TRANSMISSION RATES OF *GOSSYPIUM MUSTELINUM* AND *G. TOMENTOSUM* SNP MARKERS IN EARLY-GENERATION BACKCROSSES TO COTTON (*G. HIRSUTUM L.*)

A Thesis

by

JIALE XU

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,	Wayne Smith
	Kevin Crosby
Head of Department,	David D. Baltensperger

December 2014

Major Subject: Plant Breeding

ABSTRACT

Gossypium hirsutum L. is the most widely cultivated cotton species in the genus *Gossypium*. The genetic diversity of *G. hirsutum* is considerably restricted, so it is highly desirable to introgress germplasm of related species to expand opportunities for genetic improvement. Successful interspecific introgression efforts require the transmission of alien genes into the cultivated species and homologous recombination. Transmission distortion can restrict or preclude gene transfer, and reduced rates of homologous recombination can reduce or preclude recovery of desirable genetic products. Marker-based analysis of specific chromosome segments and loci during early generations of backcrossing can reveal general and locus-specific features of alien germplasm transmission and recombination with the recurrent parent, and help guide decisions for expanded analysis, subsequent backcrosses and analogous efforts with other donors.

Interspecific monosomic hybrids were used to localize pre-validated single nucleotide polymorphism (SNP) markers in six target chromosomes. Eventually, 67 SNPs were used to analyze transmission rates. In the present research, each "breeding situation" was defined as a unique combination of cross direction, backcross generation and cross location. Twelve *G. hirsutum* backcross populations were derived under different breeding situations, six from each of two alien donors, *G. mustelinum* and *G. tomentosum*. KASP analysis of the SNPs revealed the presence or absence of specific donor loci and segments in 784 individuals of the 12 populations, and was used to determine transmission rates.

Linkage groups were constructed based on segregation ratios in BC1F1 populations for each donor. The average transmission rate of germplasm from *G. tomentosum* was similar to but higher (3%) than *G. mustelinum*, indicating a closer relationship of *G. tomentosum* with *G. hirsutum*. Several markers exhibited strongly distorted transmission relative to other loci of the respective linkage groups in specific populations. Several loci

exhibited significant differences due to cross direction and cross location. Analysis of the BC3F1 populations revealed the crosses from field environments were relatively favorable for transmission of alien germplasm and greenhouse environments for loss of alien germplasm. Multiple comparisons based on general linear model (GLM) for effects of breeding situations on transmission rates revealed generation and location significantly affected transmission of *G. mustelinum* germplasm, whereas cross direction, location and generation affected transmission of *G. tomentosum* germplasm.

DEDICATION

I dedicate my dissertation work to my loving parents, Songgeng Xu and Huajun Lu whose words of encouragement and push for tenacity ring in my ears.

ACKNOWLEDGEMENTS

I would like to express my deepest gratitude to my major advisor, Dr. David Stelly for his excellent guidance, patience and providing me with an excellent atmosphere for doing this research program. And I would also like to thank him for helping me with presentation and writing skills. To me, he is not only an advisor on my research program, but also an advisor on other aspects of my development. I am also thankful to my two other committee members, Dr. Wayne Smith and Dr. Kevin Crosby, for their suggestions on my dissertation.

I would like to thank Dr. Fei Wang, who helped me learn experimental skills in DNA extraction and KASP analysis. She was also very helpful in SNP marker design, localization and validation. I am thankful to some other individuals who also involved in SNP development between *G. hirsutum* and the two aliens species (*G. mustelinum* and *G. tomentosum*): Dr. David Stelly (concepts, strategies and problem solving), Kevin A. Hoegenauer and Amanda Hulse Kemp (SNP alignment), myself (literature work, data interpretation, data organization and database construction).

I would like to thank Wayne Raska and Robert Vaughn for their help in planting in field and greenhouse. Their taking care of my plants is very important for my research program. As part of my dissertation, I would also thank to Wayne Raska as well as other individuals who have ever worked on development and identification of interspecific monosomic hybrids. I would also like to thank other members in our lab, Xiuting Zheng, Bo Liu, Amanda Hulse Kemp and Andrea Maeda. All of them offered a lot of help either for research or for class study. It is my greatest honor to work with them during the past two years.

Lastly, I would like to thank my parents for their encouragement and economic support for my master's degree. They always stood by me through good and bad times.

TABLE OF CONTENTS

Page

LIST OF FIGURES

Page

LIST OF TABLES

Page

CHAPTER I
INTRODUCTION

Cotton is one of the most economically important plants worldwide. Its fiber is commonly known as cotton and is the principal source of natural fibers for the textile industry. Though grown mainly for its fiber, cotton is the second most important source of plant proteins and the fifth largest source of plant oils (Lee, 1984).

The cotton genus has eight diploid genome groups: A, B, C, D, E, F, G, K (Wendel, *et al.* 2009). The *Gossypium* genus contains approximately 50 species including 45 diploid 2n=2x=26 species where n is the haploid number of chromosomes and 5 tetraploid 2n=4x=52 species, AD_1 to AD_5. At present, five distinct allotetraploid species namely *G. hirsutum, G. tomentosum, G. mustelinum, G. barbadense and G. darwinii* have been recognized (Wendel and Cornn2003). Two new ones are in the process of being recognized (Grover et al. unpublished, Wendel et al. unpublished). *G. tomentosum* is endemic to the Hawaii Islands and it has a diffuse population structure, occurring mostly as scattered individuals and small populations on several islands (Dejoode and Wendel 1992). *G. mustelinum*, another wild species, has an island-like distribution in the sense that it is an uncommon species restricted to a relatively small region of northeast Brazil. In the beginning of 2006, only four natural populations of *G. mustelinum* were known. One was located in the municipality of Caicó, in the State of Rio Grande do Norte, one in Jaguarari and two in Macururé, both municipalities in the State of Bahia. At the end of 2006, new *G. mustelinum* populations were found in Bahia (Alves *et al.* 2013). *G. hirsutum* is widely distributed in Central and South America, the Caribbean, and even reaches distant islands in the Pacific Soloman Islands, Marquesas.

G. tomentosum has strong fiber (Meyer and Meredith 1978) and is the most heat-resistant species in Gossypium (Percival and Kohel 1990). Foliage from *G. mustelinum* has the highest concentrations of the heliocides H1 and H4 and moderately high levels of gossypol. Except for gossypol, foliar concentrations of terpenoid aldehydes in the

lysigenous glands was highest in *G. mustelinum* relative to 30 species representing A, B, C, D, F, G, K, and AD genomic groups of Gossypium (Khan *et al.* 1999). Many of these allochemicals are potentially useful for improving host-plant resistance in upland cotton.

Faced with diminishing land, water and other resources, significant genetic gains will be required from all the domesticated plants to achieve greater sustainability. The genetic improvement of crop species will be a major stepping-stone to meet this demand. Due to the limited genetic diversity of domesticated *G. hirsutum* types and even the species overall, innovative exploitation of genetic resources and effective breeding methods will be key to realizing genetic improvements in sustainability. The primary gene pool of cotton includes all AD tetraploid species, which thus constitute an especially accessible reservoir of important genes for disease, pest and abiotic stress resistances as well as for the improvement of fiber quality.

Interspecific germplasm introgression can greatly expand the opportunities for crop improvement (Tanksley and Mccouch 1997). Over the past several decades, genetic diversity of many domesticated plants has been expanded by interspecific introgression. Different types of genetic populations have been constructed and some superior traits from alien species have been introgressed, including several disease and pest resistance genes. In some instances, specific crossability, reproductive or genomic homeology barriers had to be overcome. In rice, for example, genes for resistance to brown planthopper, bacterial blight and blast have been introgressed across crossability barriers from distantly related species (Brar and Khush 1997). Rust resistance of *Aegilops geniculata* was transferred to wheat (*Triticum aestivum*) by induced homologous pairing between chromosomes $5M^g$ of *Ae. geniculata* and 5D of wheat (Aghaee-Sarbarzeh *et al.* 2002).

In cotton, interspecific introgression has been applied as a tool to transfer alien genetic materials, which might be responsible for important agronomic traits, into upland cotton so that the limited gene pool can be diversified. With the construction of various

introgression lines, many high value QTLs have been mapped and more novel and beneficial gene combinations have been realized. High resistance to the reniform nematode, *Rotylenchulus reniformis* has been introgressed from *G. longicalyx* to *G. hirsutum* via tri-species hybrids and the resultant seed provided a major new tool for managing the reniform nematode in cotton, which costs U.S. producers about $100 million annually (Robinson *et al.* 2007). Near-isogenic cotton lines (NILs) derived by marker-assisted selection (MAS) from crosses between *G. barbadense* and *G. hirsutum* led to modifications of drought related traits (Levi *et al.* 2009).

Various kinds of introgression lines from *G. mustelinum* and *G. tomentosum* have been constructed. Backcross progenies in *G. mustelinum* population present improved fiber quality. For *G. tomentosum*, BC1F1 and F1 have increased fiber strength over TM-1, a cultivar of *G. hirsutum* (Gardunia 2006). Backcross-inbred families from crosses between *G. hirsutum* and *G. tomentosum* identified a total of 28 QTLs for fiber quality, including four for fiber elongation, eight for fiber fineness, four for fiber length, four for fiber strength, six for fiber uniformity, one for boll weight, and one for boll number (Zhang *et al.* 2011).

Generally, however, the level of genetic diversity in upland cotton is still low, owing to several impediments in conventional methods of interspecific introgression in cotton: (*i*) complex antagonistic relationships among important traits; (*ii*) cytogenetic differences among the species due to different ploidy levels, meiotic affinity and chromosomal structural differences including translocations and inversions; (*iii*) "linkage drag effects" leading to poor agronomic qualities; (*iv*) reduced recombination; (*v*) loss of alien genetic materials in early generations; (*vi*) sterility in the hybrids; (*vii*) complex genetic interactions such as Muller-Dobzhansky complexes and (*viii*) distorted segregation (Endrizzi *et al.* 1985).

The evolutionary consequences of introgression have been addressed at the theoretical level (Anderson 1949; Barton and Gale 1993; Rieseberg and Wendel 1993), however,

few of these theories have been tested empirically in plants. The above mentioned impediments can cause some violations of Mendel's Law and change the fate of introgressed genetic material, which in turn makes the construction of some specific genetic populations difficult and always prevent breeders from advancing the crop species to the desired direction. This process demands focused, multi-scientific efforts. Such efforts would be likely to detect the factors and their corresponding effects on transmission rate.

The system for transmission rate control is complicated. Factors related to transmission rate and their effects remained to be unveiled. In general, transmission rate can be determined by exogenous and intrinsic factors. The primary line of evidence suggesting an important role for endogenous factors on transmission rate is that distortion segregation and restricted recombination are often found in the mosaic genomes of interspecific hybrid populations (Rick, 1969; Paterson, 1990; Arnold, 1992). Both structural and genic mutations accumulated by species prior to hybridization appear to play a role in non-Mendelian inheritance (Rieseberg *et al.* 1995).

Preferential transmission of a specific allele, chromosome or genome in the advancement of generations can result from diverse and numerous phenomena, e.g. zygotic lethality (Lee, 1981) and somatic elimination (Ho and Kasha 1974). Chromosome loss / recovery may be induced by either of two types of somatic chromosome loss: (*i*) chromosome elimination or (*ii*) somatic reduction. The introgression of alien germplasm can lead to differential viability of spores, gametes or zygotes, which can distort the transmission rates.

One of the complicated mechanisms that may affect the genetic composition of a population as a consequence of meiotic events is defined as "Meiotic Drive" (Sandler and Novitski 1957). In certain genetic backgrounds, heterozygotes fail to follow Mendel's law and instead produce gametes with unequal genotypic frequencies.

Cases of meiotic drive causing preference of transmission have been demonstrated in some plants. The gametocidal " *Gc*" chromosomes or genetic factors of wheat are a typical distorter for transmission rate in wheat; these affect the viability of gametes (Nasuda *et al.* 1998). Such factors were introduced into wheat through interspecific hybridization and backcrossed to related *Aegilops* species. Only gametes with alien chromosomes carrying Gc factors can be functional and Gc factors were transmitted preferentially to the progeny (Maan 1975).

Life cycles of many crop species including cotton undergo a series of sexual processes, and many genetic studies indicate that these sexual processes are mainly controlled by nuclear genes (Johns *et al.* 1981; Kaul 1988; Okamuro *et al.* 1993). Alteration of alleles in a specific locus that is responsible for both male and female development can either lead a gene not to be expressed or expressed in an abnormal way, which results in sterility on the male or female side.

At least 30 genes are known in *Datura stramonium* that condition abortion of microspores to which they segregate (Avery *et al.* 1959). In tomato, abortion of male and female gametes is controlled by three alleles of the *Gamete eliminator* gene (*Ge*). Elimination occurs only in Ge^c/Ge^p, in which Ge^c gametes are aborted. No abortion occurs in the homozygotes Ge^c/Ge^c or in Ge^p/Ge^p (Rick 1966). A similar genetic model exits for the "pollen killer" (*Ki*) locus of wheat (Loegering and Sears 1963).

Cytoplasmic male sterility (CMS) is another form of male sterility that involves cytoplasmic organelle genes, and depending on the sterility system, may or may not involve nuclear genes that "restore fertility" (often symbolized "*Rf*"). When nuclear "restorers" are involved, these systems are sometimes dubbed "cytoplasmic-nuclear male sterility" (CNMS) systems, to more clearly highlight the interactions between cytoplasmic and nuclear genotypes. In a given CNMS system, the restoration of fertility can rely on a nuclear *Rf* gene that is expressed either sporophytically or microgametophytically, and if the latter, transmission of the *Rf* pollen and closely linked

loci is highly favored from *Rfrf* heterozygotes that contain a pollen-sterilizing cytoplasm. This system will affect some types of cells in anthers during some stages of microsporogenesis (Duvick 1965).

There are many other situations where gametophytically expressed genes might not determine pollen viability, but nevertheless affect their competitiveness. Many QTL have been mapped as related to tube growth rate and grain germinability, both of which are the main traits of pollen fitness (Sari-Gorla *et al.* 1992). Alteration of genotype in these loci is likely to change the pollen fitness and create some distortions of transmission rates, especially for linked genes.

In cotton, sexually preferential transmission has been demonstrated in previous studies. Ten monosomic addition stocks were used to make a comparison among 4 alien chromosomes C1-A, C1-B, C1-C, and C1-D for their transmission rates. It was observed that the alien chromosome C1-A was transmitted through the female gamete to more than 90% of all progenies, whereas the other monosomics averaged only 23% transmission. None of the four alien chromosomes was transmitted via male gamete (Rooney and Selly 1991). Multiple alien chromosome addition lines MACALs were developed by backcrossing F1 progeny of two hexaploid lines (2x G. hirsutum X G. australe and 2x G. hirsutum X G. sturtianum to G. hirsutum). In the BC2 MACAL families, some of the available chromosomes were preferentially inherited while some others were preferentially eliminated (Lopez-Lavalle and Brubaker 2007). Another skewed transmission rate was observed in BC3F2 plants derived from backcrossing G. barbadense to G. hirsutum, which can be best accounted for by multi-locus epistasis interactions (Jiang *et al.* 2000).

Many studies have focused attention on exogenous selection (Endler 1973; May et al. 1975; Harrison 1986). Natural selection violates assumptions for Hardy-Weinberg equilibrium, and as a consequence, genes responsible for superior traits are more likely to be conserved and passed to the next generation (Darwin 1859). It has been observed

that expression and segregation of alien genes in cotton can be influenced by environment (Sachs *et al.* 1998).

Some biological processes requisite to crop reproduction are also subject to environmental factors such as temperature and humidity. High temperature resulted in increased pollen sterility in rice, and the critical air temperature for spikelet sterility was reduced at elevated concentrations of carbon dioxide (Matsui *et al.* 1997). High temperature environments with greater than 30 C during flowering reduce boll retention and yield in cotton. And the cardinal temperature of pollen germination varies among cultivars (Kakani *et al.* 2005). At either constantly low 25% or high 90% atmospheric relative humidity, cotton *Gossypium hirsutum* L. set very few bolls because the anthers failed to dehisce. Seed cotton yields were almost zero at both 25 and 90% relative humidity, whereas yields at 40 and 65% were 48 and 164 g/plant, respectively (Hoffman and Rawlins 1970). Many studies about QTL mapping report interactions between QTL main effect and environment. Thus, it is possible that environmental effects might interfere or alter mechanisms that cause the preferential transmission.

The use of molecular markers has revolutionized the pace and precision of plant genetic analysis which in turn is facilitating the implementation of molecular breeding of crops. The last three decades have seen the development of various markers for tracking certain regions in chromosomes. Evolution of molecular markers has been driven by desire for high throughput, low cost and high reproducibility (Bernardo 2008). Depending on detection method and throughput, all the molecular markers can be divided into three major groups: (*i*) low-throughput, hybridization based markers such as restriction fragment length polymorphism RFLPs (Botstein *et al.* 1980); (*ii*) medium-throughput, PCR based markers that include random amplification of polymorphic DNA RAPD (Welsh and Mcclelland 1990), amplified fragment length polymorphism AFLP (Vos *et al.* 1995), SSRs (Jacob *et al.* 1991); (*iii*) high-throughput (HTP), sequence based markers: SNPs (Wang *et al.* 1998).

In the late eighties, RFLPs were the most widely used markers in plant molecular genetics because of reproducibility and codominance (Lander and Botstein 1989). Invention of PCR technology in the beginning of the nineties overthrew low-throughput RFLP markers and a new generation of PCR-based markers emerged. RAPD, AFLP, SSR are the major PCR-based markers. RAPDs are anonymous and the level of their reproducibility are very low due to the non-specific binding of short, random primers. Owing to the lengthy and laborious detection method, AFLP did not find widespread application in molecular breeding (Powell *et al.* 1996). Microsatellite DNA markers (SSRs) were able to eliminate all the drawbacks of the above-mentioned molecular markers, which lead them to become the most widely used markers in the beginning of the 21st century.

However, during the last six years, the hegemony of SSRs was eventually broken by SNP markers, which were discovered in the human genome and have been proved to be universal as well as the most abundant forms of genetic variation among individuals of the same species (Rafalski 2002). Although each SNP locus has less polymorphism than an SSR locus because of its bi-allelic nature, SNPs can easily compensate for this drawback by being abundant, ubiquitous and amenable to high- or ultra-high-throughout automation (Mammadov *et al.* 2012). Associated with these advantages, high-density maps can be constructed to represent genetic information across whole genomes. Such maps greatly facilitate interspecific introgression breeding programs.

The object of this project is to reveal absolute and relative rates of transmission for SNP markers located in alien DNA segments that are targeted for interspecific introgression from *G. mustelinum* and *G. tomentosum*. It is of practical and scientific interest to know if transmission rates differ by locus, species (donors), environment, backcross generation and direction of cross. For statistical analysis and discussion, we collectively refer to combinations of the environments (2), backcross generations (3) and direction of cross (2) as "breeding situations" (12). In this study, two sets of backcross populations are derived, one from each of two wild species, *G. mustelinum* and *G. tomentosum*,

respectively. TM-1, a cultivar of *G. hirsutum* is used as the recurrent parent and those two wild species are used as donor parents. The transmission rate can be detected by genotyping with SNP analysis.

