing antagonistic effects on the overall performance while still showing unidirectional effects statistically.

In the same study, another fiber fineness QTL (FF02.1) was identified in chromosome 2 with the nearest marker being pGH399a. This QTL is identified within the bounds of the same GH chromatin segment introgressed into GB in our study carrying the fiber fineness QTL qMIC.Gb.19-IH.01 (Table 5.5). This fiber fineness QTL was identified in two environments in our study, adding to its stability. While FF02.1 had a total effect of increasing fiber fineness, the one identified in our study reduced fiber fineness. This could probably be because FF02.1 was contributed by GB and the one in our study is contributed by GH in the reciprocal background. An interesting observation about this QTL is that in our companion study about fiber length parameters, this same genomic region on chromosome 2 also carried a fiber length (UHM) QTL (see Chapter 4, Discussion section), although the QTL was identified in reciprocal background (GB chromatin introgressed into GH). A more intriguing observation about this genomic region on chromosome 2 is that in the similar fashion as we identified QTLs for fiber fineness (in this study) and for fiber length (in our companion study), the authors of FF02.1 (Xavier Draye et al., 2005) also found a fiber length QTL (FL02.1) and a fiber elongation QTL (FE02.1) in the same genomic region in their two companion studies (Peng Chee et al., 2005; P. W. Chee et al., 2005). Thus, this region on chromosome 2 (spanning from 47155918 to 78271231 bp) not only seems to

have stable QTL for fiber fineness but also for fiber length and fiber elongation and is a suitable candidate to improve all three traits simultaneously.

While correspondence of individual marker-trait associations may reflect, for example, a gene for which the recurrent parent, Acala Maxxa or Pima S6, has a rare allele, non-random patterns of association across an entire genome can reflect other properties such as convergent domestication (Paterson et al., 1995). To investigate such genome wide non-random patterns of association, we also conducted an elaborate study on the correspondence of QTLs identified in our study with other previous reports, irrespective of the population type. Information on a total of 3211 QTLs related to fiber elongation, fiber strength and fiber fineness reported in 103 previous reports were downloaded from CottonGen website (https://www.cottongen.org) and the hypergeometric probability distribution was used to conduct a thorough analysis of QTL correspondence. The hypergeometric function provides a means to infer statistically whether QTLs for a trait are randomly distributed between two populations or environments. Correspondence was identified for 35 of the 105 previously reported studies for the QTLs reported in Acala Maxxa background and for 33 of the 105 previously reported studies in Pima S6 background (Table 5.9). Common QTLs were identified in most of the 67 reports listed, however, only 35 (or 33) significant P values based on the hypergeometric distribution suggest that across the genome, this correspondence is not sufficient to infer a non-random distribution of QTLs between published studies.

In silico annotation of potential candidate genes

Availability of cotton reference genomes (Paterson et al., 2012; Zhang et al., 2015) enabled us to scrutinize physical regions surrounding the identified QTLs for genes / gene families known

or suspected to affect fiber quality parameters in cotton. This investigation was limited to tightly linked regions i.e., 1 Mb on both sides of the SNP marker that is most significantly associated with the QTLs. On chromosome 5, the nearest gene of interest to an ELO QTL was a *C2HC* zinc finger superfamily protein, a member of a gene family whose gene expression level at 11 days post-anthesis has been reported to be highly correlated with UHM and which has been implicated in secondary cell wall thickening (Al-Ghazi et al., 2009). On chromosome 10, the nearest gene of interest to another fiber elongation QTL was a glucosyltransferase like *UDP-Gp* gene. The enzymatic activity of this gene increases during the period of development when cotton fiber is synthesizing massive amount of cellulose (Taliercio & Kloth, 2004; Xu et al., 2022), which like protein involved in cellulose biosynthesis during secondary cell wall formation stage.

On chromosome 25, the nearest gene of interest to a fiber strength QTL is a cytokinin-dehydrogenase related protein (*Cytokinin_DH_FAD*) which is involved in amino sugar and nucleotide sugar metabolism during cell wall biogenesis. In cotton, this gene has been found to be highly expressed during 0-1 day post anthesis and thus implicated in fiber initiation (Zeng et al., 2012). On chromosome 23, the nearest gene of interest to a fiber strength QTL is a leucine-rich repeat (*LRR*) kinase-like protein. These kinases are known to be expressed at specific fiber development stages with the highest expression during cellulose synthesis for secondary cell wall formation (Li et al., 2015). While it is premature to suggest the active roles of these genes towards fiber trait phenotypes, these genes are potential candidates for these roles given that they are so close to the most significant marker for fiber quality QTLs that have been repeatedly identified in multiple environments and studies. Improved genomic resources together with these novel populations are expected to accelerate candidate gene identification and validation for many such loci.

Conclusion

The two major species of cotton have been used in cotton breeding and improvement programs to transfer alleles from one species to another, GB being the donor in most cases due to its superiority in fiber quality parameters and partly due to high lint yield of the recipient GH. The present study demonstrates the value of GH as a source of favorable alleles for fiber quality traits in the GB background while also reiterating the reciprocal transfer of favorable alleles from GB. Using advanced backcross lines segregating for a few donor chromatins as well as a set of reciprocal near-isogenic lines, we showed a strong effect of genetic background on chromatin transfer as well as the effect of these introgressed chromatin on three important fiber quality traits. The near-isogenic genomic composition of our population provided opportunities to estimate the effects of genomic regions more precisely, but at the cost of epistatic QTL interactions. With one of the major purposes of NILs being the ability to verify the location and effects of QTLs, in addition to their identification, we were able to demonstrate the stability of a few important fiber quality QTLs identified in previous studies. Since the parents used in creating these populations are both elite lines representing the two major domesticated species of cotton, with their own specialties and differences, the populations we developed are not just a platform to identify genomic locations underpinning fiber quality traits, but also a good starting point from which study divergence, domestication and evolutionary history of cotton as well as to select superior individual lines from a pool of these immortal lines for breeding purposes as well as for commercialization.

References

Adhikari, J. (2015). *Effects of exotic genotypes and genetic backgrounds on fiber quality and plant architectural traits in Upland cotton (Gossypium hirsutum L.)* University of Georgia].

Adhikari, J., Das, S., Wang, Z., Khanal, S., Chandnani, R., Patel, J. D., Goff, V. H., Auckland, S., Rainville, L. K., & Jones, D. (2017). Targeted identification of association between cotton fiber quality traits and microsatellite markers. *213*(3), 1-15.

Al-Ghazi, Y., Bourot, S., Arioli, T., Dennis, E. S., & Llewellyn, D. J. (2009). Transcript profiling during fiber development identifies pathways in secondary metabolism and cell wall structure that may contribute to cotton fiber quality. *Plant Cell Physiol*, *50*(7), 1364-1381.

Bradow, J. M., & Davidonis, G. H. (2000). Quantitation of fiber quality and the cotton production-processing interface: a physiologist's perspective. *Journal of Cotton Science*, *4*(1), 34-64.

Broman, K. W. (2001). Resource Review of Statistical Methods for QTL Mapping in Experimental Crosses. *30*(7), 44-52.

Broman, K. W., & Sen, S. (2009). *A guide to QTL mapping with R/qtl* Springer.

Brown, N., Kumar, P., Singh, R., Lubbers, E., Campbell, B. T., Myers, G. O., Wright, R. J., Subramani, J., Jones, D., Paterson, A. H., & Chee, P. W. (2019). Evaluation of a Chromosome Segment from Gossypium barbadense Harboring the Fiber Length QTL qFL-Chr.25 in Four Diverse Upland Cotton Genetic Backgrounds. *Crop Science*, *59*(6), 2621-2633.

Chandnani, R., Kim, C., Guo, H., Shehzad, T., Wallace, J. G., He, D., Zhang, Z., Patel, J. D., Adhikari, J., Khanal, S., & Paterson, A. H. (2018). Genetic Analysis of Gossypium Fiber Quality Traits in Reciprocal Advanced Backcross Populations. *The Plant Genome*, *11*(1), 170057.

Chee, P., Draye, X., Jiang, C.-X., Decanini, L., Delmonte, T. A., Bredhauer, R., Smith, C. W., & Paterson, A. H. (2005). Molecular dissection of interspecific variation between Gossypium hirsutum and Gossypium barbadense (cotton) by a backcross-self approach: I. Fiber elongation. *Theoretical and Applied Genetics*, *111*(4), 757-763.

Chee, P. W., & Campbell, B. T. (2009). Bridging Classical and Molecular Genetics of Cotton Fiber Quality and Development. In A. H. Paterson (Ed.), *Genetics and Genomics of Cotton* (pp. 283-311). Springer US.

Chee, P. W., Draye, X., Jiang, C. X., Decanini, L., Delmonte, T. A., Bredhauer, R., Smith, C. W., & Paterson, A. H. (2005). Molecular dissection of phenotypic variation between Gossypium hirsutum and Gossypium barbadense (cotton) by a backcross-self approach: III. Fiber length [Journal Article]. *TAG Theoretical and Applied Genetics*, *111*(4), 772-781.

Draye, X., Chee, P., Jiang, C.-X., Decanini, L., Delmonte, T. A., Bredhauer, R., Smith, C. W., & Paterson, A. H. (2005). Molecular dissection of interspecific variation between Gossypium hirsutum and G. barbadense (cotton) by a backcross-self approach: II. Fiber fineness [Article]. *Theoretical & Applied Genetics*, *111*(4), 764-771.

Fang, D. (2015). Molecular Breeding. In.

Fang, D. D., Jenkins, J. N., Deng, D. D., McCarty, J. C., Li, P., & Wu, J. (2014). Quantitative trait loci analysis of fiber quality traits using a random-mated recombinant inbred population in Upland cotton (Gossypium hirsutum L.). *BMC Genomics*, *15*(1), 397.

Fang, L., Wang, Q., Hu, Y., Jia, Y., Chen, J., Liu, B., Zhang, Z., Guan, X., Chen, S., Zhou, B., Mei, G., Sun, J., Pan, Z., He, S., Xiao, S., Shi, W., Gong, W., Liu, J., Ma, J., . . . Zhang, T.

(2017). Genomic analyses in cotton identify signatures of selection and loci associated with fiber quality and yield traits. *Nature Genetics*, *49*(7), 1089-1098.

Feltus, F. A., Hart, G. E., Schertz, K. F., Casa, A. M., Kresovich, S., Abraham, S., Klein, P. E., Brown, P. J., & Paterson, A. H. (2006). Alignment of genetic maps and QTLs between inter- and intra-specific sorghum populations. *Theor Appl Genet*, *112*(7), 1295-1305.

Khanal, S. (2018). *Dissection of quantitative variation introgressed into bermudagrass and upland cotton* University of Georgia].

Kooke, R., Wijnker, E., & Keurentjes, J. J. (2012). Backcross populations and near isogenic lines. *Methods Mol Biol*, *871*, 3-16.

Li, F., Fan, G., Lu, C., Xiao, G., Zou, C., Kohel, R. J., Ma, Z., Shang, H., Ma, X., Wu, J., Liang, X., Huang, G., Percy, R. G., Liu, K., Yang, W., Chen, W., Du, X., Shi, C., Yuan, Y., . . . Yu, S. (2015). Genome sequence of cultivated Upland cotton (Gossypium hirsutum TM-1) provides insights into genome evolution. *Nature biotechnology*, *33*(5), 524-530.

Lin, Z., He, D., Zhang, X., Nie, Y., Guo, X., Feng, C., & Stewart, J. M. (2005). Linkage map construction and mapping QTL for cotton fibre quality using SRAP, SSR and RAPD [Article]. *Plant Breeding*, *124*(2), 180-187.

Monforte, A. J., & Tanksley, S. D. (2000). Development of a set of near isogenic and backcross recombinant inbred lines containing most of the Lycopersicon hirsutum genome in a L. esculentum genetic background: A tool for gene mapping and gene discovery. *Genome*, *43*(5), 803-813.

Paran, I., & Zamir, D. (2003). Quantitative traits in plants: beyond the QTL. *Trends in Genetics*, *19*(6), 303-306.

Paterson, A. H., Boman, R. K., Brown, S. M., Chee, P. W., Gannaway, J. R., Gingle, A. R., May, O. L., & Smith, C. W. (2004). Reducing the Genetic Vulnerability of Cotton. *Crop Science*, *44*(6), 1900-1901.

Paterson, A. H., Lin, Y. R., Li, Z., Schertz, K. F., Doebley, J. F., Pinson, S. R., Liu, S. C., Stansel, J. W., & Irvine, J. E. (1995). Convergent domestication of cereal crops by independent mutations at corresponding genetic Loci. *Science*, *269*(5231), 1714-1718.

Paterson, A. H., Wendel, J. F., Gundlach, H., Guo, H., Jenkins, J., Jin, D., Llewellyn, D., Showmaker, K. C., Shu, S., Udall, J., Yoo, M.-j., Byers, R., Chen, W., Doron-Faigenboim, A., Duke, M. V., Gong, L., Grimwood, J., Grover, C., Grupp, K., . . . Schmutz, J. (2012). Repeated polyploidization of Gossypium genomes and the evolution of spinnable cotton fibres. *Nature*, *492*(7429), 423-427.

Perkins Jr., H. H., Ethridge, D. E., & Bragg, C. K. (1984). Fiber. In *Cotton* (pp. 437-509).

Rong, J., Feltus, F. A., Waghmare, V. N., Pierce, G. J., Chee, P. W., Draye, X., Saranga, Y., Wright, R. J., Wilkins, T. A., May, O. L., Smith, C. W., Gannaway, J. R., Wendel, J. F., & Paterson, A. H. (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), 2577-2588.

Said, J. I., Knapka, J. A., Song, M., & Zhang, J. (2015). Cotton QTLdb: a cotton QTL database for QTL analysis, visualization, and comparison between Gossypium hirsutum and G. hirsutum × G. barbadense populations. *Mol Genet Genomics*, *290*(4), 1615-1625.

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

Said, J. I., Lin, Z., Zhang, X., Song, M., & Zhang, J. (2013). A comprehensive meta QTL analysis for fiber quality, yield, yield related and morphological traits, drought tolerance, and disease resistance in tetraploid cotton. *BMC Genomics*, *14*(1), 776.

Said, J. I., Song, M., Wang, H., Lin, Z., Zhang, X., Fang, D. D., & Zhang, J. (2015). A comparative meta-analysis of QTL between intraspecific Gossypium hirsutum and interspecific G. hirsutum × G. barbadense populations. *Mol Genet Genomics*, *290*(3), 1003-1025.

Shen, X., Guo, W., Zhu, X., Yuan, Y., Yu, J. Z., Kohel, R. J., & Zhang, T. (2005). Molecular mapping of QTLs for fiber qualities in three diverse lines in Upland cotton using SSR markers [Article]. *Molecular Breeding*, *15*(2), 169-181.

Szalma, S. J., Hostert, B. M., LeDeaux, J. R., Stuber, C. W., & Holland, J. B. (2007). QTL mapping with near-isogenic lines in maize [Article]. *Theoretical & Applied Genetics*, *114*(7), 1211-1228.

Taliercio, E., & Kloth, R. (2004). Expression and Characterization of a UDP-Glucose Pyrophosphorylase Gene in Cotton. *Journal of Cotton Science*, *8*(2), 91-98.

Wang, M., Tu, L., Lin, M., Lin, Z., Wang, P., Yang, Q., Ye, Z., Shen, C., Li, J., Zhang, L., Zhou, X., Nie, X., Li, Z., Guo, K., Ma, Y., Huang, C., Jin, S., Zhu, L., Yang, X., . . . Zhang, X. (2017). Asymmetric subgenome selection and cis-regulatory divergence during cotton domestication. *Nature Genetics*, *49*(4), 579-587.

Wang, Z., Zhang, D., Wang, X., Tan, X., Guo, H., & Paterson, A. H. (2013). A Whole-Genome DNA Marker Map for Cotton Based on the D-Genome Sequence of Gossypium raimondii L. *G3 Genes|Genomes|Genetics*, *3*(10), 1759-1767.

Xu, Z., He, J., Tehseen Azhar, M., Zhang, Z., Fan, S., Jiang, X., Jia, T., Shang, H., & Yuan, Y. (2022). UDP-glucose pyrophosphorylase: genome-wide identification, expression and functional analyses in Gossypium hirsutum. *PeerJ*, *10*, e13460.

Yu, J., Zhang, K., Li, S., Yu, S., Zhai, H., Wu, M., Li, X., Fan, S., Song, M., Yang, D., Li, Y., & Zhang, J. (2013). Mapping quantitative trait loci for lint yield and fiber quality across environments in a Gossypium hirsutum × Gossypium barbadense backcross inbred line population. *Theor Appl Genet*, *126*(1), 275-287.

Zeng, Q.-w., Qin, S., Song, S.-q., Zhang, M., Xiao, Y.-h., Luo, M., Hou, L., & Pei, Y. (2012). Molecular Cloning and Characterization of a Cytokinin Dehydrogenase Gene from Upland Cotton (Gossypium hirsutum L.). *Plant Molecular Biology Reporter*, *30*(1), 1-9.

Zhang, J. F., Percy, R. G., & McCarty, J. C. (2014). Introgression genetics and breeding between Upland and Pima cotton: a review. *Euphytica*, *198*, 1-12.

Zhang, T., Hu, Y., Jiang, W., Fang, L., Guan, X., Chen, J., Zhang, J., Saski, C. A., Scheffler, B. E., Stelly, D. M., Hulse-Kemp, A. M., Wan, Q., Liu, B., Liu, C., Wang, S., Pan, M., Wang, Y., Wang, D., Ye, W., . . . Chen, Z. J. (2015). Sequencing of allotetraploid cotton (Gossypium hirsutum L. acc. TM-1) provides a resource for fiber improvement. *Nature biotechnology*, *33*(5), 531-537.

