DEVELOPMENT AND CHARACTERIZATION OF

Gossypium hirsutum L. x *G. mustelinum* Miers ex Watts

CHROMOSOME SEGMENT SUBSTITUTION LINE POPULATION

A Thesis

by

JOHN CHRISTIAN HITZELBERGER

Submitted to the Office of Graduate and Professional Studies of
Texas A&M University
in partial fulfillment of the requirements for the degree of

MASTER OF SCIENCE

Chair of Committee:	David M. Stelly
Committee Members:	C. Wayne Smith
	Jane K. Dever
Head of Department:	David D. Baltensperger

December 2019

Major Subject: Plant Breeding

ABSTRACT

Genetic improvements of Upland cotton (*Gossypium hirsutum* L.) must be continuous if the crop is to remain biologically and economically viable. The notoriously low genetic diversity found among Upland cottons constrains opportunities for improvements by breeding based on conventional elite-by-elite crosses. Fortunately, each related wild AD-genome species harbors about 80,000 genes and so introgression of a wild species genome could significantly increase the breadth of variation available among Upland cottons. Most genes from a non-domesticated donor species would expectedly be agriculturally neutral or deleterious, thus, for multi-genic traits, it would expectedly be virtually impossible to discern the presence of a beneficial allele or gene in a donor genome until that DNA variant is selectively integrated into an Upland cotton genetic background. In this study, sub-chromosomal segments of the *G. mustelinum* genome were concomitantly introgressed and separated by developing chromosome segment substitution lines (CSSLs). Modified backcross-inbreeding and marker-based selections enabled the creation of a panel of CSSLs, each containing one to several small sub-chromosomal introgressed alien segments but otherwise isogenic to the *G. hirsutum* recurrent inbred line parent and to each other. Genotyping was based on single-nucleotide polymorphism markers (SNPs). Segment-targeted genotyping was based on PACE or KASP assays for small sets of two to several inter-spaced SNPs per segment, drawn from a recently developed genome-spanning panel of ~260 such assays. In contrast, genome-wide high-density genotyping was based on the Illumina Cotton63KSNP array for 15,000+ SNPs. Coverage and pedigree-based tracking of specific segments at BC_4 and BC_5 generations was enabled by prior CottonSNP63K-based genotyping of 18 BC_2F_1s. In 2017, 410 BC_4F_1s were backcrossed to *G. hirsutum* and selected, of which 92

were genotyped. In 2018, 378 of 933 BC_5F_1s were genotyped, self-pollinated and selected. Based on targeted analyses with spaced SNPs completed before January 2018, the CSSL panel comprises 65 BC_5F_1 plants that collectively contain approximately 50% of the *G. mustelinum* genome in a heterozygous state; these descend from 18 different BC_2F_1s. In subsequent research, each heterozygous donor segment must be recovered in homozygous form, and additional CSSLs with complementary genome coverage (~50%) must be identified to attain 100% of donor genome coverage. Available germplasm resources include BC_5S_1 seed from 378 BC_5F_1 plants and 77 $BC_{4:5}F_1$ families. To facilitate the follow-through efforts, I created well-organized computer spreadsheets that integrate relevant pedigree and SNP data; these help identify which segments to target, which pedigree to use for a given segment, and which SNPs to genotype for selection of heterozygotes and homozygotes. Recovery of homozygotes and complementary segments will be facilitated by the availability of plants and/or seed at BC_5S_1, BC_5F_1, and BC_4F_1 generations. To begin gauging if donor genes affect fiber quality traits, BC_5F_1 fiber samples were harvested on a single-plant basis within BC_2-derived families and characterized using High Volume Instrument (HVI) analysis. ANOVA of HVI data showed that differences were significant among families (α=0.05) for micronaire ($p = 0.0342$), upper half mean length ($p = 0.0004$), elongation ($p = 0.0253$), and strength ($p = 0.0224$). If substantiated, the results would reflect dominant or co-dominant effects, but insufficient experimental replication precludes conclusiveness at this time. More authoritative deductions about dominant, co-dominant and recessive genotypic effects will be possible once homozygous BC_5S_n CSSL lines are established, as these will be amenable to seed increases, use of larger experimental units, replication and multiple environments.

DEDICATION

To:

All my family, friends and people I have met along the way.

ACKNOWLEDGEMENTS

I would like to thank my Graduate Advisory Committee chair, Dr. Stelly, along with my Committee members Dr. Smith and Dr. Dever for their guidance, knowledge, and support throughout the course of my research. I would also like to thank my family for their moral and financial support; as well as my friends who made my time in College Station enjoyable.

CONTRIBUTERS AND FUNDING SOURCES

This work was supervised by a thesis committee chaired by Dr. Stelly of the Department of Genetics, and membered by Dr. Smith and Dr. Dever from the Department of Soil and Crop Sciences.

Funding for this research was provided by Cotton Inc. Previous and coinciding work related to this research was conducted by Ms. (now Dr.) Amanda Hulse-Kemp, Dr. Robert Vaughn, Mr. Wayne Raska, Mr. (now Dr.) Jiale (Jerry) Xu, Mr. Yu-Ming Lin, Mr. Luis De Santiago, and Ms. Kübra Velioğlu.

TABLE OF CONTENTS

LIST OF FIGURES

Page

LIST OF TABLES

Page

CHAPTER I

INTRODUCTION AND LITERATURE REVIEW

Twentieth century researchers established that the *Gossypium* genus includes both 2n=26 (diploid) and 2n=52 (tetraploid) species, and that the n=13 genomes were diversified into groups differed in terms of physical size, meiotic pairing and recombination (Beasley 1942, Skovsted 1934, Skovsted 1937). Observations and experiments demonstrated that 2n=52 species contained two subgenomes that were most closely related in size and meiotic affinity with the two extant Old World diploid species containing A genomes, including *G. herbaceum* and *G. arboreum*, and extant New World diploid species containing D genomes, such as *G. raimondii* (Webber 1934). Modern data, especially sequencing data, continues to improve the understanding of the origins of these genomes, including multiple rounds of paleo-polyploidization, and estimates of when key polyploidization events occurred (Paterson et al., 2012). The most recent polyploidization event is estimated to have occurred 1-2 MYA, creating a genome ancestral to modern 26-chromosome AD genomes. The AD-genome *Gossypium* species include seven currently recognized species: *G. hirsutum* (AD_1), *G. barbadense* (AD_2), *G. tomentosum* (AD_3), *G. mustelinum* (AD_4), *G. darwinii* (AD_5), and two newly recognized species, *G. ekmanianum* (AD_6) and *G. stephensii* (AD_7), formally known as 'Wake Island' cotton and classified as *G. hirsutum* (Gallagher et al., 2017).

Today, cotton is one of the most important textiles in the world. In the United States, cotton accounts for up to $120 billion in products and services annually (NCC 2018). Although cotton hectarage planted and production can fluctuate due to a multitude of factors, in 2017, US

producers planted 4.8 million hectares to cotton from coast to coast, and generated 22.7 million bales of cotton and 5.2 million metric tons of cottonseed (NCC 2018). Globally, total cotton production was an estimated 120 million bales with India, China, United States, and Brazil ranking as the largest producers (USDA Cotton Outlook 2018). However, the United States is the largest exporter of cotton with 3.37 million bales exported in 2017 (USDA 2017/2018). Upland cotton, *Gossypium hirsutum*, serves both as a fiber and oilseed crop, with most of the economic importance in fiber and accounts for 97% of the annual United States cotton crop (USDA 2016).

Genetic improvements of Upland cottons in the United States must be continuous if the crop is to remain biologically and economically viable, i.e., competitive to synthetic fibers, resistant to most diseases and pests, profitable to produce, and be environmentally sustainable. The threat of reduced arable lands, changing climates, and increasing population create additional challenges with which cotton producers and breeders must contend (Godfray et al., 2010). The notoriously low genetic diversity found among Upland cottons constrains opportunities for improvements by breeding based on conventional elite-by-elite crosses (Meredith et al., 2000; Wendel et al., 1992). When diversity within the elite germplasm base (i.e., new elite or obsolete cultivars) is insufficient, efforts to genetically address new problems require increased amounts of time and money, plus some problems may not be genetically addressable, and the opportunities to improve quantitative traits are diminished. Diversity in cultivated species can be increased and should be researched and developed to preemptively counter possible catastrophic crop or economic losses due to lack of diversity, and to create opportunities for genetic enhancements of crop performance and sustainability. Diversity has traditionally been increased through introgression and mutation, and more recently genetic

transformation and molecular gene-editing methods (Sunilkumar et al., 2006; Abdurakhmonov et al., 2016; Gao et al., 2017).

Fortunately, each related wild AD-genome species harbors about 80,000 genes and therefore introgression of wild species germplasm can significantly increase the breadth of variation available among Upland cottons. The *Gossypium* genus encompasses over 50 species (Gallagher et al., 2017). One of the wild AD-genome species, *Gossypium mustelinum* (2n=52, $(AD)_2$ genome), is a wild Brazilian cotton and belongs to one of the more isolated clades of allotetraploid cottons (Wendel et al., 2015). It has small 3- and 4-locule bolls with beaked tips (Figure 1, 2). The fiber is sparse and tan (Pickersgill et al., 1975) (Figure 3). Most traits of this wild species indicate that it has no agronomic value. However, for most genes affecting multigenic "quantitative" traits, the overwhelming additive effects of "wild" alleles and epistasis (gene interactions) make it virtually impossible to detect by direct phenotypic observation whether a potentially useful variant exists among the many loci of a wild species that influence a given trait. A more effective means to discover beneficial genetic variants in alien genomes is "divide and conquer", i.e., by systematically replacing each of a large number of known segments of the Upland cotton genome with the corresponding homologous alien segments. This kind of "substitution" can be accomplished by modified breeding methods that involve hybridization and backcrossing, coupled either to cytogenetic manipulations and/or marker-assisted selection. Once substitutions are recovered as true-breeding lines with a common adapted Upland cotton genetic background, phenotypic effects of each substitution, and thus genes contained therein, are far more readily discernible. Chromosome substitution (CS) lines have previously been created in wheat and cotton by cytogenetic methods, and subsequently tested for influences within and between traits associated with introgressions (Law et al., 1978;

Stelly et al., 2005; Jenkins et al., 2007). These CS lines illustrate the feasibility of creating population lines containing limited amounts of introgressed alien germplasm into a common background; the lines are generally stable and true-breeding, relatively easy to maintain and increase without the need of cytogenetic expertise, and facilitate replicated experimentation, breeding and distribution to researchers and breeders (Saha et al., 2004). These attributes, especially the amenability to replicated experimentation, greatly enhance evaluations needed to detect beneficial effects on quantitative traits by alien genetic components that would otherwise be obscured by the genetic background of the donor species. CS lines can be evaluated directly to detect average effects of the substitutions, and some cases, detect major effects by individual loci (Saha et al., 2004). They also have been bred to create various types of progenies (e.g., F1, F2, F3, and RIL) that enable the detection and dissection of various types of quantitative genetic effects, such as additive and dominance effects, and single-gene and epistatic interaction effects (Saha et al., 2008; Jenkins et al., 2006; Wu et al., 2009; Saha et al., 2013; Saha et al., 2011). Top crosses can be used to detect interaction effects of the alien gene(s) with variants present in elite Upland germplasm (Jenkins et al., 2006; Jenkins et al., 2007; Jenkins et al., 2012; Jenkins et al., 2017). Chromosome-specific RIL populations offer a means to conduct high-resolution quantitative analysis and localize QTL effects, and also provide a facile pathway to identify product lines suitable for germplasm release (Saha et al., 2017; Stelly et al., 2005).

Figure 1. Unopened bolls at late maturation. Harvested from greenhouse-grown plants. From left to right: *G. mustelinum, G. hirustum x G. mustelinum* (F_1), and *G. hirsutum* (TM-1).

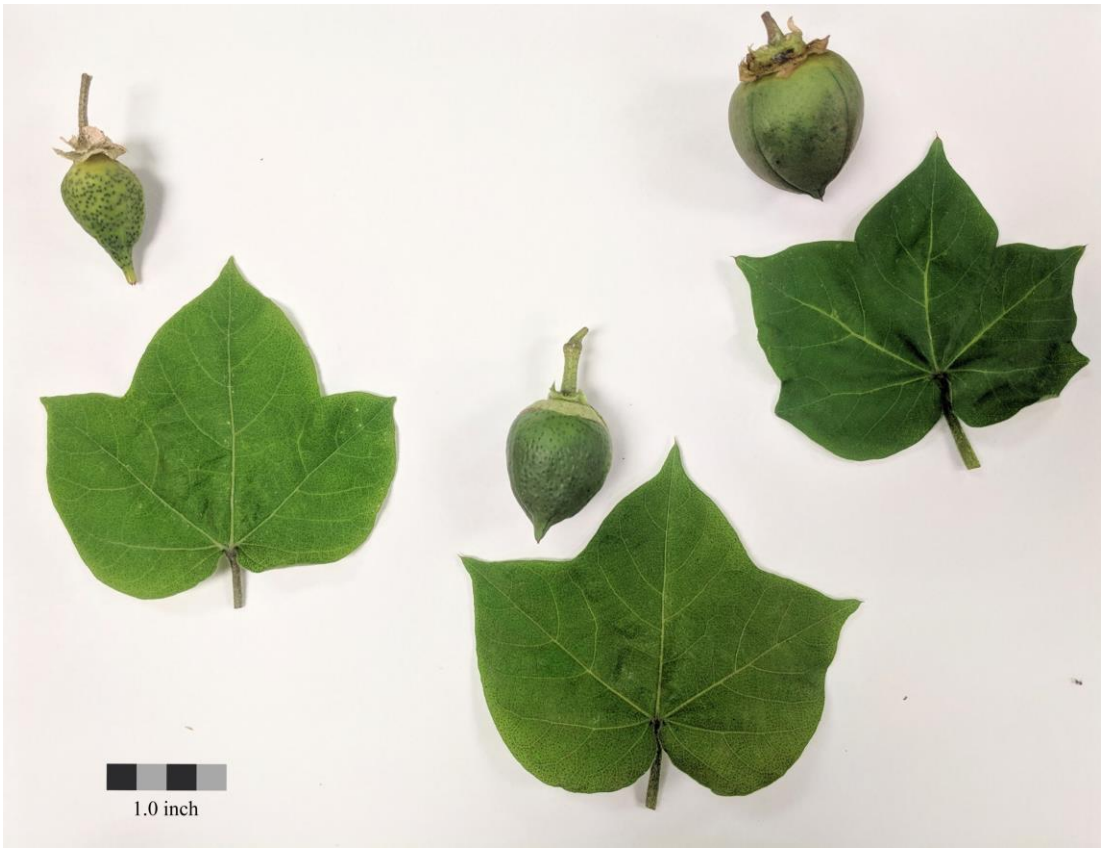


Figure 2. Unopened mature bolls with bracts removed and leaf. Harvested from greenhouse-grown plants .From left to right: *G. mustelinum, G. hirustum x G. mustelinum* (F_1), and *G. hirsutum* (TM-1).

Figure 3. Opened bolls with fiber from *G. hirsutum* TM-1 and *G. mustelinum*. *G. hirsutum* (top) and *G. mustelinum* (bottom).

A new backcross breeding method was developed to facilitate the breeding of chromosome segment substitution lines (CSSLs); the method entails modified backcross-inbreeding to create backcross introgression lines (BILs), but where marker-assisted selection is used to track introgressions through backcrossing and into the final lines, and to strategically select CSSLs into a panel that maximizes coverage of the donor genome (Eshed and Zamir, 1995; Gur and Zamir, 2004). In many instances, the final goal is to derive CSSLs that collectively contain complete or nearly complete coverage of the introgressed genome. Chromosome segment substitution populations have been developed and successfully used in rice, wheat, peanut, and tomato (Saha et al., 2013; Holtan et al., 2003; Fonceka et al., 2012; Kubo et al., 2002). Each CSSL is expected to be marker-selected to contain one to several small sub-chromosomal introgressed alien segments but otherwise be isogenic to the recurrent parent.

In developing CSSLs in cotton, we endeavor to extend the isogenic platform presently composed of CS lines that were derived by repeated backcrossing to *G. hirsutum* inbred line TM-1 (PI 607172, SA-2269). Thus, the same recurrent parent is being used for CSSL development. Overall plans include the development of panels or libraries of CSSLs for each donor genome, where each CSSL is to be strategically chosen by marker-based selection such that each CSSL harbors a unique substitution, i.e., one or more donor-genome segment(s), and that the panel of CSSLs collectively "captures" all of the alien donor genome, or most of it as is possible. It is feasible to create a CSSL panel that is initially incomplete (less than 100% coverage), but nevertheless useful, then complete it later (100% coverage) using marker-based introgression to add to it strategically. Whether complete or incomplete, the CSSLs will provide a powerful tool for introgression, characterization and utilization of *G. mustelinum* germplasm. Successful

completion will significantly expand the isogenic platform, which presently includes previously developed chromosome substitution lines (Table I). The isogenic platform creates the ability to cross multiple introgressions, from one or multiple species, together while maintaining a genetically similar, referred to as isogenic, background.