CHAPTER II
MATERIALS AND METHODS

2.1. Plant materials

2.1.1 Plant materials for backcross programs

A backcross program was launched in the summer of 2011 by reciprocally crossing greenhouse-grown F1 plants (*G. mustelinum* x *G. hirsutum*, *G. tomentosum* x *G. hirsutum*) with *G. hirsutum* to create BC1F1 seeds, which were grown for the second generation of backcross during the summer of 2012. Twenty BC1F1 hybrids of *G. mustelinum* and *G. hirsutum* and twenty-seven BC1F1 hybrids of *G. tomentosum* and *G. hirsutum* were used for additional backcrossing to create advanced generations. BC2F1 from each of these hybrids were field-grown in 2013 for the third backcross. For the third backcross using a field environment, the BC2F1 plants served as the female parent and *G. hirsutum* served as the male parent. Fifteen BC3F1 seeds sampled from each BC2F1 plant were subsequently planted in 2014.

A subset of BC2F1 populations was vegetatively maintained and used during the winter in 2013 to make additional backcrosses in a winter greenhouse environment. Twenty BC2F1 hybrids of *G. mustelinum* and *G. hirsutum* and twenty-seven BC2F1 hybrids of *G. tomentosum* and *G. hirsutum* were randomly selected and transferred from field to greenhouse for reciprocal crosses during fall/winter of 2013. Ten BC3F1 progenies were selected from each hybrid.

Eventually, Two sets of populations were derived, one set from each of the two wild species donors. In each set, 6 populations were defined by generation (BC1F1, BC2F1 and BC3F1), location (greenhouse and agriculture field) and cross direction (*G. mustelinum* X *G. hirsutum*, *G. tomentosum* X *G. hirsutum*, *G. hirsutum* X *G. mustelinum*, *G. hirsutum* X *G. tomentosum*). The whole process for the development of the two sets of populations was illustrated in Figure 2.1 and Figure 2.2. In all, 12

groups were classified as shown in Table 2.1.

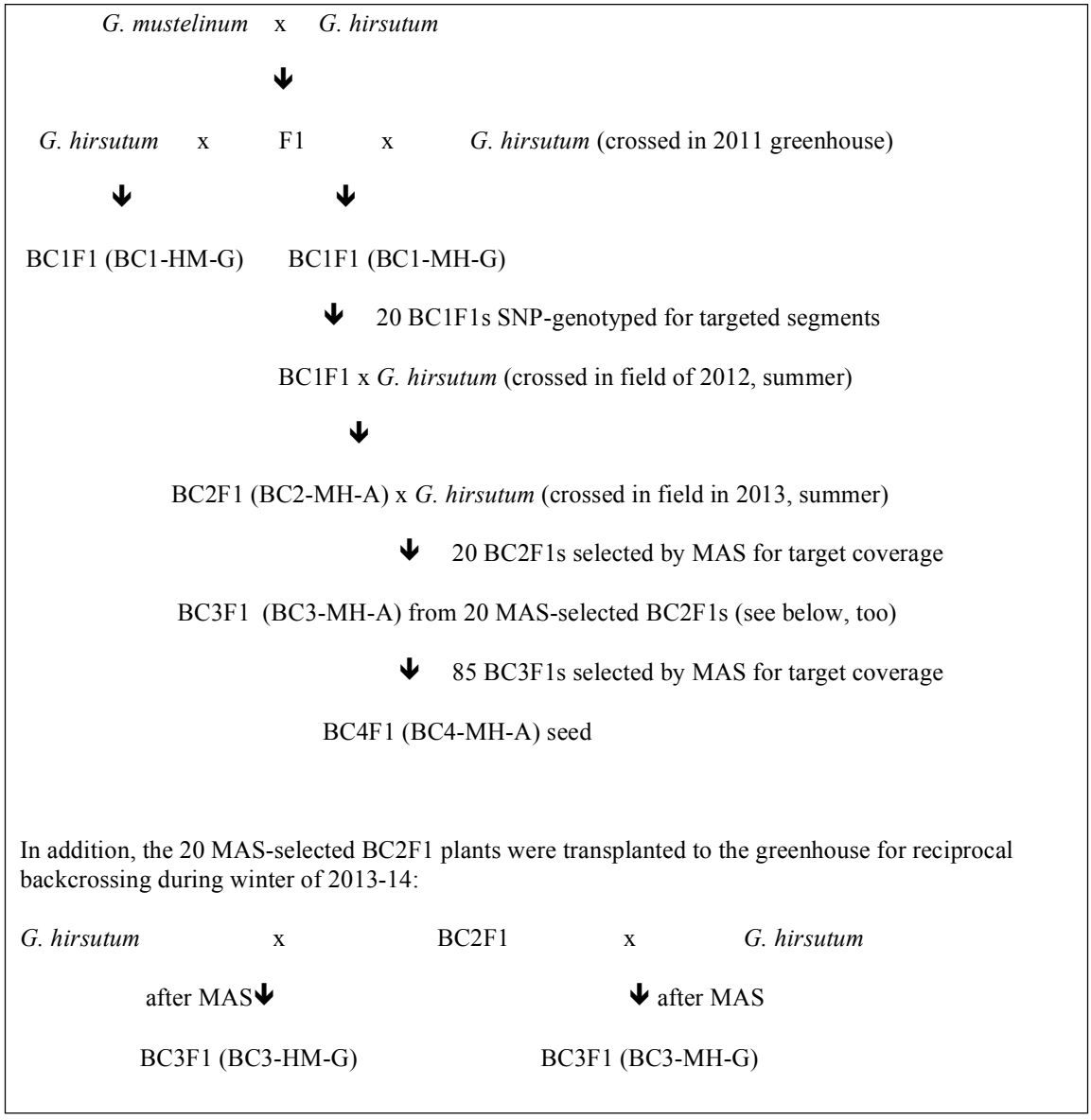

Figure 2.1 Development of *G. mustelinum*-derived populations.

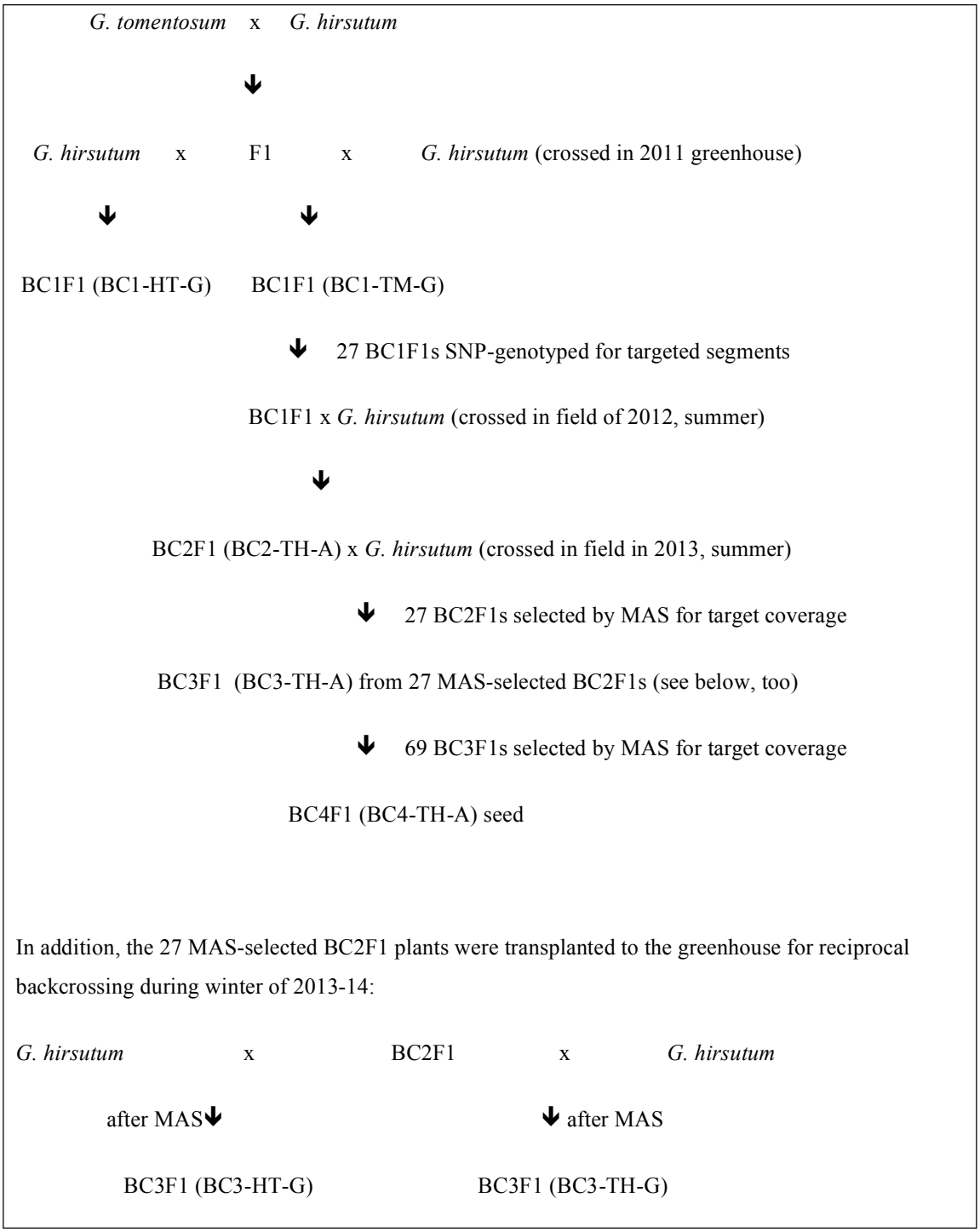

Figure 2.2 Development of *G. tomentosum*-derived populations.

TABLE 2.1
Populations created in this study and their nomenclature based on generation, cross type and location

Wild species	Population	Generation	Cross Direction	Location	Size
GM	BC1-MH-G	BC1F1	GM X GH	G	66
GM	BC1-HM-G	BC1F1	GH X GM	G	7
GM	BC2-MH-A	BC2F1	GM X GH	A	78
GM	BC3-MH-A	BC3F1	GM X GH	A	85
GM	BC3-MH-G	BC3F1	GM X GH	G	54
GM	BC3-HM-G	BC3F1	GHX GM	G	28
GT	BC1-TH-G	BC1F1	GT X GH	G	39
GT	BC1-HT-G	BC1F1	GH X GT	G	21
GT	BC2-TH-A	BC2F1	GT X GH	A	83
GT	BC3-TH-A	BC3F1	GT X GH	A	69
GT	BC3-TH-G	BC3F1	GT X GH	G	57
GT	BC3-HT-G	BC3F1	GH X GT	G	30

*GH, GM and GT are short for *G. hirsutum*, *G. mustelinum and G. tomentosum,* respectively.

2.1.2 Plant materials for linkage analysis

Plant materials for linkage analysis included all BC1F1 plants mentioned above, i.e., 20 for *G. mustelinum* and 27 for *G. tomentosum*. These populations were augmented for linkage mapping by growing additional remnant BC1F1 seed. In total, the final linkage mapping populations included 73 and 60 BC1F1s, respectively.

2.2. SNP analysis

A population of SNPs was first selected from markers shared by *G. mustelinum* and *G. tomentosum,* which means the polymorphism is identical for these two wild species relative to *G. hirsutum* probably due to the mutation in sequence in *G. hirsutum*. KASP assay primers for selected SNPs (Table 2.2) were designed according to the D5 scaffolds version 2.1 (Lin *et al.* 2010) based on the information from the CottonGen database. For each SNP, two forward primers and a shared reverse primer were designed and synthesized by a commercial provider (Integrated DNA Technologies, Inc., Coralville, Iowa).

Primer sets were pre-tested on a very small panel consisting of parental species and a pair of reciprocal F1 hybrids. The validation of successfully pre-tested SNP markers was conducted by using KBiosciences' Competitive Allele Specific PCR KASPar combined with the SNP line platform (SNP line XL; http://www.kbioscience.co.uk). Markers were considered validated if they clustered into three groups based on a test panel that included the parental species, F1 and the respective set of 20 or 27 BC1F1 plants. Validated markers were retained for chromosomal localization.

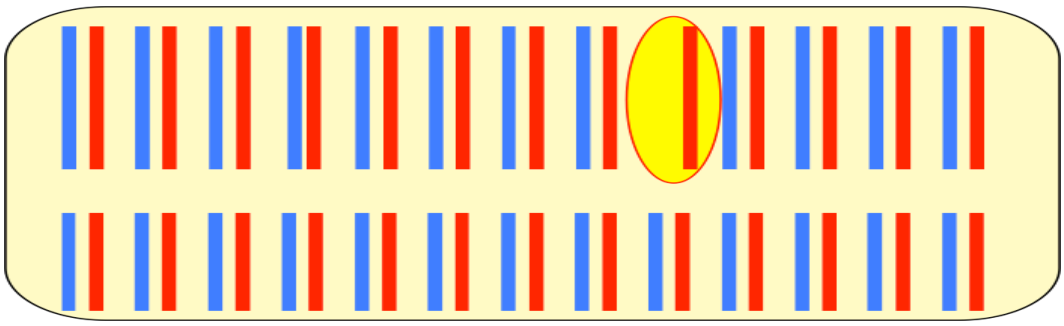

Figure 2.3 Monosomic interspecific F1 hybrid hemizygous for one alien chromosome

Hypoaneuploid *G. hirsutum* plants that lack specific chromosomes or chromosome arms have been identified based on phenotypic syndromes and conventional meiotic metaphase I configuration analysis of acetocarmine-stained microsporocytes (Saha *et al.* 2006). These stocks were utilized in our research to construct monosomic F1 hybrids by crossing the *G. hirsutum* hypoaneuploids with *G. mustelinum* and *G. tomentosum*. The resultant monosomic interspecific F1 hybrid is heterozygous for SNPs in 25 of the 26 chromosome pairs, but completely lacks one *G. hirsutum* chromosome and instead contains just one copy of the alien homologous chromosome, i.e., from *G. mustelinum* or *G. tomentosum*. Therefore, all SNP loci in that alien chromosome will be hemizygous for alien allele. Markers were selected for loci in six target chromosomes: 3, 9, 19, 20, 25, and 26.

For each species, the validated and localized markers were used for genetic mapping based on segregation in respective BC1F1 populations. Population sizes are 73 and 60 for *G. mustelinum* and *G. tomentosum,* respectively.

TABLE 2.2
Identities of SNP loci subjected to KASP primer design, and their mapped sequence locations in the reference D5 genome assembly (version 2.1)

SNP ID	Start	*Gossypium raimondii* Scaffold	Corresponding D-subgenome Chromosome	Corresponding A-subgenome Chromosome
mus001141tom000908	108618	*D05*	Chr 14	Chr 2/3
mus001140tom000906	109281	*D05*	Chr 14	Chr 2/3
mus001139tom000905	109824	*D05*	Chr 14	Chr 2/3
mus001691tom001379	123333	*D05*	Chr 14	Chr 2/3
mus015622tom013057	548039	*D05*	Chr 14	Chr 2/3
mus023303tom019318	2094854	*D05*	Chr 14	Chr 2/3
mus016050tom013383	4320302	*D05*	Chr 14	Chr 2/3
mus000792tom000613	5042851	*D05*	Chr 14	Chr 2/3
mus003477tom002946	5847254	*D05*	Chr 14	Chr 2/3

TABLE 2.2 Continued

SNP ID	Start	*Gossypium raimondii* Scaffold	Corresponding D-subgenome Chromosome	Corresponding A-subgenome Chromosome
mus015822tom013213	5849078	*D05*	Chr 14	Chr 2/3
mus016251tom013560	6840964	*D05*	Chr 14	Chr 2/3
mus004293tom003597	7732531	*D05*	Chr 14	Chr 2/3
mus010572tom008798	8674953	*D05*	Chr 14	Chr 2/3
mus000315tom000228	8863698	*D05*	Chr 14	Chr 2/3
mus004077tom003405	10003030	*D05*	Chr 14	Chr 2/3
mus001813tom001491	10394173	*D05*	Chr 14	Chr 2/3
mus005007tom004199	11973930	*D05*	Chr 14	Chr 2/3
mus000448tom000339	12074652	*D05*	Chr 14	Chr 2/3
mus001278tom001040	12225094	*D05*	Chr 14	Chr 2/3
mus008943tom007467	12887179	*D05*	Chr 14	Chr 2/3
mus000613tom000476	13128788	*D05*	Chr 14	Chr 2/3
mus016876tom014045	13128788	*D05*	Chr 14	Chr 2/3
mus014845tom012384	16029864	*D05*	Chr 14	Chr 2/3
mus004741tom003970	17821955	*D05*	Chr 14	Chr 2/3
mus014756tom012291	17822003	*D05*	Chr 14	Chr 2/3
mus000765tom000585	18619677	*D05*	Chr 14	Chr 2/3
mus015112tom012593	19152883	*D05*	Chr 14	Chr 2/3
mus003022tom002558	25152798	*D05*	Chr 14	Chr 2/3
mus005859tom004925	25602848	*D05*	Chr 14	Chr 2/3
mus002204tom001849	29003174	*D05*	Chr 14	Chr 2/3
mus010105tom008407	34531944	*D05*	Chr 14	Chr 2/3
mus015921tom013283	39205476	*D05*	Chr 14	Chr 2/3
mus001099tom000875	40835735	*D05*	Chr 14	Chr 2/3
mus004933tom004139	41040112	*D05*	Chr 14	Chr 2/3
mus003408tom002878	45808833	*D05*	Chr 14	Chr 2/3
mus008133tom006804	46638238	*D05*	Chr 14	Chr 2/3
mus010691tom008910	49797959	*D05*	Chr 14	Chr 2/3
mus016330tom013634	53362105	*D05*	Chr 14	Chr 2/3
mus003630tom003065	53533527	*D05*	Chr 14	Chr 2/3