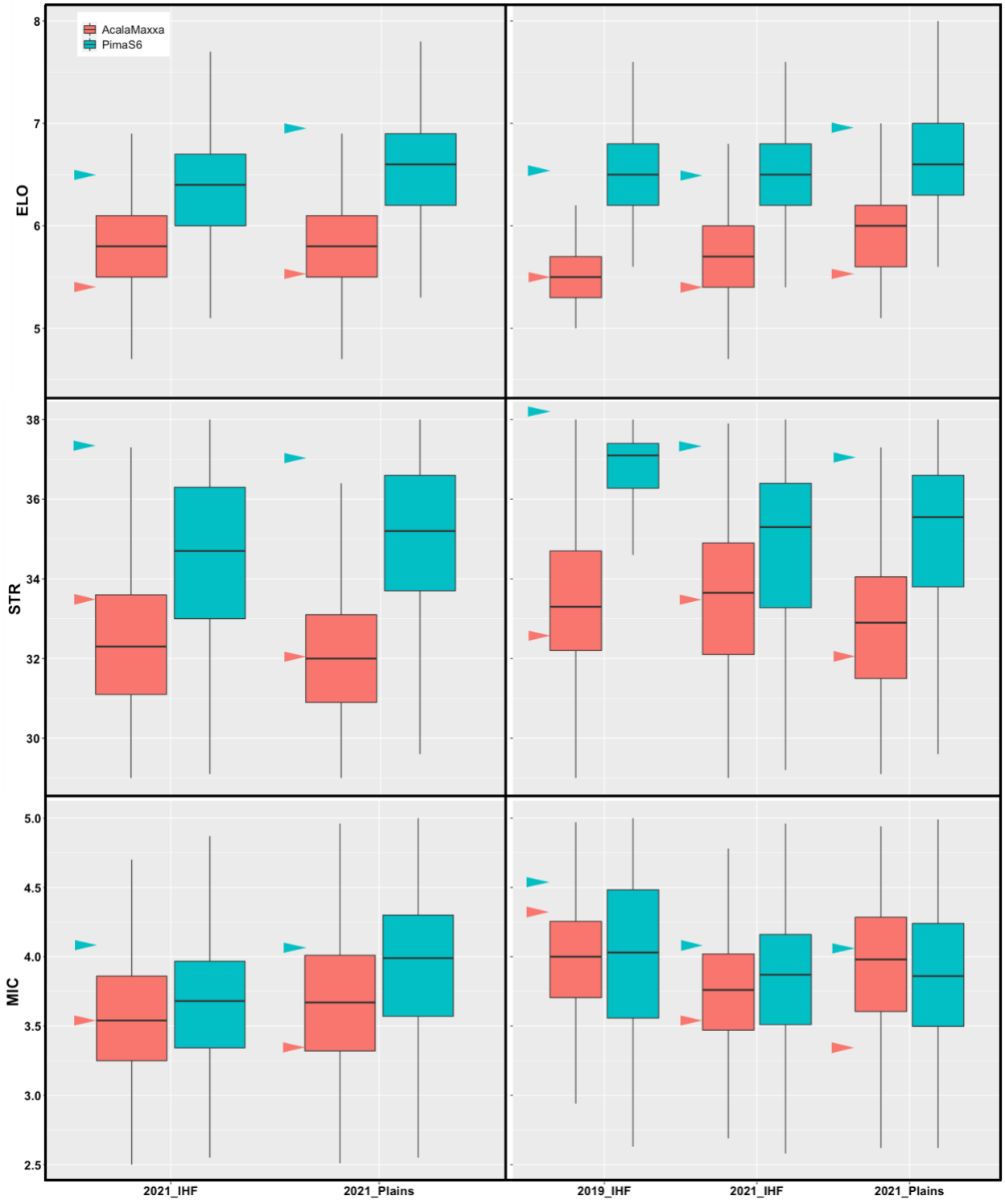

Figure 5.1. Distribution of fiber quality traits for the parents, reciprocal advanced backcrosses and NILs. Panels on the left show distribution of NILs for the two environments tested and panels on the right show distribution of the traits for advanced backcross populations. Green arrows represent Pima S6 parent and red arrows represent Acala Maxxa.

Table 5.1. Summary statistics of distribution of fiber quality traits in advanced reciprocal backcross populations (BC_5F_2) and parental lines.

Trait	Env	Parental means		Acala Maxxa Background		Pima S6 Background	
		Acala Maxxa	Pima S6	Mean (Sd)	Range	Mean (Sd)	Range
MIC	IHF 2019	04.33	04.35	03.94 (0.44)	02.51 - 04.97	04.02 (0.55)	02.63 - 05.26
	IHF 2021	03.56	04.11	03.73 (0.44)	02.45 - 04.78	03.83 (0.48)	02.47 - 05.07
	Plains 2021	03.38	04.10	03.93 (0.48)	02.40 - 04.94	03.83 (0.54)	02.40 - 04.99
ELO	IHF 2019	01.13	01.32	01.14 (0.03)	01.04 - 01.24	01.31 (0.05)	01.12 - 01.45
	IHF 2021	05.44	06.40	05.73 (0.40)	04.70 - 06.90	06.53 (0.49)	05.40 - 09.10
	Plains 2021	05.55	06.42	05.96 (0.45)	05.10 - 07.70	06.65 (0.53)	05.00 - 08.60
STR	IHF 2019	32.63	39.40	33.61 (2.01)	29.00 - 42.40	40.08 (2.45)	33.90 - 48.01
	IHF 2021	33.54	36.72	33.64 (2.28)	26.50 - 39.90	36.48 (3.56)	25.01 - 46.02
	Plains 2021	32.05	37.04	32.69 (2.25)	25.60 - 41.40	36.23 (2.95)	27.10 - 45.40
	IHF 2019	84.30	85.82	84.54 (1.15)	81.30 - 87.80	85.66 (1.25)	81.60 - 88.30

Table 5.2. Summary statistics of distribution of fiber quality traits in Near-isogenic lines (NILs) and parental lines. Sd denotes sample standard deviation.

Trait	Env	Parental means		Acala Maxxa Background		Pima S6 Background	
		Acala Maxxa	Pima S6	Mean (Sd)	Range	Mean (Sd)	Range
MIC	IHF 2021	3.56	4.11	3.53 (0.47)	2.15 - 5.01	3.65 (0.46)	02.32 - 4.87
	Plains 2021	3.38	4.10	3.65 (0.53)	2.03 - 5.21	3.91 (0.55)	02.33 - 5.35
ELO	IHF 2021	5.44	6.40	5.80 (0.48)	4.50 - 7.80	6.39 (0.52)	04.70 - 8.40
	Plains 2021	5.55	6.42	5.82 (0.46)	4.60 - 8.70	6.56 (0.54)	05.00 - 8.50
STR	IHF 2021	33.54	36.72	32.13 (2.27)	22.70 - 41.20	35.26 (3.23)	21.60 - 45.20
	Plains 2021	32.05	37.04	31.87 (1.88)	25.30 - 38.60	35.84 (2.59)	25.30 - 45.00

Table 5.3. Analysis of variance of fiber quality traits. DF denote degrees of freedom, SS denotes sum of squares, MS denotes mean sum of squares, PVE denotes percent variance explained by the factor. ABL = Advanced backcross lines, NIL = Near-isogenic lines.

			Acala Maxxa background					Pima S6 background				
Trait	Pop	Source	DF	SS	MS	F-value	PVE	DF	SS	MS	F-value	PVE
ELO	ABL	ENV	2	26.53	13.27	142.48***	18.10	2	5.40	2.70	22.07***	2.14
		REP(ENV)	3	0.15	0.05	0.53	0.10	3	3.17	1.06	8.64***	1.25
		GEN	172	53.62	0.31	3.35***	36.59	230	110.30	0.48	3.92***	43.62
		GEN*ENV	287	31.90	0.11	1.19	21.77	410	76.24	0.19	1.52***	30.15
		Error	369	34.36	0.09		23.44	472	57.73	0.12		22.83
	NIL	ENV	1	0.00	0.00	0.00	0.00	1	10.8	10.8	67.17***	2.89
		REP(ENV)	2	0.42	0.21	1.81	0.13	2	2.63	1.31	8.17***	0.70
		GEN	397	207.27	0.52	4.51***	63.83	407	223.19	0.55	3.41***	59.80
		GEN*ENV	389	40.32	0.10	0.90	12.42	369	52.53	0.14	0.89	14.07
		Error	663	76.71	0.12		23.62	523	84.09	0.16		22.53
STR	ABL	ENV	2	160.02	80.01	20.72***	3.96	2	3772.57	1886.29	263.53***	27.92
		REP(ENV)	3	9.90	3.30	0.85	0.24	3	558.75	186.25	26.02***	4.14
		GEN	172	1107.69	6.44	1.67**	27.39	230	2422.76	10.53	1.47**	17.93
		GEN*ENV	286	1341.87	4.69	1.21*	33.18	410	3391.59	8.27	1.16	25.10
		Error	369	1425.11	3.86		35.23	470	3364.11	7.16		24.90
	NIL	ENV	1	27.63	27.63	7.34**	0.44	1	79.43	79.43	10.51**	0.70
		REP(ENV)	2	29.08	14.54	3.86*	0.46	2	3.37	1.68	0.22	0.03
		GEN	397	2385.75	6.01	1.60***	37.58	407	4127.02	10.14	1.34*	36.58
		GEN*ENV	390	1402.76	3.60	0.96	22.10	369	3111.91	8.43	1.12	27.58
		Error	665	2502.63	3.76		39.42	524	3960.73	7.56		35.11
MIC	ABL	ENV	2	6.27	3.14	22.09***	3.60	2	9.74	4.87	35.24***	3.03
		REP(ENV)	3	16.44	5.48	38.59***	9.45	3	85.31	28.44	205.87***	26.55
		GEN	172	47.52	0.28	1.95***	27.31	230	72.46	0.32	1.28***	22.55
		GEN*ENV	282	52.39	0.19	1.31*	30.11	410	89.13	0.22	1.57	27.74

	Error	362	51.40	0.14		29.54	468	64.65	0.14		20.12
NIL	ENV	1	3.18	3.18	16.92***	0.87	1	23.34	23.34	131.45***	6.60
	REP(ENV)	2	5.00	2.50	13.32***	1.37	2	0.02	0.01	0.05	0.01
	GEN	397	165.04	0.42	2.21***	45.14	407	160.97	0.41	2.23***	45.50
	GEN*ENV	390	67.54	0.17	0.92	18.47	369	76.43	0.21	1.17	21.60
	Error	665	124.82	0.19		34.14	524	93.03	0.18		26.30

Table 5.4. QTLs for fiber quality traits (ELO, MIC and STR) identified in the Acala Maxxa background.

Trait	ENV	QTL	Marker	LOD	PVE	a	d	d/a
ELO	IHF 2019	qELO.Gh.19-IH.01	S3_4311227	5.13	17.39		1.33	
		qELO.Gh.19-IH.02	S10_98258409	3.61	12.81		0.55	
		qELO.Gh.19-IH.03	S12_15376394	5.55	18.41		0.71	
		qELO.Gh.19-IH.04	S12_51105721	4.83	16.28		0.59	
		qELO.Gh.19-IH.05	S15_3094439	4.87	17.37		0.53	
		qELO.Gh.19-IH.06	S15_59647072	4.37	14.26		0.73	
		qELO.Gh.19-IH.07	S24_19787425	3.12	12.11		0.57	
		qELO.Gh.19-IH.08	S25_16260844	3.53	13.35		0.56	
	IHF 2021	qELO.Gh.21-IH.01	S1_85772418	3.54	4.35	0.46		
		qELO.Gh.21-IH.02	S2_8696312	3.22	4.13	1.87	-0.63	-0.34
		qELO.Gh.21-IH.03	S5_85261315	4.21	4.85	0.41	1.65	4.02
		qELO.Gh.21-IH.04	S6_5856948	4.99	3.01	0.99		
		qELO.Gh.21-IH.05	S12_51105721	3.66	15.53		0.51	
		qELO.Gh.21-IH.06	S14_58546089	6.60	3.59	1.04		
		qELO.Gh.21-IH.07	S20_53920551	8.62	4.15	0.61		
		qELO.Gh.21-IH.08	S21_52944095	4.10	7.39		0.50	
		qELO.Gh.21-IH.09	S23_57997331	4.34	9.82	0.37		
	Plains 2021	qELO.Gh.21-PL.01	S2_8987426	5.35	6.51	2.19	1.29	0.59
		qELO.Gh.21-PL.02	S6_89095603	3.51	13.37		1.45	
		qELO.Gh.21-PL.03	S9_7254005	6.29	7.61	1.89	-0.29	-0.15
		qELO.Gh.21-PL.04	S12_15376394	4.05	15.03		0.25	
		qELO.Gh.21-PL.05	S15_3094439	4.26	5.18	0.88		
MIC	IHF 2019	qMIC.Gh.19-IH.01	S3_15940226	3.38	11.56		-0.49	
		qMIC.Gh.19-IH.02	S10_98258409	4.86	9.97		-0.49	
		qMIC.Gh.19-IH.03	S15_565263	3.61	9.38		-0.39	
		qMIC.Gh.19-IH.04	S19_6255385	5.20	17.43		-0.25	

		qMIC.Gh.19-IH.05	S22_7493686	5.39	8.77		-0.29	
	IHF 2021	qMIC.Gh.21-IH.01	S1_16078074	3.27	4.18	-0.45	-0.33	0.73
		qMIC.Gh.21-IH.02	S5_46132258	3.41	4.35	0.11	1.23	11.18
		qMIC.Gh.21-IH.03	S7_47004922	3.25	3.75	0.51	-1.32	-2.59
		qMIC.Gh.21-IH.04	S10_11070841	3.42	4.36	-1.02		
		qMIC.Gh.21-IH.05	S22_36659220	1.52	2.55		0.27	
STR	IHF 2019	qSTR.Gh.19-IH.01	S9_70264132	4.02	9.76		1.12	
		qSTR.Gh.19-IH.02	S11_84002686	4.69	21.89		6.48	
		qSTR.Gh.19-IH.03	S13_67883692	3.05	9.48		1.95	
		qSTR.Gh.19-IH.04	S18_15452322	3.30	11.97		1.33	
		qSTR.Gh.19-IH.05	S23_60016776	3.00	10.29		2.08	
		qSTR.Gh.19-IH.06	S24_17478523	8.99	26.97		2.33	
		qSTR.Gh.19-IH.07	S25_57473917	8.46	19.02		4.21	
	IHF 2021	qSTR.Gh.21-IH.01	S9_3895750	3.05	3.85	-0.39	-8.69	22.28
		qSTR.Gh.21-IH.02	S13_75310113	3.14	9.68		5.00	
		qSTR.Gh.21-IH.03	S19_4973364	4.75	7.37		2.57	
		qSTR.Gh.21-IH.04	S21_51861885	3.10	13.91		-3.28	
		qSTR.Gh.21-IH.05	S25_43210819	4.08	5.18	2.12		
	Plains 2021	qSTR.Gh.21-PL.01	S1_37032466	3.36	13.29		10.81	
		qSTR.Gh.21-PL.02	S3_4311227	3.98	14.56		10.83	
		qSTR.Gh.21-PL.03	S18_2874202	4.04	3.54	5.46	-1.83	-0.34
		qSTR.Gh.21-PL.04	S21_7645491	3.57	12.22		-3.91	
		qSTR.Gh.21-PL.05	S22_1396527	4.14	3.66	4.18		
		qSTR.Gh.21-PL.06	S26_52187750	4.21	10.07		2.69	

Table 5.5. QTLs for fiber quality traits (ELO, MIC and STR) identified in the Pima S6 background.

Trait	ENV	QTL	Marker	LOD	PVE	a	d	d/a
ELO	IHF 2019	qELO.Gb.19-IH.01	S3_84465939	4.54	12.03		0.61	
		qELO.Gb.19-IH.02	S8_39417905	4.72	7.39		-0.39	
	IHF 2021	qELO.Gb.21-IH.01	S6_14751779	4.42	6.35	-1.06	-0.51	0.48
		qELO.Gb.21-IH.02	S7_39864766	3.97	5.59	1.22	1.82	1.49
		qELO.Gb.21-IH.03	S8_66072956	3.10	3.67	-0.57	-0.33	0.58
		qELO.Gb.21-IH.04	S9_68857432	3.80	4.48	0.92	0.31	0.34
		qELO.Gb.21-IH.05	S10_71211391	3.60	5.08	-0.37	2.02	-5.46
		qELO.Gb.21-IH.06	S16_30879676	4.84	9.87		1.67	
		qELO.Gb.21-IH.07	S18_20136056	3.27	4.98	0.92	0.12	0.13
	Plains 2021	qELO.Gb.21-PL.01	S4_47743671	3.46	5.08	0.96	-0.45	-0.47
		qELO.Gb.21-PL.02	S6_14751779	3.14	4.63	-0.94	-0.48	0.51
		qELO.Gb.21-PL.03	S9_50755816	3.22	4.37	1.19	0.59	0.50
		qELO.Gb.21-PL.04	S18_20136056	3.90	3.95	1.16	0.39	0.34
		qELO.Gb.21-PL.05	S22_8187981	4.15	6.84		1.66	
		qELO.Gb.21-PL.06	S26_8664132	3.81	6.09		0.89	
MIC	IHF 2019	qMIC.Gb.19-IH.01	S2_45391696	4.85	7.74		-0.47	
		qMIC.Gb.19-IH.02	S5_13365934	2.90	8.00		-1.16	
		qMIC.Gb.19-IH.03	S11_83610252	3.48	9.37		-0.73	
		qMIC.Gb.19-IH.04	S21_61479558	4.57	6.99		-0.88	
	IHF 2021	qMIC.Gb.21-IH.01	S2_42304414	3.59	5.06	-0.71	-0.80	1.13
		qMIC.Gb.21-IH.02	S3_39931753	3.43	3.45	0.43		
		qMIC.Gb.21-IH.03	S8_11482715	4.24	3.27	0.78	-0.56	-0.72
		qMIC.Gb.21-IH.04	S10_86400950	3.68	3.81	-0.87	-0.41	0.47
	Plains 2021	qMIC.Gb.21-PL.01	S1_16503983	4.38	3.52	-0.72	0.06	-0.08
		qMIC.Gb.21-PL.02	S3_43331620	3.18	3.45	0.26		

		qMIC.Gb.21-PL.03	S8_63786462	4.15	4.99	-0.89	-0.12	0.13
		qMIC.Gb.21-PL.04	S16_14838576	3.19	9.97		-1.20	
		qMIC.Gb.21-PL.05	S23_31371484	3.59	8.57		-0.98	
STR	IHF 2019	qSTR.Gb.19-IH.01	S2_21552398	4.17	11.12		-3.02	
		qSTR.Gb.19-IH.02	S11_73481854	3.41	9.19		-2.93	
		qSTR.Gb.19-IH.03	S24_3236094	3.97	10.84		7.23	
	IHF 2021	qSTR.Gb.21-IH.01	S3_40841772	3.42	4.95	4.73		
		qSTR.Gb.21-IH.02	S12_40029401	3.21	4.65	-12.30		
		qSTR.Gb.21-IH.03	S21_36351252	3.82	9.69		-7.27	
		qSTR.Gb.21-IH.04	S24_14054412	4.44	8.42		-11.65	
	Plains 2021	qSTR.Gb.21-PL.01	S10_13522710	4.12	6.72		-3.49	
		qSTR.Gb.21-PL.02	S14_11729890	3.15	6.84		-9.13	

Table 5.6. Subgenomic distribution of fiber quality traits.