Table I. Chromosome substitution lines development status (generation of development). All introgressions occur within *G. hirsutum* TM-1; BC_5S_n lines are considered complete. (Robert Vaughn and Stelly, personal communication, 2019)

	G. barbadense			*G. mustelinum*			*G. tomentosum*		
	CS	lo	sh	CS	lo	sh	CS	lo	sh
1	BC5S5			BC5S2			BC5S5		
2	BC5S5			BC5S3			BC5S4		
3	BC5S4								
4	BC5S5			BC5S3			BC5S4		
5			BC5S5			BC5S2			BC5S4
6	BC5S5			BC5S3			BC5S4		
7	BC5S5			BC5S4			BC5S4		
8			BC5S4			BC5S3			BC5S4
9	BC5S5			BC0F1			BC4S1		
10	BC5S5			BC5S5			BC5S3		
11	BC0F1	BC5S2	BC5S5	BC0F1		BC5S3		BC5S2	BC5S4
12	BC5S5		BC5S5						
13									
14			BC5S5			BC5S2			BC5S4
15			BC5S5			BC5S3			BC5S4
16	BC5S5			BC5S3			BC5S2		
17	BC5S5			BC5S2			BC5S4		
18	BC5S5			BC5S2			BC5S4		
19									
20	BC3F1	BC5S2	BC5S2			BC5S3			
21									
22		BC5S5	BC5S5		BC5S3	BC5S3		BC5S5	BC5S4
23									
24									
25	BC5S5			BC0F1			BC5S1		
26		BC5S5			BC5S2			BC5S4	

Wide-cross introgression has classically been associated with large introgressions and negative linkage drag effects (Meredith and Bridge, 1971). Modern genetic technologies can be used to circumvent or overpower at least some of the traditional barriers traditionally associated with wide-cross introgression and germplasm characterization. The work of Eshed and Zamir (1995) in tomato exemplified their utility for strategized genome introgression, with the ultimate creation of "introgression lines", i.e., chromosome segment substitution lines.

Genetic markers have been used successfully across multiple applications within plant breeding for upwards of three decades (Antoni and Tingey, 1993; Crossa et al., 2010; Francia et al., 2005). The use of markers expedites research and production pipelines. Popular early marker technology used simple sequence repeats (SSRs) and restriction fragment length polymorphisms (RFLPs) among various others (Tanksley et al., 1989; Staub et al., 1996). In addition to previous technology, many of today's marker technology applications are based on single nucleotide polymorphisms (SNPs). SNPs provide many benefits that increase through-put of genotyping assays such as a large presence across genomes, simplicity of design and validation and compatibility across many unique assays (Mammadov et al., 2012; Hulse-Kemp et al., 2015). Two such assays employed here to develop CSSLs include the highly multiplexed Illumina CottonSNP63K array and simplex fluorescence-based polymerase chain reaction (PCR) assays. The Illumina CottonSNP63K array features surface-bound "beads" of DNA oligos to which sample DNA hybridizes and can potentially undergo single-base extension to fluoresce, where the bead-specific fluorescence is diagnostic for a given polymorphism. The simplex fluorescence-based PCR assays use a set of three primers for each locus, including two forward allele-specific primers and one common reverse primer. The binding and subsequent amplification of SNP regions in PCR leads to release of SNP-specific fluorescent dyes from

FRET-based quenching. The ratio of resulting fluoresce signals is used to determine the genotype of the locus (He et al., 2014). The use of DNA markers is an integral part of CSSL creation and marker-based selections as it can leveraged to various degrees as a selection tool during the multiple generations required to create CSSLs.

CHAPTER II

DEVELOPMENT OF CHROMOSOME SEGMENT SUBSTITUTION LINES

Expanding on Previous Research

Breeding efforts to develop *G. mustelinum* CSSLs had been advanced by previous researchers in our laboratory to the BC_4F_1 seed and plant generation (Xu 2014, Lin 2017), and a tentative high-density linkage map had been constructed based on a small BC_1F_1 population (Lin 2017). A small BC_2F_1 population was genotyped at high density with the Illumina CottonSNP63K array, but no additional genotyping was conducted beyond the BC_2 generation. In other words, the BC_3F_1 generation had been advanced without genotyping or SNP-based selection, because a genome-wide set of spaced cost-effective simplex SNP assays for targeted SNP genotyping was still too incomplete to be very useful for genome-wide selection of large non-recombinant segments (personal comm., David M. Stelly). The goals of this research included advancement of the *G. mustelinum* CSSLs to the BC_5 generation, inbreeding and genotyping to track the BC_2F_1-derived specific segments within each of the respective pedigrees, with ultimate goal of defining and maximizing donor genome coverage in the final panel of CSSLs. Concomitant with this research, a genome-wide set of spaced (~15 cM) SNP assays will be available on a collaborative basis for use in this research.

To expand on the work that was done previously and prior to genotyping new generations, which included any succeeding generations derived from the BC_2F_1, genotype data were used to establish existing introgression in the BC_2F_1 generation. SNP data from the Illumina CottonSNP63K array were collected for 18 BC_2F_1 individuals which gave rise to all

individuals in succeeding generations. The raw SNP data were converted to IUPAC codes then to ABH format. ABH format uses the letters A, B, and H to represent the state and origin of a locus' genotype, where A and B are homozygous for respective parents and H is heterozygous. The ABH data were imported in to a graphical genotyping software, GGT2.0 (van Berloo 1999), following software protocol. Using the genotype visualization output from GGT2.0, areas with missing introgressions could easily be spotted and further analysis could be done using the SNP genotype data to examine specific areas of introgression. The BC_2F_1 population was found to lack an introgression for chromosome 11, but had near complete coverage for the rest of the sought introgressions (Figure 4). These data proved useful for targeting segments through all generations of genotyping.

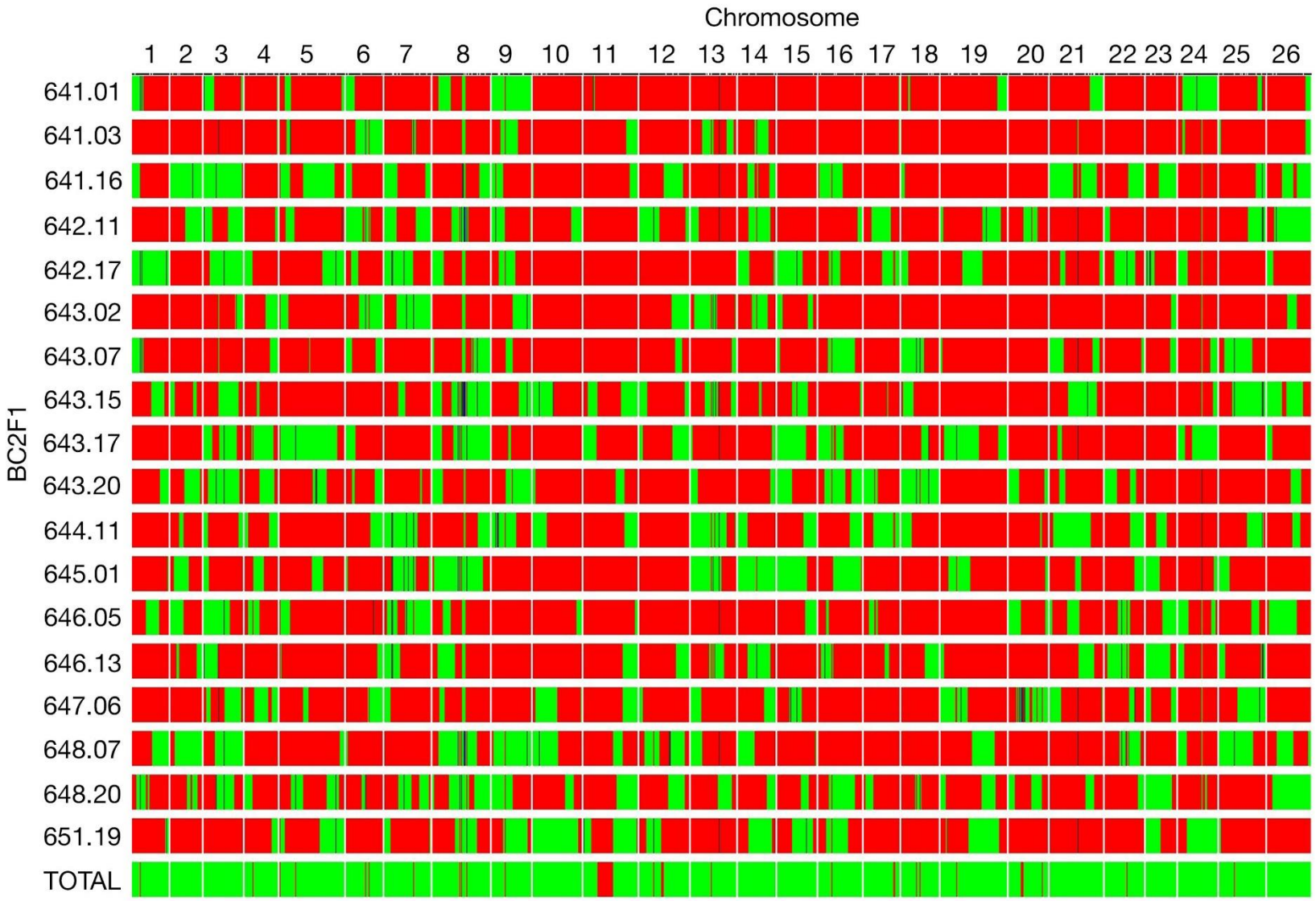


Figure 4. Graphical depiction of CottonSNP63K-based genotypes. A total of 18 BC_2F_1s containing segments heterozygous for *G. mustelinum* (green) or homozygous for *G. hirsutum* TM-1 (red). Y-axis lists BC_2F_1 identities from 2013. X-axis depicts chromosome linkage groups 1 to 26.

Defining Targets for Introgression and Segment-specific SNP Assays

A population of BC_4F_1 (*G. mustelinum*/*G. hirsutum*) plants were grown in College Station, Texas in the summer of 2017. Tissue was collected from 92 individuals representing all BC_2-derived families. One to three new unfurled leaves, less than 1 cm in length, were collected from the meristem and placed in 1.5 ml tubes labeled with the field location of the plant. DNA was extracted using Single Prep Macherey-Nagel Genomic DNA isolation kits (Macherey-Nagel, GmbH & Co. KG) following the manufacturer's protocol with the modification of isolated genomic DNA eluted in Nanopure™ water. Each sample of eluted DNA was tested for concentration in nanograms per microliter (ng/ul) using a DeNovix spectrophotometer (DeNovix Inc.). Based of spectrophotometer data, a secondary stock of DNA was created in a 96-well microplate by diluting a portion of the parent stock to 10 ng/ul. The diluted and original stock were stored in a freezer at -20 C.

Marker sets were selected to genotype the BC_4F_1 individuals based on newly developed SNP marker sets selected for equal spacing throughout the genome (Velioglu et al. 2019). To implement a cost-effective whole-genome search method to reveal the presence of any heterozygous segments, two Fluidigm 96.96 Dynamic Arrays™ (Fluidigm Corp.) were used to screen BC_4F_1 individuals. This multi-plex array allows 96 samples to be genotyped by 96 marker sets creating up to 9,216 potential data points from a single run. All 92 BC_4F_1 DNA samples along with DNA from *G. hirsutum, G. mustelinum,* and the respective F_1 were loaded in the Fluidigm Array at a concentration of 70-75 ng/ul per manufacturer's protocol (Fluidigm 2018). Multiple unique primers sets (192) were selected to screen the population. The first Fluidigm array was loaded with 96 primers representing chromosomes 1 – 13. The second

Fluidigm array was loaded with 96 primers representing chromosomes 14 – 26. All primer sets were mixed with loading reagents and loaded per manufacturer's protocol with the following adjustments to the sample mix: 17.6 ul of H_2O, and 4.4 ul of 50 mM $MgCl_2$. The first Fluidigm plate screening contained 86 samples that successfully amplified and genotyped loci. Of the total 96 primers, 63 loci were co-dominant and produced usable data that allowed the locus to be called heterozygous or homozygous. Dominant markers accounted for 15 marker sets and 18 markers sets failed to amplify. The second Fluidigm plate contained 72 samples which successfully amplified. Of 96 primer sets, 59 were co-dominant and were able to amplify loci. Across both Fluidigm plates, 122 loci of a possible 196 were amplified which equates to a 62% coverage rate. These genotype data were used to initiate a data sheet that tracked *G. mustelinum* introgression segments. When genotype data for all BC_4F_1 individuals were considered collectively, it was determined that at least 60% of the *G. mustelinum* introgression was represented in the 92 BC_4F_1 individuals. The genotyping was incomplete likely due to the selection of unconfirmed markers due to the lack of complete genome coverage at the time or low-quality DNA. Nonetheless, it was sufficient to make a limited number of direct selections, and in terms of both plant samples and genome samples it added to collective information that allowed for chromosome recovery in the subsequent generations.

In the summer of 2018, 960 BC_5F_1 seedlings were space-transplanted into research plots. After two to three weeks of growth, tissue was collected from a stand of 941 BC_5F_1 plants. Three to four unfurled leaves with a length of less than 1cm were collected in a 96 deep-well homogenization plate (OPS Diagnostics, LLC.). Homogenization plates containing tissue were stored in a -20 C freezer until processed for DNA extraction. DNA was extracted using ten 96-Well Synergy Plant DNA Extraction Kits following manufacturer's protocol (OPS Diagnostics,

LLC.) with the final elution step in Nanopure™ water. Extracted DNA was stored at -20C. Genotyping was conducted post-harvest, therefore the only individuals that produced self-pollinated seed were selected for genotyping. A total of 378 DNA samples representing the selected BC_5F_1 individuals were collected from across ten DNA 96-well stock plates and consolidated to four 96-well stock plates to expedite genotyping assays and concentration analysis. The selected 378 samples were quantified using Cytation 5 Cell Imaging Multi-Mode Reader (BioTek Instruments, Inc.). DNA concentrations were as low as 20 ng/ul and as high as 1600 ng/ul with an average of 350 ng/ul and median of 266 ng/ul. The 260/280 ratios, a measurement of DNA purity and containments by light reflectivity where DNA consistently absorbs 260 nm light and 280 nm absorbance will be affected by levels of contaminants such as phenol or proteins (Thermo Fisher Scientific 2009), ranged from 1.2 to 2.8 with an average of 1.72 and median of 1.82. Dilution plates were prepped using 96-well micro plates. The calculated volume of DNA to achieve 10 ng/ul in a total volume of 100 ul was pipetted into predetermined wells and dried for one hour at 96 C. Once the water was fully evaporated, 100 ul of Nanopure™ H_2O was added to every well. The plate was sealed, agitated, then allowed to rest at room temperature for one hour to allow the DNA pellet to dissolve. The dilution and original stock plates were stored at -20 C.

Simplex assays were used to genotype the selected BC_5F_1 individuals on a chromosome basis to track introgression segments. Some marker sets were chosen based on genotype data from the BC_2F_1 and BC_4F_1 generations; others were chosen because they involved donor segments for which coverage was known at BC_2F_1 but yet to be established at BC_4 or BC_5 generations. KASP master mix was replaced by the comparable PACE master mix (3CR Bioscience Ltd.). A total of 48, 96-well and 384-well PCR plates were used to genotype

individuals across all 26 chromosomes. The 96-well plate PCR preparations were as follows: 4 ul of 10 ng/ul sample DNA was added to each well and dried at 65 C for one hour. Master mix was made to accommodate a total reaction volume of 8 ul and was composed of 4 ul Nanopure™ H_2O, 4 ul PACE master mix, and 0.112 ul primer set. A single master mix was concocted in a 1.5 ml centrifuge tube by taking the product of a single well volume and the number of wells sharing the same primer, plus 3% to account for any pipette loss. The 384-well micro plate PCR preparation were as follows: 2.5 ul of 10 ng/ul sample DNA was added to each well and dried at 65 C for one hour. Reactions were a total volume of 5 ul. Each reaction was composed of 2 ul Nanopure™ or PCR quality water, 3 ul PACE master mix, and 0.07 ul primer set. Both plate sizes followed identical thermocycling conditions suggested by the manufacturer (PACE, 3CR Bioscience Ltd.). Fluorescent readings were taken at 40, 45, and 50 cycles using a PHERAstar Microplate Reader (BMG Labtech Inc.).

CSSL Coverage

Recovery of introgressed segments among genotyped BC5F1 generation varied by chromosome. Considering only the segments for which diagnostic SNP genotypes were determined, a majority of the genome was recovered, i.e., when the "missing genotype data" were excluded. No genotype data were recovered from chromosomes 5, 7, 13, and 14 due to poor quality of DNA extractions from some individuals, PCR errors, progeny selection, or other errors. Percentages of coverage were calculated for the minimum and maximum based on cM distance using the GHMv1 map (Lin 2017). Minimum segment coverages were calculated using only the distance between two loci confirmed through genotyping. Maximum coverages were calculated using the distance up to the next (flanking) unconfirmed SNP, to the end of the chromosome, and pedigree analysis (Table II, Figure 5).

The available germplasm and corresponding SNP genotype data made it possible to list the identified introgressed segments and their associated marker sets (Table III). The marker sets will allow expedited genotyping in future CSSL-specific screenings. The marker set for a given segment contains the SNP assays that have been verified in ongoing lab efforts of genome-wide spaced-SNP KASP/PACE assay development. In future applications of CSSLs that involve segments which span multiple SNP-assayable loci, the specific situation will determine is just one SNP or multiple segment-spanning SNPs must be genotyped to detect the presence or absence of a segment. Genotyping for just one segment-specific SNP should suffice if the parent is homozygous for the donor segment of interest, e.g., to confirm hybridity of a seed after cross-pollination, but genotyping for multiple segment-specific spaced SNPs would be desirable if from a parent heterozygous for the donor segment, e.g., for genetic "dissection".

Table II. Estimated coverage of donor (*G. mustelinum*) genome by genotyped individuals of BC_5F_1 CSSL panel. Minimum cM values are based on the presence of segment-specific combinations of SNPs and their linkage map positions in the CottonSNP63K-based genome-wide interspecific BC_1F_1 linkage map from *G. hirsutum* x (*G. hirsutum* - *G. mustelinum* F_1) by Lin (2017). Maximum cM values extend potential coverage to the nearest "negative" SNP assay(s). Percentages state results relative to overall estimated chromosome length (Lin 2017).