TABLE 2.2 Continued

SNP ID	Start	*Gossypium raimondii* Scaffold	Corresponding D-subgenome Chromosome	Corresponding A-subgenome Chromosome
mus011138tom009340	56033907	*D05*	Chr 14	Chr 2/3
mus003215tom002726	56035316	*D05*	Chr 14	Chr 2/3
mus014104tom011779	56035316	*D05*	Chr 14	Chr 2/3
mus006426tom005388	57101308	*D05*	Chr 14	Chr 2/3
mus001077tom000854	57313643	*D05*	Chr 14	Chr 2/3
mus017304tom014387	57648113	*D05*	Chr 14	Chr 2/3
mus002236tom001874	57794347	*D05*	Chr 14	Chr 2/3
mus013634tom011379	57869439	*D05*	Chr 14	Chr 2/3
mus013201tom011026	59317038	*D05*	Chr 14	Chr 2/3
mus012576tom010531	59421599	*D05*	Chr 14	Chr 2/3
mus000449tom000340	60759804	*D05*	Chr 14	Chr 2/3
mus009359tom007833	60922549	*D05*	Chr 14	Chr 2/3
mus020619tom017000	61353535	*D05*	Chr 14	Chr 2/3
mus003066tom002589	61396031	*D05*	Chr 14	Chr 2/3
mus011511tom009646	61527206	*D05*	Chr 14	Chr 2/3
mus016766tom013960	62835349	*D05*	Chr 14	Chr 2/3
mus017319tom014393	63026167	*D05*	Chr 14	Chr 2/3
mus003452tom002927	63062781	*D05*	Chr 14	Chr 2/3
mus004743tom003972	63139233	*D05*	Chr 14	Chr 2/3
mus004742tom003971	63139390	*D05*	Chr 14	Chr 2/3
mus012104tom010121	63199531	*D05*	Chr 14	Chr 2/3
mus012105tom010122	63199643	*D05*	Chr 14	Chr 2/3
mus020913tom017234	63641942	*D05*	Chr 14	Chr 2/3
mus002363tom001989	63687968	*D05*	Chr 14	Chr 2/3
mus009324tom007806	63919230	*D05*	Chr 14	Chr 2/3
mus014101tom011776	63978770	*D05*	Chr 14	Chr 2/3
mus020522tom016915	2120285	*D02*	Chr 15	Chr 1
mus010091tom008387	508444	*D01*	Chr 16	Chr 7
mus000811tom000631	166882	*D03*	Chr 17	Chr 2/3
mus009525tom007966	336797	*D03*	Chr 17	Chr 2/3
mus013025tom010880	596850	*D03*	Chr 17	Chr 2/3

TABLE 2.2 Continued

SNP ID	Start	*Gossypium raimondii* Scaffold	Corresponding D-subgenome Chromosome	Corresponding A-subgenome Chromosome
mus001315tom001084	642772	*D03*	Chr 17	Chr 2/3
mus001995tom001656	880797	*D03*	Chr 17	Chr 2/3
mus004219tom003527	2373031	*D03*	Chr 17	Chr 2/3
mus019289tom015960	2742505	*D03*	Chr 17	Chr 2/3
mus018092tom015045	3969478	*D03*	Chr 17	Chr 2/3
mus020316tom016757	5288588	*D03*	Chr 17	Chr 2/3
mus020316tom016757	5288588	*D03*	Chr 17	Chr 2/3
mus011074tom009293	5333966	*D03*	Chr 17	Chr 2/3
mus021232tom017466	8078303	*D03*	Chr 17	Chr 2/3
mus013928tom011654	8424321	*D03*	Chr 17	Chr 2/3
mus017873tom014865	8424321	*D03*	Chr 17	Chr 2/3
mus016533tom013792	8514716	*D03*	Chr 17	Chr 2/3
mus003533tom002990	8885051	*D03*	Chr 17	Chr 2/3
mus011045tom009273	10000984	*D03*	Chr 17	Chr 2/3
mus006880tom005759	15432755	*D03*	Chr 17	Chr 2/3
mus022045tom018199	15437597	*D03*	Chr 17	Chr 2/3
mus014806tom012334	15437640	*D03*	Chr 17	Chr 2/3
mus009069tom007563	19955048	*D03*	Chr 17	Chr 2/3
mus017765tom014781	21713959	*D03*	Chr 17	Chr 2/3
mus008024tom006689	29867026	*D03*	Chr 17	Chr 2/3
mus000478tom000365	30491031	*D03*	Chr 17	Chr 2/3
mus017876tom014869	30562262	*D03*	Chr 17	Chr 2/3
mus012134tom010143	30639852	*D03*	Chr 17	Chr 2/3
mus010725tom008941	33300315	*D03*	Chr 17	Chr 2/3
mus010713tom008931	34430355	*D03*	Chr 17	Chr 2/3
mus003387tom002860	35043585	*D03*	Chr 17	Chr 2/3
mus003673tom003098	35589445	*D03*	Chr 17	Chr 2/3
mus006394tom005361	35727732	*D03*	Chr 17	Chr 2/3
mus014181tom011837	37262057	*D03*	Chr 17	Chr 2/3
mus008624tom007195	40324950	*D03*	Chr 17	Chr 2/3
mus005739tom004835	40439763	*D03*	Chr 17	Chr 2/3
mus018083tom015036	40724020	*D03*	Chr 17	Chr 2/3

TABLE 2.2 Continued

SNP ID	Start	*Gossypium raimondii* Scaffold	Corresponding D-subgenome Chromosome	Corresponding A-subgenome Chromosome
mus010888tom009100	41211242	*D03*	Chr 17	Chr 2/3
mus014067tom011760	41211261	*D03*	Chr 17	Chr 2/3
mus023277tom019280	41211261	*D03*	Chr 17	Chr 2/3
mus023116tom019128	41414068	*D03*	Chr 17	Chr 2/3
mus010980tom009196	41887973	*D03*	Chr 17	Chr 2/3
mus009217tom007710	42010942	*D03*	Chr 17	Chr 2/3
mus018798tom015617	42357131	*D03*	Chr 17	Chr 2/3
mus012233tom010249	42745658	*D03*	Chr 17	Chr 2/3
mus000625tom000482	42754901	*D03*	Chr 17	Chr 2/3
mus004530tom003788	43701792	*D03*	Chr 17	Chr 2/3
mus004529tom003787	43701876	*D03*	Chr 17	Chr 2/3
mus011666tom009785	43904672	*D03*	Chr 17	Chr 2/3
mus021276tom017510	44864920	*D03*	Chr 17	Chr 2/3
mus001501tom001212	45342121	*D03*	Chr 17	Chr 2/3
mus001935tom001597	45390367	*D03*	Chr 17	Chr 2/3
mus012935tom010810	45650694	*D03*	Chr 17	Chr 2/3
mus000724tom000551	135593	*D09*	Chr 19	Chr 4/5
mus010215tom008482	412197	*D09*	Chr 19	Chr 4/5
mus003469tom002941	496010	*D09*	Chr 19	Chr 4/5
mus006467tom005430	801657	*D09*	Chr 19	Chr 4/5
mus008787tom007341	955195	*D09*	Chr 19	Chr 4/5
mus002827tom002395	997894	*D09*	Chr 19	Chr 4/5
mus002041tom001715	1232254	*D09*	Chr 19	Chr 4/5
mus023408tom019407	1232254	*D09*	Chr 19	Chr 4/5
mus005528tom004631	1380703	*D09*	Chr 19	Chr 4/5
mus020853tom017185	1883054	*D09*	Chr 19	Chr 4/5
mus022120tom018262	1883780	*D09*	Chr 19	Chr 4/5
mus020884tom017212	1883897	*D09*	Chr 19	Chr 4/5
mus006540tom005500	2205430	*D09*	Chr 19	Chr 4/5
mus013243tom011056	2457442	*D09*	Chr 19	Chr 4/5
mus006090tom005103	3109846	*D09*	Chr 19	Chr 4/5
mus000819tom000639	3221080	*D09*	Chr 19	Chr 4/5

TABLE 2.2 Continued

SNP ID	Start	*Gossypium raimondii* Scaffold	Corresponding D-subgenome Chromosome	Corresponding A-subgenome Chromosome
mus000033tom000022	3445801	*D09*	Chr 19	Chr 4/5
mus002255tom001881	3582134	*D09*	Chr 19	Chr 4/5
mus016179tom013508	3592297	*D09*	Chr 19	Chr 4/5
mus018987tom015756	3592297	*D09*	Chr 19	Chr 4/5
mus015132tom012611	3592317	*D09*	Chr 19	Chr 4/5
mus001180tom000946	3912264	*D09*	Chr 19	Chr 4/5
mus010766tom008988	4089564	*D09*	Chr 19	Chr 4/5
mus015463tom012900	4504901	*D09*	Chr 19	Chr 4/5
mus006662tom005587	4596205	*D09*	Chr 19	Chr 4/5
mus013464tom011246	4833647	*D09*	Chr 19	Chr 4/5
mus001316tom001086	4847289	*D09*	Chr 19	Chr 4/5
mus011595tom009721	4847289	*D09*	Chr 19	Chr 4/5
mus001317tom001087	4847381	*D09*	Chr 19	Chr 4/5
mus011596tom009722	4847381	*D09*	Chr 19	Chr 4/5
mus005241tom004411	4959739	*D09*	Chr 19	Chr 4/5
mus015140tom012618	4983928	*D09*	Chr 19	Chr 4/5
mus008506tom007098	5215165	*D09*	Chr 19	Chr 4/5
mus017126tom014236	5371615	*D09*	Chr 19	Chr 4/5
mus006924tom005790	5510225	*D09*	Chr 19	Chr 4/5
mus016212tom013533	5908663	*D09*	Chr 19	Chr 4/5
mus015743tom013147	6360245	*D09*	Chr 19	Chr 4/5
mus018709tom015564	6360245	*D09*	Chr 19	Chr 4/5
mus017566tom014582	6771965	*D09*	Chr 19	Chr 4/5
mus009530tom007969	6804630	*D09*	Chr 19	Chr 4/5
mus009527tom007968	6905126	*D09*	Chr 19	Chr 4/5
mus001063tom000846	7089955	*D09*	Chr 19	Chr 4/5
mus009123tom007619	7210389	*D09*	Chr 19	Chr 4/5
mus004133tom003449	7256115	*D09*	Chr 19	Chr 4/5
mus014110tom011785	7509047	*D09*	Chr 19	Chr 4/5
mus012952tom010834	7649894	*D09*	Chr 19	Chr 4/5
mus004533tom003789	7708047	*D09*	Chr 19	Chr 4/5
mus009543tom007975	8062220	*D09*	Chr 19	Chr 4/5

TABLE 2.2 Continued

SNP ID	Start	*Gossypium raimondii* Scaffold	Corresponding D-subgenome Chromosome	Corresponding A-subgenome Chromosome
mus010226tom008494	8163172	*D09*	Chr 19	Chr 4/5
mus013151tom010985	8706840	*D09*	Chr 19	Chr 4/5
mus013150tom010984	8707133	*D09*	Chr 19	Chr 4/5
mus020092tom016589	8838922	*D09*	Chr 19	Chr 4/5
mus003493tom002966	9427964	*D09*	Chr 19	Chr 4/5
mus017688tom014723	9787626	*D09*	Chr 19	Chr 4/5
mus001959tom001623	9893287	*D09*	Chr 19	Chr 4/5
mus014238tom011909	10032485	*D09*	Chr 19	Chr 4/5
mus018649tom015501	10174839	*D09*	Chr 19	Chr 4/5
mus014103tom011778	11090821	*D09*	Chr 19	Chr 4/5
mus009682tom008068	11180561	*D09*	Chr 19	Chr 4/5
mus014518tom012110	11193610	*D09*	Chr 19	Chr 4/5
mus013531tom011307	11634122	*D09*	Chr 19	Chr 4/5
mus005265tom004432	12124058	*D09*	Chr 19	Chr 4/5
mus007440tom006217	12553876	*D09*	Chr 19	Chr 4/5
mus006558tom005515	12554140	*D09*	Chr 19	Chr 4/5
mus011982tom010019	12619863	*D09*	Chr 19	Chr 4/5
mus006859tom005737	12858759	*D09*	Chr 19	Chr 4/5
mus005751tom004839	13256334	*D09*	Chr 19	Chr 4/5
mus019931tom016454	13356012	*D09*	Chr 19	Chr 4/5
mus016131tom013445	13428147	*D09*	Chr 19	Chr 4/5
mus000908tom000718	13625054	*D09*	Chr 19	Chr 4/5
mus023167tom019173	13672118	*D09*	Chr 19	Chr 4/5
mus001568tom001286	13827861	*D09*	Chr 19	Chr 4/5
mus007950tom006621	13866708	*D09*	Chr 19	Chr 4/5
mus001852tom001523	14247181	*D09*	Chr 19	Chr 4/5
mus001851tom001522	14249415	*D09*	Chr 19	Chr 4/5
mus007477tom006246	14556343	*D09*	Chr 19	Chr 4/5
mus020538tom016927	14998459	*D09*	Chr 19	Chr 4/5
mus004747tom003977	15775545	*D09*	Chr 19	Chr 4/5
mus015471tom012907	16133074	*D09*	Chr 19	Chr 4/5
mus012120tom010133	16308415	*D09*	Chr 19	Chr 4/5

TABLE 2.2 Continued

SNP ID	Start	*Gossypium raimondii* Scaffold	Corresponding D-subgenome Chromosome	Corresponding A-subgenome Chromosome
mus019907tom016435	16733631	*D09*	Chr 19	Chr 4/5
mus011849tom009920	16915453	*D09*	Chr 19	Chr 4/5
mus021704tom017888	17087565	*D09*	Chr 19	Chr 4/5
mus003570tom003016	17695063	*D09*	Chr 19	Chr 4/5
mus023155tom019155	17695063	*D09*	Chr 19	Chr 4/5
mus009177tom007674	17734998	*D09*	Chr 19	Chr 4/5
mus016768tom013961	18046328	*D09*	Chr 19	Chr 4/5
mus008915tom007431	18046609	*D09*	Chr 19	Chr 4/5
mus006519tom005481	18590053	*D09*	Chr 19	Chr 4/5
mus021617tom017802	18844057	*D09*	Chr 19	Chr 4/5
mus013994tom011712	19345682	*D09*	Chr 19	Chr 4/5
mus017658tom014688	19481418	*D09*	Chr 19	Chr 4/5
mus001183tom000949	19500029	*D09*	Chr 19	Chr 4/5
mus009062tom007558	19818082	*D09*	Chr 19	Chr 4/5
mus002470tom002082	20094178	*D09*	Chr 19	Chr 4/5
mus021946tom018107	20205767	*D09*	Chr 19	Chr 4/5
mus011154tom009351	22193514	*D09*	Chr 19	Chr 4/5
mus020405tom016836	22419642	*D09*	Chr 19	Chr 4/5
mus018022tom014982	22999428	*D09*	Chr 19	Chr 4/5
mus002389tom002010	23003969	*D09*	Chr 19	Chr 4/5
mus019502tom016117	23037707	*D09*	Chr 19	Chr 4/5
mus002303tom001932	23060513	*D09*	Chr 19	Chr 4/5
mus013748tom011495	23197909	*D09*	Chr 19	Chr 4/5
mus003053tom002577	23443067	*D09*	Chr 19	Chr 4/5
mus001536tom001254	24948642	*D09*	Chr 19	Chr 4/5
mus001260tom001020	25884141	*D09*	Chr 19	Chr 4/5
mus010812tom009031	27092903	*D09*	Chr 19	Chr 4/5
mus005581tom004672	27767144	*D09*	Chr 19	Chr 4/5
mus000058tom000040	28445009	*D09*	Chr 19	Chr 4/5
mus022610tom018687	28445009	*D09*	Chr 19	Chr 4/5
mus001200tom000967	39825320	*D09*	Chr 19	Chr 4/5
mus018352tom015261	40104820	*D09*	Chr 19	Chr 4/5

TABLE 2.2 Continued

SNP ID	Start	*Gossypium raimondii* Scaffold	Corresponding D-subgenome Chromosome	Corresponding A-subgenome Chromosome
mus013156tom010988	41522771	*D09*	Chr 19	Chr 4/5
mus001644tom001332	49649921	*D09*	Chr 19	Chr 4/5
mus000717tom000547	52712750	*D09*	Chr 19	Chr 4/5
mus023195tom019198	54311432	*D09*	Chr 19	Chr 4/5
mus008938tom007456	56385253	*D09*	Chr 19	Chr 4/5
mus001015tom000800	62930255	*D09*	Chr 19	Chr 4/5
mus012539tom010501	64405458	*D09*	Chr 19	Chr 4/5
mus007708tom006440	64539818	*D09*	Chr 19	Chr 4/5
mus009240tom007744	66183193	*D09*	Chr 19	Chr 4/5
mus000013tom000009	68796971	*D09*	Chr 19	Chr 4/5
mus010771tom008990	69171590	*D09*	Chr 19	Chr 4/5
mus010961tom009169	69372567	*D09*	Chr 19	Chr 4/5
mus005797tom004881	70107071	*D09*	Chr 19	Chr 4/5
mus007405tom006183	70227405	*D09*	Chr 19	Chr 4/5
mus016334tom013641	70506896	*D09*	Chr 19	Chr 4/5
mus003124tom002644	681059	*D11*	Chr 20	Chr 10
mus014822tom012351	702273	*D11*	Chr 20	Chr 10
mus009657tom008051	748777	*D11*	Chr 20	Chr 10
mus023127tom019134	748777	*D11*	Chr 20	Chr 10
mus015169tom012644	1047253	*D11*	Chr 20	Chr 10
mus001908tom001559	1309190	*D11*	Chr 20	Chr 10
mus018082tom015034	1567478	*D11*	Chr 20	Chr 10
mus004510tom003766	2001156	*D11*	Chr 20	Chr 10
mus001220tom000981	2220176	*D11*	Chr 20	Chr 10
mus013862tom011597	3788247	*D11*	Chr 20	Chr 10
mus014399tom012021	4124331	*D11*	Chr 20	Chr 10
mus007139tom005998	4829636	*D11*	Chr 20	Chr 10
mus018406tom015309	5154998	*D11*	Chr 20	Chr 10
mus016323tom013623	5447507	*D11*	Chr 20	Chr 10
mus000092tom000067	5824923	*D11*	Chr 20	Chr 10
mus015018tom012525	6375502	*D11*	Chr 20	Chr 10
mus010449tom008677	12455258	*D11*	Chr 20	Chr 10

TABLE 2.2 Continued

SNP ID	Start	*Gossypium raimondii* Scaffold	Corresponding D-subgenome Chromosome	Corresponding A-subgenome Chromosome
mus010836tom009051	12572283	*D11*	Chr 20	Chr 10
mus006516tom005477	13133127	*D11*	Chr 20	Chr 10
mus010997tom009205	13326998	*D11*	Chr 20	Chr 10
mus005192tom004355	16175630	*D11*	Chr 20	Chr 10
mus002144tom001795	19400436	*D11*	Chr 20	Chr 10
mus014154tom011814	21492494	*D11*	Chr 20	Chr 10
mus007876tom006556	22506348	*D11*	Chr 20	Chr 10
mus003111tom002631	22579870	*D11*	Chr 20	Chr 10
mus001414tom001152	25843092	*D11*	Chr 20	Chr 10
mus001853tom001525	26498560	*D11*	Chr 20	Chr 10
mus017790tom014802	32884060	*D11*	Chr 20	Chr 10
mus007212tom006055	41016382	*D11*	Chr 20	Chr 10
mus003701tom003130	44303716	*D11*	Chr 20	Chr 10
mus008102tom006761	47319559	*D11*	Chr 20	Chr 10
mus010245tom008513	49816274	*D11*	Chr 20	Chr 10
mus000396tom000304	49858441	*D11*	Chr 20	Chr 10
mus010243tom008510	56185372	*D11*	Chr 20	Chr 10
mus018100tom015053	58191556	*D11*	Chr 20	Chr 10
mus007745tom006463	58435160	*D11*	Chr 20	Chr 10
mus023134tom019139	60031338	*D11*	Chr 20	Chr 10
mus004943tom004148	60335071	*D11*	Chr 20	Chr 10
mus007354tom006156	61136951	*D11*	Chr 20	Chr 10
mus008794tom007345	61203789	*D11*	Chr 20	Chr 10
mus004844tom004057	62106936	*D11*	Chr 20	Chr 10
mus012144tom010153	62601484	*D11*	Chr 20	Chr 10
mus003305tom002792	341729	*D12*	Chr 22	Chr 4/5
mus005823tom004904	469487	*D12*	Chr 22	Chr 4/5
mus006237tom005224	831171	*D12*	Chr 22	Chr 4/5
mus001982tom001646	1624105	*D12*	Chr 22	Chr 4/5
mus003228tom002744	2387631	*D12*	Chr 22	Chr 4/5
mus016143tom013482	2868673	*D12*	Chr 22	Chr 4/5
mus008141tom006812	3151338	*D12*	Chr 22	Chr 4/5