	Acala Maxxa background		Pima S6 background		Overall	
Trait	At	Dt	At	Dt	At	Dt
ELO	13	9	7	5	20	14
STR	7	11	5	4	12	15
MIC	6	4	10	3	16	7
Total	26	24	22	12	48	36

Table 5.7. Chromosomal distribution of fiber quality QTLs.

	Acala Maxxa Background			Pima S6 Background		
Chr	No of QTLs	PV < 10%	Traits	No of QTLs	PV < 10%	Traits
Chr 1	3	2	ELO, MIC, STR	1	1	MIC
Chr 2	2	2	ELO	3	2	MIC, STR
Chr 3	3	0	ELO, MIC, STR	4	3	ELO, MIC, STR
Chr 4				1	1	ELO
Chr 5	2	2	ELO, MIC	1	1	MIC
Chr 6	2	1	ELO	2	2	ELO
Chr 7	1	1	MIC	1	1	ELO
Chr 8				4	4	ELO, MIC
Chr 9	3	3	ELO, STR	2	2	ELO
Chr 10	3	2	ELO, MIC	3	3	ELO, MIC, STR
Chr 11	1	0	STR	2	2	MIC, STR
Chr 12	4	0	ELO	1	1	STR
Chr 13	2	2	STR			
Chr 14	1	1	ELO	1	1	STR
Chr 15	4	2	ELO, MIC			
Chr 16				2	1	ELO, MIC
Chr 18	2	1	STR	2	2	ELO
Chr 19	2	1	MIC, STR			
Chr 20	1	1	ELO			
Chr 21	3	1	ELO, STR	2	2	MIC, STR
Chr 22	3	3	MIC, STR	1	1	ELO
Chr 23	2	1	ELO, STR	1	1	MIC
Chr 24	2	0	ELO, STR	2	1	STR
Chr 25	3	1	ELO, STR			
Chr 26	1	0	STR	1	1	ELO
Total	50	27 (54%)		37	33 (89.19%)	

Table 5.8. Clustering of fiber quality QTLs.

Cluster name	Chr	QTLs in cluster	Physical Coordinates (Mb)
		Acala Maxxa background	
cQTL.Gh.01.1	1	qMIC.Gh.21-IH.01, qSTR.Gh.21-PL.01	16.08 - 37.03
cQTL.Gh.03.1	3	qELO.Gh.19-IH.01, qSTR.Gh.21-PL.02, qMIC.Gh.19-IH.01	4.31 - 15.94
cQTL.Gh.09.1	9	qSTR.Gh.21-IH.01, qELO.Gh.21-PL.03	3.90 - 7.25
cQTL.Gh.10.1	10	qELO.Gh.19-IH.02, qMIC.Gh.19-IH.02	95.25 - 99.31
cQTL.Gh.15.1	15	qMIC.Gh.19-IH.03, qELO.Gh.21-PL.05, qELO.Gh.19-IH.05	0.57 - 3.09
cQTL.Gh.19.1	19	qMIC.Gh.19-IH.04, qSTR.Gh.21-IH.03	4.97 - 6.96
cQTL.Gh.21.1	21	qSTR.Gh.21-IH.04, qELO.Gh.21-IH.08	51.86 - 52.94
cQTL.Gh.22.1	22	qMIC.Gh.19-IH.05, qSTR.Gh.21-PL.05	1.40 - 7.49
cQTL.Gh.23.1	23	qSTR.Gh.19-IH.05, qELO.Gh.21-IH.09	58.01 - 60.02
		Pima S6 background	
cQTL.Gb.02.1	2	qSTR.Gb.19-IH.01, qMIC.Gb.21-IH.01, qMIC.Gb.19-IH.01	21.55 - 45.39
cQTL.Gb.03.1	3	qMIC.Gb.21-PL.02, qSTR.Gb.21-IH.01, qMIC.Gb.21-IH.02	39.93 - 43.33
cQTL.Gb.08.1	8	qELO.Gb.21-IH.03, qMIC.Gb.21-PL.03	63.79 - 66.07
cQTL.Gb.11.1	11	qSTR.Gb.19-IH.02, qMIC.Gb.19-IH.03	73.48 - 83.61
cQTL.Gb.16.1	16	qMIC.Gb.21-PL.04, qELO.Gb.21-IH.06	14.84 - 30.88

Table 5.9. Correspondence of fiber quality QTLs between current and previous study. P-values obtained using hypergeometric probability distribution (see methods).

	Previous Study			Common QTLs		P.value	
Trait	Pop	# QTLs	Reference	GH	GB	GH	GB
MIC	RIL	32	Liu et al., 2017	5	5	**0.009**	**0.030**
	F2	33	Deng et al., 2019	5	3	**0.011**	0.218
	F2:3	78	Diouf et al., 2018	2	4	0.129	0.197
	F2:3	2	Guo et al., 2014	1	1	0.095	0.122
	F2:3	16	He et al., 2007	2	3	0.149	0.061
	F2	3	Huang et al., 2015	2	1	**0.007**	0.172
	MAGIC	27	Huang et al., 2018	4	5	**0.027**	**0.015**
	RIL	142	Jamshed et al., 2016	5	6	0.091	**0.035**
	RIL	49	Jia et al., 2016	2	4	0.291	0.212
	F2	1	Jiang et al., 1998	1	1	0.051	0.065
	F2	4	Kumar et al., 2019	0	0	0.813	0.763
	BC1	11	Lacape et al., 2005	4	5	**0.001**	**0.000**
	RIL	26	Li et al., 2016	4	4	**0.024**	0.055
	RIL	104	Li et al., 2018	5	6	0.250	0.205
	F2:3	14	Liang et al., 2013	2	1	0.124	0.395
	RIL	160	Liu et al., 2018	8	7	0.310	**0.020**
	RIL	28	Ma et al., 2016	2	3	0.271	0.177
	F2	6	Rong et al., 2007	3	4	**0.002**	**0.000**
	F2	23	Saranga et al., 2004	6	6	**0.000**	**0.001**
	RIL	12	Shang et al., 2015	1	1	0.356	0.386
	RIL	134	Shang et al., 2016	5	6	0.133	0.063
	F2:3	6	Shen et al., 2005	1	1	0.237	0.285
	RIL	33	Shen et al., 2007	0	1	0.157	0.245
	RIL	21	Tan et al., 2015	4	4	**0.011**	**0.029**
	RIL	58	Tan et al., 2018	2	8	0.247	**0.010**
	RIL	29	Tang et al., 2015	2	3	0.278	0.186
	RIL	5	Frelichowski et al., 2006	2	2	**0.020**	**0.033**
	F2	10	Wang et al., 2015	1	2	0.327	0.111
	RIL	44	Wang et al., 2015	2	2	0.305	0.249
	BIL	16	Yu et al., 2013	2	2	0.149	0.205
	RIL	20	Yu et al., 2014	4	3	**0.009**	0.098
	RIL	6	Zheng et al., 2009	1	1	0.237	0.285
	F2:3	2	Zhang et al., 2005	1	1	0.095	0.122
	RIL	32	Zhang et al., 2015	0	0	0.167	0.096

	CSIL	34	Zhang et al., 2016	2	4	0.301	0.111
	RIL	10	Wang et al., 2007	1	1	0.327	0.367
ELO	RIL	17	Liu et al., 2017	7	4	**0.001**	**0.023**
	F2	35	Deng et al., 2019	7	4	**0.043**	0.157
	F2:3	42	Diouf et al., 2018	7	6	0.088	**0.048**
	F2	3	Huang et al., 2015	2	1	**0.031**	0.194
	MAGIC	7	Huang et al., 2018	2	1	0.142	0.337
	RIL	119	Jamshed et al., 2016	7	6	**0.004**	0.062
	RIL	34	Jia et al, 2016	8	7	**0.013**	**0.004**
	F2	6	Kumar et al., 2019	1	1	0.376	0.311
	BC1	12	Lacape et al., 2005	5	4	**0.004**	**0.007**
	RIL	22	Li et al., 2016	3	2	0.238	0.291
	RIL	88	Li et al., 2018	11	6	0.149	0.204
	F2	22	Liang et al., 2013	4	2	0.134	0.291
	RIL	3	Liu et al., 2016	1	1	0.264	0.194
	RIL	26	Ma et al., 2016	4	2	0.178	0.302
	F2	5	Rong et al., 2007	3	3	**0.010**	**0.003**
	RIL	10	Shang et al., 2015	4	1	**0.013**	0.384
	RIL	50	Shang et al., 2016	7	4	0.144	0.234
	F2:3	12	Shen et al., 2005	4	4	**0.026**	**0.007**
	RIL	4	Shen et al., 2007	1	1	0.314	0.241
	F2	10	Saranga et al., 2004	4	2	**0.013**	0.137
	RIL	10	Tan et al., 2015	4	4	**0.013**	**0.003**
	RIL	76	Tan et al., 2018	9	5	0.174	0.206
	RIL	21	Tang et al., 2015	5	5	**0.046**	**0.010**
	RIL	9	Wang et al., 2007	2	2	0.196	0.117
	RIL	64	Wang et al., 2015	5	2	0.125	0.067
	F2	67	Wang et al., 2016	7	5	0.188	0.223
	BIL	12	Yu et al., 2013	5	2	**0.004**	0.174
	RIL	12	Yu et al., 2014	4	3	**0.026**	**0.043**
	RIL	3	Zheng et al., 2009	1	1	0.264	0.194
	F2:3	2	Zhang et al., 2005	2	1	**0.012**	0.139
	CSIL	27	Zhang et al., 2016	4	5	0.188	**0.028**
STR	F2	4	Jiang et al., 1998	2	2	**0.039**	**0.010**
	F2	39	Deng et al., 2019	6	4	0.073	0.059
	F2:3	58	Diouf et al., 2018	6	4	0.189	0.162
	F2:3	2	Guo et al., 2014	1	1	0.165	0.086
	F2:3	6	He et al., 2007	1	0	0.344	0.756
	F2	10	Huang et al., 2015	2	2	0.174	0.061
	MAGIC	19	Huang et al., 2018	5	6	**0.014**	**0.000**

RIL	159	Jamshed et al., 2016	7	4	**0.000**	**0.016**
RIL	44	Jia et al, 2016	9	6	**0.005**	**0.004**
F2	6	Kumar et al., 2019	1	0	0.344	0.756
BC1	14	Lacape et al., 2005	7	3	**0.000**	**0.016**
RIL	18	Li et al., 2016	1	1	0.332	0.391
RIL	72	Li et al., 2018	9	3	0.087	0.275
F2	30	Liang et al., 2013	4	1	0.164	0.374
RIL	28	Ma et al., 2016	4	4	0.145	**0.021**
F2	2	Mei et al., 2004	1	0	0.165	0.912
F2	6	Rong et al., 2007	1	1	0.344	0.219
RIL	14	Shang et al., 2015	1	0	0.380	0.513
RIL	101	Shang et al., 2016	8	4	0.169	0.248
F2:3	10	Shen et al., 2005	3	3	**0.042**	**0.006**
RIL	18	Shen et al., 2007	2	1	0.289	0.391
F2	21	Saranga et al., 2004	10	4	**0.000**	**0.007**
RIL	33	Tan et al., 2015	8	5	**0.003**	**0.006**
RIL	104	Tan et al., 2018	9	5	0.191	0.260
RIL	26	Tang et al., 2015	4	4	0.126	**0.016**
RIL	6	Frelichowski et al., 2006	2	1	0.082	0.219
F2	54	Wang et al., 2015	7	3	0.104	0.256
RIL	13	Wang et al., 2015	1	0	0.388	0.539
BIL	8	Yu et al., 2013	2	2	0.129	**0.041**
RIL	18	Yu et al., 2014	4	0	**0.049**	0.420
RIL	4	Zheng et al., 2009	0	0	0.684	0.831
F2:3	4	Zhang et al., 2005	1	2	0.275	**0.010**
RIL	34	Zhang et al., 2015	1	0	0.121	0.180
CSIL	26	Zhang et al., 2016	4	3	0.126	0.078
RIL	134	Zhang et al., 2017	11	8	0.172	0.117
RIL	5	Wang et al., 2007	2	2	0.060	**0.016**

CHAPTER 6

GENETIC DISSECTION OF QUANTITTIVE VARIATION IN COTTON FIBER LINT PERCENTAGE IN A RECIPROCAL INTERSPECIFIC SET OF NEAR-ISOGENIC LINES

Jeevan Adhikari and Andrew H Paterson. To be submitted to *Molecular Breeding.*

Abstract

Estimation of precise locations and accurate effects of quantitative trait loci (QTLs) affecting fiber quality traits is an essential prerequisite for using these QTLs in breeding programs via marker assisted selection. Advanced backcross lines (ABLs) and/or near-isogenic lines (NILs) carrying one or few chromosomal segments introgressed from donor parent into the recipient background serve an important purpose of delineating the boundaries of a QTL to a small and rather precise genomic location, and also add to the accuracy of the effects of the QTL by reducing background noise caused by multiple segregating donor chromatin. We developed a set of reciprocal interspecific advanced backcross populations by crossing two elite cotton cultivars, Acala Maxxa (*G. hirsutum*) and Pima S6 (*G. barbadense*) representing the two major domesticated species of cotton. We selected a total of 399 individuals in the Acala Maxxa background and 423 individuals in the Pima S6 background, tiling 78.72% and 71.48% of the recipient genome respectively. These two populations (ABLs and NILs) were then used to identify genomic locations underpinning lint percentage in cotton. Phenotypic evaluation of these lines in three environments revealed a total of 32 QTLs, with one stable QTL identified on chromosome 17 at two environments - IHF-2019 and IHF-2021. We were able to identify both major QTLs (explaining > 10% of phenotypic variation) as well as small-effect QTLs (explaining < 10% of phenotypic variation) in almost equal proportions, indicating the merit of the developed populations in identifying both significant and nominal effects of QTLs. Although a couple of reciprocal QTLs with both antagonistic and similar effects were identified, limited reciprocity of majority of the QTLs in the two backgrounds shows strong influence of recipient genome, in addition to the combined consequences of epistasis, small phenotypic effects and imperfect coverage of donor chromatin in the recipient background.

Introduction

Cotton is the most important natural fiber and the leading textile crop of global economic importance. While two domesticated species of cotton - Upland cotton (*Gossypium hirsutum* L.) and Pima Cotton (*G. barbadense* L.), dominate global cotton production, collectively accounting for more than 95% of world's fiber yield, Upland cotton is the major lint producer and is known for its lint yield. On the other hand, Pima cotton is popular for its quality parameters and not so much for lint yield. Although the quality of lint fibers has gained much attention in the recent years, lint yield still remains the major factor in cotton industry. Despite stiff competition from synthetic fibers, consumer preference towards natural fiber and elevated market demands for high quality textile fibers have generated renewed interest in breeding for better fiber quality (Peng W. Chee & B. Todd Campbell, 2009). However, there is a low level of genetic differentiation in the cultivated cotton gene pool (Fang et al., 2017; Wang et al., 2017) resulting from the depletion of favorable variation due to evolutionary and breeding (domestication) bottlenecks (Adhikari, 2015a; Andrew H. Paterson et al., 2004). In addition, negative correlations between yield components and fiber quality parameters have restricted simultaneous breeding gains.

Improving fiber quality parameters simultaneously with increasing lint yield has been an important goal for cotton improvement. Cotton yield is typically influenced by a combination of complex quantitative traits like lint percentage, boll weight, boll number, seed index and lint index (Imran et al., 2012; Qin et al., 2008). Lint percentage (LP) is a key contributor to lint yield, which in turn is the decisive factor for the crops economic value. Several previous studies have shown lint percentage to have both positive as well as negative correlation with yield as well as other fiber quality traits (Adhikari et al., 2017; Imran et al., 2012; Qin et al., 2008; Wang et al., 2007; Wang

et al., 2011; Zhu et al., 2021). The negative correlation between these important traits makes the improvement of one or more of these traits simultaneously very challenging.

Genetic dissection of such complex traits components into underlying quantitative trait loci (QTL) creates opportunities to manipulate quantitative traits based on Mendelian principles, using techniques such as marker-assisted selection (MAS), QTL pyramiding, fine mapping, and map-based cloning. Utilizing suitable populations segregating for a few chromosomal segments from the donor parent and using recent genome scanning technologies to delineate donor chromatin precisely might provide a way for breaking tight linkages between genomic locations governing traits with opposite effects. Introgressive breeding has been one of the successful techniques employed not only to transfer or combine desirable traits from two or more related species but also to identify genomic locations underpinning complex traits by using segregating populations developed via introgressive hybridization.

Among the two widely cultivated cotton species, Upland cotton is known for its high yield potential while Pima cotton has superior fiber quality. Transferring desirable traits from Pima cotton to Upland cotton has been a long-standing goal for cotton breeders and geneticists. Numerous introgressive breeding programs have been conducted in cotton in the past six or seven decades to simultaneously improve fiber yield and quality by interspecific hybridization (Zhang et al., 2014). In the past two decades alone, geneticists have tagged hundreds of QTLs underlying cotton fiber yield and quality parameters (J. Rong et al., 2007; J. I. Said, J. A. Knapka, et al., 2015; J. I. Said et al., 2013; J. I. Said, M. Song, et al., 2015). With progress in genetic mapping technologies, more QTLs are being tagged and published, but the use of these QTLs in marker-assisted breeding for cotton improvement has been limited (Khanal, 2018; J. I. Said, J. A. Knapka, et al., 2015). Stability (consistency) of QTL expression across environments and among

generations and/or genetic backgrounds, variable effects and low precision in mapped locations are among the major concerns (Fang, 2015). Elevated noise from multiple chromosomal segments segregating independently, some of which span entire chromosomes, often result in over- or underestimation of the effects of mapped QTLs and thus in lack of reproducible effects across environments and/or generations. Experimental populations with reduced background noise and smaller size of introgressed chromatin segments facilitate estimation of the location and effects of these genomic benchmarks more precisely. Near-isogenic lines (NILs), with only one chromatin segment from the donor parent, or advanced backcross lines with only a few donor segments segregating in the recipient genome, can serve the purpose of eliminating much of the background noise present in early generation populations.