Chr.	Minimum (cM)	Maximum (cM)	Total (cM)	Min%	Max%
1	103.5	127.3	134.1	77%	95%
2	28.6	83.2	117.2	24%	71%
3	108.8	132.6	146.2	74%	91%
4	83.2	122.3	122.3	68%	100%
5	n/a	n/a	n/a	n/a	n/a
6	110.3	115.4	134.1	82%	86%
7	n/a	n/a	n/a	n/a	n/a
8	66.2	128.9	213.9	31%	60%
9	136.1	144.7	144.7	94%	100%
10	146.0	180.1	180.1	81%	100%
11	89.7	195.5	200.3	45%	98%
12	45.8	86.6	185.0	25%	47%
13	n/a	n/a	n/a	n/a	n/a
14	n/a	n/a	n/a	n/a	n/a
15	69.5	135.8	146.0	48%	93%
16	94.9	137.5	161.3	59%	85%
17	59.4	71.4	132.6	45%	54%
18	11.7	76.3	137.5	9%	55%
19	158.3	233.1	246.6	64%	95%
20	10.0	55.5	144.5	7%	38%
21	76.4	128.1	197.1	39%	65%
22	23.9	40.8	144.4	17%	28%
23	95.0	112.1	112.1	85%	100%
24	112.9	146.0	146.0	77%	100%
25	71.3	144.4	169.9	42%	85%
26	93.3	159.5	162.9	57%	98%

A.

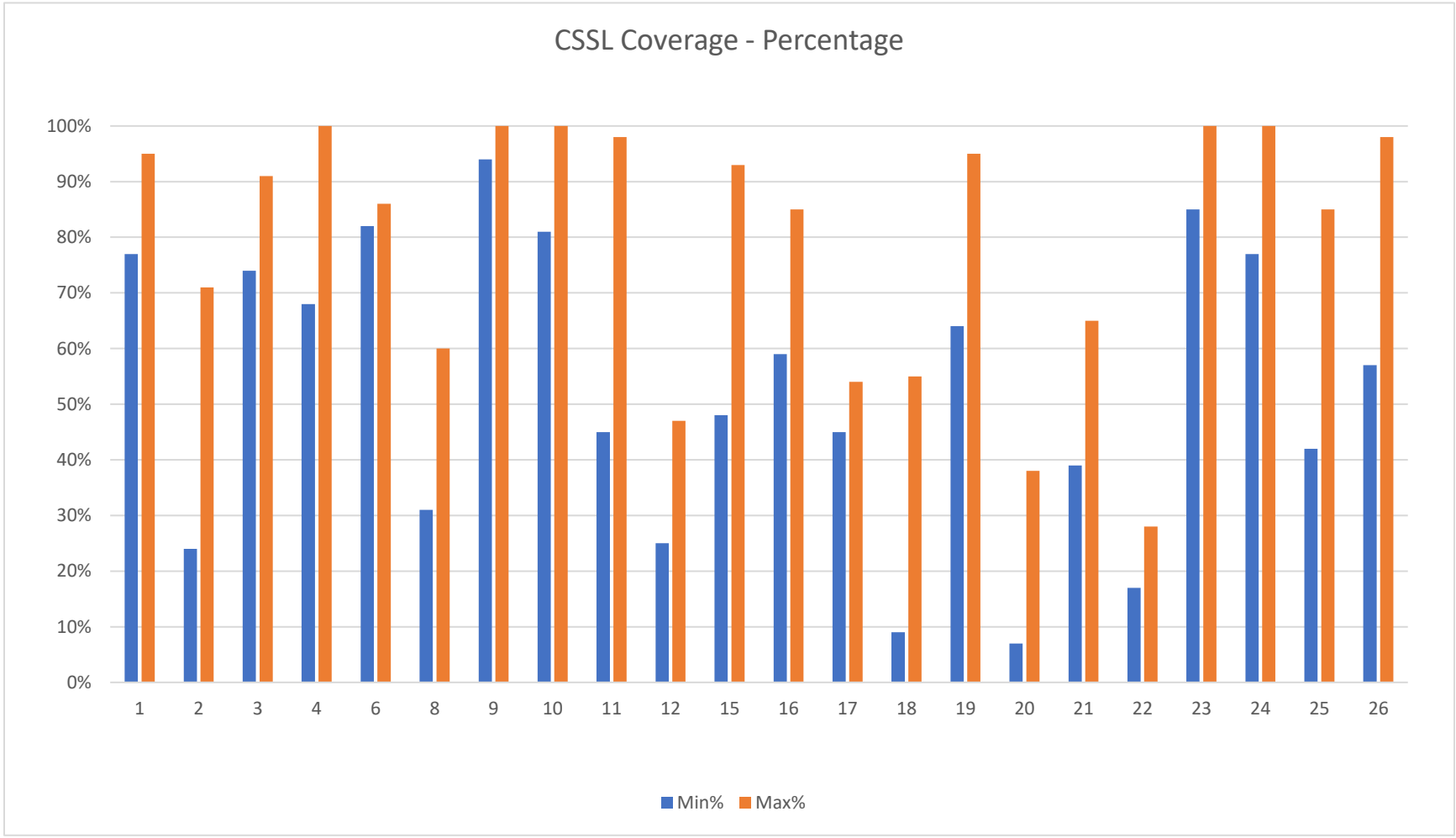


Figure 5. Estimated chromosome-specific coverage provided by current CSSL panel. Histograms depict estimated minimum and maximum coverages of donor (*G. mustelinum*) chromosomes among genotyped individuals of BC_5F_1 CSSL panel. Estimates for each chromosome are shown as A) a percentage (Top) or B) in proportion (Bottom) to the overall map length of the respective chromosome in the CottonSNP63K-based genome-wide interspecific BC_1F_1 linkage map from *G. hirsutum* x (*G. hirsutum* - *G. mustelinum* F1) by Lin (2017).

B.

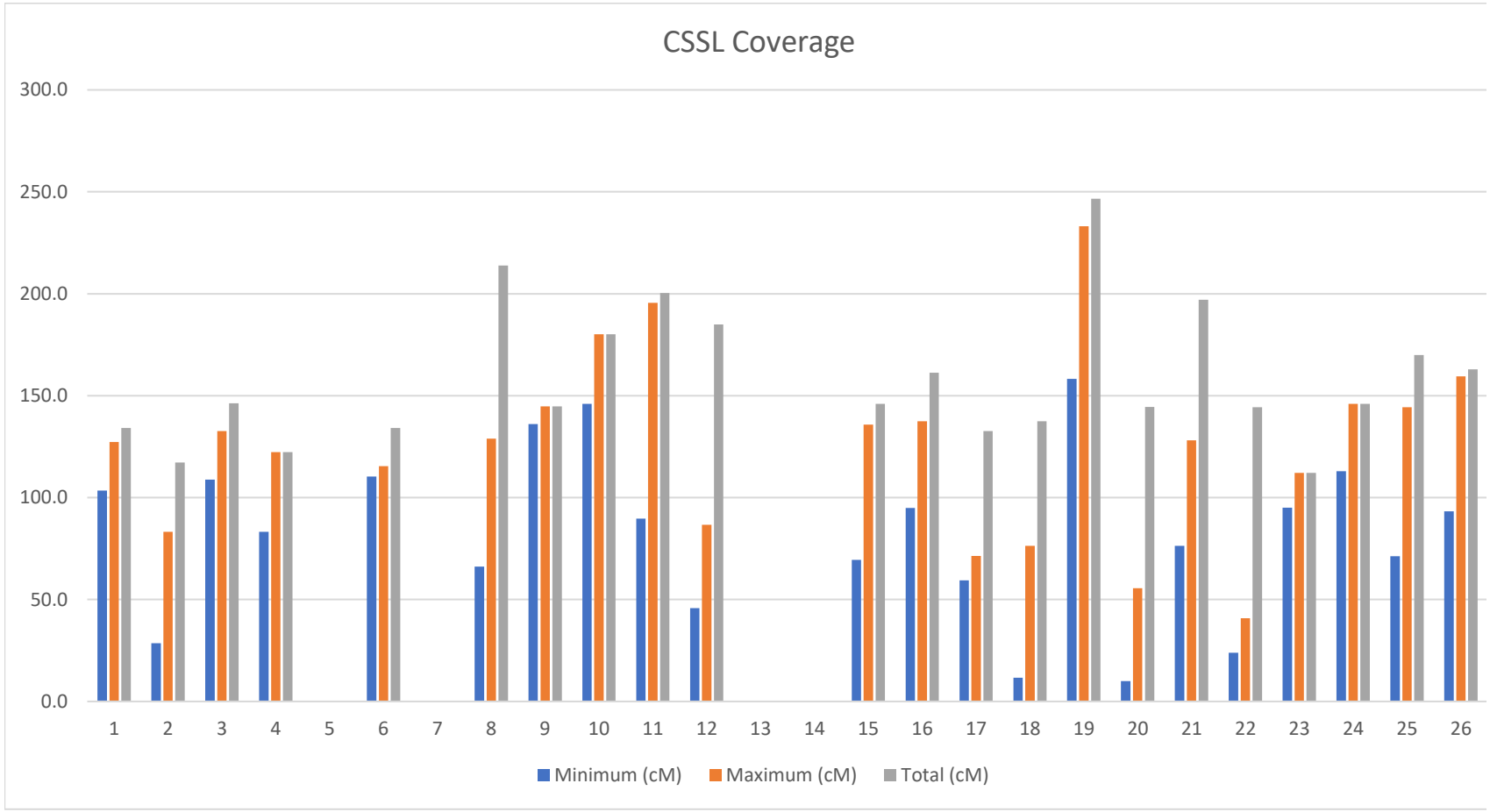


Figure 5 Continued

Table III. Recovered segment-specific SNP groups.

Chromosome: Segment	SNP	cM
1:1	i61225Gt	1.7
	i23982Gh	18.7
1:2	i23982Gh	18.7
	i51041Gb	39.0
	i27626Gh	49.2
	i34731Gh	62.8
	i45642Gh	66.2
1:3	i51041Gb	39.0
	i27626Gh	49.2
	i34731Gh	62.8
	i45642Gh	66.2
	i47972Gh	69.6
	i02391Gh	98.5
	i50196Gb	105.2
2:1	i52423Gb	20.4
2:2	i01044Gh	45.9
	i43496Gh	52.7
	i61080Gt	59.5
2:3	i63139Gt	105.3
	i38489Gh	110.4
3:1	i05735Gh	8.5
3:2	i53555Gb	28.8
	i00178Gh	54.4
	i05459Gh	61.1
	i14041Gh	76.5
	i14878Gh	88.4
	i52480Gb	91.8
	i14000Gh	117.3
	i43612Gh	132.6
4:1	i25377Gh	25.5
4:2	i25377Gh	25.5
	i49276Gh	34.0
	i29272Gh	54.4
	i32552Gh	61.2
	i47058Gh	62.9
	i52022Gb	91.7
4:3	i52022Gb	91.7
	i53431Gb	108.7
6:1	i11399Gh	20.4
	i11312Gh	42.5
	i39865Gh	61.1
	i10945Gh	86.6

Chromosome: Segment	SNP	cM
6:2	i55732Gb	64.5
	i10945Gh	86.6
	i39178Gh	90.0
	i62896Gt	113.7
	i52154Gb	134.1
8:1	i30796Gh	50.9
	i31486Gh	66.2
8:2	i61072Gt	163.0
	i61692Gt	213.9
9:1	i52286Gb	1.7
	i41274Gh	17.0
	i06016Gh	39.2
9:2	i06016Gh	39.2
	i06087Gh	57.9
	i50477Gb	78.2
	i51975Gb	120.8
9:3	i51975Gb	120.8
	i51086Gb	136.1
10:1	i12245Gh	34.1
	i17601Gh	46.0
10:2	i12245Gh	34.1
	i17601Gh	46.0
	i51066Gb	64.7
10:3	i51066Gb	64.7
	i00406Gh	81.6
	i11901Gh	91.8
	i48101Gh	100.3
10:4	i39597Gh	154.6
	i50887Gb	176.7
10:5	i12245Gh	34.1
	i17601Gh	46.0
	i51066Gb	64.7
	i00406Gh	81.6
	i11901Gh	91.8
	i48101Gh	100.3
	i39597Gh	154.6
11:1	i00366Gh	5.2
11:2	i52531Gb	62.9
	i52034Gb	103.6
	i03278Gh	117.2
	i52558Gb	120.6
	i57155Gb	122.2
	i19367Gh	142.6

Table III Continued

Chromosome: Segment	SNP	cM
11:3	i29860Gh	191.9
	i50720Gb	191.9
12:1	i49602Gh	28.9
	i16251Gh	37.3
	i07954Gh	47.5
	i52800Gb	62.8
12:2	i52800Gb	62.8
	i50197Gb	73.0
15:1	i02942Gh	20.4
15:2	i50860Gb	40.7
	i02862Gh	42.4
	i14720Gh	52.6
	i50187Gb	57.7
	i02576Gh	90.0
	i35660Gh	93.4
	i18386Gh	105.2
16:1	i51425Gb	1.7
	i43172Gh	13.6
16:2	i47062Gh	32.3
	i58337Gb	39.1
	i24697Gh	54.3
16:3	i44284Gh	78.1
	i01703Gh	96.8
	i14326Gh	134.1
16:4	i51754Gb	161.3
17:1	i53982Gb	0.0
17:2	i03371Gh	71.4
	i63519Gm	93.5
	i39989Gh	95.2
	i60810Gt	125.8
18:1	i13068Gh	18.7
	i34219Gh	25.4
18:2	i50412Gb	83.1
19:1	i10421Gh	25.4
	i26414Gh	58.1
	i10126Gh	58.1
19:2	i10421Gh	25.4
	i26414Gh	58.1
	i10126Gh	58.1
	i45829Gh	73.3
	i09898Gh	88.6
	i09626Gh	115.8

Chromosome: Segment	SNP	cM
19:3	i09626Gh	115.8
	i09526Gh	131.1
	i09244Gh	178.7
19:4	i08803Gh	246.6
20:1	i17673Gh	0.0
20:2	i11470Gh	144.5
21:1	i50608Gb	1.7
21:2	i52139Gb	39.0
	i50923Gb	71.3
	i16137Gh	156.3
21:3	i16137Gh	156.3
	i52000Gb	195.4
22:1	i12895Gh	120.5
	i12939Gh	144.4
23:1	i35476Gh	23.8
	i61300Gt	35.7
23:2	i05794Gh	0.0
	i35476Gh	23.8
	i61300Gt	35.7
	i15725Gh	49.3
	i62777Gt	57.8
	i23862Gh	73.1
23:3	i15725Gh	49.3
	i62777Gt	57.8
	i23862Gh	73.1
	i34362Gh	90.0
23:4	i06632Gh	108.7
24:1	i54758Gb	0.0
24:2	i04738Gh	23.8
24:3	i35512Gh	32.3
24:4	i04615Gh	39.1
	i49724Gh	66.2
	i04257Gh	91.7
	i50008Gb	103.6
	i03855Gh	130.7
	i03844Gh	132.4
24:5	i03844Gh	132.4
	i50055Gb	146.0

Table III Continued

Chromosome: Segment	SNP	cM
25:1	i53437Gb	25.5
	i11316Gh	35.6
25:2	i33715Gh	64.5
	i17265Gh	69.6
25:3	i17265Gh	69.6
	i52068Gb	93.4
25:4	i10803Gh	137.6
	i17168Gh	141.0
	i60789Gt	166.5
25:5	i60789Gt	166.5
26:1	i40991Gh	1.7
26:2	i59951Gb	34.0
	i50506Gb	54.3
	i47604Gh	73.0
26:4	i59951Gb	34.0
	i47604Gh	73.0
	i08115Gh	84.9
	i16420Gh	118.8
	i55509Gb	125.6

Comparison of Tentative CSSL Coverage to CS Coverage

The advent of facile molecular genotyping has enabled chromosome segment substitution by marker-based backcross-inbred development; it also can be used to facilitate hypoaneuploid-based chromosome substitution. Prior to marker-assisted selection, the process of chromosome substitution by cytogenetic manipulation of cotton was largely based on plant phenotyping and meiotic metaphase I analysis. Chromosome substitution (CS) lines provide an additional format for dissection and discovery of wild germplasm. A partial set of CS lines has previously been developed using *G. mustelinum* (CS-M). Although still in development, the CS-M panel has a majority of chromosomes introgressed. The current panel of CS-M contains all or parts of 19 of 26 chromosomes (Table I).

CS and CSS panels are conceptually similar in that piecemeal introgression of the donor species' genome is sought across multiple unique lines, but CS and CSS differ in the target size of introgressions and population size. Existing CS lines are mainly of two types -- have either a complete chromosome substitution or a partial chromosome (~arm) substitution; most are derived by repeated backcrossing to isogenic or quasi-isogenic one monosomic and one ("short" or "long" arm) or both ("short" *and* "long" arm) acrocentric ("monotelodisomic") Upland cotton cytogenetic stocks. Upon completion, BC_5S_n CS lines are high-density genotyped to confirm homozygosity of the targeted disomic substitution, and to detect inadvertent substitutions in other chromosomes, and their status (homozygous or heterozygous). For CSSLs, coverage of a given chromosomes would typically involve multiple isogenic lines, each homozygous for one or more “random” donor segments. For similar reasons, CSS lines should also similarly genotyped and characterized, i.e., at completion, e.g., after marker-based selection at the BC5Sn.