TABLE 2.2 Continued

SNP ID	Start	*Gossypium raimondii* Scaffold	Corresponding D-subgenome Chromosome	Corresponding A-subgenome Chromosome
mus002504tom002108	4444669	*D12*	Chr 22	Chr 4/5
mus012665tom010606	4607566	*D12*	Chr 22	Chr 4/5
mus009908tom008272	4759496	*D12*	Chr 22	Chr 4/5
mus017128tom014239	4864333	*D12*	Chr 22	Chr 4/5
mus014837tom012376	5162792	*D12*	Chr 22	Chr 4/5
mus007418tom006198	5284082	*D12*	Chr 22	Chr 4/5
mus015372tom012813	6123178	*D12*	Chr 22	Chr 4/5
mus021711tom017895	7516002	*D12*	Chr 22	Chr 4/5
mus007667tom006400	9191532	*D12*	Chr 22	Chr 4/5
mus000847tom000664	11905334	*D12*	Chr 22	Chr 4/5
mus003378tom002855	12694071	*D12*	Chr 22	Chr 4/5
mus003133tom002651	13715017	*D12*	Chr 22	Chr 4/5
mus012288tom010298	14043037	*D12*	Chr 22	Chr 4/5
mus019654tom016233	15729442	*D12*	Chr 22	Chr 4/5
mus009405tom007871	16703755	*D12*	Chr 22	Chr 4/5
mus011272tom009450	16704045	*D12*	Chr 22	Chr 4/5
mus004522tom003780	22392873	*D12*	Chr 22	Chr 4/5
mus015568tom013003	22392873	*D12*	Chr 22	Chr 4/5
mus003592tom003034	23690716	*D12*	Chr 22	Chr 4/5
mus016744tom013946	23690716	*D12*	Chr 22	Chr 4/5
mus006822tom005702	24020722	*D12*	Chr 22	Chr 4/5
mus003587tom003031	24495400	*D12*	Chr 22	Chr 4/5
mus022967tom019004	24495400	*D12*	Chr 22	Chr 4/5
mus001947tom001609	25137662	*D12*	Chr 22	Chr 4/5
mus011735tom009831	25138351	*D12*	Chr 22	Chr 4/5
mus000960tom000757	25141099	*D12*	Chr 22	Chr 4/5
mus000961tom000758	25141346	*D12*	Chr 22	Chr 4/5
mus010576tom008803	26243587	*D12*	Chr 22	Chr 4/5
mus021940tom018105	28641954	*D12*	Chr 22	Chr 4/5
mus021941tom018106	28642038	*D12*	Chr 22	Chr 4/5
mus020043tom016544	29080520	*D12*	Chr 22	Chr 4/5
mus008306tom006933	29995338	*D12*	Chr 22	Chr 4/5

TABLE 2.2 Continued

SNP ID	Start	*Gossypium raimondii* Scaffold	Corresponding D-subgenome Chromosome	Corresponding A-subgenome Chromosome
mus014662tom012208	31001321	*D12*	Chr 22	Chr 4/5
mus006292tom005278	32303944	*D12*	Chr 22	Chr 4/5
mus012170tom010189	32304175	*D12*	Chr 22	Chr 4/5
mus022253tom018375	32304175	*D12*	Chr 22	Chr 4/5
mus006449tom005413	32902102	*D12*	Chr 22	Chr 4/5
mus002475tom002089	34554440	*D12*	Chr 22	Chr 4/5
mus022348tom018463	35325579	*D12*	Chr 22	Chr 4/5
mus018007tom014958	1234895	*D06*	Chr 23	Chr 9
mus000086tom000060	1312673	*D06*	Chr 23	Chr 9
mus007474tom006242	2278181	*D06*	Chr 23	Chr 9
mus001243tom000999	13647845	*D06*	Chr 23	Chr 9
mus017429tom014461	16270682	*D06*	Chr 23	Chr 9
mus020181tom016650	16270682	*D06*	Chr 23	Chr 9
mus003065tom002587	16391851	*D06*	Chr 23	Chr 9
mus000278tom000204	19036183	*D06*	Chr 23	Chr 9
mus010326tom008567	22358559	*D06*	Chr 23	Chr 9
mus020387tom016826	31692338	*D06*	Chr 23	Chr 9
mus000361tom000278	32103058	*D06*	Chr 23	Chr 9
mus012291tom010303	33803123	*D06*	Chr 23	Chr 9
mus012292tom010304	33803251	*D06*	Chr 23	Chr 9
mus020749tom017109	33814047	*D06*	Chr 23	Chr 9
mus002624tom002215	33880774	*D06*	Chr 23	Chr 9
mus001949tom001611	34141395	*D06*	Chr 23	Chr 9
mus011741tom009842	34498890	*D06*	Chr 23	Chr 9
mus018310tom015231	34674143	*D06*	Chr 23	Chr 9
mus009913tom008274	35975682	*D06*	Chr 23	Chr 9
mus008327tom006950	36081476	*D06*	Chr 23	Chr 9
mus013175tom011009	36214034	*D06*	Chr 23	Chr 9
mus011450tom009592	38601936	*D06*	Chr 23	Chr 9
mus003614tom003052	39589590	*D06*	Chr 23	Chr 9
mus006883tom005760	40592553	*D06*	Chr 23	Chr 9
mus021702tom017887	40612534	*D06*	Chr 23	Chr 9

TABLE 2.2 Continued

SNP ID	Start	*Gossypium raimondii* Scaffold	Corresponding D-subgenome Chromosome	Corresponding A-subgenome Chromosome
mus006251tom005232	40612628	*D06*	Chr 23	Chr 9
mus001165tom000927	40825897	*D06*	Chr 23	Chr 9
mus017006tom014136	41051501	*D06*	Chr 23	Chr 9
mus006513tom005472	41548877	*D06*	Chr 23	Chr 9
mus011908tom009966	42172196	*D06*	Chr 23	Chr 9
mus015813tom013210	42200673	*D06*	Chr 23	Chr 9
mus019201tom015899	42200673	*D06*	Chr 23	Chr 9
mus021847tom018026	42545392	*D06*	Chr 23	Chr 9
mus011623tom009747	43335970	*D06*	Chr 23	Chr 9
mus000616tom000477	43492849	*D06*	Chr 23	Chr 9
mus018413tom015323	44339716	*D06*	Chr 23	Chr 9
mus002559tom002143	44709825	*D06*	Chr 23	Chr 9
mus020841tom017171	45035283	*D06*	Chr 23	Chr 9
mus016718tom013930	45809496	*D06*	Chr 23	Chr 9
mus012450tom010430	46357890	*D06*	Chr 23	Chr 9
mus010330tom008570	46617637	*D06*	Chr 23	Chr 9
mus008083tom006748	47090652	*D06*	Chr 23	Chr 9
mus016551tom013804	47463872	*D06*	Chr 23	Chr 9
mus007031tom005895	47561355	*D06*	Chr 23	Chr 9
mus014129tom011797	47911126	*D06*	Chr 23	Chr 9
mus000282tom000207	47962461	*D06*	Chr 23	Chr 9
mus000281tom000206	47962524	*D06*	Chr 23	Chr 9
mus004790tom004013	48125859	*D06*	Chr 23	Chr 9
mus018233tom015167	48327901	*D06*	Chr 23	Chr 9
mus009242tom007746	48661571	*D06*	Chr 23	Chr 9
mus003400tom002867	48873472	*D06*	Chr 23	Chr 9
mus006288tom005272	49148306	*D06*	Chr 23	Chr 9
mus013188tom011022	49352595	*D06*	Chr 23	Chr 9
mus022383tom018487	49352595	*D06*	Chr 23	Chr 9
mus013962tom011687	49360216	*D06*	Chr 23	Chr 9
mus005304tom004466	49838336	*D06*	Chr 23	Chr 9
mus004928tom004134	50749600	*D06*	Chr 23	Chr 9

TABLE 2.2 Continued

SNP ID	Start	*Gossypium raimondii* Scaffold	Corresponding D-subgenome Chromosome	Corresponding A-subgenome Chromosome
mus016936tom014081	50860860	*D06*	Chr 23	Chr 9
mus001583tom001296	320867	*D10*	Chr 25	Chr 6
mus013449tom011225	320898	*D10*	Chr 25	Chr 6
mus012716tom010652	3150292	*D10*	Chr 25	Chr 6
mus021932tom018097	4349650	*D10*	Chr 25	Chr 6
mus001374tom001128	5276156	*D10*	Chr 25	Chr 6
mus000088tom000061	6734021	*D10*	Chr 25	Chr 6
mus022477tom018558	7080140	*D10*	Chr 25	Chr 6
mus016403tom013702	7130173	*D10*	Chr 25	Chr 6
mus020816tom017160	7978136	*D10*	Chr 25	Chr 6
mus006927tom005791	8209715	*D10*	Chr 25	Chr 6
mus008067tom006730	8502038	*D10*	Chr 25	Chr 6
mus011498tom009636	11815648	*D10*	Chr 25	Chr 6
mus021161tom017418	12394417	*D10*	Chr 25	Chr 6
mus006639tom005574	12676342	*D10*	Chr 25	Chr 6
mus013035tom010886	13708858	*D10*	Chr 25	Chr 6
mus007161tom006014	17153084	*D10*	Chr 25	Chr 6
mus000329tom000241	18667891	*D10*	Chr 25	Chr 6
mus017937tom014908	20116330	*D10*	Chr 25	Chr 6
mus001964tom001627	20595773	*D10*	Chr 25	Chr 6
mus017966tom014928	22786741	*D10*	Chr 25	Chr 6
mus000990tom000780	27209925	*D10*	Chr 25	Chr 6
mus009740tom008132	32124135	*D10*	Chr 25	Chr 6
mus023084tom019102	32125031	*D10*	Chr 25	Chr 6
mus002787tom002363	49416484	*D10*	Chr 25	Chr 6
mus006746tom005653	53642480	*D10*	Chr 25	Chr 6
mus002345tom001976	54412682	*D10*	Chr 25	Chr 6
mus011501tom009638	57827312	*D10*	Chr 25	Chr 6
mus007750tom006471	57841823	*D10*	Chr 25	Chr 6
mus017268tom014362	58112088	*D10*	Chr 25	Chr 6
mus017721tom014752	58554654	*D10*	Chr 25	Chr 6
mus022210tom018331	59313910	*D10*	Chr 25	Chr 6

TABLE 2.2 Continued

SNP ID	Start	*Gossypium raimondii* Scaffold	Corresponding D-subgenome Chromosome	Corresponding A-subgenome Chromosome
mus006515tom005476	59989200	*D10*	Chr 25	Chr 6
mus005678tom004761	60134500	*D10*	Chr 25	Chr 6
mus013284tom011088	61696112	*D10*	Chr 25	Chr 6
mus014417tom012043	61775347	*D10*	Chr 25	Chr 6
mus016003tom013332	4474932	*D08*	Chr 26	Chr 12
mus020414tom016844	4474932	*D08*	Chr 26	Chr 12
mus002489tom002095	5745082	*D08*	Chr 26	Chr 12
mus009996tom008337	10811871	*D08*	Chr 26	Chr 12
mus009997tom008338	10812663	*D08*	Chr 26	Chr 12
mus007916tom006597	11719943	*D08*	Chr 26	Chr 12
mus019662tom016241	15212477	*D08*	Chr 26	Chr 12
mus002749tom002326	15656805	D08	Chr 26	Chr 12
mus013727tom011475	26905047	D08	Chr 26	Chr 12
mus015045tom012541	27811057	D08	Chr 26	Chr 12
mus000781tom000606	34562919	D08	Chr 26	Chr 12
mus000552tom000443	35710400	D08	Chr 26	Chr 12
mus004954tom004157	37003483	D08	Chr 26	Chr 12
mus004325tom003623	41331141	D08	Chr 26	Chr 12
mus021453tom017667	42539765	D08	Chr 26	Chr 12
mus000884tom000692	45138869	D08	Chr 26	Chr 12
mus010232tom008496	47240232	D08	Chr 26	Chr 12
mus014410tom012036	48883736	D08	Chr 26	Chr 12
mus001106tom000880	49063174	D08	Chr 26	Chr 12
mus001091tom000867	50035104	D08	Chr 26	Chr 12
mus002553tom002137	50035104	D08	Chr 26	Chr 12
mus010882tom009088	51131279	D08	Chr 26	Chr 12
mus009664tom008056	53747637	D08	Chr 26	Chr 12
mus000775tom000598	54150000	D08	Chr 26	Chr 12
mus008501tom007092	54786676	D08	Chr 26	Chr 12
mus004458tom003723	54854569	D08	Chr 26	Chr 12
mus001752tom001439	55016092	D08	Chr 26	Chr 12
mus010199tom008474	55644188	D08	Chr 26	Chr 12

TABLE 2.2 Continued

SNP ID	Start	*Gossypium raimondii* Scaffold	Corresponding D-subgenome Chromosome	Corresponding A-subgenome Chromosome
mus021741tom017933	56620114	D08	Chr 26	Chr 12
mus003509tom002975	57002569	D08	Chr 26	Chr 12

High quality DNA samples were extracted from young leaves of plants with DNeasy Plant mini kit (Qiagen) and/or NucleoSpin Plant II kit (Macherey-Nagel Inc.). For each DNA sample, concentration and wavelength ratio were estimated by NanoDrop2000 (Supplementary Material 1). The tested DNA samples were then diluted to 10 ng/ul for genotyping. SNP analysis was carried out with the same method for marker validation. Amplification of DNA was performed with the Hydrocycler^{TM} (LGC), i.e., a type of thermo-cycler for PCR.

2.3. Comparative analysis for transmission rate

The significance of differences due to potential effects of the wild species on marker transmission rates was judged according to t-tests that compared the two population sets. For each set of populations within a species, variance component analysis of transmission rates was conducted for three factors, namely location, cross direction and generation.

A general linear model (GLM) was used to analyze the SNP data in SAS 9.3. The GLM model was:

$$R_{ijkl} = \mu + G_i + D_j + L_k + \varepsilon_{ijkl}$$

The transmission rate for each SNP marker and the overall mean is denoted with R_{ijkl} and μ, respectively. G_i is the effect of generation with three levels, i.e., BC1F1, BC2F1 or

BC3F1. D_j is the effect of cross direction and it has two levels (i.e. *G. mustelinum* x *G. hirsutum*, *G. hirsutum* x *G. mustelinum*, *G. tomentosum* x *G. hirsutum*, *G. hirsutum* x *G. tomentosum*). L_k symbolizes the effects of location and it has two levels (i.e. field and greenhouse). ε_{ijkl} is the random error in each measure. The model does not account for possible interactions between these variables. Another assumption for this model is that some location-dependent effects remain constant across years.

CHAPTER III
RESULTS AND DISCUSSION

3.1. Marker localization and validation

SNP markers were validated by population-based genotyping. Polymorphism of the marker was confirmed when three clearly clustered groups were generated from BC1F1 individuals, parents and F1 hybrids as shown in Figure 3.1. Eventually, 67 SNP markers were developed and validated for our target chromosomes.

SNP markers were localized by comparative analysis of monosomic interspecific F1 hybrids as exemplified in Figure 3.2. In the graph, the red circle at "F11" (5th row, 11th column) depicts genotype of the monosomic F1 hybrid that lacks chromosome 9 of *G. hirsutum* but contains one copy of chromosome 9 from *G. mustelinum*. The expectation is that for any SNP locus for which a monosomic F1 hybrid differs from the euploid F1 hybrids and instead has the same genotype as *G. mustelinum* parent, the tested SNP marker is likely localized on chromosome 9. This kind of localization analysis has sometimes been called "deficiency analysis" or "loss of heterozygosity (LOH)" (Gutiérrez *et al.* 2009).

The results of validation and localization for each marker are summarized in Table 3.1. Detailed information about these markers is provided in Supplementary Material 2.

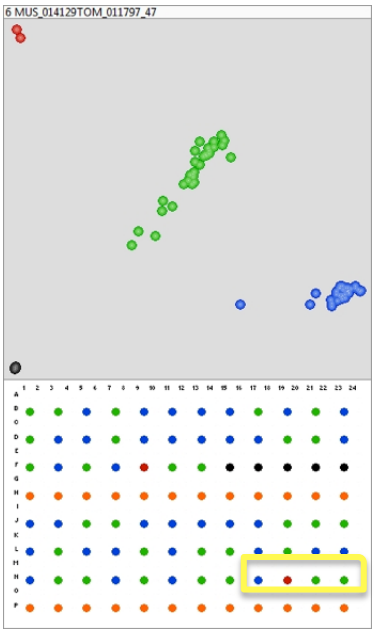

Figure 3.1 Validation of SNP marker mus014129tom011797 by KASP analysis of parents, reciprocal F1 hybrids (left-to-right in box) and 63 BC1F1 segregates. Black dots denote non-template (water) controls. Strong clustering confirmed this marker as robust.

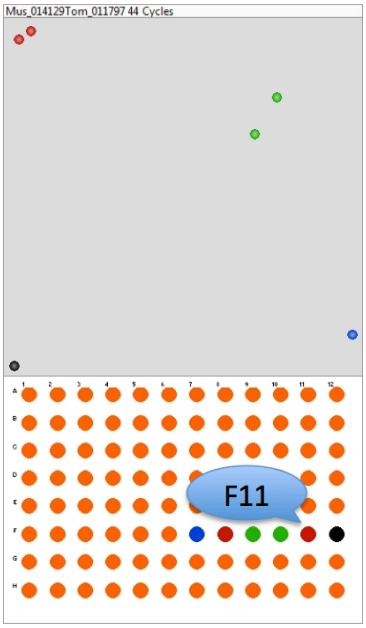

Figure 3.2 Localization of SNP marker mus014129tom011797 by KASP analysis of (left to right) both parents, two reciprocal F1 hybrids and a monosomic interspecific F1 hybrid that lacks chromosome 9 (coordinate "F11") of *G. hirsutum* and is hemizygous for SNPs of *G. tomentosum*.