Near isogenic lines (NILs), with only one chromosomal segments from a donor parent in an homogeneous genetic background of a recurrent parent, can serve as a good resource for both genetic mapping and breeding (R. Kooke et al., 2012). Since there are few (ideally one) introgressed segment(s) in each NIL, phenotypes due to QTLs on the segment(s) are rendered much more discrete than in F_2 or backcross populations, often behaving as simple Mendelian factors (Paran & Zamir, 2003). QTL mapping based on NILs can thus increase accuracy of QTL position and detect small effect QTLs that might otherwise be obscured by larger-effect genes in more complex populations. In addition, because of the fixed genotype of NILs, they can be replicated in different environments to test interaction between genetic and environmental factors (Monforte & Tanksley, 2000). By crossing NILs to the recurrent parent, fine mapping of specific QTLs toward their cloning is facilitated.

To better understand the genetic architecture of lint percentage in populations consisting of chromatin segments reciprocally transmitted between the two mostly cultivated species of

cotton, we constructed a set of reciprocal interspecific advanced backcross populations using Acala Maxxa (*G. hirsutum*) and Pima S6 (*G. barbadense*) as parents, tiling 71.48% of the Acala Maxxa genome in Pima S6 background (hereafter GB background) and 78.72% of the Pima S6 genome in the Acala Maxxa background (hereafter GH background). With advanced backcross populations and NILs, we found 32 QTLs for lint percentage.

Materials and Methods

Population development

Plant materials used in this study were developed from a set of reciprocal crosses between *Gossypium hirsutum* acc. Acala Maxxa and *G. barbadense* acc. Pima S6 (both inbred lines). These genotypes have been extensively used to produce several molecular tools and resources including BAC libraries and Illumina genome sequences. Reciprocal advanced backcross populations were developed by first crossing the parents in a two-way cross (Acala Maxxa (♀)× Pima S6 (♂) – hereafter referred to as *G. hirsutum* background; and Pima S6 (♀) × Acala Maxxa (♂) – hereafter referred to as *G. barbadense* background), then independently backcrossing F_1 plants to the respective female parent to create 300 to 400 BC_1 progenies for each cross. The backcrossing scheme included planting only one seed from each preceding backcross to generate the next generation (Figure 3.1).

After five generation of backcrossing, 173 BC_5F_1 plants from the *G. hirsutum* background and 231 BC_5F_1 plants from the *G. barbadense* background were self-pollinated and a total of 404 BC_5F_2 families were grown at Iron Horse Farm (IHF), Watkinsville, Georgia in 2019 and 2021 and at Southwest Georgia Research Station (Plains), Plains, Georgia in 2021 under cultural conditions consistent with commercial irrigated cotton production. Individual BC_5F_2 plants that

contained only one introgressed segment from the donor parent were deemed as NILs. We identified a total of 397 NILs in the Acala Maxxa background and a total of 423 NILs in the Pima S6 background. Selfed seeds of these NILs (BC_5F_3 seeds) were grown at IHF and Plains in 2021 under cultural conditions consistent with commercial irrigated cotton production.

Phenotypic evaluation and data analysis

Two replications of each BC_5F_2 families and NILs were planted in a randomized complete block design (RCBD) in three environments (IHF-2019, IHF-2021 and Plains-2021). Six replications each of the two parents were included in all three environments. Bolls were hand harvested and fiber samples were weighed to get total seed weight. Fiber samples were then ginned in laboratory gin to separate fiber and seeds. Ginning was done in the same gin to obtain consistent separation of lint from seeds. Lint percentage was then calculated as:

$$Lint\ percentage = \frac{Ginned\ fiber\ weight}{Total\ fiber\ weight} \times\ 100\%$$

All statistical analyses were conducted in R programming language. Single marker analyses were done in R/qtl (Broman & Sen, 2009). The significance threshold was set to LOD of 3, to mitigate the multiple-comparison problem. Filtration of significant markers adopts the method proposed by Szalma et al. (2007b). If several markers on the same introgressed segment show significant association with phenotype, the most significant one was reported. For the co-segregation of multiple introgressions, the QTL location is examined as follows. First, if multiple families show significance for the trait and carry overlapping introgression, the introgression is

considered to carry QTL. Second, if the co-segregation of introgressions is in single families, the most significant introgression is considered to carry QTL.

Phenotypic variance explained by each locus was reported by taking the most significant marker as independent variable and phenotypic value as dependent variable in R (R Core Team 2016). Additive effects were estimated by half the difference of phenotypic values between the lines carrying the homozygous introgression and lines not carrying the introgression. Dominance effects were estimated by the difference of phenotypic values between the lines carrying the heterozygous introgression and the remaining lines that do not carry the introgression. If multiple or overlapping introgressions were present at both homozygous and heterozygous state, the estimation of additive effects utilized the lines carrying the introgression at homozygous state only and, the estimation of dominance utilized the lines carrying the introgression at heterozygous state only.

Gene actions for the QTLs were determined by calculating the degree of dominance (absolute values) for each QTL. The degree of dominance is the ratio of dominance effect to additive effect (d/a) of the QTL and based on this ratio, gene action of the QTLs can be categorized as (i) additive ($0 < d/a < 0.2$) (ii) partially dominant ($0.2 < d/a < 0.8$) (iii) dominant ($0.8 < d/a < 1.2$ and (iv) over-dominant ($d/a > 1.2$). QTLs with dominant and over-dominant effects are considered to have heterotic effect or heterozygous advantage.

Identification of common QTLs

Common QTL is defined as either the same marker is detected in the two reciprocal populations, or two different markers are detecting exactly the same introgression(s) in each population. Correspondence between QTLs for a trait across the entire genome is inferred using

the hypergeometric probability function. The model was adopted from (Feltus et al., 2006): *p* is the probability of non-random correspondence of QTLs being compared for a given trait, *n* is the number of comparable intervals which is calculated by dividing the total genome size by average introgression size in both populations; *m* is the number of common QTLs; *l* is the number of QTLs in the GH background; *s* is the number of QTLs in the GB background. The same model was also adopted to detect correspondence between QTLs reported in this study with those previously published. In this case, *l* is the total number of QTLs identified in the larger sample (study reporting higher number of QTLs) and *s* is the number of QTLs identified in the smaller sample

$$p = \frac{\binom{l}{m}\binom{n-1}{s-m}}{\binom{n}{s}}$$

Candidate gene identification

In silico annotation was performed on the identified QTLs to look for candidate genes related to flowering habit in cotton. For each QTL identified in the study, the genomic region spanning 50 kb on each side of the most significantly associated marker was used for in silico analysis. The DNA sequence from this tightly linked region was used to look for *G. hirsutum* genes in the CottonGen database and these genes were then analyzed for biological functions, with particular focus on fiber growth and development.

Results

Genomic composition of NILs and ABLs

The distribution of markers in the reciprocal populations is presented in Table 2.1. Genomic distribution of the introgressed chromosomal segments in ABLs is shown in Table 2.2

and Figure 2.3 and 2.4, while coverage of donor genome by the NILs is shown in Table 3.1 and Figures 3.2 and 3.3. Relevant information about the genomic composition of NILs and ABLs are presented in the Results section of chapters 2 and 3.

Phenotypic performance ABLs and NILs

The phenotypic performance of the two parents, reciprocal backcross populations, and reciprocal NIL populations is shown in Figure 6.1. The distribution of lint percentage was approximately normal (Shapiro and Wilk test; $p > 0.05$) and typical of quantitative inheritance. Acala Maxxa slightly outperformed Pima S6 in all three environments (Table 6.1, Figure 6.1). The average performance of the two reciprocal populations was not significantly different, but both population types performed differently in the three environments tested. Transgressive segregation was seen in all populations and environments in both directions (superior as well as inferior to both parents) (Figure 6.1).

To identify the effect of genotypes and environment in the overall performance, we conducted analysis of variance (ANOVA) keeping all variables as fixed factors. Results showed significant effects of both genotype (GEN), and genotype-by-environment (GXE). GEN captured the most variation for all traits in both population types (Table 6.2). GXE also captured significant amount of variation in the phenotypes and thus precluded the use of combined phenotypic values in identification of QTLs for these traits. Thus, marker trait association and identification of fiber quality QTLs for the three traits under study was performed separately for each environment tested.

Marker trait association and overview of QTLs

A total of 32 QTLs were identified for lint percentage (17 in the Acala Maxxa background (Table 6.3) and 15 in the Pima S6 background (Table 6.4)). Phenotypic variances explained by these QTLs ranged from 2.81% to 31.63%. Among the 33 QTLs, 19 were major QTLs, explaining > 10% of total phenotypic variation and 14 were small-effect QTLs, explaining <10% of total phenotypic variation.

A total of 17 QTLs were identified for fiber length in the Acala Maxxa background, explaining 2.81 % to 19.86% of the total phenotypic variation (Table 6.3). Of the 17 QTLs, 10 were in the At subgenome and seven were in the Dt subgenome. Ten of the 18 QTLs were small effect QTLs while the remaining seven were major effect QTLs. Except for four QTLs, one each on chromosomes 1 and 4, and two on chromosome 17, all others reduced lint percentage, consistent with the donor parent phenotype. The QTL on chromosome 17 which increased lint percentage was identified in two environments, IHF-2019 and IHF-2021.

In the Pima S6 background, a total of 15 QTLs were identified for lint percentage, explaining 4.16% to 31.63% of total phenotypic variation (Table 6.4). Four of the 15 QTLs were small effect QTLs while the remaining 11 were major QTLs, explaining more than 10% of total phenotypic variance. Among the 15 QTLs, 12 were identified in the At subgenome and only three were identified in the Dt subgenome. All except three QTLs increased lint percentage, consistent with the phenotype of the donor parent.

In total, of the 17 QTLs identified in GH background, 10 were identified in the At subgenome and seven in the Dt subgenome. In the reciprocal background, 12 of the 15 QTLs were identified in the At subgenome and only three were identified in the Dt subgenome. Chromosome

1 carried the most QTLs (3) in the GH background while chromosomes 2, 3 and 21 each carried the most QTLs (two each) in the GB background.

Discussion

Interspecific introgression has been a successful method of introducing favorable genetic variation for broadening the genetic base Upland cotton. Introgressive breeding approaches not only introduce a preponderance of novel allelic variation into cultivated gene pools, but interspecific populations developed using these approaches can be widely used for molecular dissection of complex fiber yield and quality parameters. Much of the study done in the past have focused on investigating the effects of GB chromatin segments introgressed into GH genome, however, the reverse has not been routinely studied. In the current study, we developed a reciprocal set of advanced backcross lines and NILs selected from among the selfed progenies of these advanced backcross lines and tested these populations to assess the effects of reciprocal chromatin transfer on lint percentage in cotton and identified a total of 32 marker-trait associations in the two reciprocal backgrounds. This study adds to the resources and observations available to study the quantitative nature of fiber quality traits reciprocally in two elite cotton backgrounds.

Performance of reciprocal populations

ABLs and NILs in both backgrounds showed average phenotypes consistent with their recurrent parent (Figure 6.1). Albeit the average performance of these two populations behaved like their recurrent parents, which is expected given the exceptionally large proportion of their genome coming from the recurrent parent, a discretely large amount of variation was observable for all three traits. Presence of transgressive segregants on both directions for all three traits

suggests that the chromatin segments introgressed from the donor parents has effects that could significantly alter the performance of individual lines. In fact, these alterations and their effects is shown by the phenotypic performance (QTL effects) of the lines carrying respective introgression from the donor parent.

Effect of species background

Similar total number of QTLs were identified in the two reciprocal backgrounds, suggesting the involvement of similar number of genes controlling lint percentage in respective backgrounds. However, differences in locations and effects of these QTLs indicate that species background affects the way chromatin introgressed from the donor parent influences certain traits. Reciprocity of QTLs between the two backgrounds is the first thing that comes to mind when we talk about reciprocal mapping populations.

Reciprocal QTLs were identified, albeit few, with effects both antagonistic as well as in the same direction, suggesting how different regions in the genome handle traits very differently (Tables 6.3 and 6.4). A QTL in chromosome 2 decreased LP in the GH background, while reciprocal QTL in the same location increased LP in the GB background. The antagonistic effects shown by the reciprocal QTLs align with parental phenotypes and indicates this genomic location in chromosome 2 is a potential candidate for identifying genes polymorphic for LP as well as lint yield. However, another set of reciprocal QTLs on chromosome 4 show a different shade of QTL effects. These QTLs increased LP in both backgrounds, suggesting that reciprocal donor chromatin do not always have antagonistic effects on the recipient genome. This is a good example to show that sometimes alleles coming from genotypes notorious for some traits might actually leverage these traits in a recipient background, possibly due to interaction between these donor alleles and

alleles present in the recipient genome, thus complimenting each other for more favorable phenotypic outputs.

Apart from these two sets of reciprocal QTLs, no other sets of reciprocal QTLs were identified in the study. On chromosome 3, QTLs were identified in both backgrounds but the genomic location of the most significant marker for these QTLs do not align and given the genetic architecture of our lines with small chunks of donor chromatin, no correspondence (p-value > 0.05 using the hypergeometric probability function – see methods), was found for these sets of chromosomes. Such limited correspondence of identified QTLs in the two backgrounds could be a result of several factors. One factor may be the small phenotypic effects of most of the identified QTL, increasing the likelihood that one or both members of a reciprocal pair elude detection (Broman, 2001). Another intriguing factor that could account for some failures to identify correspondence of QTLs in the reciprocal backgrounds, especially in advanced backcross lines with multiple introgressed chromosomal segments, is epistasis. Interaction between introgressed loci might result in underestimation of their effects which might have resulted in some QTLs failing to reach the biometric thresholds required to declare them as QTLs per se.

Similarity with QTLs previously reported

Since the first report of DNA maker-based genetic mapping in cotton (Reinisch et al., 1994a), several interspecific genetic maps have been constructed and used in mapping of loci controlling various ago-morphological traits. Several thousands of QTLs have been reported for fiber quality traits (J. Rong et al., 2007; J. I. Said et al., 2013; J. I. Said, M. Song, et al., 2015) and over 400 QTLs have been already reported for LP spanning all 26 chromosomes in the cotton genome (Niu et al., 2022). The hypergeometric probability distribution function (Feltus et al.,

2006) provides a means to infer statistically whether QTLs for a trait are randomly distributed between two populations. We performed an elaborate analysis of correspondence of QTLs identified in our study with those previously reported by using the hypergeometric probability distribution function (see methods for details).

A total of 565 LP QTLs reported in 17 previous studies were used to look at non-random correspondence of QTLs identified in our study (Table 6.5). Several QTLs were identified in genomic regions close to the ones identified in our study, however the P values based on the hypergeometric distribution show that QTLs identified in close proximity to the ones we reported corresponded non-randomly in only 3 of 17 previous studies in both GH and GB backgrounds. These results suggest that across the genome, this correspondence is not sufficient to infer non-random distribution. Nevertheless, those QTLs that correspond between studies with statistical significance for non-random distribution serve as a potential genomic region for further exploration of candidate genes for LP.

In silico annotation for candidate genes

Availability of cotton reference genomes (Paterson et al., 2012; Zhang et al., 2015) enabled us to scrutinize physical regions surrounding the identified QTLs for genes / gene families known or suspected to affect fiber quality parameters in cotton. This investigation was limited to tightly linked regions i.e., 1 Mb on both sides of the SNP marker that is most significantly associated with the QTLs. In chromosome 2, the nearest gene of interest is Inositol-3-phosphate synthase (*INO1*), which catalyzes the majority of myo-inositol synthesis required for plant growth and development. In cotton this gene has been shown to be expressed highly during and being involved in fiber

elongation during initial growth stages by enhancing ethylene biosynthesis and thus improving overall fiber biomass (Dou et al., 2022).

Another gene closest to a LP QTL on chromosome 3 is a sensor for the *INO1* gene on chromosome 2. This gene *PHO1* is a phosphate transporter associated with transcription factor regulating inorganic phosphate in cell walls (Wang et al., 2008). On chromosome 17, the gene closest to a LP QTL is the *COBRA* like protein 4, which is anchored component of cell membrane and is responsible for cell growth and cellulose microfibril organization. The members of this gene family have been proven to be key regulators in the orientation of cell expansion and cellulose crystallinity and in cotton, they have been shown to have significantly positive correlation with fiber quality traits, indicting their important roles in fiber development (Niu et al., 2015), While it is premature to suggest the candidacy of these genes, improved genomic resources together with these new populations are expected to accelerate candidate gene identification and validation for many loci.

Conclusion

The present study demonstrates the value of GH as a source of favorable alleles for fiber quality traits in the GB background while also reiterating the reciprocal transfer of favorable alleles from GB. Using advanced backcross lines segregating for a few donor chromatins as well as a set of reciprocal near-isogenic lines, we showed a strong effect of genetic background on chromatin transfer as well as the effect of these introgressed chromatin on three important fiber length. The near-isogenic genomic composition of our population provided opportunities to estimate the effects of genomic regions more precisely, but at the cost of epistatic QTL interactions. With one of the major purposes of NILs being the ability to verify the location and effects of QTLs, in

addition to their identification, we were able to demonstrate the stability of a few important fiber quality QTLs identified in previous studies. Since the parents used in creating these populations are both elite lines representing the two major domesticated species of cotton, with their own specialties and differences, the populations we developed are not just a platform to identify genomic locations underpinning fiber quality traits, but also a good starting point from which study divergence, domestication and evolutionary history of cotton as well as to select superior individual lines from a pool of these immortal lines for breeding purposes as well as for commercialization.

Reference

Abdurakhmonov, I. Y. (2012). Genetic Diversity in Gossypium genus. In: IntechOpen.

Adhikari, J. (2015a). *Effects of exotic genotypes and genetic backgrounds on fiber quality and plant architectural traits in Upland cotton (Gossypium hirsutum L.)* University of Georgia].