Available and genotyped CS (BC_5S_n) and CSS (BC_5F_1) were compared on a chromosome-specific basis (Table I). Chromosome 1 is fully substituted in the CS-M panel, while approximately 82% of *G. mustelinum* is accounted for between three introgressions. Chromosome 2 is fully substituted in the CS-M panel and approximately 37% of is accounted for between three individuals. No part of chromosome 3 is currently represented in the CS-M panel, however 83% of is accounted for on within the CSSL panel between two individuals. Chromosome 4 is fully introgressed in the CS-M panel and approximately 88% is accounted for between three individuals within the CSSL panel. Chromosome 5 is represented by a short arm introgression in the CS-M panel. The amount of introgression on chromosome 5 in the CSSL panel is unknown due to genotype failure and for introgression purposes was labeled as having 0% coverage. Chromosome 6 is fully introgressed in the CS-M panel, while approximately 86% is accounted for between three individuals in the CSSL panel. Chromosome 7 is fully introgressed in the CS-M panel. The amount of introgression on chromosome 7 in the CSSL panel is unknown due to genotype failure and for introgression purposes was labeled as having 0% coverage. The short arm of chromosome 8 is introgressed in the CS-M panel, and approximately 34% of is accounted for on chromosome 8 between two individuals. A F_1 has been created with the entire introgression of chromosome 9 in the CS-M panel, the CSSL panel also has approximately 100% of chromosome 9 introgressed between three individuals. Chromosome 10 is also entirely introgressed in the CS-M and the CSSL panel. A F_1 has been created with a long arm introgression of chromosome 11 in the CS-M panel, and approximately 59% is accounted for in the CSSL panel between three individuals. The CS-M panel lacks introgression for both chromosomes 12 and 13. Approximately 39% of chromosome 12 is represented in the CSSL panel, chromosome 13 is not accounted for due to genotyping failure.

Chromosome 14 is also not accounted for in the CSSL panel, but is represented in the CS-M panel with a long arm introgression. Chromosome 15 is accounted for in the CS-M with a long arm introgression, and approximately 53% is represented in the CSSL panel. Chromosomes 16, 17, and 18 are fully introgressed in the CS-M panel, with approximately 67%, 53%, and 21%, respectively, represented in the CSSL panel. Chromosome 19 is not accounted for in the CS-M panel, but approximately 78% is represented in the CSSL panel. Chromosome 20 is represented by a long arm introgression in the CS-M panel, while approximately 30% is accounted for in the CSSL panel. Chromosome 21 is not accounted for in the CS-M panel but is accounted for in the CSSL panel with approximately 54% introgression. Chromosome 22 is fully or nearly fully introgressed in the CS-M population between two lines, one short arm introgression and one long arm introgression. Only approximately 13% of chromosome 22 is accounted for in the CSSL panel. Chromosomes 23 and 24 lack introgressions in the CS-M panel, but are approximately 89% and 100%, respectfully, represented in the CSSL panel. A F_1 has been created to establish a full substitution of chromosome 25 in the CS-M panel. Approximately 52% of chromosome 25 is accounted for in the CSSL panel. Chromosome 26 is accounted for as a long arm introgression in the CS-M panel, and approximately 64% is accounted for in the CSSL panel.

Advancing Program to BC_5S_1

A population of BC_4F_1 (*G. mustelinum* x *G. hirsutum*) plants was grown in College Station, Texas in the summer of 2017. BC_4F_1 individuals were backcrossed to *G. hirsutum* Texas-Marker 1 (TM-1) without regard to BC_4F_1 genotype. BC_4F_1 individuals were emasculated and used as females. In the evenings, candles were emasculated using forceps by first removing the petals, then by carefully removing all anthers to prevent self-pollination. The style was then covered with a straw to prevent unwanted crosses. The candles of the male parent, TM-1, were tied shut with a twist tie to prevent contamination of pollen. The following morning, flowers were collected from TM-1 plants to serve as pollen donors. Flowers were untied and used immediately to cross or placed in an ice cube tray with a small amount of water under sun light or a lamp to reach maximum anthesis potential. To create a controlled cross-pollination, the straw covering the stigma and style was temporarily removed, and anthers of the male parent's flower were rubbed against the bare stigma, before replacing the straw. Crosses were tagged with paper tags immediately after pollination with plant identification numbers (female x male) and the date. A single pollen donor flower was used to pollinate up to 15 female plants depending on the amount of pollen available in the donor flower. Crosses were harvested over a month as bolls matured. Bolls were harvested individually by putting both the seed cotton and crossing tag in small envelopes. The parents of the cross were later transcribed to the outside of the envelope for ease of sorting. Once all crosses were harvested and dried for up to one week at room temperature, the BC_5F_1 seed were ginned on a roller gin. Ginned BC_5F_1 seed were stored in a cool and dry environment.

Selection of BC_5F_1 seeds were based on BC_4F_1 and BC_2F_1 genotype data. The genotype data were incomplete, so segments that were not represented were traced from specific BC_2F_1 individuals to BC_4F_1 descendants that served as parents to BC_5F_1 seed. A population of BC_5F_1 seeds that covered multiple BC_3F_1 and BC_4F_1 parents was created, and a total of 87 bolls harvested from BC_4F_1 were chosen for planting into progeny rows. Overall, the progeny of the selected 87 BC_4F_1 represented 36 BC_3F_1s and all 18 BC_2F_1s. In addition to the selected progeny, two chromosome substitution lines were selected to recover missing introgression regions on chromosome 11.

In the summer of 2018, 960 BC_5F_1 seedlings were transplanted into research plots in College Station, Texas. After three weeks the final stand count was 941 BC_5F_1 plants. All plants were self-pollinated to create progeny segregating for homozygous *G. mustelinum* introgressions. Plants were self-pollinated by securing a twist tie around the peduncle then around candles with a final secure twist at the tip. This method ensured no unwanted cross-pollination would occur. The twist ties were left on post-pollination and were eventually pushed away as the boll matured. Self-pollinated bolls were easily identified by a twist tie still secured to the peduncle. Bolls were harvested on an individual basis. Seed cotton was collected from each boll and placed into a small envelope labeled with the plant identification number and labeled as self-pollinated. Only 374 plants produced controlled self-pollinated bolls due to boll drop or other environmental factors (e.g. small non-productive plants). The 374 BC_5F_1 plants represented 18 BC_2F_1s, 35 BC_3F_1s and 85 BC_4F_1s. The BC_5S_1 seed was organized by plant identification number and stored in a cool and dry storage room.

Creation of Facile and Amendable Segment Library

To allow for tentative inclusion of segments prospectively covered by CSSLs still in development a library is necessary. Using Microsoft Excel, a graphical library was created to track segments found in the BC_5F_1. This library can easily be updated and referenced in the future. The library contains segments with their respective position on a chromosome, pedigree information, and generation.

The *G. hirsutum x G. mustelinum* linkage map data were imported into Microsoft excel. The SNP name as appearing on the Illumina CottonSNP63K Array and chromosome position (cM) based on GHMv1 map were used to create SNP bins. Bins were created by using one SNP to represent one map position, which was often shared by up to eight or more SNPs in the non-binned map. By creating bins, the number of rows in the excel library was reduced from 15,826 to 1,810 across all 26 chromosomes. The elimination of redundant data for the purposes of a graphical library to track introgression segments was necessary to facilitate an efficient and effective user interface. The detailed SNP data are still available for in-depth analysis or if the bin representative should need to be changed to match a SNP used in simplex assay development. Each chromosome has a separate sheet within the Excel file; within the sheet the SNP bins are listed with the respective map position. If a SNP is accounted for by a usable marker set within the genome-wide simplex assay system the cell of that SNP is green. The introgression segments are represented across different columns by a heavy-weight black outline labeled with the line number or name (Figure 6).

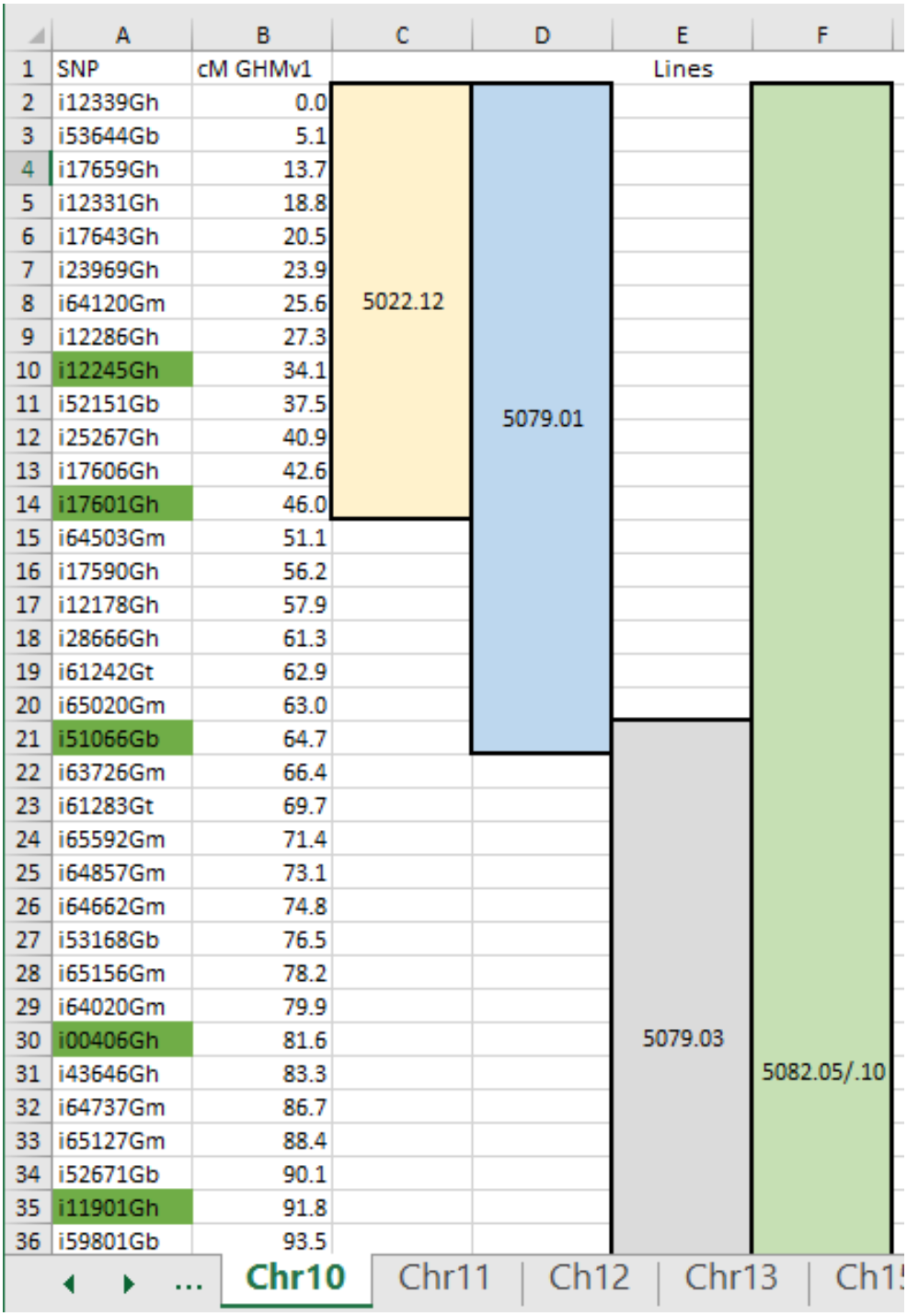


Figure 6. Graphical depiction of spreadsheet (Microsoft Excel) resource for tracking tentative introgression segments and segment-specific assays. Green-highlighted SNPs in column-1 are available as simplex assays. Various segments are represented by differently colored boxes.

The BC_5F_1s were genotyped by simplex assays, however to achieve cost-effectiveness, the resolution of the spaced SNP markers is low. To more accurately resolve the introgression segment(s) and characterize the remainder of the genome in a prospectively homozygous CSS line, high-density genotyping is necessary in the future. The available low-resolution data were used to create probable introgression sizes based on previous high-density genotyping in the BC_2F_1. For instance, a segment may exceed the boundaries of which current simplex SNP markers lie on the map or in the low probability event a double crossover could have occurred between simplex SNP markers. For the former, if the BC_2F_1 showed the segment extended to a position not covered by SNP the segment was extended in the tentative library to that map position. The likelihood of a crossover between SNPs spaced 10-15cM apart is low but possible. If free of crossover interference, a double crossover between two SNPs 10cM apart has the probability of $f(.10^2) = f(0.01)$; it would be lower in segments subjected to positive crossover interference exists. While the segments within the CSS library are likely correct, they must be considered tentative and must be confirmed and defined more accurately as to absolute segment length and position.

In addition to the graphical representation of tentative segments, a pedigree is included for each line listed in the CSSL panel on a separate sheet that can be indexed and sorted (Figure 7). This library can easily be shared due to the availability of Microsoft software and related cloud programs. As generations progress the graphical representations of segments and pedigree information can readily and effortlessly be changed to reflect the most recent data.

	A	B	E	F	G	H
1	Line	Latest Generation	BC5F1	BC4F1	BC3F1	BC2F1
2	1.1	BC5S1	125.08-18	1700346.19	1400541.13	641.16-13
3	1.1	BC5S1	125.08-18	1700346.19	1400541.13	641.16-13
4	1.2	BC5S1	111.09-18	1700352.09	1400542.07	642.17-13
5	1.3	BC5S1	111.01-18	1700352.07	1400542.07	642.17-13
6	2.1	BC5S1	126.08-18	1700347.11	1400541.15	641.16-13
7	2.2	BC5S1	118.11-18	1700431.03	1400544.05	645.01-13
8	2.2	BC5S1	118.09-18	1700431.03	1400544.05	645.01-13
9	2.3	BC5S1	127.17-18	1700348.11	1400542.01	642.11-13
10	2.3	BC5S1	127.18-18	1700348.11	1400542.01	642.11-13
11	3.1	BC5S1	125.08-18	1700346.15	1400541.13	641.16-13
12	3.2	BC5S1	124.02-18	1700345.17	1400541.11	641.16-13
13	4.1	BC5S1	111.09-18	1700352.09	1400542.07	642.17-13
14	4.1	BC5S1	164.17-18	1700352.09	1400542.07	642.17-13
15	4.2	BC5S1	137.20-18	1700363.15	1400543.13	643.17-13
16	4.3	BC5S1	118.01-18	1700366.17	1400544.01	644.11-13
17	6.1	BC5S1	113.12-18	1700347.15	1400541.17	642.11-13
18	6.2	BC5S1	112.14-18	1700343.15	1400541.07	641.03-13
19	8.1	BC5S1	152.11-18	1700443.05	1400545.09	648.07-13
20	8.1	BC5S1	151.15-18	1700442.11	1400545.09	648.07-13

... | Chr17 | Chr18 | Ch19 | Chr20 | Chr21 | Chr22 | Chr23 | Chr24 | Chr25 | Chr26 | Pedigree

Figure 7. Partial display of spreadsheet that describes pedigreed lineages of chromosome segment substitution (CSS) library. Spreadsheet describes advanced BC_5S_1 families, BC_5F_1 parents and pedigrees through BC_2F_1 (each BC_2F_1 plant was high-density genotyped with CottonSNP63K Array).

Plan for Genome Recovery

The power of the CSSLs derives from several factors, one being the representation of the entire genome; others include isogenicity, association of each CSSL donor segment SNP-based markers and high amenability to replicated experimentation. The total genome coverage of *G. mustelinum* introgression across all BC_5F_1 lines was approximately 51% including chromosomes with missing genotype data. Excluding missing genotype data, the total genome coverage of *G. mustelinum* introgression across all BC_5F_1 lines was approximately 62%. Recovering the approximately 49% of missing introgressions is integral to the completion of the CSSL panel. The following will detail plans for future researchers to recover the missing genome.

Incomplete introgression occurred on chromosomes 2, 5, 7, 8, 12, 13, 14, 18, 20, and 22 due to technical errors or insufficient selection of progeny. A common error that rendered PCR reactions incapable of producing genotype calls was contaminated DNA determined by 260/280 and 260/230 ratios. The contaminants were assumed to be polyphenols which are common in cotton tissue; polyphenols interact with the chemistry and DNA within the PCR and suppress target amplification. It is also possible a genotype was not carried into the current generation and therefore the genotype was not recovered.

To rectify the missing segments, a reverse pedigree search was used to find seed that may contain the missing segments. Searches were initiated from the BC_2F_1 genotype data to find potential lines that would contain the segments being sought. Once one or two candidate BC_2F_1 plants were selected as a prospective source to recover a given segment, a list the BC_3F_1, BC_4F_1, and BC_5F_1 descendants was compiled. The lists were then used to select the most current generation of potentially useful seeds, while also striving to sample diversely across BC_3F_1

parents. When considering potential selections with low probability of containing any specifically targeted introgression segment, genotype data on earlier generations were also used to reduce collective redundancy. This initial introgression recovery focuses on the chromosomes with the largest gaps in introgression. Later recovery seeking to fill smaller gaps can be pursued based on high-density genotyping that shows conclusive sizes of introgressions. The chromosomes selected below were chosen due to the obviousness of missing introgression as determined by low-density simplex genotyping.

Based on presence and absence of spaced SNPs, BC_5F_1 coverage of chromosome two was estimate to be between 24 to 71% (minimum to maximum) of the BC_1F_1 linkage map. Without additional genotyping to more finely resolve donor segment lengths, efforts to complete coverage will thus require identifying a small number of BC5 individuals that collectively containing 29 to 76% of the chromosome. To complete coverage of chromosome two, four BC_5F_1 bolls were chosen from the BC_2F_1 parents 1300642.16 and 1300641.11 which had complete or partial coverage for chromosome two. The four selections each came from a unique BC_3F_1 parent to maximize potential to recover the missing segments; 1400541.11, 1400541.13, 1400541.15, and 1400541.17.