TABLE 3.1
The SNP marker localization based on hemizygosity of hypoaneuploid interspecific F1 hybrids

SNP identity	*Gossypium raimondii* scaffold	D-genome chromosome	A-genome homolog	Inferred position
mus003022tom002558	D05	Chr 14	Chr 2/3	3
mus004933tom004139	D05	Chr 14	Chr 2/3	3
mus011138tom009340	D05	Chr 14	Chr 2/3	3
mus016330tom013634	D05	Chr 14	Chr 2/3	3
mus014756tom012291	D05	Chr 14	Chr 2/3	3
mus003387tom002860	D03	Chr 17	Chr 2/3	3
mus012233tom010249	D03	Chr 17	Chr 2/3	3
mus001949tom001611	D06	Chr 23	Chr 9	9
mus011741tom009842	D06	Chr 23	Chr 9	N/A
mus006513tom005472	D06	Chr 23	Chr 9	9
mus018007tom014958	D06	Chr 23	Chr 9	9
mus001243tom000999	D06	Chr 23	Chr 9	9
mus020749tom017109	D06	Chr 23	Chr 9	9
mus006883tom005760	D06	Chr 23	Chr 9	9
mus010330tom008570	D06	Chr 23	Chr 9	9
mus014129tom011797	D06	Chr 23	Chr 9	9
mus009242tom007746	D06	Chr 23	Chr 9	9
mus022383tom018487	D06	Chr 23	Chr 9	9
mus006288tom005272	D06	Chr 23	Chr 9	9
mus013188tom011022	D06	Chr 23	Chr 9	9
mus000717tom000547	D09	Chr 19	Chr 4/5	19
mus010961tom009169	D09	Chr 19	Chr 4/5	19
mus016334tom013641	D09	Chr 19	Chr 4/5	19
mus007418tom006198	D12	Chr 22	Chr 4/5	22
mus004844tom004057	D11	Chr 20	Chr 10	20Lo
mus007354tom006156	D11	Chr 20	Chr 10	20Lo
mus007745tom006463	D11	Chr 20	Chr 10	20Lo
mus010243tom008510	D11	Chr 20	Chr 10	20
mus000396tom000304	D11	Chr 20	Chr 10	20
mus003701tom003130	D11	Chr 20	Chr 10	20
mus007876tom006556	D11	Chr 20	Chr 10	20
mus003111tom002631	D11	Chr 20	Chr 10	20
mus018406tom015309	D11	Chr 20	Chr 10	20
mus007139tom005998	D11	Chr 20	Chr 10	20sh
mus001220tom000981	D11	Chr 20	Chr 10	20sh
mus004510tom003766	D11	Chr 20	Chr 10	20sh
mus018082tom015034	D11	Chr 20	Chr 10	20sh

TABLE 3.1 Continued

SNP identity	*Gossypium raimondii* scaffold	D-genome chromosome	A-genome homolog	Inferred position
mus015169tom012644	D11	Chr 20	Chr 10	20sh
mus003124tom002644	D11	Chr 20	Chr 10	20sh
mus001414tom001152	D11	Chr 20	Chr 10	20
mus001583tom001296	D10	Chr 25	Chr 6	25
mus012716tom010652	D10	Chr 25	Chr 6	25
mus021932tom018097	D10	Chr 25	Chr 6	25
mus001374tom001128	D10	Chr 25	Chr 6	25
mus022477tom018558	D10	Chr 25	Chr 6	25
mus021161tom017418	D10	Chr 25	Chr 6	25
mus006639tom005574	D10	Chr 25	Chr 6	25
mus007161tom006014	D10	Chr 25	Chr 6	25
mus017966tom014928	D10	Chr 25	Chr 6	25
mus000990tom000780	D10	Chr 25	Chr 6	25
mus009740tom008132	D10	Chr 25	Chr 6	25
mus011501tom009638	D10	Chr 25	Chr 6	25
mus013284tom011088	D10	Chr 25	Chr 6	25
mus014417tom012043	D10	Chr 25	Chr 6	25
mus023084tom019102	D10	Chr 25	Chr 6	25
mus009997tom008338	D08	Chr 26	Chr 12	12
mus010091tom008387	D01	Chr 16	Chr 7	16
mus002489tom002095	D08	Chr 26	Chr 12	26
mus013727tom011475	D08	Chr 26	Chr 12	26
mus004954tom004157	D08	Chr 26	Chr 12	26
mus000552tom000443	D08	Chr 26	Chr 12	26
mus001106tom000880	D08	Chr 26	Chr 12	26
mus008501tom007092	D08	Chr 26	Chr 12	26
mus010199tom008474	D08	Chr 26	Chr 12	26
mus003509tom002975	D08	Chr 26	Chr 12	26
mus020522tom016915	D02	Chr 15	Chr 1	/
mus008915tom007431	D09	Chr 19	Chr 4/5	4/19

*Lo and **sh are for short long and short chromosome arms, respectively.

3.2. Genetic group mapping

Linkage groups of validated markers were constructed for each species with JoinMap based on the segregation in respective BC1F1 populations. Genetic distributions of validated SNP markers across the genomes of *G. mustelinum* and *G. tomentosum* are presented in Figure 3.3 and Figure 3.4, respectively. The linkage maps were constructed with the LOD score set at 3.

The average marker distances of the linkage maps for *G. mustelinum* and *G. tomentosum* are 12.5 cM and 13.95 cM, respectively. Tables 3.2 and 3.3 report the coverage of each linkage group on related chromosomes, relative to the estimated chromosome sizes from the previous constructed genetic map involving *G. hirsutum* x *G. barbadense* (Yu *et al.* 2011).

TABLE 3.2
Estimates of *G. mustelinum* chromosome coverage by seven segments targeted for marker-based introgression, based on comparisons between seven linkage maps of targeted *G. hirsutum* - *G. mustelinum* markers versus lengths of six previously reported *G. hirsutum* - *G. barbadense* linkage groups

Linkage group	Localized chromosome	Estimated chromosome size (cM)	No. of loci	Genetic coverage	
				Length (cM)	Percentage (%)
A	Chr 3	162.0	4	6.9	4.26
B	Chr 9	187.0	3	27.5	14.71
C	Chr 9	187.0	9	136.4	75.78
D	Chr 19	243.4	3	45.4	18.65
E	Chr 20	160.9	14	124.3	77.25
F	Chr 25	154.0	11	90	58.44
G	Chr 26	123.7	8	121.4	98.14

TABLE 3.3
Estimates of *G. tomentosum* chromosome coverage by eight segments targeted for marker-based introgression, based on comparisons between eight linkage maps of targeted *G. hirsutum* - *G. tomentosum* markers versus lengths of six previously reported *G. hirsutum* - *G. barbadense* linkage groups

Linkage group	Localized chromosome	Estimated chromosome size	No. of loci	Genetic coverage	
				Length (cM)	Percentage (%)
A	Chr 3	162.0	5	72.6	44.81
B	Chr 9	187.0	6	86.4	46.20
C	Chr 19	243.4	3	32	13.65
D	Chr 20	160.9	8	70.3	43.69
E	Chr 20	160.9	6	40.5	25.17
F	Chr 25	154.0	11	171.1	111.10
G	Chr 26	123.7	4	28.3	22.88
H	Chr 26	123.7	4	28.9	23.36

Markers of *G. mustelinum* clustered into seven groups, whereas those of *G. tomentosum* clustered into seven groups. Some markers were so close to each other that they were co-localized (no recombinants observed), e.g. mus011138tom009240, mus004933tom004139 and mus016330tom013834. Nine and fifteen of the 67 markers of *G. mustelinum* and *G. tomentosum*, respectively, were not statistically linked to any other marker and were therefore classified as "independent", although they were localized by hypoaneuploid tests to the same chromosome as one of the linkage groups noted above. Marker mus011741tom009842 was not localized on any chromosome by using monosomic interspecific F1 hybrids. However, in the linkage map for *G. mustelinum*, it was grouped with two other markers, both of which were localized on chromosome 9.

The mean transmission rate was calculated for each linkage group in each population. Among the populations of *G. mustelinum*, the highest transmission rate (~0.9) was observed for group D of BC3-MH-G, while the lowest transmission rate (0.1) was observed for group A of BC3-HM-G (Figure 3.5). Similarly, the lowest rate for group C was also observed in BC3-HM-G. In contrast, group B exhibited its highest rate in that population. Comparatively, the rate of group F was stable across different populations.

For *G. tomentosum*, the highest transmission rate was observed for group H of BC3-TH-G. The lowest transmission rate was observed for groups A and B, as they were transmitted to BC3-HT-G (Figure 3.6). For group A, B and E, the lowest rate was in BC3-HT-G. In contrast, group D underwent its highest rate of transmission in the same population.

The above variation in transmission rates could be caused by the interactions between the linkage groups and the breeding situations used to develop the related populations. For example, in the *G. mustelinum* genome, the cross direction (i.e. hybrid used as paternal parent) and the cross location (i.e. greenhouse) for BC3-HM-G development was observed to negatively affect the transmission of group A and C while promoting group transmission of B. Therefore, specific breeding situations affect transmission rates of different linkage groups in idiosyncratic fashion, which might provide a new avenue for breeders to promote (reduce) the transmission of desirable (undesirable) chromosome segments or factors.

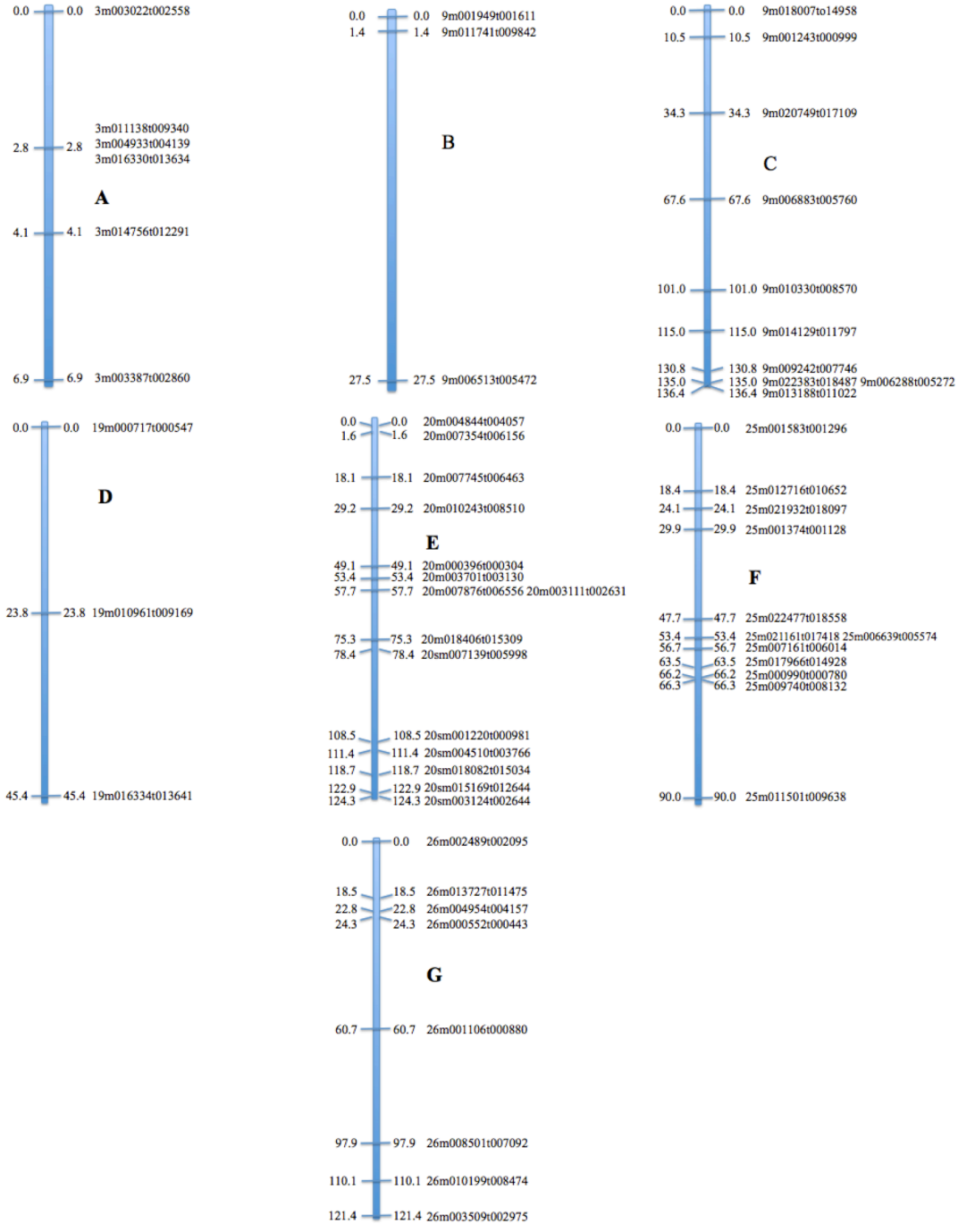

Figure 3.3 Linkage maps of SNPs of targeted *G. mustelinum* segments.

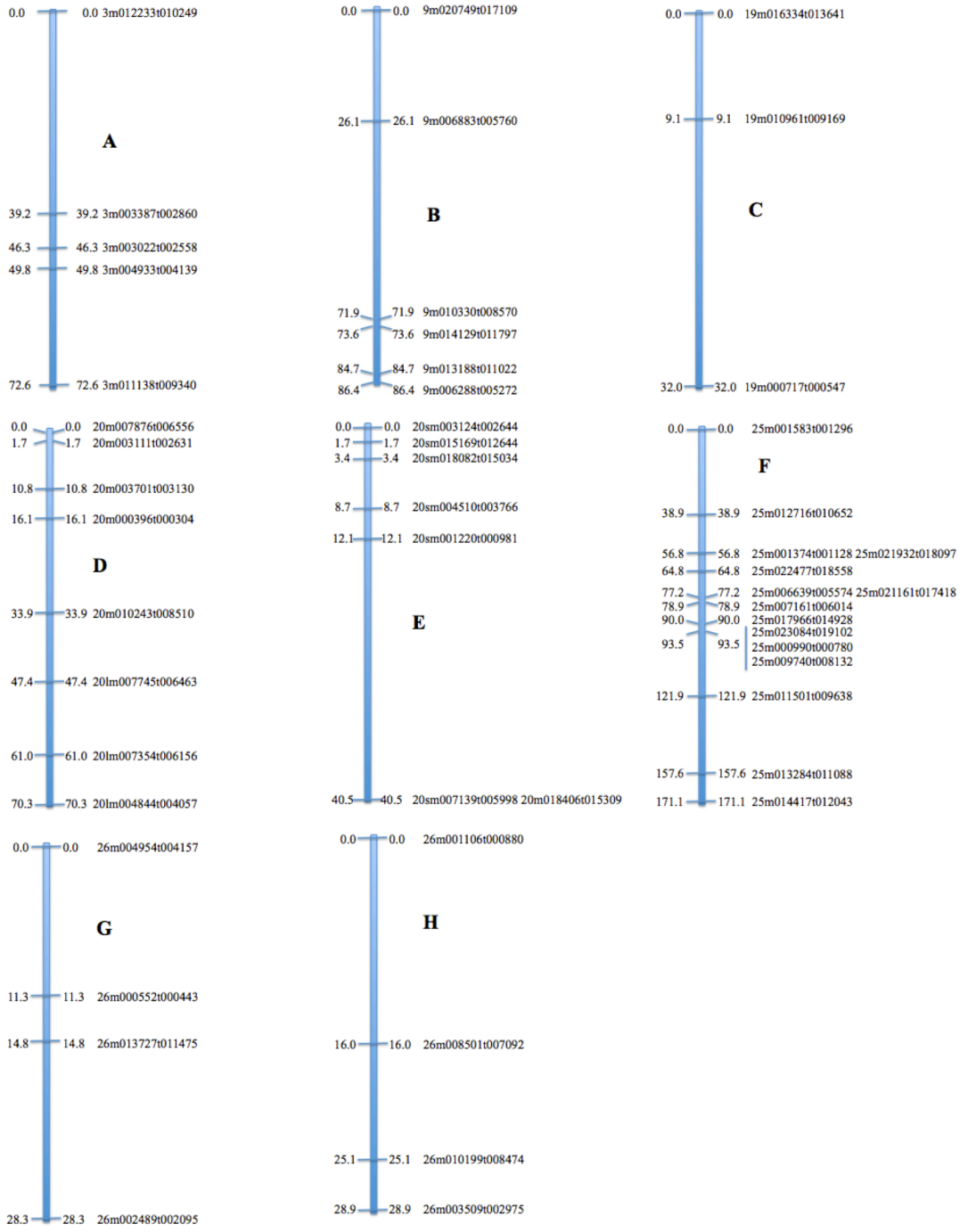

Figure 3.4 Linkage maps of SNPs of targeted *G. tomentosum* segments.

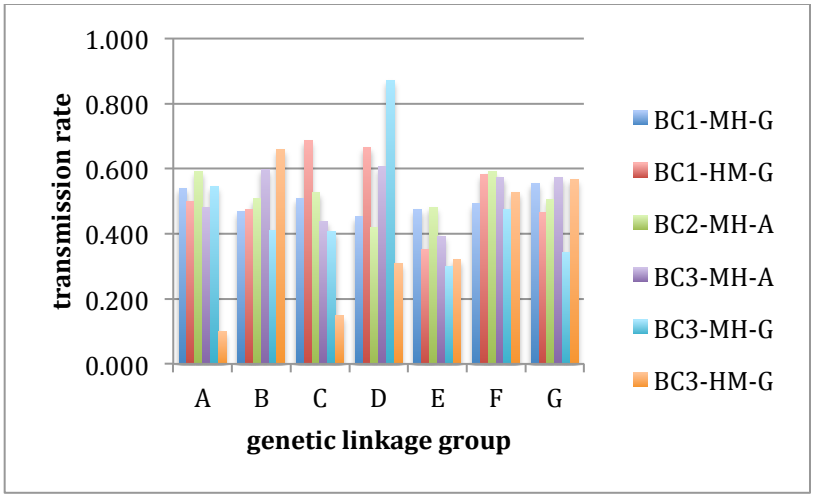

Figure 3.5 Mean transmission rates of the seven *G. mustelinum* segments (linkage groups) across different backcross populations.

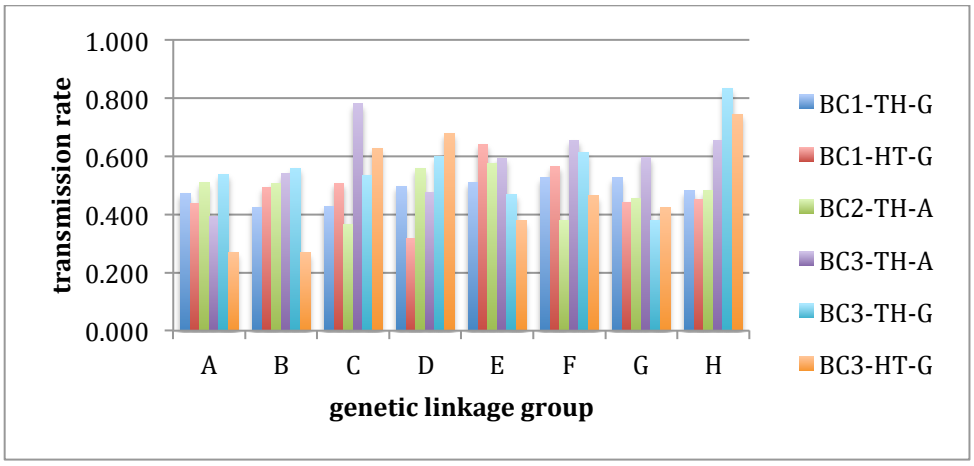

Figure 3.6 Mean transmission rates of the eight *G. tomentosum* segments (linkage groups) across different backcross populations.