Adhikari, J., Das, S., Wang, Z., Khanal, S., Chandnani, R., Patel, J. D., Goff, V. H., Auckland, S., Rainville, L. K., & Jones, D. (2017). Targeted identification of association between cotton fiber quality traits and microsatellite markers. *213*(3), 1-15. https://search.ebscohost.com/login.aspx?direct=true&AuthType=ip,shib&db=edsbl&AN=RN383166005&site=eds-live&custid=uga1

Al-Ghazi, Y., Bourot, S., Arioli, T., Dennis, E. S., & Llewellyn, D. J. (2009). Transcript profiling during fiber development identifies pathways in secondary metabolism and cell wall structure that may contribute to cotton fiber quality. *Plant Cell Physiol*, *50*(7), 1364-1381. https://doi.org/10.1093/pcp/pcp084

Anderson, E. (1949). *Introgressive hybridization* [BibliographiesNon-fiction]. J. Wiley. https://search.ebscohost.com/login.aspx?direct=true&AuthType=ip,shib&db=cat06564a&AN=uga.993366963902959&site=eds-live&custid=uga1

Andolfatto, P., Davison, D., Erezyilmaz, D., Hu, T. T., Mast, J., Sunayama-Morita, T., & Stern, D. L. (2011). Multiplexed shotgun genotyping for rapid and efficient genetic mapping. *Genome Research*, *21*(4), 610-617. https://doi.org/10.1101/gr.115402.110

Baack, E., Melo, M. C., Rieseberg, L. H., & Ortiz-Barrientos, D. (2015). The origins of reproductive isolation in plants [https://doi.org/10.1111/nph.13424]. *New Phytologist*, *207*(4), 968-984. https://doi.org/https://doi.org/10.1111/nph.13424

Basra, A. S., & Malik, C. P. (1984). Development of the Cotton Fiber. In J. F. D. G.H. Bourne & K. W. Jeon (Eds.), *International Review of Cytology* (Vol. Volume 89, pp. 65-113). Academic Press. https://doi.org/http://dx.doi.org/10.1016/S0074-7696(08)61300-5

Beasley, J. O. (1942). Meiotic chromosome behaviour in species, species hybrids, haploids, and induced polyploids of Gossypium [Journal Article]. *Genetics*, *27*, 25-54. https://search.ebscohost.com/login.aspx?direct=true&AuthType=ip,shib&db=ldh&AN=19431600225&site=eds-live&custid=uga1

Beavis, W. (1997). QTL Analyses. In *Molecular Dissection of Complex Traits* (pp. 146-162). CRC Press. https://doi.org/doi:10.1201/9781420049381.ch1010.1201/9781420049381.ch10

Bradow, J. M., & Davidonis, G. H. (2000). Quantitation of fiber quality and the cotton production-processing interface: a physiologist's perspective. *Journal of Cotton Science*, *4*(1), 34-64. http://proxy-remote.galib.uga.edu/login?url=http://search.ebscohost.com/login.aspx?direct=true&db=lah&AN=20000707210&site=eds-live

Broman, K. W. (2001). Resource Review of Statistical Methods for QTL Mapping in Experimental Crosses. *30*(7), 44-52. https://search.ebscohost.com/login.aspx?direct=true&AuthType=ip,shib&db=edsbl&AN=RN098883447&site=eds-live&custid=uga1

Broman, K. W., & Sen, S. (2009). *A guide to QTL mapping with R/qtl*

Brown, N., Kumar, P., Khanal, S., Singh, R., Suassuna, N. D., McBlanchett, J., Lubbers, E., Jones, D., Paterson, A. H., & Chee, P. W. (2020). Registration of eight upland cotton (Gossypium hirsutum L.) germplasm lines with qFL-Chr.25, a fiber-length QTL introgressed from

Gossypium barbadense [https://doi.org/10.1002/plr2.20009]. *Journal of Plant Registrations*, *14*(1), 57-63. https://doi.org/https://doi.org/10.1002/plr2.20009

Brown, N., Kumar, P., Singh, R., Lubbers, E., Campbell, B. T., Myers, G. O., Wright, R. J., Subramani, J., Jones, D., Paterson, A. H., & Chee, P. W. (2019). Evaluation of a Chromosome Segment from Gossypium barbadense Harboring the Fiber Length QTL qFL-Chr.25 in Four Diverse Upland Cotton Genetic Backgrounds [https://doi.org/10.2135/cropsci2019.05.0321]. *Crop Science*, *59*(6), 2621-2633. https://doi.org/https://doi.org/10.2135/cropsci2019.05.0321

Brubaker, C. L., Paterson, A. H., & Wendel, J. F. (1999). Comparative genetic mapping of allotetraploid cotton and its diploid progenitors. *42*(2), 184-203. https://search.ebscohost.com/login.aspx?direct=true&AuthType=ip,shib&db=edsbl&AN=RN061567240&site=eds-live&custid=uga1

Chandnani, R., Kim, C., Guo, H., Shehzad, T., Wallace, J. G., He, D., Zhang, Z., Patel, J. D., Adhikari, J., Khanal, S., & Paterson, A. H. (2018). Genetic Analysis of Gossypium Fiber Quality Traits in Reciprocal Advanced Backcross Populations [https://doi.org/10.3835/plantgenome2017.06.0057]. *The Plant Genome*, *11*(1), 170057. https://doi.org/https://doi.org/10.3835/plantgenome2017.06.0057

Chandnani, R., Wang, B., Draye, X., Rainville, L. K., Auckland, S., Zhuang, Z., & Lubbers, E. L. (2017). Segregation distortion and genome-wide digenic interactions affect transmission of introgressed chromatin from wild cotton species [Report]. *Theoretical and Applied Genetics*, *130*(10), 2219. https://doi.org/10.1007/s00122-017-2952-y

Chee, P., Bredhauer, R., Smith, C. W., Paterson, A. H., Delmonte, T. A., Draye, X., Jiang, C. X., & Decanini, L. (2005). Molecular dissection of interspecific variation between Gossypium

hirsutum and Gossypium barbadense (cotton) by a backcross-self approach. I. Fiber elongation. *Theoretical and Applied Genetics*, *111*(4), 757-763. http://proxy-remote.galib.uga.edu/login?url=http://search.ebscohost.com/login.aspx?direct=true&db=agr&AN=IND43787564&site=eds-live

Chee, P., Draye, X., Jiang, C.-X., Decanini, L., Delmonte, T. A., Bredhauer, R., Smith, C. W., & Paterson, A. H. (2005). Molecular dissection of interspecific variation between Gossypium hirsutum and Gossypium barbadense (cotton) by a backcross-self approach: I. Fiber elongation. *Theoretical and Applied Genetics*, *111*(4), 757-763. https://doi.org/10.1007/s00122-005-2063-z

Chee, P. W., & Campbell, B. T. (2009). Bridging Classical and Molecular Genetics of Cotton Fiber Quality and Development. In A. H. Paterson (Ed.), *Genetics and Genomics of Cotton* (pp. 283-311). Springer US. https://doi.org/10.1007/978-0-387-70810-2_12

Chee, P. W., & Campbell, B. T. (2009). Bridging Classical and Molecular Genetics of Cotton Fiber Quality and Development. *3*, 283-312.

Chee, P. W., Draye, X., Jiang, C. X., Decanini, L., Delmonte, T. A., Bredhauer, R., Smith, C. W., & Paterson, A. H. (2005). Molecular dissection of phenotypic variation between Gossypium hirsutum and Gossypium barbadense (cotton) by a backcross-self approach: III. Fiber length [Journal Article]. *TAG Theoretical and Applied Genetics*, *111*(4), 772-781. https://doi.org/10.1007/s00122-005-2062-0

Cheng, H.-Q., Han, L.-B., Yang, C.-L., Wu, X.-M., Zhong, N.-Q., Wu, J.-H., Wang, F.-X., Wang, H.-Y., & Xia, G.-X. (2016). The cotton MYB108 forms a positive feedback regulation loop with CML11 and participates in the defense response against Verticillium dahliae infection. *Journal of Experimental Botany*, *67*(6), 1935-1950.

Desai, A., Chee, P. W., Junkang, R., May, O. L., Paterson, A. H., & Gustafson, J. P. (2006). Chromosome structural changes in diploid and tetraploid A genomes of Gossypium [Article]. *Genome*, *49*(4), 336-345. https://doi.org/10.1139/G05-116

Dou, L., Li, Z., Wang, H., Li, H., Xiao, G., & Zhang, X. (2022). The hexokinase Gene Family in Cotton: Genome-Wide Characterization and Bioinformatics Analysis [Original Research]. *Frontiers in Plant Science*, *13*. https://doi.org/10.3389/fpls.2022.882587

Draye, X., Chee, P., Jiang, C.-X., Decanini, L., Delmonte, T. A., Bredhauer, R., Smith, C. W., & Paterson, A. H. (2005). Molecular dissection of interspecific variation between Gossypium hirsutum and G. barbadense (cotton) by a backcross-self approach: II. Fiber fineness [Article]. *Theoretical & Applied Genetics*, *111*(4), 764-771.

Draye, X., Chee, P., Jiang, C. X., Decanini, L., Delmonte, T. A., Bredhauer, R., Smith, C. W., & Paterson, A. H. (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*, *111*(4), 764-771. http://proxy-remote.galib.uga.edu/login?url=http://search.ebscohost.com/login.aspx?direct=true&db=lah&AN=20053182677&site=eds-live

El-Sharkawy, I., Sherif, S., Mila, I., Bouzayen, M., & Jayasankar, S. (2009). Molecular characterization of seven genes encoding ethylene-responsive transcriptional factors during plum fruit development and ripening. *Journal of Experimental Botany*, *60*(3), 907-922. https://doi.org/10.1093/jxb/ern354

Endrizzi, J. E., Turcotte, E. L., & Kohel, R. J. (1985). Genetics, Cytology, and Evolution of Gossypium. In E. W. Caspari & J. G. Scandalios (Eds.), *Advances in Genetics* (Vol. 23, pp. 271-375). Academic Press.

Eshed, Y., & Zamir, D. (1994). A genomic library of Lycopersicon pennellii in L. esculentum: a tool for fine mapping of genes. *Euphytica*, *79*(3), 175-179. http://proxy-remote.galib.uga.edu/login?url=http://search.ebscohost.com/login.aspx?direct=true&db=lah&AN=19951604527&site=eds-live

Eshed, Y., & Zamir, D. (1995). An introgression line population of Lycopersicon pennellii in the cultivated tomato enables the identification and fine mapping of yield-associated QTL. *Genetics*, *141*.

Fang, D. (2015). Molecular Breeding. In. https://doi.org/10.2134/agronmonogr57.2013.0027

Fang, D. D., Jenkins, J. N., Deng, D. D., McCarty, J. C., Li, P., & Wu, J. (2014). Quantitative trait loci analysis of fiber quality traits using a random-mated recombinant inbred population in Upland cotton (Gossypium hirsutum L.). *BMC Genomics*, *15*(1), 397. https://doi.org/10.1186/1471-2164-15-397

Fang, L., Wang, Q., Hu, Y., Jia, Y., Chen, J., Liu, B., Zhang, Z., Guan, X., Chen, S., Zhou, B., Mei, G., Sun, J., Pan, Z., He, S., Xiao, S., Shi, W., Gong, W., Liu, J., Ma, J., . . . Zhang, T. (2017). Genomic analyses in cotton identify signatures of selection and loci associated with fiber quality and yield traits. *Nature Genetics*, *49*(7), 1089-1098.

Fankhauser, C., & Chory, J. (1997). Light control of plant development. *Annual review of cell and developmental biology*, *13*, 203-229. https://doi.org/10.1146/annurev.cellbio.13.1.203

Feltus, F. A., Hart, G. E., Schertz, K. F., Casa, A. M., Kresovich, S., Abraham, S., Klein, P. E., Brown, P. J., & Paterson, A. H. (2006). Alignment of genetic maps and QTLs between inter- and intra-specific sorghum populations. *Theor Appl Genet*, *112*(7), 1295-1305. https://doi.org/10.1007/s00122-006-0232-3

Feng, J. X., Ji, S. J., Shi, Y. H., Xu, Y., Wei, G., & Zhu, Y. X. (2004). Analysis of five differentially expressed gene families in fast elongating cotton fiber. *Acta Biochim Biophys Sin (Shanghai)*, *36*(1), 51-56. https://doi.org/10.1093/abbs/36.1.51

Flagel, L. E., & Wendel, J. F. (2010). Evolutionary rate variation, genomic dominance and duplicate gene expression evolution during allotetraploid cotton speciation [research-article]. *The New Phytologist*, *186*(1), 184-193.

Fletcher, R. S., Mullen, J. L., Yoder, S., Bauerle, W. L., Reuning, G., Sen, S., Meyer, E., Juenger, T. E., & McKay, J. K. (2013). Development of a next-generation NIL library in Arabidopsis thaliana for dissecting complex traits. *BMC Genomics*, *14*(1), 655. https://doi.org/10.1186/1471-2164-14-655

Fryxell, P. A., & Craven, L. A. (1992). A Revision of Gossypium Sect. Grandicalyx (Malvaceae), Including the Description of Six New Species [research-article]. *Systematic Botany*(1), 91. https://doi.org/10.2307/2419068

Fu, Y., Dong, C., Wang, J., Wang, Y., & Li, C. (2019). Genome-wide association study reveals the genetic control underlying node of the first fruiting branch and its height in upland cotton (Gossypium hirsutum L.). *Euphytica*, *215*(2), 35.

Ghosal, S., Zheng, B., Chapman, S. C., Potgieter, A. B., Jordan, D. R., Wang, X., Singh, A. K., Singh, A., Hirafuji, M., Ninomiya, S., Ganapathysubramanian, B., Sarkar, S., & Guo, W. (2019). A Weakly Supervised Deep Learning Framework for Sorghum Head Detection and Counting. *Plant Phenomics*, *2019*, 1525874. https://doi.org/10.34133/2019/1525874

Glaubitz, J. C., Casstevens, T. M., Lu, F., Harriman, J., Elshire, R. J., Sun, Q., & Buckler, E. S. (2014). TASSEL-GBS: A High Capacity Genotyping by Sequencing Analysis Pipeline [Article]. *PLoS ONE*, *9*(2), 1-11. https://doi.org/10.1371/journal.pone.0090346

Grant, V. (1981). *Plant speciation* (2nd ed.)

Grover, C. E., Gallagher, J. P., & Wendel, J. F. (2015). Candidate Gene Identification of Flowering Time Genes in Cotton [https://doi.org/10.3835/plantgenome2014.12.0098]. *The Plant Genome*, *8*(2), plantgenome2014.2012.0098. https://doi.org/https://doi.org/10.3835/plantgenome2014.12.0098

Grover, C. E., Yoo, M.-J., Lin, M., Murphy, M. D., Harker, D. B., Byers, R. L., Lipka, A. E., Hu, G., Yuan, D., Conover, J. L., Udall, J. A., Paterson, A. H., Gore, M. A., & Wendel, J. F. (2020). Genetic Analysis of the Transition from Wild to Domesticated Cotton (Gossypium hirsutum L.). *G3 Genes|Genomes|Genetics*, *10*(2), 731-754. https://doi.org/10.1534/g3.119.400909

Grover, E. B., & Hamby, D. S. (1960). *Handbook of textile testing and quality control*

Guan, X., Song, Q., & Chen, Z. J. (2014). Polyploidy and small RNA regulation of cotton fiber development. *Trends in Plant Science*, *19*(8), 516-528. https://doi.org/10.1016/j.tplants.2014.04.007

Guo, Y., McCarty, J. C., Jenkins, J. N., An, C., & Saha, S. (2009). Genetic detection of node of first fruiting branch in crosses of a cultivar with two exotic accessions of upland cotton. *Euphytica*, *166*(3), 317-329. https://doi.org/10.1007/s10681-008-9809-z

Guo, Y. F., McCarty, J. C., Jenkins, J. N., & Saha, S. (2008). QTLs for node of first fruiting branch in a cross of an upland cotton, Gossypium hirsutum L., cultivar with primitive accession Texas 701 [Journal Article]. *Euphytica*, *163*(1), 113-122. https://doi.org/10.1007/s10681-007-9613-1

He, X., Wang, T., Xu, Z., Liu, N., Wang, L., Hu, Q., Luo, X., Zhang, X., & Zhu, L. (2018). The cotton HD-Zip transcription factor GhHB12 regulates flowering time and plant architecture

via the GhmiR157-GhSPL pathway. *Communications Biology*, *1*(1), 229. https://doi.org/10.1038/s42003-018-0234-0

Heiser, C. B. (1979). Hybrid Populations of Helianthus divaricatus and H. microcephalus after 22 Years [research-article]. *Taxon*, *28*(1/3), 71-75. https://doi.org/10.2307/1219560

Hinchliffe, D. J., Meredith, W. R., Yeater, K. M., Kim, H. J., Woodward, A. W., Chen, Z. J., & Triplett, B. A. (2010). Near-isogenic cotton germplasm lines that differ in fiber-bundle strength have temporal differences in fiber gene expression patterns as revealed by comparative high-throughput profiling. *TAG. Theoretical And Applied Genetics. Theoretische Und Angewandte Genetik*, *120*(7), 1347-1366.

Hodgkin, T., & Hajjar, R. (2007). Using crop wild relatives for crop improvement: trends and perspectives. Wallingford, UK.

Holland, J. B. (2004, September 26 - October 1, 2004). Implementation of molecular markers for quanititative traits in a breeding program - challenges and opportunities. New directions for a diverse plannet: Proceedings for the 4th International Crop Scinece Congress, Brisbane, Australia.

Hu, Y., Chen, J., Fang, L., Zhang, Z., Ma, W., Niu, Y., Ju, L., Deng, J., Zhao, T., Lian, J., Baruch, K., Fang, D., Liu, X., Ruan, Y.-l., Rahman, M.-u., Han, J., Wang, K., Wang, Q., Wu, H., . . . Zhang, T. (2019). Gossypium barbadense and Gossypium hirsutum genomes provide insights into the origin and evolution of allotetraploid cotton. *Nature Genetics*, *51*(4), 739-748. https://doi.org/10.1038/s41588-019-0371-5

Huijser, P., & Schmid, M. (2011). The control of developmental phase transitions in plants. *138*(19), 4117-4130.