To complete coverage of chromosome 5, four BC_5F_1 bolls were chosen from the BC_2F_1 parent 1300643.17, which had near complete coverage. The BC_2F_1 parent was only represented in the pedigree by only one BC_3F_1, plant 1400543.13. To complete coverage of chromosome 7, six BC_5F_1 bolls were chosen from two BC_2F_1 parents 1300646.05 and 1300644.11. The six selections were progeny of two BC_3F_1 parents, 1400544.05 and 1400544.01. To complete coverage for chromosome 8, four BC_5F_1 bolls were chosen from one BC_2F_1 parent, 1300645.01, and three BC_3F_1 parents: 1400544.03, 1400544.05, and 1400544.07. To complete coverage for

chromosome 12, three BC_5F_1 bolls were chosen from one BC_2F_1 parent, 1300643.02, and three BC_3F_1 parents: 1400542.13, 1400542.15, and 1400542.17. To complete coverage of chromosome 13, one BC_5F_1 boll was chosen from one BC_2F_1 parents. To complete coverage for chromosome 14, ten BC_5S_1 bolls were chosen from one BC_2F_1 parent, 1300645.01. The BC_5S_1 bolls were progeny of four BC_3F_1 parents: 1400544.03, 1400544.05, 1400544.07, and 1400544.09. To complete coverage for chromosome 18, eight BC_5F_1 bolls were chosen from one BC_2F_1, 1300643.20. The eight BC_5F_1 bolls were progeny of one BC_3F_1, 1400543.19. To complete coverage for chromosome 20, two BC_5F_1 bolls were chosen from one BC_2F_1, 1300647.06. The two BC_5F_1 bolls were progeny of two BC_3F_1 parents: 1400545.01 and 1400545.03. To complete coverage for chromosome 22, three BC_5F_1 bolls were chosen from three BC_2F_1 parents: 1300642.17, 1300643.20, and 1300646.13. The three selections represented three BC_3F_1 plants: 1400542.11, 1400543.19, and 1400544.19 (Table IV).

The selections above will provide a calculated starting point for genotyping and segment discovery based on pedigree and existing genotype information. If a segment is not found within a certain set of selections specific for one chromosome, it would be beneficial to screen selections made for other chromosomes that share the same progenitors. Many selections share progenitors and should be screened with simplex assays for multiple chromosomes. CSSLs selected with simplex SNP assays can be more accurately characterized using a genome-wide high-density genotyping platform such as the CottonSNP63K to provide robust data related to elucidate accurate sizes and location introgression segments; but the relatively high expense of genotyping many samples may necessitate use of alternative genotyping methods, even if less informative. A segment may be larger or smaller than what can be inferred from a lower density

genotyping method such as PACE simplex assays. More or less screening may be needed one current introgression is completely elucidated and recovered in a homozygous state.

Table IV. Pedigree for introgression recovery. Lines are named Recovery 'Rec' followed by chromosome number.

Line	BC_5F_1	BC_4F_1	BC_3F_1	BC_2F_1
Rec_02		1700345.11	1400541.11	1300641.16
Rec_02		1700346.11	1400541.13	1300641.16
Rec_02		1700347.13	1400541.15	1300641.16
Rec_02		1700348.05	1400541.17	1300642.11
Rec_05		1700363.11	1400543.13	1300643.17
Rec_05		1700363.03	1400543.13	1300643.17
Rec_05		1700363.13	1400543.13	1300643.17
Rec_05		1700363.17	1400543.13	1300643.17
Rec_07		1700432.17	1400544.11	1300646.05
Rec_07		1700368.11	1400544.01	1300644.11
Rec_07		1700368.05	1400544.01	1300644.11
Rec_07		1700366.19	1400544.01	1300644.11
Rec_07		1700366.15	1400544.01	1300644.11
Rec_07		1700366.11	1400544.01	1300644.11
Rec_08		1700432.11	1400544.05	1300645.01
Rec_08		1700431.17	1400544.07	1300645.01
Rec_08		1700431.03	1400544.05	1300645.01
Rec_08		1700368.13	1400544.03	1300645.01
Rec_12		1700354.09	1400542.13	1300643.02
Rec_12		1700354.15	1400542.15	1300643.02
Rec_12		1700355.13	1400542.17	1300643.02
Rec_13		1700366.17	1400544.01	1300644.11
Rec_14	1800144.04	1700368.13	1400544.03	1300645.01
Rec_14	1800145.06	1700431.09	1400544.05	1300645.01
Rec_14	1800144.19	1700368.17	1400544.03	1300645.01
Rec_14	1800146.01	1700432.07	1400544.09	1300645.01
Rec_14	1800146.02	1700432.07	1400544.09	1300645.01
Rec_14	1800146.04	1700432.07	1400544.09	1300645.01
Rec_14	1800146.05	1700432.07	1400544.09	1300645.01
Rec_14	1800146.07	1700432.07	1400544.09	1300645.01

Table IV Continued

Line	BC_5F_1	BC_4F_1	BC_3F_1	BC_2F_1
Rec_14	1800164.02	1700325.03	1400544.07	1300645.01
Rec_14	1800164.03	1700325.03	1400544.07	1300645.01
Rec_18		1700365.13	1400543.19	1300643.20
Rec_18		1700365.19	1400543.19	1300643.20
Rec_18		1700365.03	1400543.19	1300643.20
Rec_18		1700365.01	1400543.19	1300643.20
Rec_20		1700437.15	1400545.01	1300647.06
Rec_20		1700438.17	1400545.03	1300647.06
Rec_22		1700353.07	1400542.11	1300642.17
Rec_22		1700364.19	1400543.19	1300643.20
Rec_22		1700435.09	1400544.19	1300646.13

CHAPTER III

FIBER ANALYSIS OF EARLY BACKCROSS GENERATIONS

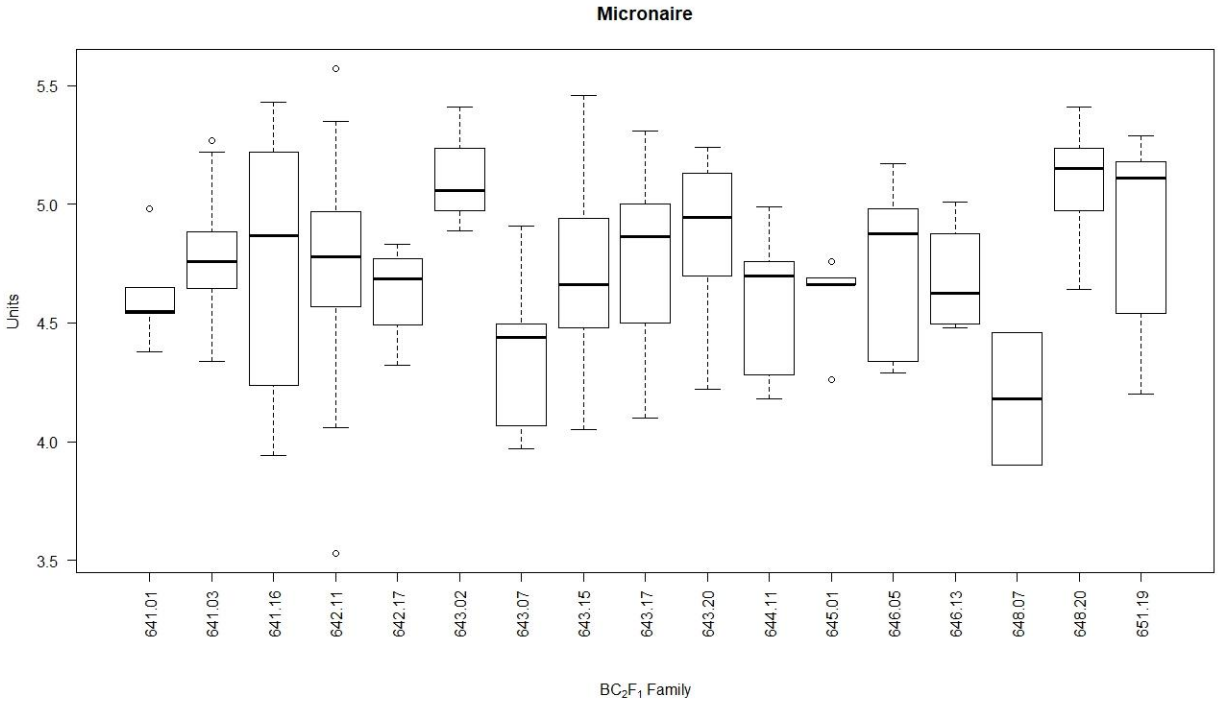
Fiber is the most important product produced by cultivated cotton plants and therefore improvement should be a continual process. Although BC_5F_1 plants were grown in 2018 mainly to advance breeding objectives, the fiber of BC_5F_1 individuals was assessed to elucidate possible individual and family genetic effects of *G. mustelinum* introgressions. Seedcotton was collected manually on a single-plant basis from all bolls from the bottom 3 to 4 branches and transferred to a paper bag with each plant's field identification number. A total of 228 plants were sampled from the breeding plots in the summer of 2018 which represented 17 BC_2F_1-derived families. Individual seedcotton samples were ginned on a single laboratory saw gin. Fiber samples were analyzed using High Volume Instrument (HVI) at the Cotton, Inc. Fiber Processing Lab in Cary, North Carolina. A total of 210 samples had sufficient lint to obtain fiber quality data.

Statistical analyses of the fiber data were conducted using JMP Pro 14 (SAS Institute Inc.). Analysis of variance (ANOVA) indicated differences among BC_2F_1 families of BC_5F_1 individuals: Micronaire (Mic) ($p = 0.0342$), upper half mean length (UHM) ($p = 0.0004$), strength (Str) ($p = 0.0224$), and elongation (Elo) ($p = 0.0253$) (Figure 8, Figure 9). Box-plots graphically revealed differences in variation among BC_2F_1-derived families from individual BC_5F_1 plants. Mic varied across all samples, from 3.5 to 5.6, with a median of 4.8. UHM ranged about 5.1 mm, from 24.1 mm to 29 mm with a median of 26.4 mm. Uniformity index (UI) ranged from 81.0 to 84.7 with a median of 82.45. Strength (Str) ranged from 231.3 to 305.8

kN m kg^{-1} with a median of 263.6 kN m kg^{-1}. Elongation (Elo) ranged from 7% to 8.6% with a median of 8.1%. Short fiber index (SFI) ranged from 7.7 to 9.4 with a median of 8.1.

The BC5F1 families and plants in this study were grown in a single field plot and single year and field positions were not randomized; related plants of a given family were generally grown in adjacent or nearby hills, and therefore field positional effects and other sources of error were confounded with individual and family genetic effects. Another important consideration is that any direct or interaction genetic effect due to *G. mustelinum* introgressions would be due to dominant or co-dominant effects, not recessive ones, i.e., because all donor alleles would have been heterozygous, not homozygous. Although the HVI data do not provide a firm basis for conclusions, the observed trait variations certainly do not discount the possibility of effects by genetic variation from the donor, nor the possibility of future improvements from introgression. The gene effects seen in early generation backcrosses are not necessarily representative of the final effects because of their heterozygous state. Further testing of lines homozygous for introgressions will be facilitated by ease of the seed increases needed for replicated experimentation, and include the ability to measure both direct and interaction effects of recessive donor alleles.

A.



B.

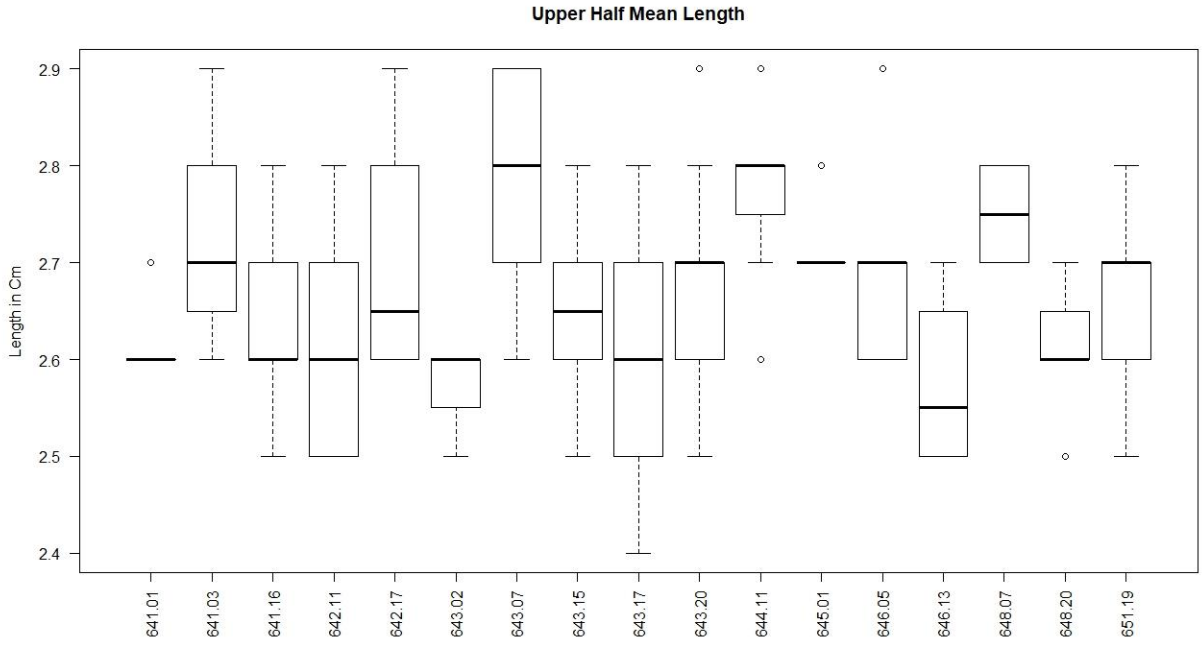
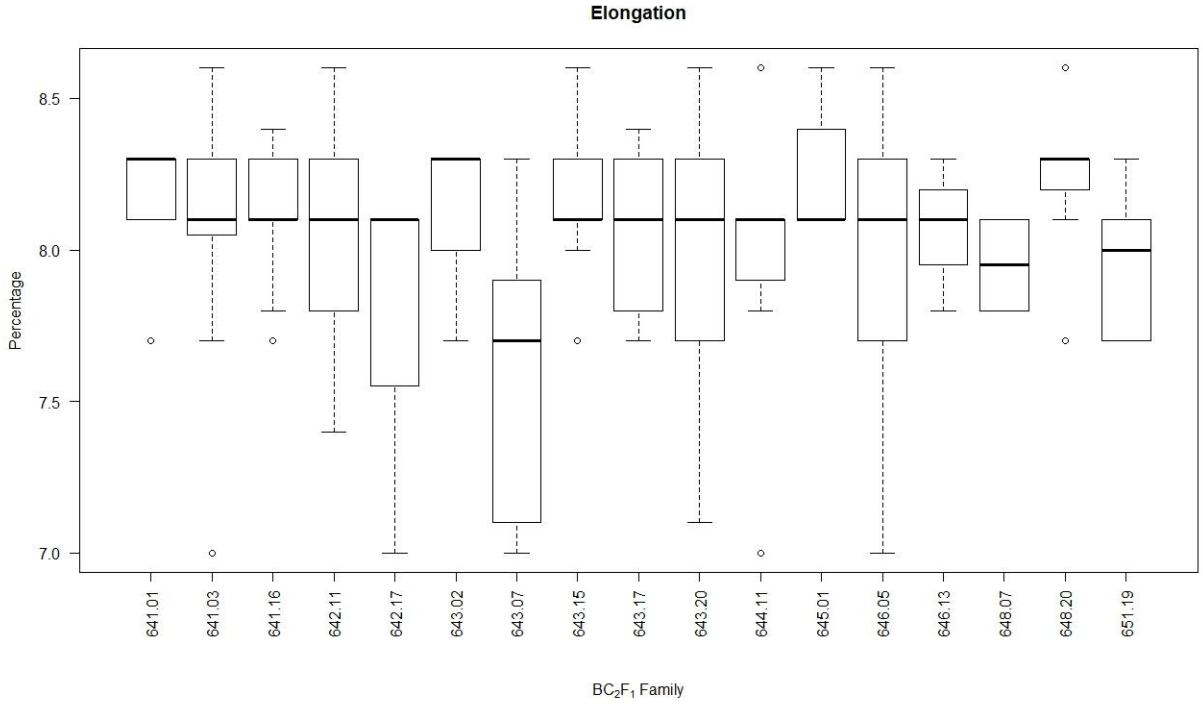


Figure 8. Box-plot distributions of fiber quality variation across BC_2F_1-derived BC_5F_1s. A. Micronaire (Mic) (ANOVA p = 0.0342), B. Upper half mean length (UHM) variances (ANOVA p = 0.0004), C. Elongation (Elo) variances (ANOVA p = 0.0253), D. Strength (Str) variances (ANOVA p = 0.0224).

C.



D.