3.3. Distorted transmission rate for individual markers

The transmission rates of some markers were extremely distorted in specific populations. Two linked markers in linkage group D of *G. mustelinum*, mus000717tom000547 and mus010961tom009169, were transmitted at a 100% rate to BC3-MH-G. Another marker transmitted at a 100% rate is mus001374tom001128 from linkage group F in BC3-HM-G, i.e., in the reciprocal cross. Marker mus004510tom003766 of linkage group E, or its *G. hirsutum* allele, may be of special interest as the alien allele was missing in all the three BC3F1 populations of *G. mustelinum*. Given its presence in 2 BC2F1 plants, the effective BC3F1 population size was 8 (P=0.0039 of zero occurrences, if randomly transmitted). Marker mus008915tom007431 behaved in a very abnormal way in *G. mustelinum*-derived populations. Although polymorphism for this marker exists, it cannot be transmitted in any hybrids (i.e., F1s and BCs). A hypothesis for this observation is that the *G. hirsutum* SNP marker is dominant, which means the donor allele is recessive and the amplicon for this allele cannot be produced in heterozygotes. It was verified by a complementary experiment in which DNA samples of *G. hirsutum* and *G. mustelinum* were combined in a 1:1 mixture and found to exhibit same KASP type as *G. hirsutum*.

Complete transmission was also observed in *G. tomentosum*-derived populations. An independent marker mus006513tom005472 in BC3-HT-G and marker mus018406tom015309 of linkage group E in BC3-TH-A underwent 100% transmission. Although there were some cases of low transmission rates among the tested loci, a 0% transmission rate was not observed in the population of *G. tomentosum*, which could reflect a higher level of genetic compatibility of *G. hirsutum* with *G. tomentosum* than *G. mustelinum*.

For both of the two donor species, significant differences of transmission rate among markers within a same linkage group were observed in some specific populations. For *G. mustelinum*, marker mus007161tom006014 of group F was observed with a

significantly higher transmission rate (0.833) than the mean rate (0.474) for its linkage group in BC3-MH-G. Its two closely linked markers (mus006639tom005574 and mus017966tom014928), which are positioned on both sides with genetic distances of 3.3 cM and 6.8 cM, respectively, have their transmission rate of 0.429 and 0.419, respectively (Figure 3.7.a). Similarly, marker mus008501tom007092 of group G in BC3-HM-G exhibited a transmission rate of 0.833, which was also much higher than the rate for other markers in the same linkage group (Figure 3.7.b). In contrast, marker mus018007tom014958 of group C was detected with a very low transmission rate (0.286) in BC1-HM-G while all of other markers in the same linkage group had their transmission rates above 0.5 and the overall rate was 0.687 (Figure 3.7.c).

Analogous cases occurred in the populations of *G. tomentosum*. In BC3-HT-G, compared to its closely linked (7.1 cM) marker (mus003022tom002558), mus003387tom002860 of group A was observed with a significantly higher transmission rate (0.545), which is also very different from the mean transmission rate of the group (0.269) (Figure 3.7.d). In BC3-TH-G, the transmission rate for marker mus001106tom000880 of group H was 0.4 while the overall rate for the group was more than 0.8; group H spanned less than 30 cM (Figure 3.7.e).

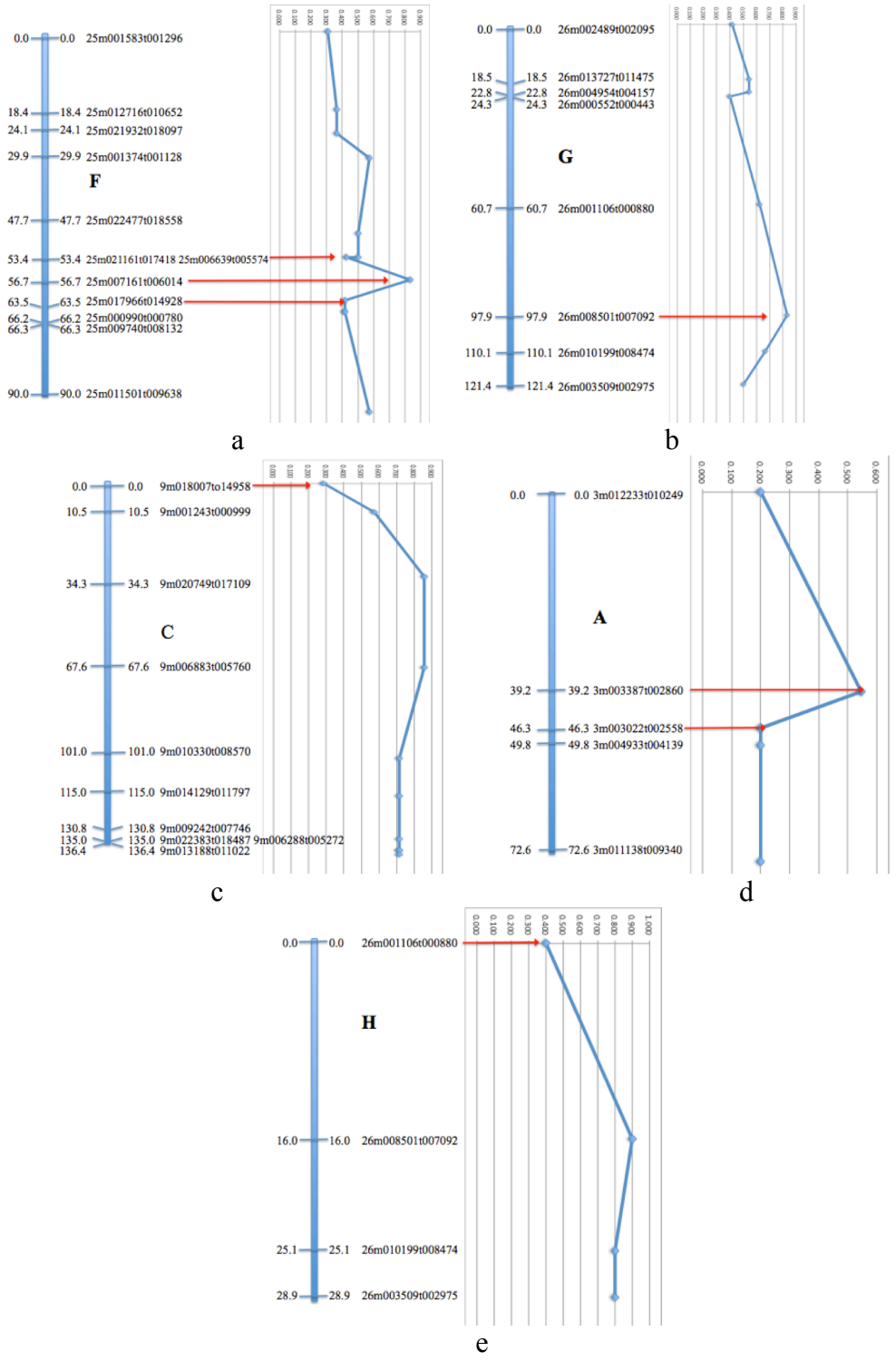

Figure 3.7 Markers with significantly different transmission rates within a linkage group. (a-c): *G. mustelinum* linkage groups F, G and C. (d-e): *G. tomentosum* linkage groups A and H.

For certain markers, the transmission rates varied widely across different populations. In *G. mustelinum*-derived populations, all six markers in group A transmitted at a decreased rate (0.1) into the last population (Figure 3.8), indicating that the breeding situation for that population (BC3-HM-G) could negatively affect the transmission rates of these markers, possibly indicating pollen-based selection for a nearby locus. The rate in the only other population with *G. mustelinum* germplasm transmitted via the pollen parent was BC1-HM-G, in which the rate of transmission for this marker was over 40% (3 of 7), but the small size of that population (n=7) precludes robust inferences. Three of the six markers of this group (A) were observed with their highest rates (>0.60) in BC2-MH-A, but rates for these loci were not especially high in BC3-MH-A. For *G. tomentosum*-derived populations, four markers from group B underwent similar patterns of transmission rates (Figure 3.9) and the lowest transmission rates for all of those markers also resulted from the breeding situations for BC3-HT-G, i.e., where the pollen parent was heterozygous for donor germplasm.

Major differences in transmission rates occurred for some markers when the cross direction changed, i.e., when donor germplasm was transmitted via the seed versus the pollen parent. As shown in Figure 3.10, two markers (mus010961tom009169 and mus007161tom006014) transmitted at much higher rates (100% and 83.3%) via seed parents, while two other markers (mus001374tom001128 and mus008501tom007092) were far more likely to be transmitted via pollen parents (100% and 83.3%). In addition, preferential transmission could also come from changes in crossing location (Figure 3.11). All of the first three markers in the graph showed significantly higher transmission rates in the field than in the greenhouse while the other two markers had increased transmission rates in the greenhouse. For the last marker, the transmission rate was 100% under the greenhouse location.

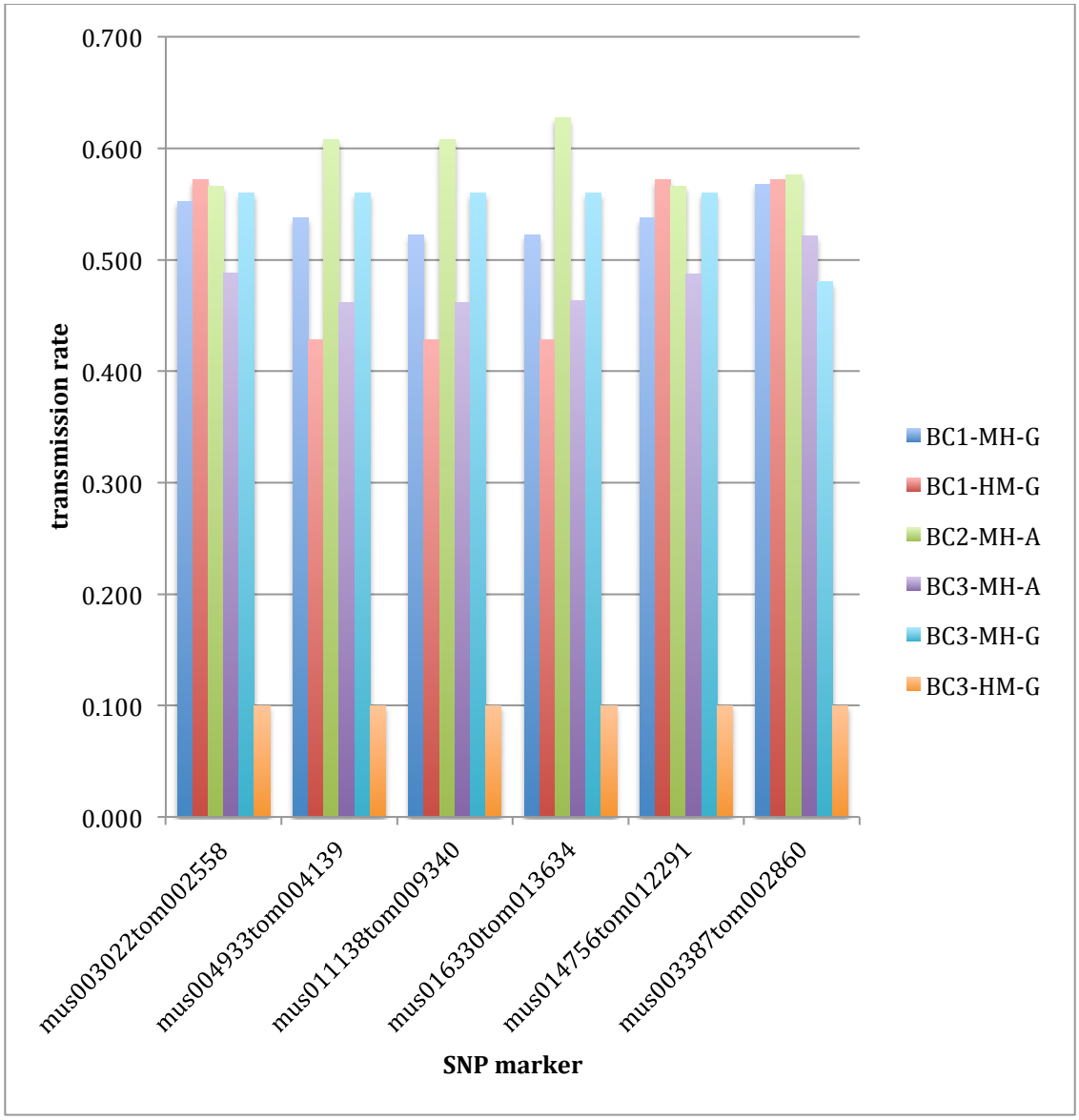

Figure 3.8 Transmission rates of markers in group A across populations of *G. mustelinum*.

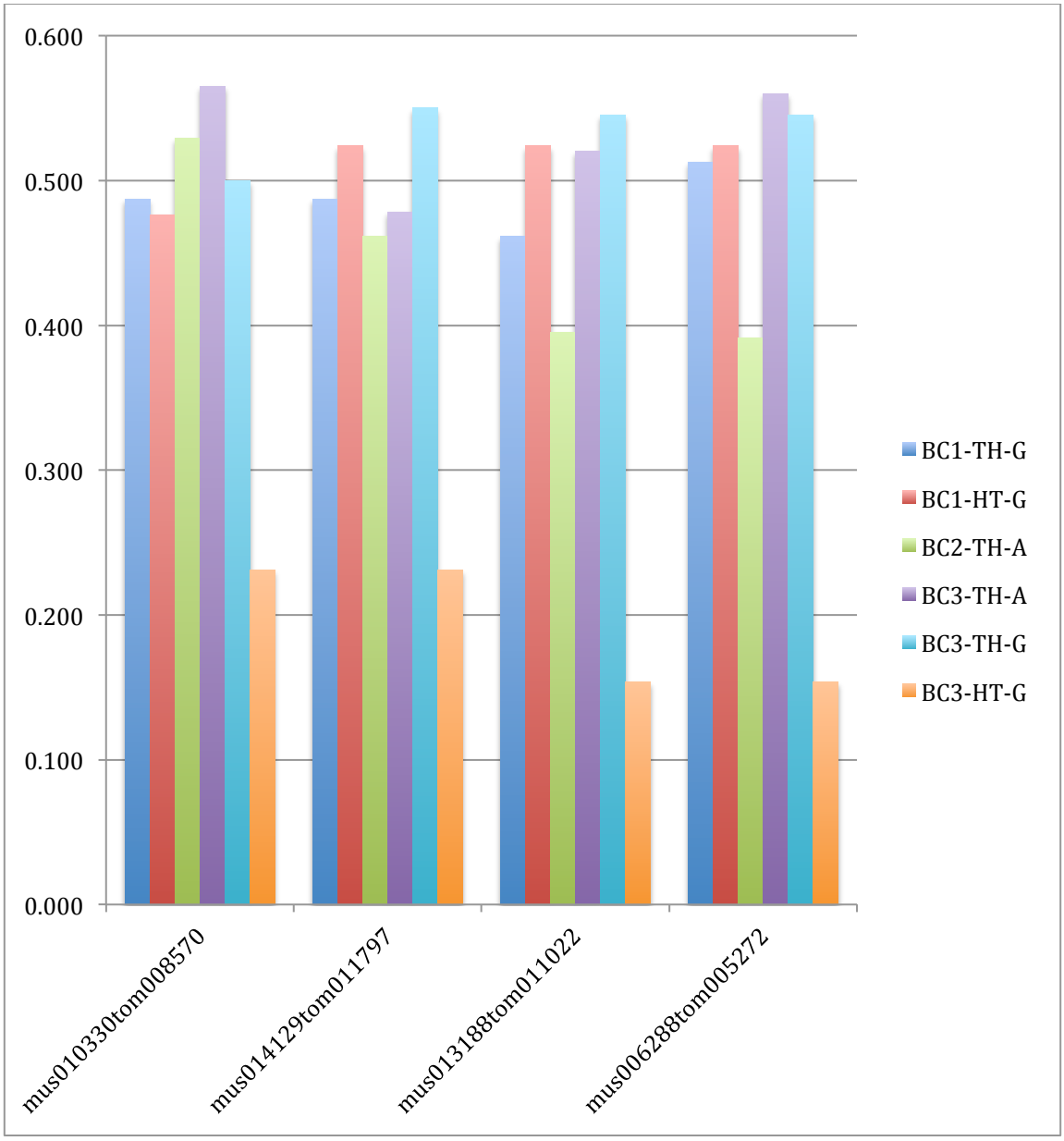

Figure 3.9 Transmission rates of markers in linkage group B across populations of *G. tomentosum*.

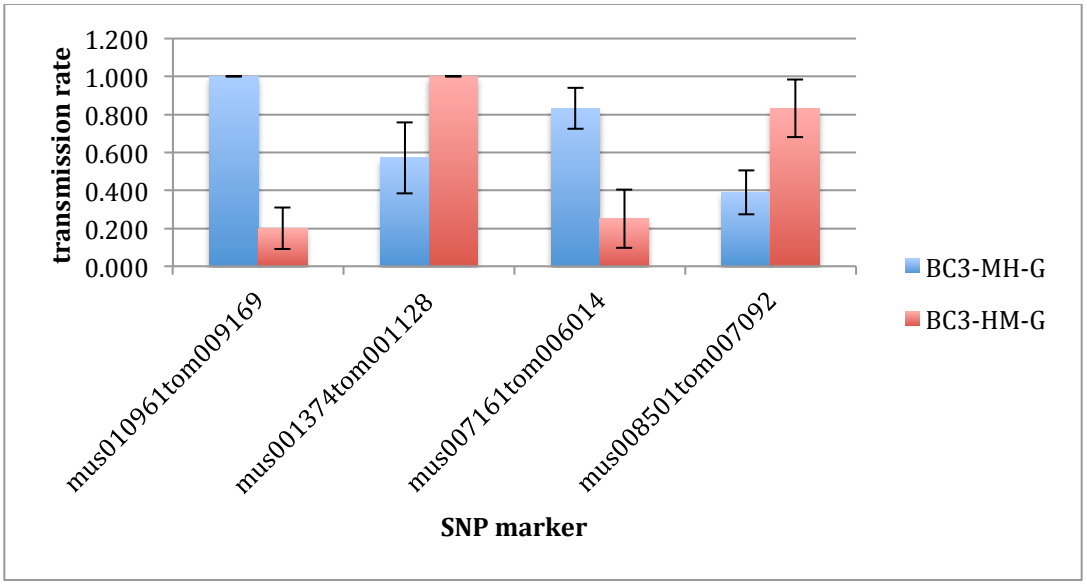

Figure 3.10 Effects of cross direction on transmission rates of certain markers from *G. mustelinum*.