Hutchinson, J. B. (1951). Intra-specific differentiation in Gossypium hirsutum. *Heredity*, *5*(2), 161-193. https://doi.org/10.1038/hdy.1951.19

Imran, M., Shakeel, A., Azhar, F. M., Farooq, J., Saleem, M. F., Saeed, A., Nazeer, W., Riaz, M., Naeem, M., & Javaid, A. (2012). Combining ability analysis for within-boll yield components in upland cotton (Gossypium hirsutum L.). *Genet Mol Res*, *11*(3), 2790-2800. https://doi.org/10.4238/2012.August.24.4

James Mc, D. S. (1975). Fiber Initiation on the Cotton Ovule (Gossypium Hirsutum). *American Journal of Botany*, *62*(7), 723-730. http://www.jstor.org/stable/2442061

Jasdanwala, R. T., Singh, Y. D., & Chinoy, J. J. (1977). Auxin Metabolism in Developing Cotton Hairs. (106), 1111.

Jia, X., Pang, C., Wei, H., Wang, H., Ma, Q., Yang, J., Cheng, S., Su, J., Fan, S., Song, M., Wusiman, N., & Yu, S. (2016). High-density linkage map construction and QTL analysis for earliness-related traits in Gossypium hirsutum L. *BMC Genomics*, *17*(1), 909. https://doi.org/10.1186/s12864-016-3269-y

Jiang, C., Wright, R. J., El-Zik, K. M., & Paterson, A. H. (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), 4419-4424.

Jiang, C.-x., Chee, P. W., Draye, X., Morrell, P. L., Smith, C. W., & Paterson, A. H. (2000). MULTILOCUS INTERACTIONS RESTRICT GENE INTROGRESSION IN INTERSPECIFIC POPULATIONS OF POLYPLOID GOSSYPIUM (COTTON) [https://doi.org/10.1111/j.0014-3820.2000.tb00081.x]. *Evolution*, *54*(3), 798-814. https://doi.org/https://doi.org/10.1111/j.0014-3820.2000.tb00081.x

Jiang, Y., Li, C., & Paterson, A. H. (2016). High throughput phenotyping of cotton plant height using depth images under field conditions. *Computers and Electronics in Agriculture, 130*, 57-68. https://doi.org/https://doi.org/10.1016/j.compag.2016.09.017

Jiang, Y., Li, C., Sun, S., Paterson, A. H., & Xu, R. (2017, 2017/01/01/). GPHENOVISION: A GROUND MOBILE SYSTEM WITH MULTI-MODAL IMAGING FOR FIELD-BASED HIGH THROUGHPUT PHENOTYPING OF COTTON.

Jiang, Y., Snider, J. L., Li, C., Rains, G. C., & Paterson, A. H. (2020). Ground Based Hyperspectral Imaging to Characterize Canopy-Level Photosynthetic Activities [Article]. *Remote Sensing*, *12*(2), 315-315. https://doi.org/10.3390/rs12020315

Kaeppler, S. M. (1997). Quantitative trait locus mapping using sets of near-isogenic lines: relative power comparisons and technical considerations. *Theoretical and Applied Genetics*, *95*(3), 384-392. https://doi.org/10.1007/s001220050574

Kearsey, M. J., & Luo, Z. W. (2003). Mapping, characterization and deployment of quantitative trait loci. In H. J. Newbury (Ed.), *Plant Molecular Breeding* (pp. 1-29). Blackwell Publishing Ltd.

Kelly, S., Huihui, L., Navarro, J. A. R., Dong, A., Maria Cinta, R., Sarah, H., Charlotte, A., Jeffrey, C. G., Sharon, M., Robert, J. E., Edward, S. B., & Peter, J. B. (2014). Novel Methods to Optimize Genotypic Imputation for Low-Coverage, Next-Generation Sequence Data in Crop Plants [article]. *The Plant Genome*, *7*(3).

Keurentjes, J. J. B., Bentsink, L., Alonso-Blanco, C., Hanhart, C. J., Blankestijn-De Vries, H., Effgen, S., Vreugdenhil, D., & Koornneef, M. (2007). Development of a Near-Isogenic Line Population of Arabidopsis thaliana and Comparison of

Mapping Power With a Recombinant Inbred Line Population [10.1534/genetics.106.066423]. *Genetics*, *175*(2), 891.

Khanal, S. (2018). *Dissection of quantitative variation introgressed into bermudagrass and upland cotton* University of Georgia].

Kim, C., Guo, H., Kong, W., Chandnani, R., Shuang, L.-S., & Paterson, A. H. (2016, 2016/01/01/). Application of genotyping by sequencing technology to a variety of crop breeding programs. Netherlands.

Kim, H. J., & Triplett, B. A. (2001). Cotton Fiber Growth in Planta and in Vitro. Models for Plant Cell Elongation and Cell Wall Biogenesis. *127*(4), 1361-1366. https://search.ebscohost.com/login.aspx?direct=true&AuthType=ip,shib&db=edsbl&AN=RN106489208&site=eds-live&custid=uga1

Kim, H. J., & Triplett, B. A. (2001). Cotton Fiber Growth in Planta and in Vitro. Models for Plant Cell Elongation and Cell Wall Biogenesis. *Plant Physiology*, *127*(4), 1361-1366. https://doi.org/10.1104/pp.010724

Kohel, R. J., Richmond, T. R., & Lewis, C. F. (1974). Genetics of flowering response in cotton. VI. Flowering behavior of Gossypium hirsutum L. and G. barbadense L. hybrids [Journal Article]. *Crop Science*, *14*(5), 696-699.

Komeda, Y. (2004). GENETIC REGULATION OF TIME TO FLOWER IN ARABIDOPSIS THALIANA [Article]. *Annual Review of Plant Biology*, *55*(1), 521-535. https://doi.org/10.1146/annurev.arplant.55.031903.141644

Kooke, R., Wijnker, E., & Keurentjes, J. J. (2012). Backcross populations and near isogenic lines. *Methods Mol Biol*, *871*, 3-16. https://doi.org/10.1007/978-1-61779-785-9_1

Kooke, R., Wijnker, E., & Keurentjes, J. J. B. (2012). Backcross Populations and Near Isogenic Lines. In S. A. Rifkin (Ed.), *Quantitative Trait Loci (QTL): Methods and Protocols* (pp. 3-16). Humana Press. https://doi.org/10.1007/978-1-61779-785-9_1

Kumar, P., Singh, R., Lubbers, E. L., Shen, X., Paterson, A. H., Campbell, B. T., Jones, D. C., & Chee, P. W. (2019). Genetic Evaluation of Exotic Chromatins from Two Obsolete Interspecific Introgression Lines of Upland Cotton for Fiber Quality Improvement [https://doi.org/10.2135/cropsci2018.12.0745]. *Crop Science*, *59*(3), 1073-1084. https://doi.org/https://doi.org/10.2135/cropsci2018.12.0745

Kushanov, F. N., Buriev, Z. T., Shermatov, S. E., Turaev, O. S., Norov, T. M., Pepper, A. E., Saha, S., Ulloa, M., Yu, J. Z., Jenkins, J. N., Abdukarimov, A., & Abdurakhmonov, I. Y. (2017). QTL mapping for flowering-time and photoperiod insensitivity of cotton Gossypium darwinii Watt [Journal Article]. *PLoS ONE*, *12*(10), e0186240.

Kushanov, F. N., Turaev, O. S., Ernazarova, D. K., Gapparov, B. M., Oripova, B. B., Kudratova, M. K., Rafieva, F. U., Khalikov, K. K., Erjigitov, D. S., Khidirov, M. T., Kholova, M. D., Khusenov, N. N., Amanboyeva, R. S., Saha, S., Yu, J. Z., & Abdurakhmonov, I. Y. (2021). Genetic Diversity, QTL Mapping, and Marker-Assisted Selection Technology in Cotton (Gossypium spp.) [Review]. *Frontiers in Plant Science*, *12*.

Lacape, J. M., Trung Bieu, N., Courtois, B., Belot, J. L., Giband, M., Gourlot, J. P., Gawryziak, G., Roques, S., & Hau, B. (2005). QTL analysis of cotton fiber quality using multiple Gossypium hirsutum×Gossypium barbadense backcross generations. *Crop Science*, *45*(1), 123-140.

Lai, D., Li, H., Fan, S., Song, M., Pang, C., Wei, H., Liu, J., Wu, D., Gong, W., & Yu, S. (2011). Generation of ESTs for Flowering Gene Discovery and SSR Marker Development in Upland Cotton. *PLoS ONE*, *6*(12), e28676. https://doi.org/10.1371/journal.pone.0028676

Landi, P., Sanguineti, M. C., Salvi, S., Giuliani, S., Bellotti, M., Maccaferri, M., Conti, S., & Tuberosa, R. (2005). Validation and characterization of a major QTL affecting leaf ABA concentration in maize. *Molecular Breeding*, *15*(3), 291-303.

Leal Valentim, F., Mourik, S. v., Posé, D., Kim, M. C., Schmid, M., van Ham, R. C. H. J., Busscher, M., Sanchez-Perez, G. F., Molenaar, J., Angenent, G. C., Immink, R. G. H., & van Dijk, A. D. J. (2015). A Quantitative and Dynamic Model of the Arabidopsis Flowering Time Gene Regulatory Network [Article]. *PLoS ONE*, *10*(2), 1-18.

Levi, A., Ovnat, L., Paterson, A. H., & Saranga, Y. (2009). Photosynthesis of cotton near-isogenic lines introgressed with QTLs for productivity and drought related traits [Article]. *Plant Science*, *177*(2), 88-96. https://doi.org/10.1016/j.plantsci.2009.04.001

Lewis, C. F., & Richmond, T. R. (1960). The genetics of flowering response in cotton. II. Inheritance of flowering response in a Gossypium barbadense cross [Journal Article]. *Genetics*, *45*, 79-85.

Li, C., Wang, C., Dong, N., Wang, X., Zhao, H., Converse, R., Xia, Z., Wang, R., & Wang, Q. (2012). QTL detection for node of first fruiting branch and its height in upland cotton (Gossypium hirsutum L.). *Euphytica*, *188*(3), 441-451. https://doi.org/10.1007/s10681-012-0720-2

Li, C., Wang, X., Dong, N., Zhao, H., Xia, Z., Wang, R., Converse, R. L., & Wang, Q. (2013). QTL analysis for early-maturing traits in cotton using two upland cotton (Gossypium hirsutum L.) crosses. *Breeding Science*, *63*(2), 154-163.

Li, C., Wang, X., Dong, N., Zhao, H., Xia, Z., Wang, R., Converse, R. L., & Wang, Q. (2013). QTL analysis for early-maturing traits in cotton using two upland cotton (Gossypium hirsutum L.) crosses. *63*(2), 154-163.

Li, C., Zhang, J., Hu, G., Fu, Y., & Wang, Q. (2016). Association mapping and favorable allele mining for node of first fruiting/sympodial branch and its height in Upland cotton (Gossypium hirsutum L.). *Euphytica*, *210*(1), 57-68. https://doi.org/10.1007/s10681-016-1697-z

Li, F., Fan, G., Lu, C., Xiao, G., Zou, C., Kohel, R. J., Ma, Z., Shang, H., Ma, X., Wu, J., Liang, X., Huang, G., Percy, R. G., Liu, K., Yang, W., Chen, W., Du, X., Shi, C., Yuan, Y., . . . Yu, S. (2015). Genome sequence of cultivated Upland cotton (Gossypium hirsutum TM-1) provides insights into genome evolution. *Nature biotechnology*, *33*(5), 524-530. https://doi.org/10.1038/nbt.3208

Li, L., Zhao, S., Su, J., Fan, S., Pang, C., Wei, H., Wang, H., Gu, L., Zhang, C., Liu, G., Yu, D., Liu, Q., Zhang, X., & Yu, S. (2017). High-density genetic linkage map construction by F2 populations and QTL analysis of early-maturity traits in upland cotton (Gossypium hirsutum L.). *PLoS ONE*, *12*(8), e0182918. https://doi.org/10.1371/journal.pone.0182918

Li, Y., Jiang, J., Du, M. L., Li, L., Wang, X. L., & Li, X. B. (2013). A cotton gene encoding MYB-like transcription factor is specifically expressed in pollen and is involved in regulation of late anther/pollen development. *Plant Cell Physiol*, *54*(6), 893-906.

Li, Y. L., Sun, J., & Xia, G. X. (2005). Cloning and characterization of a gene for an LRR receptor-like protein kinase associated with cotton fiber development. *Mol Genet Genomics*, *273*(3), 217-224. https://doi.org/10.1007/s00438-005-1115-z

Li, Z., Wang, P., You, C., Yu, J., Zhang, X., Yan, F., Ye, Z., Shen, C., Li, B., Guo, K., Liu, N., Thyssen, G. N., Fang, D. D., Lindsey, K., Zhang, X., Wang, M., & Tu, L. (2020). Combined GWAS and eQTL analysis uncovers a genetic regulatory network orchestrating the initiation of secondary cell wall development in cotton [Report]. *New Phytologist*, *226*(6), 1738.

Lin, Y., Schertz, K. F., & Paterson, A. H. (1995). Comparative analysis of QTLs affecting plant height and maturity across the Poaceae, in reference to an interspecific sorghum population [Journal Article]. *Genetics*, *141*(1), 391-411.

Lin, Z., He, D., Zhang, X., Nie, Y., Guo, X., Feng, C., & Stewart, J. M. (2005). Linkage map construction and mapping QTL for cotton fibre quality using SRAP, SSR and RAPD [Article]. *Plant Breeding*, *124*(2), 180-187.

Liu, D., Liu, F., Shan, X., Zhang, J., Tang, S., Fang, X., Liu, X., Wang, W., Tan, Z., Teng, Z., Zhang, Z., & Liu, D. (2015). Construction of a high-density genetic map and lint percentage and cottonseed nutrient trait QTL identification in upland cotton (Gossypium hirsutum L.) [Article]. *Molecular Genetics & Genomics*, *290*(5), 1683-1700.

Lubbers, E. L., & Chee, P. W. (2009). The Worldwide Gene Pool of G. hirsutum and its Improvement. *3*, 23-52.

Ma, Q., Wu, M., Pei, W., Li, H., Li, X., Zhang, J., Yu, J., & Yu, S. (2014). Quantitative phosphoproteomic profiling of fiber differentiation and initiation in a fiberless mutant of cotton. *BMC Genomics*, *15*(1), 466. https://doi.org/10.1186/1471-2164-15-466

May, O. L. (1999). Producing quality cotton by conventional breeding, marker assisted selection, and transgenic methods [Journal Article]. *Journal of New Seeds*, *1*(2), 65-82.

https://search.ebscohost.com/login.aspx?direct=true&AuthType=ip,shib&db=lah&AN=20001607149&site=eds-live&custid=uga1

May, O. L., & Green, C. C. (1994). Genetic variation for fiber properties in elite Pee Dee cotton populations [Journal Article]. *Crop Science*, *34*(3), 684-690. https://doi.org/10.2135/cropsci1994.0011183X003400030016x

Mei, M., Syed, N. H., Gao, W., Thaxton, P. M., Smith, C. W., Stelly, D. M., & Chen, Z. J. (2004). Genetic mapping and QTL analysis of fiber-related traits in cotton (Gossypium). *TAG. Theoretical And Applied Genetics. Theoretische Und Angewandte Genetik*, *108*(2), 280-291.

Meinert, M. C., & Delmer, D. P. (1977). Changes in biochemical composition of the cell wall of the cotton fiber during development. *Plant Physiology*, *59*(6), 1088-1097.

Melchinger, A. E., Utz, H. F., & Schon, C. C. (2000). From Mendel to Fisher: The power and limits of QTL mapping for quantitative traits. . *Vortr. Pflanzenzuchtg.*, *48*, 132-142.

Melchinger, A. E., Utz, H. F., & Schön, C. C. (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*(1), 383.

Meyn, O., & Emboden, W. A. (1987). Parameters and Consequences of Introgression in Salvia apiana x S. mellifera (Lamiaceae) [research-article]. *Systematic Botany*, *12*(3), 390-399.

Monforte, A. J., & Tanksley, S. D. (2000). Development of a set of near isogenic and backcross recombinant inbred lines containing most of the Lycopersicon hirsutum genome in a L. esculentum genetic background: A tool for gene mapping and gene discovery. *Genome*, *43*(5), 803-813. https://doi.org/10.1139/g00-043

Newbury, H. J. (2003). *Plant molecular breeding*. Oxford.

Ninh, T. T., Gao, W., Trusov, Y., Zhao, J.-R., Long, L., Song, C.-P., & Botella, J. R. (2021). Tomato and cotton G protein beta subunit mutants display constitutive autoimmune responses [https://doi.org/10.1002/pld3.359]. *Plant Direct*, *5*(11), e359.

Niu, E., Shang, X., Cheng, C., Bao, J., Zeng, Y., Cai, C., Du, X., & Guo, W. (2015). Comprehensive Analysis of the COBRA-Like (COBL) Gene Family in Gossypium Identifies Two COBLs Potentially Associated with Fiber Quality. *PLoS ONE*, *10*(12), e0145725. https://doi.org/10.1371/journal.pone.0145725

Niu, H., Ge, Q., Shang, H., & Yuan, Y. (2022). Inheritance, QTLs, and Candidate Genes of Lint Percentage in Upland Cotton. *Front Genet*, *13*, 855574.

Paran, I., & Zamir, D. (2003). Quantitative traits in plants: beyond the QTL. *Trends in Genetics*, *19*(6), 303-306. https://doi.org/https://doi.org/10.1016/S0168-9525(03)00117-3

Paterson, A. H., Boman, R. K., Brown, S. M., Chee, P. W., Gannaway, J. R., Gingle, A. R., May, O. L., & Smith, C. W. (2004). Reducing the Genetic Vulnerability of Cotton [https://doi.org/10.2135/cropsci2004.1900]. *Crop Science*, *44*(6), 1900-1901.

Paterson, A. H., Boman, R. K., Brown, S. M., Chee, P. W., Gannaway, J. R., Gingle, A. R., May, O. L., & Smith, C. W. (2004). Reducing the genetic vulnerability of cotton. Madison, USA.

Paterson, A. H., Brubaker, C. L., & Wendel, J. F. (1993). A rapid method for extraction of cotton (Gossypium spp.) genomic DNA suitable for RFLP or PCR analysis. *Plant Molecular Biology Reporter*, *11*(2), 122-127.