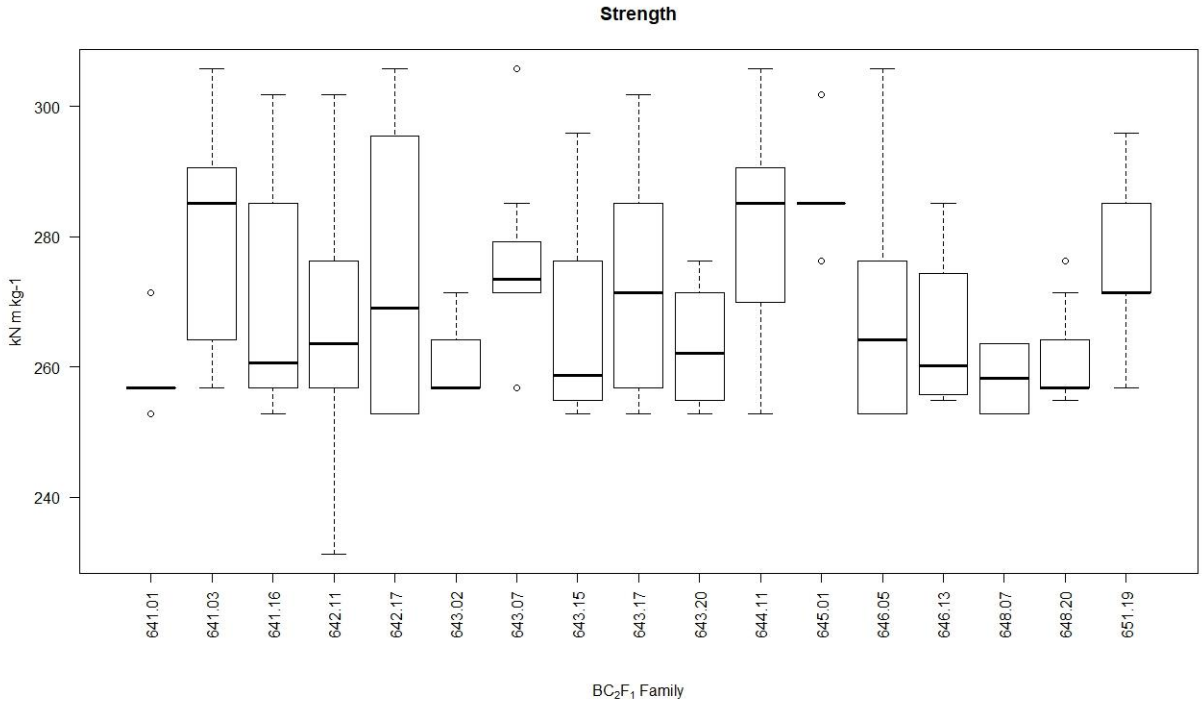
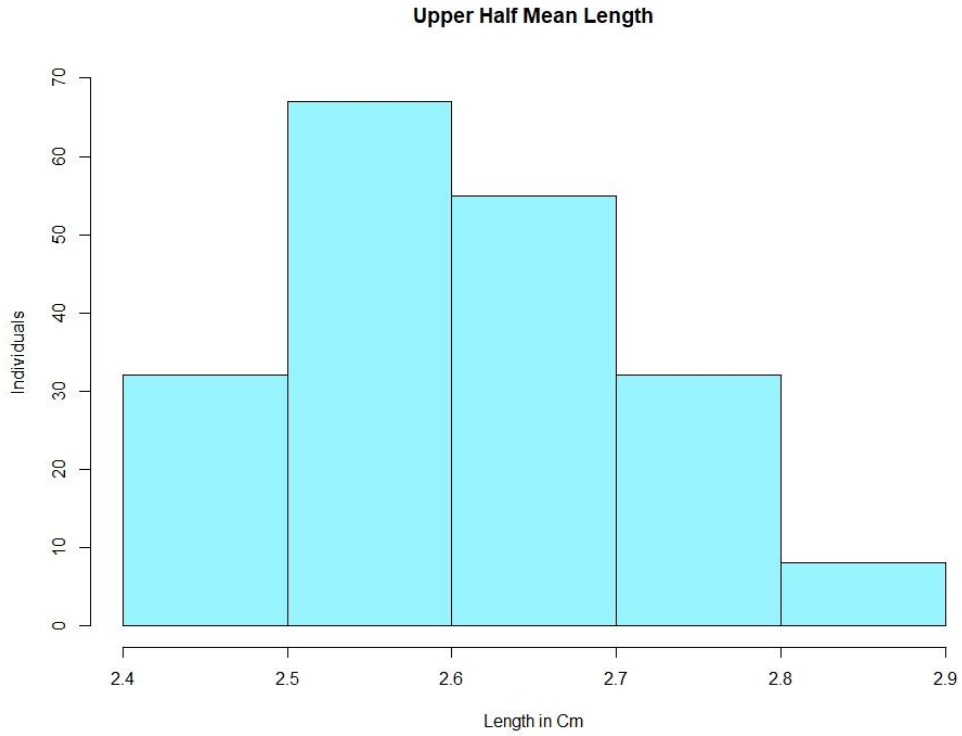


Figure 8 Continued

A.



B.

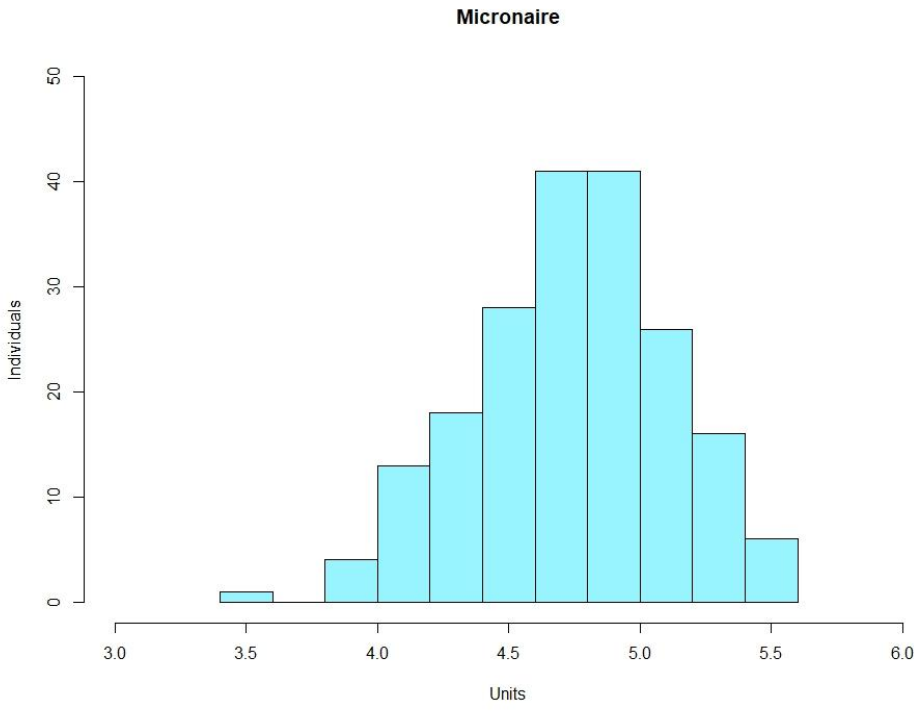
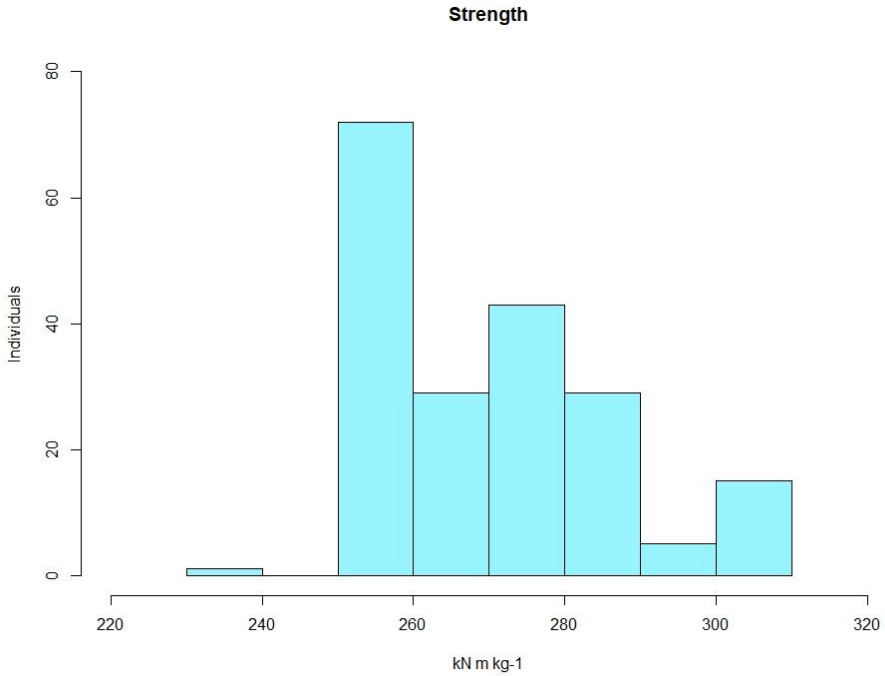


Figure 9. Histogram of HVI fiber trait distributions for seedcotton samples from individual fiber BC_5F_1 plants. Distributions across all samples for (A) upper half mean length, (B) micronaire, (C) strength, and (D) elongation.

C.



D.

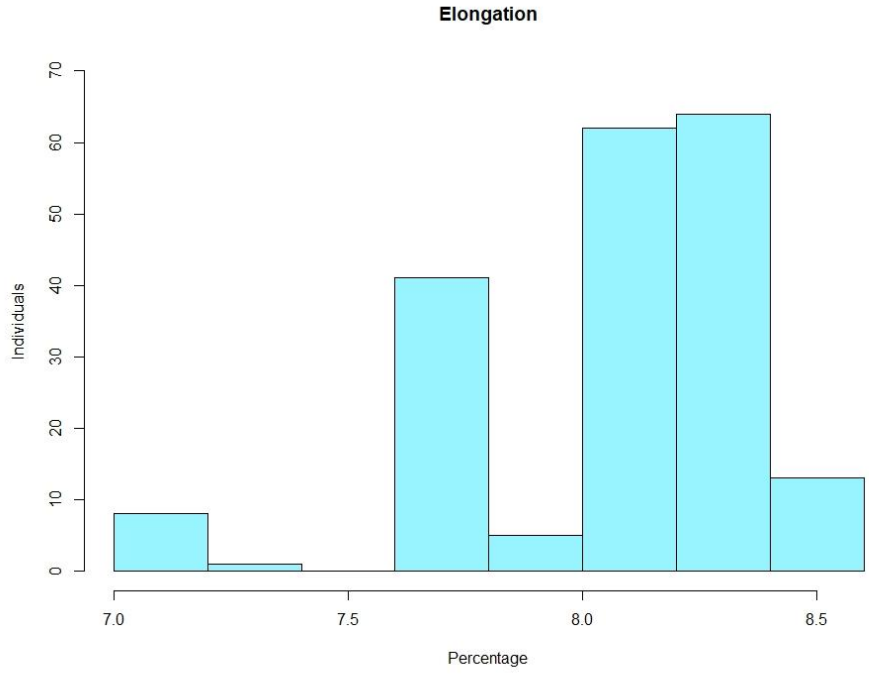


Figure 9 Continued

CHAPTER IV

CONCLUSION AND DISCUSSION

Chromosome segment substitution lines facilitate the methodical analysis of a wild genome in the search for beneficial genetic diversity that will increase the rate of improvement in the currently plateau-like state of current cultivated cotton improvement (*Gossypium hirsutum* L.). Both CSSLs and CS lines constitute substantial research resources for the discovery of genetic variants that exert significant beneficial effects in the genetic background of the cultivated species. The power of such lines is augmented by their isogenicity, facilitating direct comparisons. Genetic analyses can be expanded by incorporating methodical introgression from multiple species, creating intercrosses, topcrosses, RILs and other types of family or population structures.

Introgressing multiple species into organized donor-specific panels that share the same genetic background enhances the range of potential applications of the isogenic platform. The initial searches for phenotypic variation in a panel of CSSL or CS lines can identify variance within the panel and identify overall effects of individual segments or chromosomes, which can be further localized and genetically dissected. Furthermore, due to the isogenic background, crosses can be made between CSSL and/or CS panels to incorporate multiple introgressions from the same or multiple species to observe the effects of interactions between introgressions while reducing the effects of genetic background noise. The potential for interactions can be expanded through more complex intermatings -- planned or random. Isogenicity of the platform enables the creation of families and populations that uniquely enable robust statistical analyses that would

otherwise be difficult or impossible to achieve. CSSL panels can also be used to create recombinant inbred lines (RILs) or similar population studies, or used in conjunction with them.

A common criticism of these CSSL panels is the genetic background in which they are developed, i.e., Texas Marker-1 (TM-1), the recurrent parent. TM-1 is a multiple generation selfed line derived from an obsolete cultivar that is useful as a genetic standard. Converting all past and new CS and CSSLs to a newer cultivar standard would require years, by which time the new standard would also be obsolete. Moreover, the main purposes are to introgress, discover desirable genetic variants, and render them amenable to marker-based breeding. Additional advantages of TM-1 *G. hirsutum* are that multiple high-density maps and quality sequence assemblies are available.

A significant limitation of TM-1 is that it is a single genotype, and donor alleles may interact differently with various Upland cotton genotypes, i.e., a beneficial effect from an introgression may not have the same effect if moved into today's elite cotton cultivars, and some positive effects with other backgrounds will be missed because they may not occur in the TM-1 background. Studies using multiple elite cultivars top-crossed to an individual or multiple CS lines that have shown positive trait improvement and elucidated epistatic interactions within different genetic backgrounds and for overall improvement. Overall, the isogenic platform is a powerful tool for trait discovery, examining epistatic interaction between species, population structure studies, and a multitude of other possibilities.

To further develop the *G. mustelinum* CSSL panel described herein, additional selfing and genotyping will be needed to create and select the best lines, and the library of segments will have to be updated at each stage where finalized lines are created and high-density genotype data are created for detailed characterization. The current prediction of segment size and therefore

total introgression percentage are approximate due to reliance on the low-density genotyping at the BC_5F_1 generation. The actual amount of *G. mustelinum* that is introgressed could be greater than what is discussed in previous chapters due to the inability to define the exact proximal and distal points of the introgression -- each end of an introgression segment either lies between two SNP assay markers, or is distal to the most distal SNP marker used at that end of the chromosome. This inaccuracy can be reduced once final selections of homozygous individuals have been made and genotyped using high-density SNP genotyping methods such as the Illumina Cotton63KSNP array. Upon gathering high-density genotype data, a more accurate introgression percentage can be calculated and any lines that have introgression redundancies can be removed from the population. Any lines that contain a large amount of background introgression can also be backcrossed or selfed depending on the zygosity of the introgression to create one line with less introgression or two lines with separate introgressions.

The final goal is not to create the CSSL population, it is just one step in a path to find beneficial alleles to improve the breadth of alleles available for genetic cotton progression. The next step after completing the panel is phenotyping for all traits important for agronomic, economic, ecological or other reasons. Early generation fiber tests discussed in previous chapters were not conclusive, but they lend hope to the discovery of alleles that confer beneficial effects.

Given the currently accepted limited genetic diversity of the modern elite cotton germplasm pool, the infusion of wild germplasm seems critical for the success of future breeding. The importance of natural diversity is often a talking point echoed by breeders, but the work to successfully dissect and explore the genetic diversity possible in wild cotton remains a desired goal. Wild species are often brushed aside due to the amount of time inherently

associated with creating something commercially viable. It is easy and much "safer" for a breeder to look towards a short-term investment with a higher probability of a relatively quick return on invested time. However, it is equally important to create and expanded the germplasm base that allows those short-term investments to happen by looking towards long-term investments that have a higher risk (e.g. exotic germplasm breeding). Although most of the exotic germplasm can be assumed to be deleterious for agronomic traits, finding one beneficial allele can be worth millions of dollars to the industry's supply chain. The process of creating populations (e.g. CSSL and CS) are laborious and time consuming, but every bit as critical as elite-cultivar breeding programs.

REFERENCES

(NCC), National Cotton Council. Economics of Cotton. Accessed from: https://www.cotton.org/pubs/cottoncounts/fieldtofabric/economics.cfm. Accessed 2018.

Abdurakhmonov, I. Y., M. S. Ayubov, K. A. Ubaydullaeva, Z. T. Buriev, S. E. Shermatov, H. S. Ruziboev, A. E. Pepper. 2016. RNA Interference for Functional Genomics and Improvement of Cotton (Gossypium sp.). Front Plant Science, 7, 202. doi:10.3389/fpls.2016.00202

Alves M.F., P.A. Barroso, A.Y. Ciampi, L.V. Hoffmann, V.C. Azevedo, U. Cavalcante. 2013. Diversity and genetic structure among subpopulations of Gossypium mustelinum (Malvaceae). Genet Mol Res, 12(1), 597-609. doi:10.4238/2013

Beasley, J. O. 1942. Meiotic Chromosome Behavior In Species, Species Hybrids, Haploids, and Induced Polyploids of Gossypium. Genetics, 27(1), 25-54.

Campbell, B.T., V. E. Williams, W. Park. 2009. Using molecular markers and field performance data to characterize the Pee Dee cotton germplasm resources. Euphytica, 169, 285-301. doi: 10.1007/s10681-009-9917-4

Crossa, J., G. Campos, P. Pérez, D. Gianola, J. Burgueño, J.L. Araus, H.J. Braun. 2010. Prediction of Genetic Values of Quantitative Traits in Plant Breeding Using Pedigree and Molecular Markers. Genetics, 186(2), 713. doi:10.1534/genetics.110.118521

Ebitani, T., Y. Takeuchi, Y. Nonoue, T. Yamamoto, K. Takeuchi, M. Yano. 2005. Construction and Evaluation of Chromosome Segment Substitution Lines Carrying Overlapping Chromosome Segments of indica Rice Cultivar in a Genetic Background of japonica Elite Cultivar. Breeding Science, 55(1), 65-73. doi:10.1270/jsbbs.55.65

Eshed, Y., D. Zamir. 1995. An introgression line population of Lycopersicon pennellii in the cultivated tomato enables the identification and fine mapping of yield-associated QTL. Genetics, 141(3), 1147-1162.