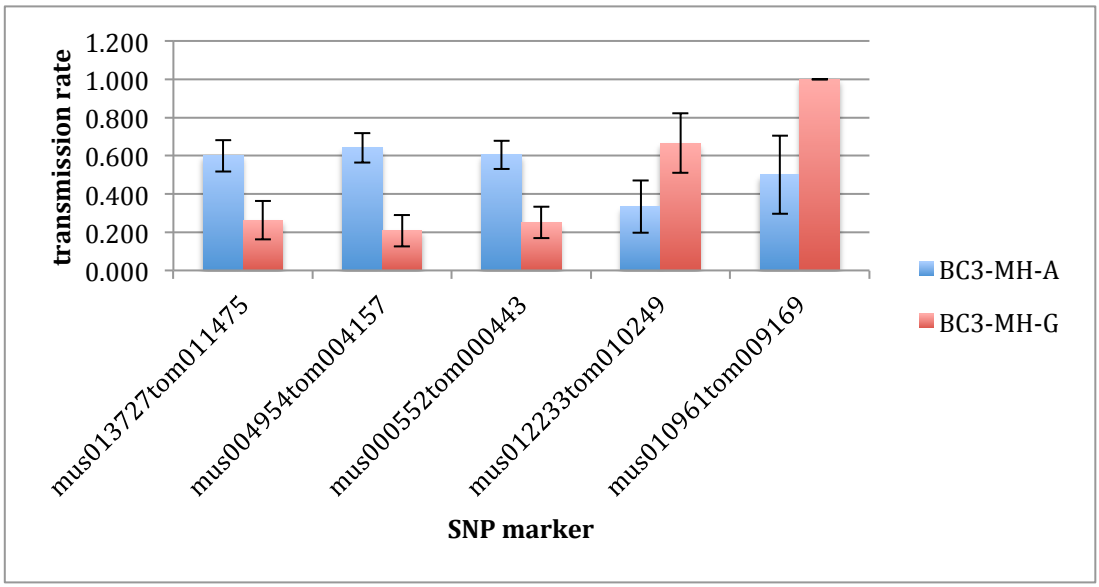

Figure 3.11 Effects of location during pollination on transmission rates of certain markers from *G. mustelinum*.

Similarly in *G. tomentosum*-derived populations, transmission rates for some markers were distorted differentially in one cross direction or the other and one location or the other. Figure 3.12 shows the effects of cross direction on transmission rates for specific markers. Transmission rates for the left-most three markers, which comprise a terminal segment (~11 cM) of linkage group D, were higher (>80%) via the pollen parent. In contrast, the transmission rates for the right-most three markers, all completely linked at an interstitial position of group F, were affected by the cross direction in an opposite way, i.e., higher (>68%) via the seed parent. The effect of location on transmission rate was marker dependent (Figure 3.13). The four left-most markers were from three linkage groups, and were transmitted at a higher rate in field crosses, while the two right-most markers, both in group D but distant from each other (~37 cM), were more frequently transmitted in greenhouse crosses (0.7, 0.8).

Significantly different transmission rates among markers within a linkage group could result from differences in linkage to nearby chromosomal regions that have significant effects on transmission. It is possible that a marker with a significantly low transmission rate might be closely linked with an allele that has deleterious effects on either or both gametophytes, the endosperm or the zygote. If so, the remarkable differences of transmission rates among the linked markers could indicate that the recombination rates differed among parental generations as backcrossing advanced. These discoveries of markers that undergo strongly non-random transmission offer opportunities to discover specific genes that govern transmission or fitness, by one means or another, e.g., meiotic drive or resilience to abiotic stress, such as temperature. So, in addition to influencing the results of introgressive breeding efforts, these could affect traits important to breeders. By associating non-randomly inherited regions with markers, we can impose MAS to assure recovery of desired types and/or recombinants in the next generation.

The transmission rates of certain markers can be affected idiosyncratically by specific "breeding situations". A given breeding situation could have different effects on the transmission process depending on the specific marker. Thus, breeders can select the breeding situations that favor/disfavor transmission of specific markers or segments. This could facilitate introgression by MAS.

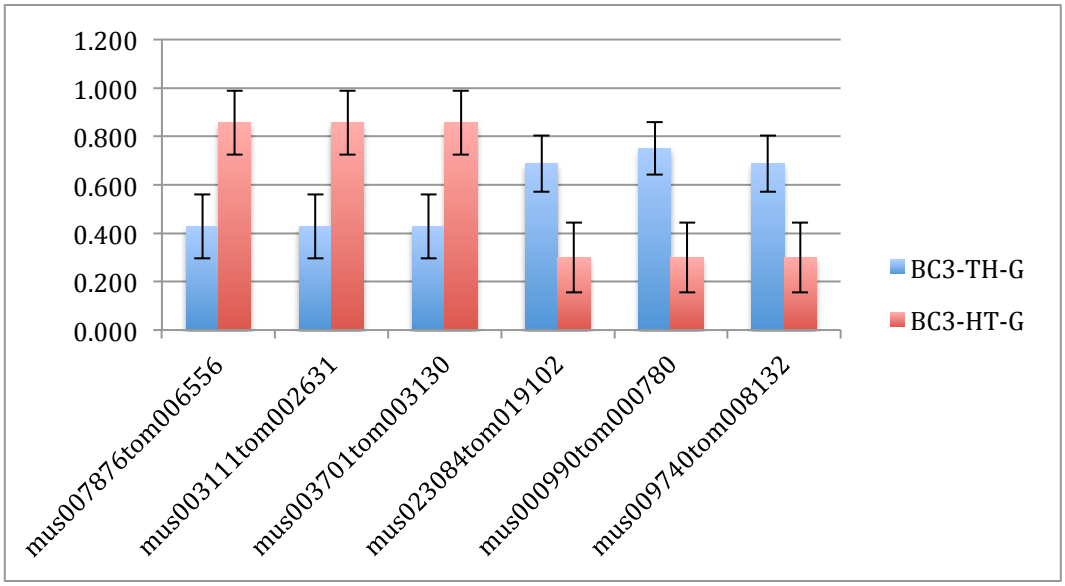

Figure 3.12 Effects of cross direction on transmission rates of different markers.

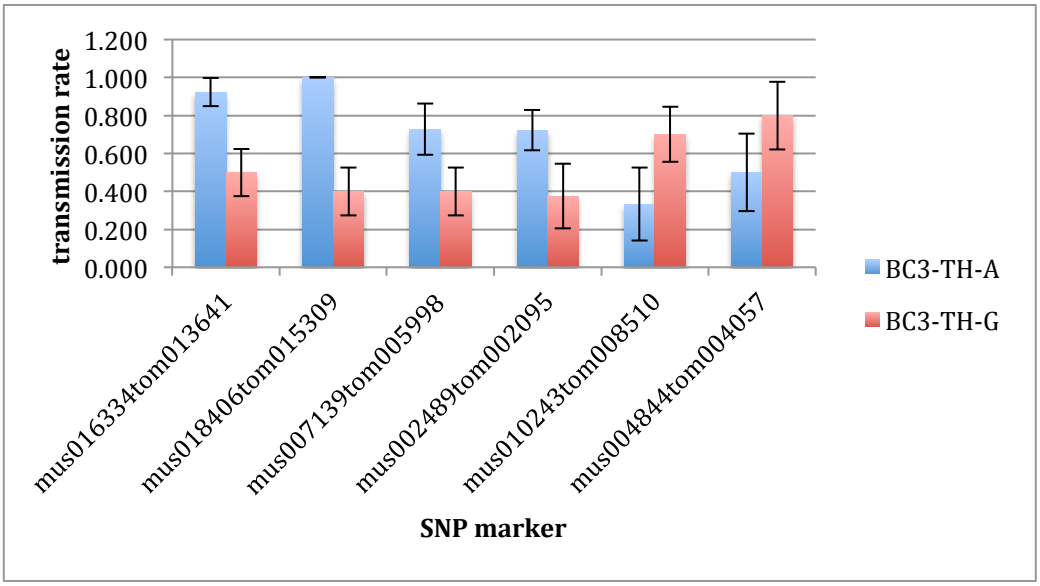

Figure 3.13 Effects of location during pollination on transmission rates of different markers.

3.4. Genetic constitution of individual plants in BC3F1 generations

For each individual plant in the BC3F1 generation, the genetic constitution was detected and the percentage of heterozygous loci in our target region was calculated. In a backcross program, half of the randomly inherited heterozygous loci are expected to be rendered homozygous in each backcross generation. Therefore, the proportion of heterozygous is 0.125. Sizes of the BC3F1 populations derived from *G. mustelinum* and *G. tomentosum* were 167 and 158, respectively. A two-tailed confidence interval (CI) with α equal to 0.05 for the theoretical value was constructed according to the population size, both of which are close to 0.125±0.050. In our research, four ranges 0~0.03, 0.03~0.07, 0.07~0.17 and 0.17~0.5 about percentage of heterozygous loci were set regarding the CI, which were then used to classify individual plants in each population. The resultant proportion for each of the four ranges within a population was quantified (Figure 3.14 and Figure 3.15).

In the 3 BC3F1 populations of *G. mustelinum*, about half of the plants were in the CI. In BC3-HM-G, the proportion of plants with less than 3% of the sampled loci in a heterozygous state was 25%, i.e. higher than the percentage in the other two populations. On the other hand, BC3-MH-A had the largest proportion of plants with high percentages of heterozygous loci.

As to the *G. tomentosum*-derived BC3F1 populations, a relatively small proportion (36.84%) of plants containing the expected range of heterozygous loci were observed in BC3-TH-G. This population contained the largest proportion (17.54%) of plants with percentages of heterozygous loci below 3%.

Similar to the transmission rate of linkage groups, the differential proportions for each population could also be related to various breeding situations. In general, *G. mustelinum* derived populations contained more plants with low heterozygosity levels (< 3%), indicating an effect on the transmission rate caused by the wild donor species. It should also be noted that populations developed in the greenhouse tended to have more plants with the low percentages of heterozygous loci, which provide some hints about the relationship between the planting location and the genetic constitution of the plants.

Information on the genetic constitution of individual plants, as presented above, is important to know for breeding purposes. On one hand, its important to isolate specific alien segments in different backcross products, and secondly, they should collectively represent the entire targeted donor contribution (often the entire genome). In developing NILs, the timeline for breeding plants with low heterozygosity levels will prospectively be shortened, which will save a considerable amount of labor and economic cost. The marker information is also useful in terms of retaining a comprehensive set of the targeted alien germplasm during backcrossing, e.g., when genome-wide introgression is desired.

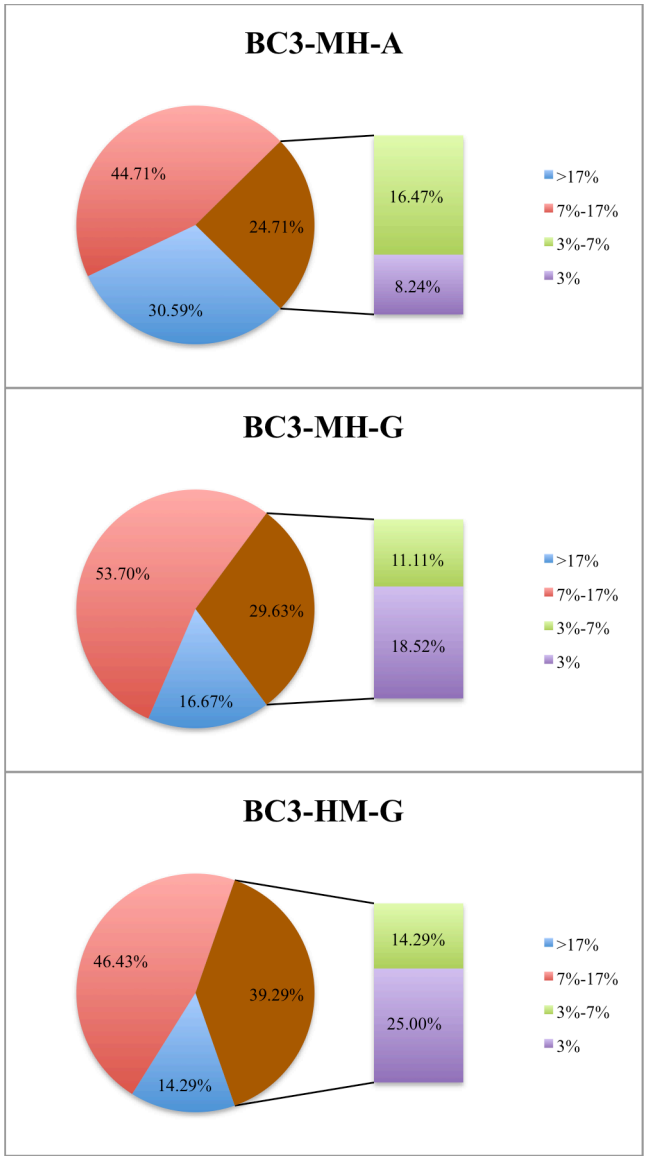

Figure 3.14 Diagrammatic depiction of heterozygosity in *G. mustelinum*-derived populations, where each plant is categorized according to its percentage of heterozygous sampled loci.

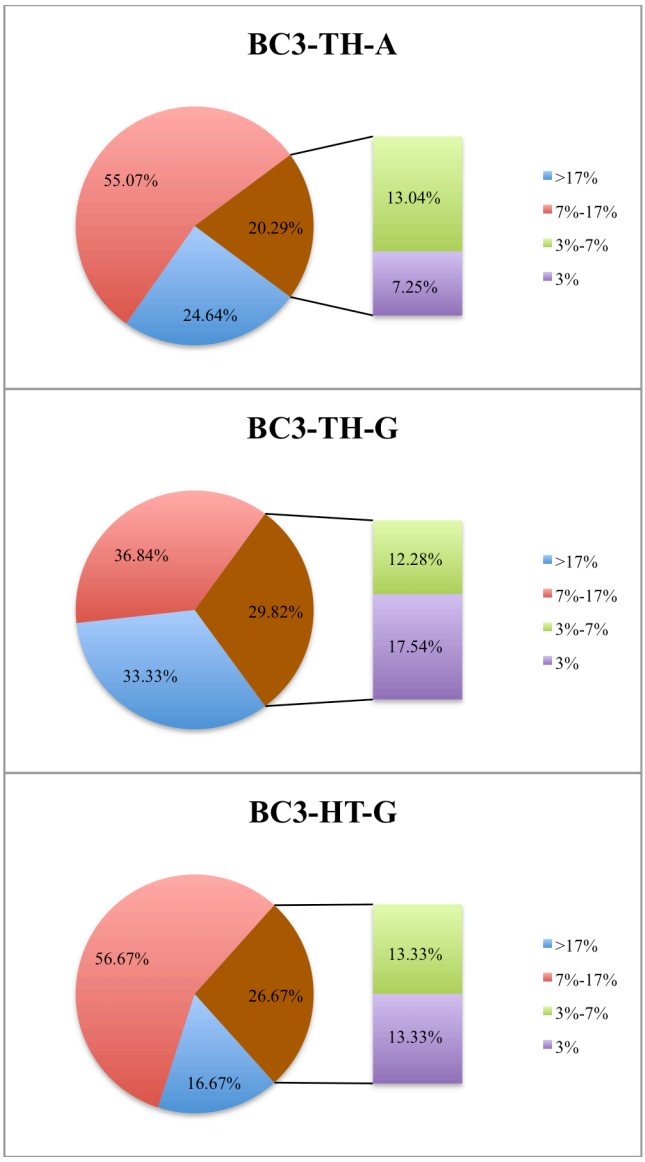

Figure 3.15 Diagrammatic depiction of heterozygosity in *G. tomentosum*-derived populations, where each plant is categorized according to its percentage of heterozygous sampled loci.

3.5. Statistical analysis of transmission rates

3.5.1 General statistics for each population

The transmission rates for marker-defined chromosome segments were determined for each population. The overall mean for each population is presented in Table 3.4. The means of all the 12 populations generally fluctuated around 0.5, ranging from 0.3308 to 0.5546. The standard error of transmission rates for alien segments was lower through the female (MH, TH) than through male gametes (HM, HT), i.e., 0.0241875 versus 0.03835, which indicates a more stable way to pass on alien chromosome segments. Transmission rates from the 3 populations (BC3-HM-G, BC3-HT-G, BC3-MH-G) were significantly lower than the rates from the other 9 populations. Possible factors causing the variation of transmission rate among different populations will be discussed below.

TABLE 3.4
The quantified mean transmission rate of each population.

Population	Mean ± 1.96 X standard error
BC1-MH-G	0.4952 ± 0.0192
BC1-HM-G	0.5117 ± 0.0452
BC1-TH-G	0.4952 ± 0.0192
BC1-HT-G	0.4940 ± 0.0261
BC2-MH-A	0.5150 ± 0.0191
BC2-TH-A	0.4820 ± 0.0192
BC3-MH-A	0.5119 ± 0.0248
BC3-TH-A	0.5546 ± 0.0285
BC3-HM-G	0.3308 ± 0.0402
BC3-HT-G	0.4113 ± 0.0419
BC3-MH-G	0.4419 ± 0.0327
BC3-TH-G	0.5401 ± 0.0308

Results from ANOVA to detect effects of donor species, i.e., *G. mustelinum* versus *G. tomentosum*, are shown in Table 3.5. The P-value ($0.0039 < 0.05$) indicates a significant difference in transmission rate due to the donor parents. The lsmean (least squares mean) for the population of *G. mustelinum* was 0.47314 while that for the population of *G. tomentosum* was 0.5057, which means the transmission rate was generally higher by 3.2% when *G. tomentosum* was the donor, rather than *G. mustelinum* (Table 3.6). For equivalent breeding purposes, slightly larger populations would be desirable for *G. mustelinum* than for *G. tomentosum*.

An explanation for this phenomenon is the genetic constitution of these two wild species, which results in distinct behaviors for the tested markers. More genome differences were detected at cytogenetic levels between *G. hirsutum* and *G. mustelinum* than between *G. hirsutum* and *G. tomentosum* (Hasenkampf and Menzel 1980). A phylogenetic tree based on molecular marker (RAPD) analysis was constructed, in which *G. tomentosum* clustered with *G. hirsutum* in a 0.78 Nei's similarity while *G. mustelinum* is clustered with *G. hirsutum* in a 0.71 Nei's similarity (Khan *et al.* 2000). A similar result was also reported based on simple matching of isozyme banding patterns and nonmetric multidimensional scaling analysis (Saha and Zipf 1997). Another indicator of relatively closer relationship between *G. hirsutum* and *G. tomentosum* is the greater average distance of linkage groups as reported in section 3.2.

The higher transmission rate indicates fewer incompatibilities between *G. tomentosum* and *G. hirsutum* genomes, which is consistent with previous phylogenetic analysis. Compared to *G. mustelinum*, the closer relationship of *G. tomentosum* with *G. hirsutum* genomes will expectedly facilitate both germplasm introgression and disruption of undesirable linkages through homologous recombination.

TABLE 3.5
ANOVA analyses for transmission rate between the two wild species.

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	1	0.20795587	0.20795587	8.40	0.0039
Error	782	19.36399413	0.02476214		
Corrected Total	783	19.57195000			

A general linear model GLM for populations derived from each set of populations is constructed and the sources of variance in transmission rate are released. GLM analysis is performed for each set of population separately as below.