Paterson, A. H., Lin, Y. R., Li, Z., Schertz, K. F., Doebley, J. F., Pinson, S. R., Liu, S. C., Stansel, J. W., & Irvine, J. E. (1995). Convergent domestication of cereal crops by independent mutations at corresponding genetic Loci. *Science*, *269*(5231), 1714-1718.

Paterson, A. H., Saranga, Y., Menz, M., Jiang, C. X., & Wright, R. J. (2003). QTL analysis of genotype × environment interactions affecting cotton fiber quality [Article]. *Theoretical & Applied Genetics*, *106*(3), 384-396.

Paterson, A. H., Wendel, J. F., Gundlach, H., Guo, H., Jenkins, J., Jin, D., Llewellyn, D., Showmaker, K. C., Shu, S., Udall, J., Yoo, M.-j., Byers, R., Chen, W., Doron-Faigenboim, A., Duke, M. V., Gong, L., Grimwood, J., Grover, C., Grupp, K., . . . Schmutz, J. (2012). Repeated polyploidization of Gossypium genomes and the evolution of spinnable cotton fibres. *Nature*, *492*(7429), 423-427. https://doi.org/10.1038/nature11798

Perkins Jr., H. H., Ethridge, D. E., & Bragg, C. K. (1984). Fiber. In *Cotton* (pp. 437-509). https://doi.org/https://doi.org/10.2134/agronmonogr24.c12

Poland, J. A., & Rifeb, T. W. (2012). Genotyping-by-sequencing for plant breeding and genetics. *Plant Genome*, *5*(3), 92-102.

Qin, H., Guo, W., Zhang, Y. M., & Zhang, T. (2008). QTL mapping of yield and fiber traits based on a four-way cross population in Gossypium hirsutum L. *Theor Appl Genet*, *117*(6), 883-894. https://doi.org/10.1007/s00122-008-0828-x

Ray, L. L., & Richmond, T. R. (1966). Morphological measures of earliness of crop maturity in cotton [Journal Article]. *Crop Science*, *6*, 527-531.

Reinisch, A. J., Dong, J. M., Brubaker, C. L., Stelly, D. M., Wendel, J. F., & Paterson, A. H. (1994a). A detailed RFLP map of cotton, Gossypium hirsutum x Gossypium barbadense: chromosome organization and evolution in a disomic polyploid genome. *Genetics*, *138*(3), 829-847. https://doi.org/10.1093/genetics/138.3.829

Reinisch, A. J., Dong, J. M., Brubaker, C. L., Stelly, D. M., Wendel, J. F., & Paterson, A. H. (1994b). A detailed RFLP map of cotton, Gossypium hirsutum×Gossypium barbadense:

chromosome organization and evolution in a disomic polyploid genome. *Genetics*, *138*(3), 829-847.

Rieseberg, L. H., Kim, M. J., & Seiler, G. J. (1999). Introgression between the Cultivated Sunflower and a Sympatric Wild Relative, Helianthus petiolaris (Asteraceae). *International Journal of Plant Sciences*, *160*(1), 102-108. https://doi.org/10.1086/314107

Rong, J., Abbey, C., Bowers, J. E., Brubaker, C. L., Chang, C., Chee, P. W., Delmonte, T. A., Ding, X., Garza, J. J., & Marler, B. S. (2004). A 3347-locus genetic recombination map of sequence-tagged sites reveals features of genome organization, transmission and evolution of cotton (Gossypium). *166*(1), 389-417.

Rong, J., Feltus, F. A., Waghmare, V. N., Pierce, G. J., Chee, P. W., Draye, X., Saranga, Y., Wright, R. J., Wilkins, T. A., May, O. L., Smith, C. W., Gannaway, J. R., Wendel, J. F., & Paterson, A. H. (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 [Article]. *Genetics*, *176*(4), 2577-2588.

Rong, J., Feltus, F. A., Waghmare, V. N., Pierce, G. J., Chee, P. W., Draye, X., Saranga, Y., Wright, R. J., Wilkins, T. A., May, O. L., Smith, C. W., Gannaway, J. R., Wendel, J. F., & Paterson, A. H. (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), 2577-2588.

Rong, J., Wang, X., Schulze, S. R., Compton, R. O., Williams-Coplin, T. D., Goff, V., Chee, P. W., & Paterson, A. H. (2012). Types, levels and patterns of low-copy DNA sequence divergence, and phylogenetic implications, for Gossypium genome types. *Heredity (Edinb)*, *108*(5), 500-506. https://doi.org/10.1038/hdy.2011.111

Said, J., Song, M., Wang, H., Lin, Z., Zhang, X., Fang, D., & Zhang, J. (2015). A comparative meta-analysis of QTL between intraspecific Gossypium hirsutum and interspecific G. hirsutum × G. barbadense populations [Article]. *Molecular Genetics & Genomics*, *290*(3), 1003-1025. https://doi.org/10.1007/s00438-014-0963-9

Said, J. I., Knapka, J. A., Song, M., & Zhang, J. (2015). Cotton QTLdb: a cotton QTL database for QTL analysis, visualization, and comparison between Gossypium hirsutum and G. hirsutum × G. barbadense populations. *Mol Genet Genomics*, *290*(4), 1615-1625.

Said, J. I., Lin, Z., Zhang, X., Song, M., & Zhang, J. (2013). A comprehensive meta QTL analysis for fiber quality, yield, yield related and morphological traits, drought tolerance, and disease resistance in tetraploid cotton. *BMC Genomics*, *14*(1), 776.

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

Said, J. I., Song, M., Wang, H., Lin, Z., Zhang, X., Fang, D. D., & Zhang, J. (2015). A comparative meta-analysis of QTL between intraspecific Gossypium hirsutum and interspecific G. hirsutum × G. barbadense populations. *Mol Genet Genomics*, *290*(3), 1003-1025.

Saranga, Y., Menz, M., Jiang, C. X., Wright, R. J., Yakir, D., & Paterson, A. H. (2001). Genomic dissection of genotype × environment interactions conferring adaptation of cotton to arid conditions. *Genome Research*, *11*(12), 1988-1995.

Seagull, R. W., Oliveri, V., Murphy, K., Binder, A., & Sushma, K. (2000). Cotton fiber growth and development 2. Changes in cell diameter and wall birefringence. *Journal of Cotton Science*, *4*(2), 97-104.

Shappley, Z. W., Jenkins, J. N., Watson, C. E., Jr., Kahler, A. L., & Meredith, W. R. (1996). Establishment of molecular markers and linkage groups in two F2 populations of Upland cotton. *Theoretical and Applied Genetics*, *92*(8), 915-919. http://proxy-remote.galib.uga.edu/login?url=http://search.ebscohost.com/login.aspx?direct=true&db=lah&AN=19961607728&site=eds-live

Shen, X., Cao, Z., Singh, R., Lubbers, E. L., Xu, P., Smith, C. W., Paterson, A. H., & Chee, P. W. (2011). Efficacy of qFL-chr1, a Quantitative Trait Locus for Fiber Length in Cotton (Gossypium spp.) [https://doi.org/10.2135/cropsci2010.11.0653]. *Crop Science*, *51*(5), 2005-2010. https://doi.org/https://doi.org/10.2135/cropsci2010.11.0653

Shen, X., Guo, W., Zhu, X., Yuan, Y., Yu, J. Z., Kohel, R. J., & Zhang, T. (2005). Molecular mapping of QTLs for fiber qualities in three diverse lines in Upland cotton using SSR markers [Article]. *Molecular Breeding*, *15*(2), 169-181.

Shi, Y., Li, W., Li, A., Ge, R., Zhang, B., Li, J., Liu, G., Li, J., Liu, A., Shang, H., Gong, J., Gong, W., Yang, Z., Tang, F., Liu, Z., Zhu, W., Jiang, J., Yu, X., Wang, T., & Wang, W. (2015). Constructing a high-density linkage map for Gossypium hirsutum × Gossypium barbadense and identifying QTLs for lint percentage [Article]. *Journal of Integrative Plant Biology*, *57*(5), 450-467. https://doi.org/10.1111/jipb.12288

Stephens, S. G. (1949). The cytogenetics of speciation in Gossypium I. Selective elimination of the donor parent genotype in interspecific backcrosses [Journal Article]. *Genetics 1949.*, *34*, 627-637.

Stuber, C. W., Polacco, M., & Lynn, M. (1999). Synergy of empirical breeding, marker-assisted selection, and genomics to increase crop yield potential. *Crop Science*, *39*(6), 1571.

Su, J., Pang, C., Wei, H., Li, L., Liang, B., Wang, C., Song, M., Wang, H., Zhao, S., Jia, X., Mao, G., Huang, L., Geng, D., Wang, C., Fan, S., & Yu, S. (2016). Identification of favorable SNP alleles and candidate genes for traits related to early maturity via GWAS in upland cotton. *BMC Genomics*, *17*(1), 687. https://doi.org/10.1186/s12864-016-2875-z

Sun, S., Li, C., Chee, P. W., Paterson, A. H., Jiang, Y., Xu, R., Robertson, J. S., Adhikari, J., & Shehzad, T. (2020). Three-dimensional photogrammetric mapping of cotton bolls in situ based on point cloud segmentation and clustering. *ISPRS Journal of Photogrammetry and Remote Sensing*, *160*, 195-207.

Szalma, S. J., Hostert, B. M., LeDeaux, J. R., Stuber, C. W., & Holland, J. B. (2007a). QTL mapping with near-isogenic lines in maize. *Theoretical and Applied Genetics*, *114*(7), 1211-1228. https://doi.org/10.1007/s00122-007-0512-6

Szalma, S. J., Hostert, B. M., LeDeaux, J. R., Stuber, C. W., & Holland, J. B. (2007b). QTL mapping with near-isogenic lines in maize [Article]. *Theoretical & Applied Genetics*, *114*(7), 1211-1228. https://doi.org/10.1007/s00122-007-0512-6

Taliercio, E., & Kloth, R. (2004). Expression and Characterization of a UDP-Glucose Pyrophosphorylase Gene in Cotton. *Journal of Cotton Science*, *8*(2), 91-98.

Tan, C., Li, C., He, D., & Song, H. (2021). *Anchor-free deep convolutional neural network for plant and plant organ detection and counting* 2021 ASABE Annual International Virtual Meeting, St. Joseph, MI. https://elibrary.asabe.org/abstract.asp?aid=52460&t=5

Tang, S., Teng, Z., Zhai, T., Fang, X., Liu, F., Liu, D., Zhang, J., Liu, D., Wang, S., Zhang, K., Shao, Q., Tan, Z., Paterson, A., & Zhang, Z. (2015). Construction of genetic map and QTL analysis of fiber quality traits for Upland cotton (Gossypium hirsutum L.) [Article]. *Euphytica*, *201*(2), 195-213. https://doi.org/10.1007/s10681-014-1189-y

Tanksley, S. D., & Nelson, J. C. (1996a). Advanced backcross QTL analysis: a method for the simultaneous discovery and transfer of valuable QTLs from unadapted germplasm into elite breeding lines. *92*(2), 191-203.

Tanksley, S. D., & Nelson, J. C. (1996b). Advanced backcross QTL analysis: a method for the simultaneous discovery and transfer of valuable QTLs from unadapted germplasm into elite breeding lines. *Theoretical and Applied Genetics*, *92*(2), 191-203.

Tuinstra, M. R., Ejeta, G., & Goldsbrough, P. B. (1997). Heterogeneous inbred family (HIF) analysis: a method for developing near-isogenic lines that differ at quantitative trait loci. *Theoretical and Applied Genetics*, *95*(5), 1005-1011.

Ulloa, M., Meredith Jr, W. R., Shappley, Z. W., & Kahler, A. L. (2002). RFLP genetic linkage maps from four F2.3 populations and a joinmap of Gossypium hirsutum L [Article]. *Theoretical & Applied Genetics*, *104*(2/3), 200-208.

Ulloa, M., & Meredith, W. R., Jr. (2000). Genetic linkage map and QTL analysis of agronomic and fiber quality traits in an intraspecific population. *Journal of Cotton Science*, *4*(3), 161-170.

USDA. (2016, May 26, 2016). *Cotton and Wool*. USDA.

Utz, H. F., Melchinger, A. E., & Schön, C. C. (2000). Bias and Sampling Error of the Estimated Proportion of Genotypic Variance Explained by Quantitative Trait Loci Determined From Experimental Data in Maize Using Cross Validation and Validation With Independent Samples. *Genetics*, *154*(4), 1839-1849.

Vollesen, K. (1987). The Native Species of Gossypium (Malvaceae) in Africa, Arabia and Pakistan [research-article]. *Kew Bulletin*, *42*(2), 337-349. https://doi.org/10.2307/4109688

Waddle, B. M., Lewis, C. F., & Richmond, T. R. (1961). The genetics of flowering response in cotton. III. Fruiting behaviour of Gossypium hirsutum race lati-folium in a cross with a variety of cultivated American upland cotton [Journal Article]. *Genetics*, *46*, 427-437.

Waghmare, V. N., Rong, J., Rogers, C. J., Bowers, J. E., Chee, P. W., Gannaway, J. R., Katageri, I., & Paterson, A. H. (2016). Comparative transmission genetics of introgressed chromatin in Gossypium (cotton) polyploids [research-article]. *American Journal of Botany*, *103*(4), 719-729.

Waghmare, V. N., Rong, J., Rogers, C. J., Pierce, G. J., Wendel, J. F., & Paterson, A. H. (2005). Genetic mapping of a cross between Gossypium hirsutum (cotton) and the Hawaiian endemic, Gossypium tomentosum [Article]. *Theoretical & Applied Genetics*, *111*(4), 665-676. https://doi.org/10.1007/s00122-005-2032-6

Wang, B., Guo, W., Zhu, X., Wu, Y., Huang, N., & Zhang, T. (2007). QTL mapping of yield and yield components for elite hybrid derived-RILs in upland cotton. *J Genet Genomics*, *34*(1), 35-45. https://doi.org/10.1016/s1673-8527(07)60005-8

Wang, C. C. N., Pei-Chun, C., Ka-Lok, N., Chun-Ming, C., Sheu, P. C. Y., & Tsai, J. J. P. (2014). A model comparison study of the flowering time regulatory network in Arabidopsis [Article]. *BMC Systems Biology*, *8*(1), 1-25. https://doi.org/10.1186/1752-0509-8-15

Wang, F., Gong, Y., Zhang, C., Liu, G., Wang, L., Xu, Z., & Zhang, J. (2011). Genetic effects of introgression genomic components from Sea Island cotton (Gossypium barbadense L.) on fiber related traits in upland cotton (G. hirsutum L.). *Euphytica*, *181*, 41-53.

Wang, G. L., Dong, J. M., & Paterson, A. H. (1995). The distribution of Gossypium hirsutum chromatin in G. barbadense germ plasm: molecular analysis of introgressive plant

breeding. *Theoretical and Applied Genetics*, *91*(6), 1153-1161. https://doi.org/10.1007/BF00223934

Wang, M., Tu, L., Lin, M., Lin, Z., Wang, P., Yang, Q., Ye, Z., Shen, C., Li, J., Zhang, L., Zhou, X., Nie, X., Li, Z., Guo, K., Ma, Y., Huang, C., Jin, S., Zhu, L., Yang, X., . . . Zhang, X. (2017). Asymmetric subgenome selection and cis-regulatory divergence during cotton domestication. *Nature Genetics*, *49*(4), 579-587. https://doi.org/10.1038/ng.3807

Wang, Y., Secco, D., & Poirier, Y. (2008). Characterization of the PHO1 gene family and the responses to phosphate deficiency of Physcomitrella patens. *Plant Physiol*, *146*(2), 646-656. https://doi.org/10.1104/pp.107.108548

Wang, Z., Zhang, D., Wang, X., Tan, X., Guo, H., & Paterson, A. H. (2013). A Whole-Genome DNA Marker Map for Cotton Based on the D-Genome Sequence of Gossypium raimondii L. *G3 Genes|Genomes|Genetics*, *3*(10), 1759-1767. https://doi.org/10.1534/g3.113.006890

Wendel, J. F. (1989). New World Tetraploid Cottons Contain Old World Cytoplasm [research-article]. *Proceedings of the National Academy of Sciences of the United States of America*, *86*(11), 4132-4136.

Wendel, J. F., & Albert, V. A. (1992). Phylogenetics of the Cotton Genus (Gossypium): Character-State Weighted Parsimony Analysis of Chloroplast-DNA Restriction Site Data and Its Systematic and Biogeographic Implications [research-article]. *Systematic Botany*(1), 115. https://doi.org/10.2307/2419069

Wendel, J. F., Schnabel, A., & Seelanan, T. (1995). Bidirectional Interlocus Concerted Evolution Following Allopolyploid Speciation in Cotton (Gossypium) [research-article]. *Proceedings of the National Academy of Sciences of the United States of America*, *92*(1), 280-284.

Wendel, J. F., Small, R. L., Cronn, R. C., & Brubaker, C. L. Genes, jeans, and genomes: reconstructing the history of cotton.

Xu, R., & Li, C. (2022). A modular agricultural robotic system (MARS) for precision farming: Concept and implementation [Report]. *Journal of Field Robotics*, *39*(4), 387.

Xu, Z., He, J., Tehseen Azhar, M., Zhang, Z., Fan, S., Jiang, X., Jia, T., Shang, H., & Yuan, Y. (2022). UDP-glucose pyrophosphorylase: genome-wide identification, expression and functional analyses in Gossypium hirsutum. *PeerJ*, *10*, e13460. https://doi.org/10.7717/peerj.13460

Yano, M., Harushima, Y., Nagamura, Y., Kurata, N., Minobe, Y., & Sasaki, T. (1997). Identification of quantitative trait loci controlling heading date in rice using a high-density linkage map. *95*(7), 1025-1032.

Yu, J., Yu, S., Lu, C., Wang, W., Fan, S., Song, M., Lin, Z., Zhang, X., & Zhang, J. (2007). High-density Linkage Map of Cultivated Allotetraploid Cotton Based on SSR, TRAP, SRAP and AFLP Markers [https://doi.org/10.1111/j.1744-7909.2007.00459.x]. *Journal of Integrative Plant Biology*, *49*(5), 716-724.

Yu, J., Zhang, K., Li, S., Yu, S., Zhai, H., Wu, M., Li, X., Fan, S., Song, M., Yang, D., Li, Y., & Zhang, J. (2013). Mapping quantitative trait loci for lint yield and fiber quality across environments in a Gossypium hirsutum × Gossypium barbadense backcross inbred line population. *Theor Appl Genet*, *126*(1), 275-287.