Fluidigm Corporation. 2018. SNP Genotyping User Guide (PN 68000098 Q1). Accessed from: https://www.fluidigm.com/binaries/content/assets/fluidigm/snp-gt_analysis_ug_68000098.pdf

Fonceka, D., H.A. Tossim, R. Rivallan, H. Vignes, E. Lacut, F. de Bellis, J.F. Rami. 2012. Construction of Chromosome Segment Substitution Lines in Peanut (Arachis hypogaea L.) Using a Wild Synthetic and QTL Mapping for Plant Morphology. PLOS ONE, 7(11), e48642. doi:10.1371/journal.pone.0048642

Gallagher J.P., C.E. Grover, K. Rex, M. Moran, J.F. Wendel. 2017. A New Species of Cotton from Wake Atoll, Gossypium stephensii (Malvaceae). Systematic Botany, 42(1), 115-123.

Gao, W., L. Long, X. Tian, F. Xu, J. Liu, P.K. Singh, C. Song. 2017. Genome Editing in Cotton with the CRISPR/Cas9 System. Plant Science. 8(1364). doi:10.3389/fpls.2017.01364

Godfray, H.C.J., J.R. Beddington, I.R. Crute, L. Haddad, D. Lawrence, J.F. Muir, C. Toulmin. 2010. Food Security: The Challenge of Feeding 9 Billion People. Science. 327(5967), 812-818. doi:10.1126/science.1185383

Guang C. 2006. Genetic Diversity of Source Germplasm of Upland Cotton in China as Determined by SSR Marker Analysis. Acta Genetica Sinica, 33(8), 733-745.

Gur, A., & D. Zamir, D. 2004. Unused Natural Variation Can Lift Yield Barriers in Plant Breeding. PLOS Biology, 2(10), e245. doi:10.1371/journal.pbio.0020245

He, C., J. Holme, J. Anthony. 2014. SNP Genotyping: The KASP Assay. Crop Breeding, 1145, 75-86.

Holtan, H. E., & S. Hake. 2003. Quantitative trait locus analysis of leaf dissection in tomato using Lycopersicon pennellii segmental introgression lines. Genetics, 165(3), 1541-1550.

Jenkins, J. N., J. Wu, J.C. McCarty, S. Saha, O. Gutierrez, R. Hayes, & D.M. Stelly. 2006. Genetic Effects of Thirteen Gossypium barbadense L. Chromosome Substitution Lines in Topcrosses with Upland Cotton Cultivars: I. Yield and Yield Components. Crop Science, 2006 v.46 no.3(no. 3), pp. 1169-1178. doi:10.2135/cropsci2005.08-0269

Jenkins, J. N., J.C. McCarty, J. Wu, S. Saha, O. Gutierrez, R. Hayes, & D.M. Stelly. 2007. Genetic Effects of Thirteen Gossypium barbadense L. Chromosome Substitution Lines in Topcrosses with Upland Cotton Cultivars: II. Fiber Quality Traits. Crop Science, 47(2), 561-570. doi:10.2135/cropsci2006.06.0396

Jenkins, J. N., J.C. McCarty, R. Wu, J. Jixiang & D. Hayes. 2012. Genetic effects of nine Gossypium barbadense L. chromosome substitution lines in top crosses with five elite Upland cotton G. hirsutum L. cultivars. Euphytica, 2012 v.187, pp. 161-173. doi:10.1007/s10681-011-0580-1

Jenkins, J., B.T. Campbell, R.W. Hayes, J. Wu, S. Saha, D.M. Stelly. 2017. Genetic Effects of Chromosome 1, 4, and 18 from Three Tetraploid Gossypium Species in Topcrosses with Five Elite Cultivars. Crop Science, 57, 1338-1346.

Johnson J., S. MacDonald, L. Meyer, L. Stone. 2018. United States Department of Agriculture, Agricultural Outlook Forum. The World and United States Cotton Outlook.

Kubo, T., Y. Aida, K. Nakamura, H. Tsunematsu, K. Doi, A. Yoshimura. 2002. Reciprocal Chromosome Segment Substitution Series Derived from Japonica and Indica Cross of Rice (Oryza sativa L.). Breeding Science, 52(4), 319-325. doi:10.1270/jsbbs.52.319

Law, C. N., C.F. Young, J.W.S. Brown, J.W. Snape, A.J. Worland. 1978. The study of grain protein control in wheat using whole chromosome substitution lines. International Atomic Energy Agency, Vienna, Austria. 483-502.

Li, X., W. Wang, Z. Wang, K. Li, Y.P. Lim, Z Piao. 2015. Construction of chromosome segment substitution lines enables QTL mapping for flowering and morphological traits in Brassica rapa. Frontiers In Plant Science, 6, 432-432. doi:10.3389/fpls.2015.00432

Lin, Y.-M. 2017. High-Density SNP Genotyping Applied to Interspecific Germplasm in Upland Cotton (Gossypium Hirsutum L.): (I.) CSB17 Chromosome-Specific RIL Analysis and (II.) G. Mustelinum (Miers Ex Watt) Linkage Mapping. Texas A&M University, Retrieved from http://hdl.handle.net/1969.1/173253

Mammadov J., R. Aggarwal, R. Buyyarapu, S. Kumpatla. 2012. SNP Markers and Their Impact on Plant Breeding. International Journal of Plant Genomics. 2012. Article ID 728398. http://dx.doi.org/10.1155/2012/728398

Meredith, W. R., R.R. Bridge. 1971. Breakup of Linkage Blocks in Cotton, Gossypium hirsutum L.1. 11(5), 695-698. doi:10.2135/cropsci1971.0011183X001100050027x

Meredith, W.R. 2000. Cotton Yield Progress - Why Has It Reached a Plateau? Better Crops, 84(4).

Paterson, A. H., Wendel, J. F., Gundlach, H., Guo, H., Jenkins, J., Jin, D., . . . Schmutz, J. (2012). Repeated polyploidization of Gossypium genomes and the evolution of spinnable cotton fibres. Nature, 492, 423. doi:10.1038/nature11798

Pickersgill, B., Spencer, C. H. B., & de Andrade-Lima, D. (1975). Wild Cotton in Northeast Brazil. Biotropica, 7(1), 42-54. doi:10.2307/2989799

Rabinowicz, P.D., R. Citek, M.A. Budiman, A. Nunberg, J.A. Bedell, N. Lakey, A.L. O'Shaughnessy, L.U. Nasciemento, W.R. McCombie, R.A. Marienssen. 2005. Differential methylation of genes and repeats in land plants. Genome Research, 15, 1431-1440. doi:10.1101/gr.4100405

Rafalski, J. A. and S. V. Tingey. 1993. Genetic diagnostics in plant breeding: RAPDs, microsatellites and machines. Trends in Genetics, 9(8), 275-280. doi:https://doi.org/10.1016/0168-9525(93)90013-8

Saha, S., J. Wu, J.N. Jenkins, J.C McCarty, O.A. Gutierrez, D.M. Stelly, R.G. Percy, D.A. Raska. 2004. Effect of chromosome substitutions from Gossypium barbadense L. 3-79 into G. hirsutum L. TM-1 on agronomic and fiber traits. Journal of Cotton Science, 2004 v.8 no.3, pp. 6-9.

Saha, S., J.N. Jenkins, J. Wu, J.C. McCarty, D.M. Stelly. 2008. Genetic analysis of agronomic and fibre traits using four interspecific chromosome substitution lines in cotton. Plant Breeding, 127(6), 612-618. doi:10.1111/j.1439-0523.2008.01532.x

Saha, S., J. Wu, J.N. Jenkins, J.C. McCarty, R. Hayes, D.M. Stelly. 2011. Delineation of interspecific epistasis on fiber quality traits in Gossypium hirsutum by ADAA analysis of intermated G. barbadense chromosome substitution lines. Theoretical And Applied Genetics, 122(7), 1351-1361. doi:10.1007/s00122-011-1536-5

Saha, S., D.M. Stelly, D.A. Raska, J. Wu, J.N. Jenkins, J.C. McCarty, A. Makamov, V. Gotmare, I.Y. Abdurakhmonov, B.T. Campbell. 2012. Chromosome Substitution Lines: Concept,

Development and Utilization in the Genetic Improvement of Upland Cotton. Plant Breeding. doi: 10.5772/35585.

Saha, S., J. Wu, J.N. Jenkins, J.C. McCarty, D.M. Stelly. 2013. Interspecific chromosomal effects on agronomic traits in Gossypium hirsutum by AD analysis using intermated G. barbadense chromosome substitution lines. Theoretical And Applied Genetics, 126(1), 109-100. doi:10.1007/s00122-012-1965-9

Saha, S., J.N. Jenkins, J.C. McCarty, R.W. Hayes, D.M. Stelly, B.T. Campbell. 2017. Four Chromosome-Specific (Gossypium barbadense Chromosome 5sh) Upland Cotton RILs with Improved Elongation. 11(2), 165-167. doi:10.3198/jpr2015.09.0060crg

Skovsted, A. 1934. Cytological Studies In Cotton II. Two Interspecific Hybrids Between Asiatic and New World Cottons. Genetics, 28(3), 407-424.

Skovsted, A. 1937. Cytological Studies In Cotton IV. Chromosome Conjugation In Interspecific Hybrids. Genetics, 34(1), 97-134.

Staub J., F. Serquen, M. Gupta. 1996. Genetic Markers, Map Construction, and Their Application in Plant Breeding. Hort Science, 31(5), 729-741.

Stelly, D. M., S. Saha, D.A. Raska, J.N. Jenkins, J.C. McCarty. O.A. Gutierrez. 2005. Registration of 17 Upland (gossypium hirsutum) cotton germplasm lines disomic for different G. barbadense chromosome or arm substitutions. Crop Science, 45(6), 2663-2665.

Sunilkumar, G., L.M. Campbell, L. Puckhaber, R.D. Stipanovic, K.S. Rathore. 2006. Engineering cottonseed for use in human nutrition by tissue-specific reduction of toxic gossypol. Proceedings of the National Academy of Sciences, 103(48), 18054-18059. doi:10.1073/pnas.0605389103

Tanksley, S. D., N.D. Young, A.H. Paterson, M.W. Bonierbale. 1989. RFLP mapping in piant breeding: New tools for an old science. Bio/Technology, 7(3), 257-264. doi:10.1038/nbt0389-257

Thermo Fisher Scientific. (2009). T042 NanoDrop Spectrophotometers Nucleic Acid Purity Ratios. Accessed from: https://www.nhm.ac.uk/content/dam/nhmwww/our-science/dpts-facilities-staff/Coreresearchlabs/nanodrop.pdf

USDA. 2017. Economic Research Service. Retrieved from: https://www.ers.usda.gov/data-products/cotton-wool-and-textile-data/cotton-and-wool-yearbook/

USDA. 2018. National Agricultural Statistics Service. Retrieved from https://www.nass.usda.gov/Statistics_by_Subject/result.php?CDC03D5A-502B-344F-9014-F3F3729512C1&sector=CROPS&group=FIELD%20CROPS&comm=COTTON

van Berloo, R. 2008. GGT 2.0: Versatile Software for Visualization and Analysis of Genetic Data. Journal of Heredity, 99(2), 232-236. doi:10.1093/jhered/esm109

Velioglu, S.K. 2019. Genome-Wide Spaced Simplex SNP Assays for Marker-Based Interspecific Germplasm Introgression and Genetic Manipulation in Cotton. Cotton Beltwide Conferences. New Orleans, Louisiana. January 8-10, 2019.

Webber, J. M. (1934). Cytogenetic Notes On Cotton and Cotton Relatives. Science, 80(2073), 268-269.

Wendel, J. F., C.L. Brubaker, A.E. Percival. 1992. Genetic Diversity in Gossypium hirsutum and the Origin of Upland Cotton. American Journal of Botany, 79(11), 1291-1310. doi:10.2307/2445058

Wendel, J. F., R. Rowley, J.M. Stewart. 1994. Genetic diversity in and phylogenetic relationships of the Brazilian endemic cotton, Gossypium mustelinum (Malvaceae). 192(1), 49-59. doi:10.1007/bf00985907

Wendel, J.F., R.C. Cronn. 2003. Polyploidy and the Evolutionary History of Cotton. Advances in Agronomy, 78, 139-186.

Wendel, J. F., C. L. Brubaker, I. Alvarez, R. Cronn, and J. M. Stewart, 2009: Evolution and natural history of the cotton genus. In: A. H. Paterson (ed.), Genetics and Genomics of Cotton, 3, 12. Springer Publisher, New York, NY, USA.

Wu, J., O.A. Gutierrez, J.N. Jenkins, J.C. McCarty, J. Zhu. 2009. Quantitative analysis and QTL mapping for agronomic and fiber traits in an RI population of Upland cotton. Euphytica, 165(2), 231-245.

Xu, J. (2014). Transmission Rates of Gossypium mustelinum and G. tomentosum SNP Markers in Early-generation Backcrosses to Cotton. Texas A&M University, Retrieved from https://oaktrust.library.tamu.edu/handle/1969.1/153869

APPENDIX A

PARTIAL PEDIGREE – PLANTING AND FIELD NUMBERS

BC_2F_1	BC_4F_1 2017 Plant #	BC_4F_1 2017 Field #	BC_5F_1 2018 Planting #	BC_5F_1 2018 Field #
1300641.01	3056_001	341.01	5002.01	111.11
1300641.01	3056_001	341.01	5002.02	111.12
1300641.01	3056_001	341.01	5002.06	111.15
1300641.01	3056_001	341.01	5002.09	111.17
1300641.01	3056_004	341.07	5003.05	112.05
1300641.01	3056_004	341.07	5003.09	112.07
1300641.01	3056_004	341.07	5003.10	112.08
1300641.01	3056_004	341.07	5003.12	112.10
1300641.03	3057_005	343.15	5004.03	112.13
1300641.03	3057_005	343.15	5004.04	112.14
1300641.03	3057_005	343.15	5004.05	112.15
1300641.03	3057_005	343.15	5004.08	112.18
1300641.03	3057_005	343.15	5004.10	112.20
1300641.16	3058_014	323.15	5005.03	113.01
1300641.16	3058_014	323.15	5005.04	113.02
1300641.16	3058_014	323.15	5005.05	113.03
1300641.16	3058_014	323.15	5005.06	113.04
1300641.16	3058_014	323.15	5005.07	113.05
1300641.16	3058_014	323.15	5005.12	113.09
1300641.16	3058_014	323.15	5005.15	165.08
1300641.16	3058_004	345.17	5024.01	124.01
1300641.16	3058_004	345.17	5024.02	124.02
1300641.16	3058_004	345.17	5024.04	124.04
1300641.16	3058_004	345.17	5024.07	124.07
1300641.16	3058_005	345.19	5025.05	124.12
1300641.16	3058_005	345.19	5025.06	124.13
1300641.16	3058_005	345.19	5025.08	124.15
1300641.16	3058_005	345.19	5025.10	124.17
1300641.16	3058_005	345.19	5025.11	124.18
1300641.16	3058_005	345.19	5025.12	124.19
1300641.16	3058_005	345.19	5025.13	124.20
1300641.16	3058_015	346.15	5026.04	125.03
1300641.16	3058_015	346.15	5026.09	125.08
1300641.16	3058_015	346.15	5026.14	167.02
1300641.16	3058_017	346.19	5027.04	125.13

BC_2F_1	BC_4F_1 2017 Plant #	BC_4F_1 2017 Field #	BC_5F_1 2018 Planting #	BC_5F_1 2018 Field #
1300641.16	3058_017	346.19	5027.09	125.16
1300641.16	3058_017	346.19	5027.11	125.18
1300641.16	3058_024	347.11	5028.02	126.01
1300641.16	3058_024	347.11	5028.06	126.05
1300641.16	3058_024	347.11	5028.09	126.08
1300641.16	3058_024	347.11	5028.13	167.05
1300641.16	3058_024	347.11	5028.15	167.03
1300642.11	3059_001	347.15	5006.02	113.12
1300642.11	3059_001	347.15	5006.06	113.16
1300642.11	3059_001	347.15	5006.07	113.17
1300642.11	3059_001	347.15	5006.09	113.19
1300642.11	3059_006	348.05	5007.01	114.01
1300642.11	3059_006	348.05	5007.02	114.02
1300642.11	3059_006	348.05	5007.07	114.06
1300642.11	3059_006	348.05	5007.09	114.08
1300642.11	3059_006	348.05	5007.11	114.09
1300642.11	3059_003	347.19	5029.02	126.12
1300642.11	3059_003	347.19	5029.04	126.14
1300642.11	3059_003	347.19	5029.05	126.15
1300642.11	3059_003	347.19	5029.09	126.18
1300642.11	3059_003	347.19	5029.11	126.20
1300642.11	3059_003	347.19	5029.12	167.07
1300642.11	3059_003	347.19	5029.14	167.06
1300642.11	3059_008	348.07	5030.01	127.01
1300642.11	3059_008	348.07	5030.05	127.04
1300642.11	3059_008	348.07	5030.09	127.07
1300642.11	3059_008	348.07	5030.10	127.08
1300642.11	3059_008	348.07	5030.11	127.09
1300642.11	3059_008	348.07	5030.12	127.10
1300642.11	3059_008	348.07	5030.13	167.08
1300642.11	3059_013	348.11	5031.01	127.11
1300642.11	3059_013	348.11	5031.02	127.12
1300642.11	3059_013	348.11	5031.07	127.14
1300642.11	3059_013	348.11	5031.08	127.15
1300642.11	3059_013	348.11	5031.10	127.17
1300642.11	3059_013	348.11	5031.11	127.18
1300642.11	3059_013	348.11	5031.12	127.19
1300642.11	3059_013	348.11	5031.14	167.10