Yu, Y., Yuan, D., Liang, S., Li, X., Wang, X., Lin, Z., & Zhang, X. (2011). Genome structure of cotton revealed by a genome-wide SSR genetic map constructed from a BC1 population between gossypium hirsutum and G. barbadense. *BMC Genomics*, *12*(1), 15.

Yufang, G., McCarty, J. C., Saha, S., Jenkins, J. N., & Jixiang, W. (2007/01//). QTL ANALYSES OF FLOWERING TIME IN A PRIMITIVE COTTON ACCESSION (GOSSYPIUM HIRSUTUM L.).

Zeng, L. H., Meredith, W. R., Boykin, D. L., & Taliercio, E. (2007). Evaluation of an exotic germplasm population derived from multiple crosses among Gossypium tetraploid species. *Journal of Cotton Science*, *11*(3), 118-127.

Zeng, Q.-w., Qin, S., Song, S.-q., Zhang, M., Xiao, Y.-h., Luo, M., Hou, L., & Pei, Y. (2012). Molecular Cloning and Characterization of a Cytokinin Dehydrogenase Gene from Upland Cotton (Gossypium hirsutum L.). *Plant Molecular Biology Reporter*, *30*(1), 1-9.

Zhang, J., Jia, X., Guo, X., Wei, H., Zhang, M., Wu, A., Cheng, S., Cheng, X., Yu, S., & Wang, H. (2021). QTL and candidate gene identification of the node of the first fruiting branch (NFFB) by QTL-seq in upland cotton (Gossypium hirsutum L.). *BMC Genomics*, *22*(1), 882.

Zhang, J., Zhang, P., Huo, X., Gao, Y., Chen, Y., Song, Z., Wang, F., & Zhang, J. (2021). Comparative Phenotypic and Transcriptomic Analysis Reveals Key Responses of Upland Cotton to Salinity Stress During Postgermination [Original Research]. *Frontiers in Plant Science*, *12*. https://doi.org/10.3389/fpls.2021.639104

Zhang, J. F., Percy, R. G., & McCarty, J. C. (2014). Introgression genetics and breeding between Upland and Pima cotton: a review. *Euphytica*, *198*, 1-12.

Zhang, Q., Zhang, H., Sun, L., Fan, G., Ye, M., Jiang, L., Liu, X., Ma, K., Shi, C., Bao, F., Guan, R., Han, Y., Fu, Y., Pan, H., Chen, Z., Li, L., Wang, J., Lv, M., Zheng, T., . . . Cheng, T. (2018). The genetic architecture of floral traits in the woody plant Prunus mume. *Nature Communications*, *9*(1), 1702. https://doi.org/10.1038/s41467-018-04093-z

Zhang, T., Hu, Y., Jiang, W., Fang, L., Guan, X., Chen, J., Zhang, J., Saski, C. A., Scheffler, B. E., Stelly, D. M., Hulse-Kemp, A. M., Wan, Q., Liu, B., Liu, C., Wang, S., Pan, M., Wang, Y., Wang, D., Ye, W., . . . Chen, Z. J. (2015). Sequencing of allotetraploid cotton (Gossypium hirsutum L. acc. TM-1) provides a resource for fiber improvement. *Nature biotechnology*, *33*(5), 531-537.

Zhang, T., Yuan, Y., Yu, J., Guo, W., & Kohel, R. J. (2003). Molecular tagging of a major QTL for fiber strength in Upland cotton and its marker-assisted selection [Article]. *Theoretical & Applied Genetics*, *106*(2), 262-268.

Zhen, Z., Haihong, S., Yuzhen, S., Long, H., Junwen, L., Qun, G., Juwu, G., Aiying, L., Tingting, C., Dan, W., Yanling, W., Palanga, K. K., Muhammad, J., Weijie, L., Quanwei, L., Xiaoying, D., Yunna, T., Weiwu, S., Juan, C., & Pengtao, L. (2016). Construction of a high-density genetic map by specific locus amplified fragment sequencing (SLAF-seq) and its application to Quantitative Trait Loci (QTL) analysis for boll weight in upland cotton (Gossypium hirsutum.) [Article]. *BMC Plant Biology*, *16*, 1-18.

Zheng, L., Wu, H., Qanmber, G., Ali, F., Wang, L., Liu, Z., Yu, D., Wang, Q., Xu, A., & Yang, Z. (2020). Genome-Wide Study of the GATL Gene Family in Gossypium hirsutum L. Reveals that GhGATL Genes Act on Pectin Synthesis to Regulate Plant Growth and Fiber Elongation. *Genes (Basel)*, *11*(1). https://doi.org/10.3390/genes11010064

Zhu, G., Hou, S., Song, X., Wang, X., Wang, W., Chen, Q., & Guo, W. (2021). Genome-wide association analysis reveals quantitative trait loci and candidate genes involved in yield components under multiple field environments in cotton (Gossypium hirsutum). *BMC Plant Biology*, *21*(1), 250. https://doi.org/10.1186/s12870-021-03009-2

Zhu, L. L., & Kuraparthy, V. (2014). Molecular genetic mapping of the major effect photoperiod response locus in Pima cotton (Gossypium barbadense L.) [Journal Article]. *Crop Science, 54*(6), 2492-2498.

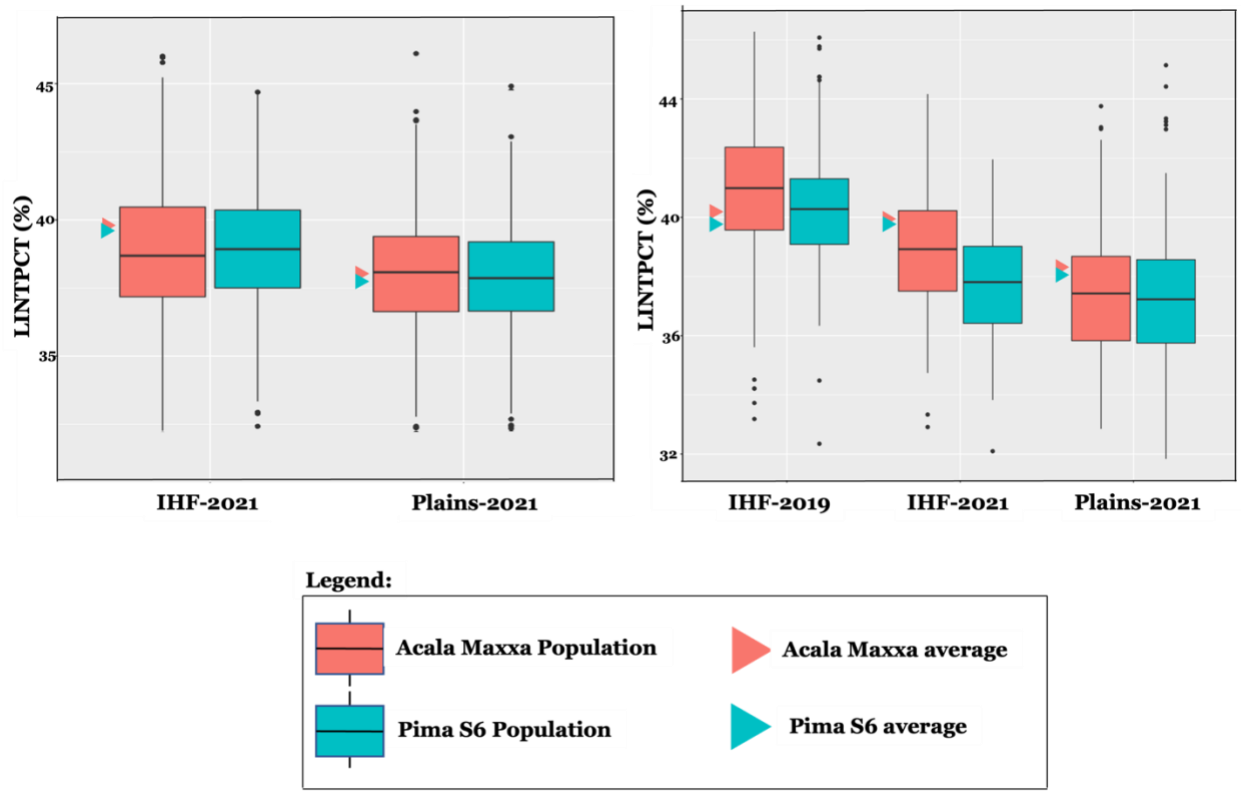

Figure 6.1. Distribution of lint percentage in NILs and advanced backcross populations.

Table 6.1. A summary statistics of fiber quality traits in advanced backcross populations (ABL), Near-isogenic lines (NIL), and parental lines. Sd denotes standard deviation, Pop denotes population type, Env denotes environment in which the populations were evaluated.

Pop	Env	Parental means		Acala Maxxa Background		Pima S6 Background	
		Acala Maxxa	Pima S6	Mean (Sd)	Range	Mean (Sd)	Range
ABL	IHF 2019	40.53	39.08	40.83 (2.16)	33.18 - 46.28	40.25 (1.72)	32.35 - 46.08
	IHF 2021	39.98	39.09	38.95 (1.96)	32.91 - 44.17	37.76 (1.83)	32.09 - 41.96
	Plains 2021	38.27	38.08	37.41 (2.12)	32.85 - 43.76	37.14 (2.19)	31.83 - 45.14
NIL	IHF 2021	39.98	39.09	38.77 (2.53)	31.15 - 46.01	38.88 (2.15)	31.74 - 44.69
	Plains 2021	38.27	38.08	37.99 (2.14)	31.39 - 46.11	37.94 (2.07)	31.56 - 46.68

Table 6.2. Analysis of variance of lint percentage in reciprocal backgrounds. DF denote degrees of freedom, SS denotes sum of squares, MS denotes mean sum of squares, PVE denotes percent variance explained by the factor. ABL = Advanced backcross lines, NIL = Near-isogenic lines.

	Source	DF	SS	MS	F-value	PVE
			Acala Maxxa Background			
ABL	ENV	2	1545.17	772.58	250.84***	29.57
	REP(ENV)	3	139.82	46.61	15.13***	2.68
	GEN	172	1302.55	7.57	2.46***	24.93
	GEN*ENV	285	1156.30	4.06	1.32**	22.13
	Error	351	1081.06	3.08	0	20.69
NIL	ENV	1	217.30	217.30	52.52***	2.68
	REP(ENV)	2	7.83	3.91	0.95	0.10
	GEN	397	3727.22	9.39	2.27***	46.02
	GEN*ENV	388	1448.62	3.73	0.9	17.89
	Error	652	2697.45	4.14		33.31
			Pima S6 Background			
ABL	ENV	2	2112.47	1056.23	417.85***	34.60
	REP(ENV)	3	9.46	3.15	1.25	0.15
	GEN	230	1547.19	6.73	2.66***	25.34
	GEN*ENV	404	1281.73	3.17	1.26**	20.99
	Error	457	1155.20	2.53		18.92
NIL	ENV	1	227.48	227.48	98.24***	3.77
	REP(ENV)	2	152.22	76.11	32.87***	2.52
	GEN	407	3440.74	8.45	3.65***	56.96
	GEN*ENV	368	1025.18	2.79	1.2*	16.97
	Error	516	1194.79	2.32		19.78

Table 6.3. QTLs for lint percentage in Acala Maxxa background.

ENV	Marker	Chr	Position (Mb)	LOD	PVE	a	d	d/a
IHF 2019	S1_37032466	1	37.03	5.93	19.86		-11.32	
	S5_83952793	5	83.95	3.93	14.33		-2.26	
	S5_85261315	5	85.26	4.08	15.17		-4.74	
	S13_16475595	13	16.48	4.81	15.45		-3.43	
	S17_47349753	17	47.35	3.38	7.65		3.54	
	S19_6882699	19	6.88	3.54	11.71		-2.09	
IHF 2021	S1_14476996	1	14.48	4.02	4.64	-1.42	-2.25	
	S2_73855069	2	73.86	7.96	5.01	-3.39		
	S3_97806275	3	97.81	3.49	13.22		-5.15	
	S13_5758744	13	5.76	13.05	3.81	-1.95		
	S16_7675496	16	7.68	15.85	4.21	-4.27		
	S17_49781200	17	49.78	14.88	3.73	0.43		
	S24_62607540	24	62.61	4.19	9.09		-4.83	
	S25_9916155	25	9.92	12.86	2.97	-1.42		
	S26_59137353	26	59.14	4.87	18.79		-8.46	
Plains 2021	S1_99771502	1	99.77	4.28	2.81	4.23		
	S4_60291193	4	60.29	4.43	5.42	5.43	1.62	0.29
	S9_56212591	9	56.21	3.76	11.66		-8.39	

Table 6.4. QTLs for lint percentage in Pima S6 background.

ENV	Marker	Chr	Position (Mb)	LOD	PVE	a	d	d/a
IHF 2019	S3_28485593	3	28.49	3.17	8.56		2.47	
	S15_57446890	15	57.45	3.53	5.79		2.58	
	S21_45230676	21	45.23	4.00	5.67		-2.26	
IHF 2021	S2_11802922	2	11.80	8.26	11.34	-6.83	-4.66	0.68
	S3_40841772	3	40.84	8.59	12.02	4.67		
	S5_43471494	5	43.47	4.56	15.35		11.52	
	S13_66569822	13	66.57	3.09	10.69		9.58	
Plains 2021	S2_73250497	2	73.25	3.97	12.93		9.27	
	S4_53563508	4	53.56	4.10	13.34		10.03	
	S6_1941831	6	1.94	10.90	31.63		12.31	
	S7_54915076	7	54.92	3.81	4.16	-1.84	-1.79	0.97
	S8_64595541	8	64.60	6.92	21.46		9.63	
	S10_73434173	10	73.43	9.90	12.84	10.09	-0.71	-0.07
	S11_46995320	11	47.00	3.89	12.01		7.44	
	S21_15093792	21	15.09	5.28	16.83		9.13	

Table 6.5. Correspondence of QTLs identified in the present study with those previously reported. P-value is based on hypergeometric function. Values in "bold" indicate correspondence of QTLs in the respective background (GB/GH) with those from previous studies.

Previous Study			Common QTLs		P.value	
Population	# QTLs	Reference	GH	GB	GH	GB
F2	13	Huang et al., 2015	5	5	**0.002**	**0.001**
RIL	23	Shen et al., 2007	2	1	0.301	0.315
F2	63	Deng et al., 2019	6	4	0.196	0.216
F2:3	14	Wang et al., 2014	4	2	**0.018**	0.207
F2:3	28	Diouf et al., 2018	5	3	0.051	0.212
RIL	22		3	3	0.183	0.152
F2	4	Wang et al., 2015	1	1	0.264	0.241
RIL	40	Wang et al., 2015	4	4	0.219	0.194
RIL	35	Li et al., 2016	5	4	0.100	0.157
RIL	104	Liu et al., 2018	5	5	0.031	0.071
F2	4	Rong et al., 2007	1	0	0.264	0.730
BIL	10	Yu et al., 2013	2	2	0.162	0.137
CSIL	128	Zhang et al., 2016	6	5	**0.009**	**0.010**
RIL	2	Yu et al., 2014	0	0	0.837	0.855
F2:3	5	Zhang et al., 2005	1	2	0.303	**0.043**
RIL	57	Liu et al., 2015	4	3	0.207	0.190
RIL	13	Wang et al., 2007	2	3	0.219	0.052

CHAPTER 7

SUMMARY

In this study, we reveal the genetic architecture of quantitative variation in cotton phenological traits as well as in various fiber yield and quality traits through QTL mapping and discover the nature and pattern of reciprocal chromatin transfer in the two mostly cultivated species of cotton. In cotton, transferring favorable alleles from Pima cotton to Upland cotton has been a long-standing interest in introgressive breeding and a lot of studies and breeding efforts have been done previously in achieving genetic gains and improving fiber quality through the introgression of alleles or allelic combinations from Pima cotton to Upland cotton. While GB has always been seen as a potential source of favorable alleles, GH has not much been investigated from similar viewpoint. Here, we developed a reciprocal set of advanced backcross lines, each set containing a few segregating chromosomal segments derived from the donor parents, Acala Maxxa and Pima S6, in the reciprocal background and investigated on the nature and pattern of reciprocal chromatin transfer in this set of two advanced backcross populations. The two parents used in our study, albeit both elite, are highly diverse in various characteristics like fiber yield, quality, flower color, flowering response, growth habit and so on. In addition, these two genotypes have been extensively used to produce molecular tools and resources including BAC libraries and Illumina genome sequences. Thus, they not only serve as potential sources of selectable genetic variation but also as references to perform various molecular genetic analyses.

Results of this study suggest that species background has a significant effect on both the amount and the location of donor chromatin introgression. While the proportion of donor alleles in the recipient background was higher than expected for the given population structure, certain genomic regions spread throughout the cotton genome were found to be recalcitrant for donor chromatin. This typical characteristic of recipient genome being impermeable to donor chromatin is not merely a random chance but because of strong background influences as we provided statistical evidence to support this hypothesis and also reiterated previous studies with similar reporting that support our findings. The pattern of such intriguing observation was seen not only for the GB chromatin introgressed into GH but also in the reciprocal transfer as well, suggesting that this genetic variation in transmission of GH chromatin into GB background might hold a promise in introducing favorable alleles or allelic combinations from GH to GB. To test this hypothesis, we selected NILs with single introgressions introgressed from GB to GH and in the reciprocal background and evaluated these lines (as well as the advanced backcross lines) for six important fiber quality traits, five cotton phenological traits, and one fiber yield trait.

Results suggest that GH is also a source of favorable alleles in the GB background, although, not as many favorable QTLs were identified in the GB background as in the GH background. Nevertheless, these results are promising in that they open up discussions and further need of explorations of GH as a source of favorable alleles in the GB background. The near-isogenic composition of our population provided opportunities to estimate the effect of the QTLs more precisely but at the cost of identifying epistatic interactions due to a few segregating multi-locus combinations in these populations. Limited QTL correspondence in reciprocal backgrounds suggest strong influence of genetic background. Identification of a significant number of QTLs with smaller effects signifies the power of our populations to reduce background noise while also

revealing the potential of these populations to accurately identify QTLs with major effects. In addition to the identification of stable QTLs in multiple environments, we were able to identify QTLs in the same introgressed region that were previously reported in similar populations signifying the role our populations in validating QTLs previously reported.