BC_2F_1	BC_4F_1 2017 Plant #	BC_4F_1 2017 Field #	BC_5F_1 2018 Planting #	BC_5F_1 2018 Field #
1300642.11	3059_018	348.19	5032.03	128.02
1300642.11	3059_018	348.19	5032.04	128.03
1300642.11	3059_018	348.19	5032.07	128.05
1300642.11	3059_018	348.19	5032.09	128.07
1300642.11	3059_019	351.01	5033.01	128.11
1300642.11	3059_019	351.01	5033.06	128.14
1300642.11	3059_019	351.01	5033.09	128.17
1300642.11	3059_019	351.01	5033.10	128.18
1300642.11	3059_019	351.01	5033.12	128.20
1300642.11	3059_019	351.01	5033.13	167.14
1300642.11	3059_020	351.03	5034.02	131.02
1300642.11	3059_020	351.03	5034.04	131.04
1300642.11	3059_020	351.03	5034.05	131.05
1300642.11	3059_020	351.03	5034.07	131.07
1300642.11	3059_020	351.03	5034.08	131.08
1300642.11	3059_020	351.03	5034.09	131.09
1300642.11	3059_021	351.05	5035.04	131.14
1300642.11	3059_021	351.05	5035.05	131.15
1300642.11	3059_021	351.05	5035.06	131.16
1300642.11	3059_021	351.05	5035.08	131.18
1300642.11	3059_021	351.05	5035.14	167.17
1300642.11	3059_021	351.05	5035.15	167.18
1300642.11	3059_023	351.09	5036.01	132.01
1300642.11	3059_023	351.09	5036.02	132.02
1300642.11	3059_023	351.09	5036.03	132.03
1300642.11	3059_023	351.09	5036.06	132.05
1300642.11	3059_023	351.09	5036.11	132.10
1300642.11	3059_024	351.11	5037.01	132.11
1300642.11	3059_024	351.11	5037.05	132.14
1300642.11	3059_024	351.11	5037.12	132.12
1300642.11	3059_024	351.11	5037.15	168.01
1300642.11	3059_025	351.13	5038.04	133.03
1300642.11	3059_025	351.13	5038.08	133.07
1300642.11	3059_025	351.13	5038.09	133.08
1300642.11	3059_025	351.13	5038.10	133.09
1300642.11	3059_025	351.13	5038.11	133.10
1300642.11	3059_025	351.13	5038.12	168.02
1300642.11	3059_025	351.13	5038.13	168.03

BC_2F_1	BC_4F_1 2017 Plant #	BC_4F_1 2017 Field #	BC_5F_1 2018 Planting #	BC_5F_1 2018 Field #
1300642.11	3059_009	323.19	5039.02	133.12
1300642.11	3059_009	323.19	5039.04	133.14
1300642.11	3059_009	323.19	5039.06	133.16
1300642.11	3059_009	323.19	5039.08	133.18
1300642.11	3059_009	323.19	5039.10	133.20
1300642.11	3059_009	323.19	5039.15	168.05
1300642.17	3060_004	352.07	5001.01	111.01
1300642.17	3060_004	352.07	5001.03	111.03
1300642.17	3060_004	352.07	5001.06	111.05
1300642.17	3060_004	352.07	5001.08	111.06
1300642.17	3060_004	352.07	5001.09	111.07
1300642.17	3060_005	352.09	5001.11	111.09
1300642.17	3060_005	352.09	5001.13	164.16
1300642.17	3060_005	352.09	5001.14	164.17
1300642.17	3060_005	352.09	5001.15	164.15
1300642.17	3060_001	352.01	5008.01	114.11
1300642.17	3060_001	352.01	5008.14	165.13
1300643.02	3061_013	355.05	5009.12	165.15
1300643.02	3061_013	355.05	5009.13	165.16
1300643.02	3061_013	355.05	5009.15	165.17
1300643.02	3061_014	355.07	5010.07	115.17
1300643.02	3061_014	355.07	5010.08	115.18
1300643.02	3061_014	355.07	5010.09	115.19
1300643.02	3061_014	355.07	5010.11	115.20
1300643.02	3061_014	355.07	5010.12	165.18
1300643.02	3061_014	355.07	5010.13	165.19
1300643.02	3061_014	355.07	5010.15	165.20
1300643.07	3062_003	356.13	5011.04	116.01
1300643.07	3062_019	324.07	5014.01	117.01
1300643.07	3062_019	324.07	5014.04	117.03
1300643.07	3062_019	324.07	5014.06	117.05
1300643.07	3062_019	324.07	5014.10	117.06
1300643.07	3062_019	324.07	5014.11	117.07
1300643.07	3062_019	324.07	5014.12	117.08
1300643.07	3062_019	324.07	5014.14	117.10
1300643.15	3063_001	358.07	5012.04	116.08
1300643.15	3063_001	358.07	5012.08	116.12
1300643.15	3063_001	358.07	5012.15	166.03

BC_2F_1	BC_4F_1 2017 Plant #	BC_4F_1 2017 Field #	BC_5F_1 2018 Planting #	BC_5F_1 2018 Field #
1300643.15	3063_002	358.09	5040.11	168.06
1300643.15	3063_002	358.09	5040.12	168.07
1300643.15	3063_003	358.11	5041.04	134.13
1300643.15	3063_003	358.11	5041.08	134.17
1300643.15	3063_003	358.11	5041.13	168.08
1300643.15	3063_003	358.11	5041.15	168.09
1300643.15	3063_005	358.15	5042.04	135.04
1300643.15	3063_005	358.15	5042.05	135.05
1300643.15	3063_005	358.15	5042.08	135.07
1300643.15	3063_005	358.15	5042.10	135.09
1300643.15	3063_005	358.15	5042.15	168.11
1300643.15	3063_011	361.07	5043.01	135.11
1300643.15	3063_011	361.07	5043.03	135.13
1300643.15	3063_011	361.07	5043.04	135.14
1300643.15	3063_011	361.07	5043.06	135.16
1300643.15	3063_016	361.17	5044.01	136.01
1300643.15	3063_016	361.17	5044.04	136.04
1300643.15	3063_016	361.17	5044.07	136.07
1300643.15	3063_022	362.09	5045.02	136.12
1300643.15	3063_022	362.09	5045.05	136.15
1300643.15	3063_022	362.09	5045.08	136.18
1300643.15	3063_022	362.09	5045.09	136.19
1300643.17	3064_001	362.17	5013.03	116.16
1300643.17	3064_001	362.17	5013.04	116.17
1300643.17	3064_001	362.17	5013.07	116.20
1300643.17	3064_001	362.17	5013.09	166.05
1300643.17	3064_001	362.17	5013.10	166.06
1300643.17	3064_001	362.17	5013.12	166.07
1300643.17	3064_001	362.17	5013.13	166.08
1300643.17	3064_009	363.09	5046.02	137.02
1300643.17	3064_009	363.09	5046.03	137.03
1300643.17	3064_009	363.09	5046.06	137.06
1300643.17	3064_009	363.09	5046.10	137.10
1300643.17	3064_009	363.09	5046.12	168.16
1300643.17	3064_009	363.09	5046.14	137.04
1300643.17	3064_013	363.15	5047.10	137.19
1300643.17	3064_013	363.15	5047.11	137.20
1300643.17	3064_013	363.15	5047.13	168.17

BC_2F_1	BC_4F_1 2017 Plant #	BC_4F_1 2017 Field #	BC_5F_1 2018 Planting #	BC_5F_1 2018 Field #
1300643.17	3064_015	363.19	5048.06	138.06
1300643.17	3064_015	363.19	5048.09	138.09
1300643.20	3065_021	366.07	5015.03	117.12
1300643.20	3065_021	366.07	5015.04	117.13
1300643.20	3065_021	366.07	5015.06	117.15
1300643.20	3065_021	366.07	5015.09	117.17
1300643.20	3065_021	366.07	5015.10	117.18
1300643.20	3065_021	366.07	5015.12	117.20
1300643.20	3065_003	364.13	5049.03	138.13
1300643.20	3065_003	364.13	5049.05	138.15
1300643.20	3065_003	364.13	5049.06	138.16
1300643.20	3065_003	364.13	5049.09	138.19
1300643.20	3065_003	364.13	5049.10	138.20
1300643.20	3065_003	364.13	5049.13	168.19
1300643.20	3065_004	364.15	5050.03	141.02
1300643.20	3065_004	364.15	5050.05	141.04
1300643.20	3065_004	364.15	5050.06	141.05
1300643.20	3065_006	364.17	5051.01	141.11
1300643.20	3065_006	364.17	5051.06	141.16
1300643.20	3065_006	364.17	5051.08	141.20
1300643.20	3065_013	365.11	5052.09	142.09
1300643.20	3065_013	365.11	5052.10	142.10
1300643.20	3065_016	365.17	5053.04	142.13
1300643.20	3065_016	365.17	5053.05	142.14
1300643.20	3065_016	365.17	5053.08	142.17
1300643.20	3065_018	366.01	5054.05	143.05
1300643.20	3065_018	366.01	5054.06	143.06
1300643.20	3065_018	366.01	5054.10	143.10
1300643.20	3065_019	366.03	5055.06	143.16
1300644.11	3066_004	366.17	5016.01	118.01
1300644.11	3066_004	366.17	5016.05	118.04
1300644.11	3066_004	366.17	5016.08	118.07
1300644.11	3066_004	366.17	5016.09	118.08
1300644.11	3066_004	366.17	5016.10	118.09
1300644.11	3066_004	366.17	5016.11	118.10
1300644.11	3066_004	366.17	5016.12	118.03
1300644.11	3066_004	366.17	5016.15	118.05
1300644.11	3066_004	366.17	5088.01	163.11

BC_2F_1	BC_4F_1 2017 Plant #	BC_4F_1 2017 Field #	BC_5F_1 2018 Planting #	BC_5F_1 2018 Field #
1300644.11	3066_004	366.17	5088.12	163.17
1300644.11	3066_004	366.17	5088.14	163.19
1300644.11	3066_004	366.17	5088.15	163.20
1300645.01	3067_006	431.03	5017.01	118.11
1300645.01	3067_006	431.03	5017.05	118.15
1300645.01	3067_006	431.03	5017.06	118.16
1300645.01	3067_006	431.03	5017.08	118.18
1300645.01	3067_006	431.03	5017.09	118.19
1300645.01	3067_001	368.13	5056.04	144.04
1300645.01	3067_001	368.13	5056.10	144.10
1300645.01	3067_003	368.17	5057.09	144.19
1300645.01	3067_003	368.17	5057.10	144.20
1300645.01	3067_009	431.09	5058.03	145.03
1300645.01	3067_009	431.09	5058.04	145.04
1300645.01	3067_009	431.09	5058.06	145.06
1300645.01	3067_023	432.07	5060.01	146.01
1300645.01	3067_023	432.07	5060.02	146.02
1300645.01	3067_023	432.07	5060.04	146.04
1300645.01	3067_023	432.07	5060.05	146.05
1300645.01	3067_023	432.07	5060.07	146.07
1300645.01	3067_019	325.03	5089.02	164.02
1300645.01	3067_019	325.03	5089.03	164.03
1300645.01	3067_019	325.03	5089.07	164.06
1300645.01	3067_019	325.03	5089.10	164.09
1300646.05	3068_002	432.17	5018.01	121.01
1300646.05	3068_002	432.17	5018.03	121.03
1300646.05	3068_002	432.17	5018.04	121.04
1300646.05	3068_002	432.17	5018.05	121.05
1300646.05	3068_002	432.17	5018.08	121.08
1300646.05	3068_002	432.17	5018.09	121.09
1300646.05	3068_002	432.17	5018.10	121.10
1300646.05	3068_007	433.01	5019.02	121.12
1300646.05	3068_007	433.01	5019.03	121.13
1300646.05	3068_007	433.01	5019.07	121.17
1300646.05	3068_007	433.01	5019.10	121.18
1300646.05	3068_007	433.01	5019.12	121.20
1300646.13	3069_005	435.17	5023.01	123.15
1300646.13	3069_005	435.17	5023.05	123.11

BC_2F_1	BC_4F_1 2017 Plant #	BC_4F_1 2017 Field #	BC_5F_1 2018 Planting #	BC_5F_1 2018 Field #
1300646.13	3069_005	435.17	5023.14	166.19
1300646.13	3069_005	435.17	5023.15	166.20
1300647.06	3070_002	437.13	5061.01	146.11
1300647.06	3070_002	437.13	5061.04	146.14
1300647.06	3070_002	437.13	5061.05	146.15
1300647.06	3070_002	437.13	5061.09	146.18
1300647.06	3070_003	437.15	5062.04	147.04
1300647.06	3070_003	437.15	5062.05	147.05
1300647.06	3070_003	437.15	5062.06	147.06
1300647.06	3070_003	437.15	5062.07	147.07
1300647.06	3070_006	438.01	5063.01	147.11
1300647.06	3070_006	438.01	5063.05	147.15
1300647.06	3070_006	438.01	5063.10	147.20
1300647.06	3070_018	441.01	5065.01	148.01
1300647.06	3070_018	441.01	5065.04	148.03
1300647.06	3070_018	441.01	5065.06	148.05
1300647.06	3070_018	441.01	5065.10	148.08
1300648.07	3071_002	441.09	5020.01	122.02
1300648.07	3071_002	441.09	5020.04	122.04
1300648.07	3071_002	441.09	5020.07	122.07
1300648.07	3071_002	441.09	5020.10	122.10
1300648.07	3071_002	441.09	5020.12	166.13
1300648.07	3071_002	441.09	5020.14	166.14
1300648.07	3071_007	441.17	5066.05	148.15
1300648.07	3071_007	441.17	5066.10	148.20
1300648.07	3071_010	442.03	5067.01	151.01
1300648.07	3071_010	442.03	5067.02	151.02
1300648.07	3071_010	442.03	5067.03	151.03
1300648.07	3071_010	442.03	5067.04	151.04
1300648.07	3071_010	442.03	5067.06	151.06
1300648.07	3071_010	442.03	5067.08	151.08
1300648.07	3071_010	442.03	5067.10	151.10
1300648.07	3071_014	442.11	5068.05	151.15
1300648.07	3071_014	442.11	5058.09	145.09
1300648.07	3071_018	442.19	5069.03	152.02
1300648.07	3071_018	442.19	5069.04	152.03
1300648.07	3071_018	442.19	5069.06	152.05
1300648.07	3071_022	443.05	5070.01	152.11

BC_2F_1	BC_4F_1 2017 Plant #	BC_4F_1 2017 Field #	BC_5F_1 2018 Planting #	BC_5F_1 2018 Field #
1300648.07	3071_022	443.05	5070.07	152.17
1300648.20	3072_003	443.11	5021.03	122.13
1300648.20	3072_003	443.11	5021.04	122.14
1300648.20	3072_003	443.11	5021.06	122.16
1300648.20	3072_003	443.11	5021.07	122.17
1300648.20	3072_003	443.11	5021.08	122.18
1300648.20	3072_003	443.11	5021.09	122.19
1300648.20	3072_003	443.11	5021.10	122.20
1300648.20	3072_001	443.07	5071.06	153.06
1300648.20	3072_001	443.07	5071.08	153.08
1300648.20	3072_006	443.17	5072.02	153.12
1300648.20	3072_006	443.17	5072.07	153.17
1300648.20	3072_006	443.17	5072.08	153.18
1300648.20	3072_007	443.19	5073.06	154.06
1300648.20	3072_018	445.01	5074.01	154.11
1300648.20	3072_018	445.01	5074.11	154.18
1300648.20	3072_021	445.05	5075.04	155.04
1300648.20	3072_021	445.05	5075.08	155.07
1300648.20	3072_024	445.11	5076.04	155.14
1300648.20	3072_008	444.01	5085.08	162.08
1300651.19	3073_003	445.15	5022.01	123.01
1300651.19	3073_003	445.15	5022.04	123.04
1300651.19	3073_003	445.15	5022.06	123.06
1300651.19	3073_003	445.15	5022.07	123.07
1300651.19	3073_003	445.15	5022.08	123.08
1300651.19	3073_003	445.15	5022.12	166.17
1300651.19	3073_008	446.01	5077.04	156.04
1300651.19	3073_011	446.05	5078.04	156.14
1300651.19	3073_011	446.05	5078.07	156.17
1300651.19	3073_012	446.07	5079.01	157.01
1300651.19	3073_012	446.07	5079.03	157.03
1300651.19	3073_012	446.07	5079.06	157.06
1300651.19	3073_012	446.07	5079.08	157.08
1300651.19	3073_012	446.07	5079.09	157.09
1300651.19	3073_013	446.09	5080.01	157.11
1300651.19	3073_013	446.09	5080.03	157.13
1300651.19	3073_013	446.09	5080.06	157.16
1300651.19	3073_013	446.09	5080.08	157.18

BC_2F_1	BC_4F_1 2017 Plant #	BC_4F_1 2017 Field #	BC_5F_1 2018 Planting #	BC_5F_1 2018 Field #
1300651.19	3073_023	447.07	5081.02	158.02
1300651.19	3073_023	447.07	5081.09	158.09
1300651.19	3073_024	447.09	5082.02	158.11
1300651.19	3073_024	447.09	5082.03	158.12
1300651.19	3073_024	447.09	5082.05	158.14
1300651.19	3073_024	447.09	5082.09	158.17
1300651.19	3073_024	447.09	5082.10	158.18
1300651.19	3073_024	447.09	5082.13	158.20
1300651.19	3073_025	447.11	5083.01	161.01
1300651.19	3073_025	447.11	5083.02	161.02
1300651.19	3073_025	447.11	5083.07	161.06
1300651.19	3073_025	447.11	5083.09	161.08
1300651.19	3073_025	447.11	5083.11	161.10
1300651.19	3073_026	447.13	5084.02	161.12
1300651.19	3073_026	447.13	5084.04	161.14
1300651.19	3073_026	447.13	5084.05	161.20
1300651.19	3073_026	447.13	5084.09	161.19