TABLE 3.6
Transmission rate least squares means for donor species, generation, direction of cross and location

G. mustelinum				*G. tomentosum*			
0.4731b*				0.5057a			
BC1F1	BC2F1	BC3F1	P-value	BC1F1	BC2F1	BC3F1	P-value
0.5427a	0.4579b	0.4448b	0.0003	0.5209a	0.4339b	0.5204a	0.0020
A	G			A	G		
0.5219a	0.4417b		0.0002	0.5184a	0.4651b		0.0202
HM	MH			HT	TH		
0.4674	0.4962		0.1595	0.4674b	0.5161a		0.0053

* Common (different) letters indicate that differences are not (are) significant (Padiff test, $p<0.05$).

3.5.2 Effect of backcross generation on transmission rate

Comparisons among least squares means for transmission rates into populations of *G. mustelinum* revealed that generation effects were highly significant ($p < 0.01$), which suggests that the backcross generation significantly affected transmission rates in the population of *G. mustelinum* (Table 3.6). The least squares mean of the transmission rate from the F1 to the BC1F1 was 0.5427, which is significantly higher than least squares means of transmission to the BC2F1 and BC3F1 generations (0.4579b, 0.4448b, respectively).

Highly significant effects by backcross generation were also observed for least squares means for the populations of *G. tomentosum*, but the relative levels were somewhat different. Transmission rates to the BC1 and BC3 generations were higher than to the BC2 generation.

With exception of transmission to the *G. tomentosum* BC2, the least squares means for alien germplasm transmission decreased with increasing backcross generation, possibly due to impeding effects of increasingly homozygous levels of the *G. hirsutum* genetic background. A possible reason for this phenomenon is epistasis. Genetic interactions between loci during the process of interspecific introgression could lead to events like DMI and render some individual plants to be unviable and sterile (i.e. post-zygotic isolation) so that segregation for some genotypes would be distorted (Mittelbach *et al.* 2007). DMI-like effects presumably affect gametophytes and endosperm, too, and in fact, would logically be especially potent at the haploid phase, particularly in pollen. More complex epistasis has been reported in the form of genotype and/or allelic marker transmission distortion in *Solanum* by using double introgression lines {Moyle and Nakazato 2009).

In our research, with the accumulation of homozygous chromosome segments from the recurrent parent (*G. hirsutum*) each backcross generation, transmission of some alien alleles might be progressively negatively impacted. Therefore, the rate was turned down

as the generation advanced. The opposite may have happened in transmission to tracked SNPs in the BC2F1 *G. tomentosum* population, i.e., the increasing frequency of homozygosity in one or more loci, may have preferably favored transmission of certain *G. tomentosum* alleles over the *G. hirsutum* alternatives.

3.5.3 Effect of location on transmission rate

Another source for the deviation of the transmission rate was the cross location. Hybrids from *G. mustelinum* planted in field (F) possessed a least squares mean transmission rate of 0.5219, 8% higher than ones grown in greenhouse (G) by 8%. Similar results were observed for *G. tomentosum* segments during backcrossing. They showed significant difference in transmission rates for field-based crosses (0.5184) over greenhouse-based crosses (0.4615).

Differential rates of transmission for the two locations in our experiment might be caused by any of multiple environmental factors that varied between greenhouse and open field (Table 3.7). While the exact ranges differ among years, an approximate sense of the environmental differences between the two locations is provided by temperature, humidity and daylength data in Table 3.5.4 during the periods of pollination during 2013. Differences in these factors could alter the transmission rates through pollen, ovules, endosperm and zygotic products, and affect some related genetic mechanisms. Variations in these factors across years would be confounded with generations. It is conceivable that confounding effects and interactions with genotypic factors account for the seemingly disparate rates of transmission into the BC2-TH-A population, which arose from cross-pollinations made in June and July of 2012.

Homologous recombination in arabidopsis plants was found to depend on temperature and day length (Boyko *et al.* 2005). The recombination rate was higher in plants grown at suboptimal temperatures, whether lower (4 C) or higher (32 C) as compared to the optimum (22 C). On the other hand, when grown at different day lengths (8-24 h), recombination rates were minimal in plants grown in the longest day (24 h) conditions,

and highest in the plants grown in the shortest day (8 h) conditions. In our backcross breeding program, changes in recombination rate, either somatic or meiotic, could affect the elimination and /or recurrence of involved alleles in the next generation by altering linkage relationships with loci with alleles that are subject to differential selection.

From a physiological perspective, alteration in environment can affect many processes involved in germplasm transmission between two generations. For example, changes in temperature can influence pollen germination and pollen tube growth, for which cardinal temperature varies among cultivars (Kakani *et al.* 2005). Burke (2011) noted strong sensitivities of cotton pollen and genetic varation for such sensitivities to various envirnonmental conditions, including humidity. The relatively higher transmission rates displayed in field crosses suggest an underline promotion for the fitness of the pollen with more alien alleles.

As discussed before, loss or recovery of genetic material can be influenced by epistasis. Thus, extrinsic factors that affect epistasis would eventually lead to distorted transmission rates at specific loci across the genome. In studies involving interpopulation crosses of the copepod *Tigriopus californicus*, Willett and Burton (2003) presented a dramatic example. Influences of temperature and light environment led to selection at the CYC (cytochrome c) locus, which markedly affected genotypic frequencies at many other loci, via numerous epistatic interactions, and led to a large-scale distortion on transmission as a consequence.

TABLE 3.7
Environmental differences of greenhouse and field locations during pollinations in 2013

Factors	Greenhouse	Field
	December 2013	July 2013
Temperature (°C)	23.9-29.4	23.4-35.4
Humidity (%)	70	42-94
Day length (h)	10	14

3.5.4 Effect of cross direction on transmission rate

Cross direction did not lead to significant difference in the population of *G. mustelinum*, indicating that overall transmission rates were not affected by the direction of cross. On the other hand, significant effects were observed in *G. tomentosum* populations, in which alien alleles tended to be transmitted at a higher rate through maternal gametes (0.5161) than paternal gametes (0.4674).

In *G. tomentosum*, differential transmission rates for the two cross directions might have been due to reduced pollen fitness, which could have been caused by introgressed germplasm from the alien parent (*G. tomentosum*). This phenomenon is consistent with the observation that certain markers were transmitted at higher rates through the seed parent. A possible explanation for such bias is that transmission via pollen entails higher levels of gametophytic gene expression than does transmission via megagametophytes. Given higher expression in pollen, selection pressure would also be higher.

The degree of transmission distortion due to cross direction is determined by the linkage intensities between each marker and the neighboring locus that is subject to direct selection in a cross-specific manner (male versus female parent). For both of the two donor species, certain markers underwent skewed transmission rates via the pollen parent or seed parent, indicating some markers were linked to loci favored through paternal transmission and others were linked to loci favored through maternal transmission.

CHAPTER IV
CONCLUSION

Using recently developed SNP markers, a combination of bioinformatic and experimental methods was used to identify SNPs in chromosomes or chromosome segments targetted for marker-assisted interspecific introgression from *G. tomentosum* and *G. mustelinum*. We developed linkage maps for the linked loci and proceeded to study the transmission and recombination chraracteristics of markers and linked segments during early generations of backcrossing (BC1, BC2 and BC3), so that the breeding behavior of alien germplasm might be better understood. According to relative "lengths", the SNP-based linkage groups used in this analysis covered previously reported respective linkage groups for the targetted chromosomes varied from about 5% to nearly 100%.

We noted that some of the selected SNPs unexpectedly mapped as independent markers, rather than exhibiting linkage to other markers from the same chromosome. These results indicated some discordance between the resulting linkage maps among the SNPs and previous inferences on location using monosomic interspecific hybrids.

The mean transmission rate for each linkage group varied across the 12 different populations. Some SNPs exhibited extremely distorted transmission rates in specific populations. When viewed in terms of the numbers of plants categorized according to alien SNP retention percentage, the 6 BC3F1 populations were found to differ considerably. Those, which were created from greenhouse crosses, had more plants with low percentages of alien germplasm, at least in the SNP-targeted regions.

Variation of the rate of SNP transmission into BC1, BC2 and BC3 backcross populations was analyzed by General Linear Model (GLM) to analyze effects of some factors for each species of the donor parent: (*i*) generation, (*ii*) location and (*iii*) direction of cross. Generation, location and direction effects differed between the two population sets (species). In the *G. mustelinum* populations, only two factors, generation and location

significantly distorted the transmission rates, while in the *G. tomentosum* populations, all three factors significantly influenced the inheritance of the alien chromosomal segments.

In the GLM analysis, least squares means of transmission rates were calculated for all the 67 SNP markers. The differences among factor levels were significant but spanned a small range, ~10% or less. In contrast, rates for specific chromosomal segments varied by as much as from 20% to 80%. Therefore, more factors should be considered and integrated into our model.

In general, the research provides a view of the variation of transmission in early backcross-based germplasm introgression, as well as its related causal agents. Knowledge of the preferential transmission rates related to specific levels of individual factors, e.g., BC2, could used by breeders to influence the inheritance of genes and agronomical traits. In a backcross-breeding program, it is desirable for breeders to introgress alien germplasm into cultivated species with a genetic background nearly isogenic to the recipient parent. When using MAS for backcross introgression, breeders can opt for specific breeding situation that favors the transmission of germplasm from recipient parent, such that near-isogenic lines containing the targeted genes will be expectedly developed within a relatively short time. In construction of chromosome segment substitution lines (CSSLs), introgression of alien germplasm is always required in early generations while in later generations, one often seeks to reduce or eliminate inadvertant alien gene content from the genetic background. Results from our research suggest that CSSL construction could benefit by altering the breeding situations in a segment-specific manner. The markers with exceptionally distorted transmission rates may require special attention by breeders, i.e., to break some linkages between them and other genes. They also may be special interest for construction of transgenes that are preferentially transmitted (or not) through a pollen or ovule parent.

For our future investigation, experiments for specific factors, e.g., locations, should be conducted in greater detail. Genome-wide genetic maps for *G. mustelinum and*

G. tomentosum with high marker density should be constructed by processing marker development. Markers with extremely high or low transmission rates should be characterized more extensively.

REFERENCES

Aghaee-Sarbarzeh, M., M. Ferrahi, S. Singh, H. Singh, B. Friebe *et al.*, 2002 Ph I-induced transfer of leaf and stripe rust-resistance genes from *Aegilops triuncialis* and *Ae. geniculata* to bread wheat. Euphytica 127**:** 377-382.

Alves, M., P. Barroso, A. Ciampi, L. Hoffmann, V. Azevedo *et al.*, 2013 Diversity and genetic structure among subpopulations of *Gossypium mustelinum* (Malvaceae). Genet Mol Res 12**:** 597-609.

Anderson, E., 1949 Introgressive hybridization. Introgressive hybridization.

Arnold, M. L., 1992 Natural hybridization as an evolutionary process. Annual review of Ecology and Systematics**:** 237-261.

Avery, A. G., S. Satina and J. Rietsema, 1959 Blakeslee: the genus Datura. Blakeslee: the genus Datura.

Barton, N. H., and K. S. Gale, 1993 Genetic analysis of hybrid zones. Hybrid zones and the evolutionary process**:** 13-45.

Bernardo, R., 2008 Molecular Markers and Selection for Complex Traits in Plants: Learning from the Last 20 Years. Crop Sci. 48**:** 1649-1664.

Botstein, D., R. L. White, M. Skolnick and R. W. Davis, 1980 Construction of a genetic linkage map in man using restriction fragment length polymorphisms. American journal of human genetics 32**:** 314.

Boyko, A., J. Filkowski and I. Kovalchuk, 2005 Homologous recombination in plants is temperature and day-length dependent. Mutation Research/Fundamental and Molecular Mechanisms of Mutagenesis 572**:** 73-83.

Brar, D., and G. Khush, 1997 Alien introgression in rice, pp. 35-47 in Oryza: From Molecule to Plant, edited by T. Sasaki, Moore, Graham. Springer.

Burke, J. J., 2011 Cotton flowers: pollen and petal humidity sensitivities determine reproductive competitiveness in diverse environments. Stress Physiology in Cotton**:** 25.

Darwin, C., 1859 On the origin of species by means of natural selection, or, The preservation of favoured races in the struggle for life. J. Murray, London.

Dejoode, D. R., and J. F. Wendel, 1992 Genetic Diversity and Origin of the Hawaiian-Islands Cotton, *Gossypium-Tomentosum*. American Journal of Botany 79**:** 1311-1319.

Duvick, D. N., 1965 Cytoplasmic pollen sterility in corn. Advances in genetics 13**:** 1-56.

Endler, J. A., 1973 Gene Flow and Population Differentiation Studies of clines suggest that differentiation along environmental gradients may be independent of gene flow. Science 179**:** 243-250.

Endrizzi, J., E. Turcotte and R. Kohel, 1985 Genetics, cytology and evolution of *Gossypium*. Advances in genetics 23**:** 271-375.

Gardunia, B. W., 2006 Introgression from *Gossypium mustelinum* and *G. tomentosum* into upland cotton, *G. hirsutum*, pp. Texas A&M University.

Gutiérrez, O., D. Stelly, S. Saha, J. Jenkins, J. McCarty, Jr. *et al.*, 2009 Integrative placement and orientation of non-redundant SSR loci in cotton linkage groups by deficiency analysis. Molecular Breeding 23**:** 693-707.

Harrison, R. G., 1986 Pattern and process in a narrow hybrid zone. Heredity 56**:** 337-349.

Hasenkampf, C. A., and M. Y. Menzel, 1980 Incipient genome differentiation in Gossypium. II. Comparison of 12 chromosomes in *G. hirsutum*, *G. mustelinum* and *G. tomentosum* using heterozygous translocations. Genetics 95**:** 971-983.

Ho, K. M., and K. J. Kasha, 1974 Differential Chromosome Contraction at Pachytene Stage of Meiosis in Alfalfa (Medicago-Sativa L). Chromosoma 45**:** 163-172.

Hoffman, G., and S. Rawlins, 1970 Infertility of cotton flowers at both high and low relative humidities. Crop science 10**:** 721-723.

Jacob, H. J., K. Lindpaintner, S. E. Lincoln, K. Kusumi, R. K. Bunker *et al.*, 1991 Genetic mapping of a gene causing hypertension in the stroke-prone spontaneously hypertensive rat. Cell 67**:** 213-224.

Jiang, C. x., P. W. Chee, X. Draye, P. L. Morrell, C. W. Smith *et al.*, 2000 Multilocus interactions restrict gene introgression in interspecific populations of polyploid Gossypium (cotton). Evolution 54**:** 798-814.

Johns, C., X. Delannay and R. Palmer, 1981 Structural sterility controlled by nuclear mutations in angiosperms. Nucleus.

Kakani, V., K. Reddy, S. Koti, T. Wallace, P. Prasad *et al.*, 2005 Differences in in vitro pollen germination and pollen tube growth of cotton cultivars in response to high temperature. Annals of botany 96**:** 59-67.

Kaul, M. L., 1988 Male sterility in higher plants. Springer-Verlag.

Khan, M. A., J. M. Stewart and J. B. Murphy, 1999 Evaluation of the *Gossypium* Gene Pool for Foliar Terpenoid Aldehydes. Crop Sci. 39**:** 253-258.

Khan, S., D. Hussain, E. Askari, J. M. Stewart, K. Malik *et al.*, 2000 Molecular phylogeny of *Gossypium* species by DNA fingerprinting. Theoretical and Applied Genetics 101**:** 931-938.

Lander, E. S., and D. Botstein, 1989 Mapping mendelian factors underlying quantitative traits using RFLP linkage maps. Genetics 121**:** 185-199.

Lee, J. A., 1981 Genetics of D3 Complementary Lethality in *Gossypium-Hirsutum* and *Gossypium-Barbadense*. Journal of Heredity 72**:** 299-300.

Lee, J. A., 1984 Cotton as a World Crop, pp. 1-25 in *Cotton*, edited by R. J. Kohel and C. F. Lewis. American Society of Agronomy, Crop Science Society of America, Soil Science Society of America.

Levi, A., A. H. Paterson, V. Barak, D. Yakir, B. Wang *et al.*, 2009 Field evaluation of cotton near-isogenic lines introgressed with QTLs for productivity and drought related traits. Molecular Breeding 23**:** 179-195.

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

Loegering, W. Q., and E. R. Sears, 1963 Distorted inheritance of stem-rust resistance of Timstein wheat caused by a pollen-killing gene. Canadian Journal of Genetics and Cytology 5**:** 65-72.

Lopez-Lavalle, L. A. B., and C. L. Brubaker, 2007 Frequency and fidelity of alien chromosome transmission in *Gossypium* hexaploid bridging populations. Genome 50**:** 479-491.

Maan, S. S., 1975 Exclusive Preferential Transmission of an Alien Chromosome in Common Wheat. Crop Science 15**:** 287-292.

Mammadov, J., R. Aggarwal, R. Buyyarapu and S. Kumpatla, 2012 SNP markers and their impact on plant breeding. International journal of plant genomics 2012: 728398.

Matsui, T., O. S. Namuco, L. H. Ziska and T. Horie, 1997 Effects of high temperature and CO_2 concentration on spikelet sterility in indica rice. Field Crops Research 51**:** 213-219.

May, R. M., J. A. Endler and R. E. McMurtrie, 1975 Gene frequency clines in the presence of selection opposed by gene flow. American Naturalist 109**:** 659-676.

Meyer, V., and W. Meredith, 1978 New germplasm from crossing Upland cotton (*Gossypium hirsutum*) with G. tomentosum. Journal of Heredity 69**:** 183-187.

Mittelbach, G. G., D. W. Schemske, H. V. Cornell, A. P. Allen, J. M. Brown *et al.*, 2007 Evolution and the latitudinal diversity gradient: speciation, extinction and biogeography. Ecology letters 10**:** 315-331.

Moyle, L. C., and T. Nakazato, 2009 Complex epistasis for Dobzhansky–Muller hybrid incompatibility in Solanum. Genetics 181**:** 347-351.

Nasuda, S., B. Friebe and B. S. Gill, 1998 Gametocidal genes induce chromosome breakage in the interphase prior to the first mitotic cell division of the male gametophyte in wheat. Genetics 149**:** 1115-1124.

Okamuro, J. K., B. Den Boer and K. D. Jofuku, 1993 Regulation of Arabidopsis flower development. The Plant Cell 5**:** 1183.

Paterson, A. H., J. W. DeVerna, B. Lanini and S. D. Tanksley, 1990 Fine mapping of quantitative trait loci using selected overlapping recombinant chromosomes, in an interspecies cross of tomato. Genetics 124**:** 735-742.

Percival, A. E., and R. J. Kohel, 1990 Distribution, collection, and evaluation of Gossypium. Adv. Agron 44**:** 225-256.

Powell, W., G. C. Machray and J. Provan, 1996 Polymorphism revealed by simple sequence repeats. Trends in plant science 1**:** 215-222.

Rafalski, J. A., 2002 Novel genetic mapping tools in plants: SNPs and LD-based approaches. Plant science 162**:** 329-333.

Rick, C. M., 1966 Abortion of male and female gametes in the tomato determined by allelic interaction. Genetics 53**:** 85.

Rick, C. M., 1969 Controlled introgression of chromosomes of Solanum pennellii into Lycopersicon esculentum: segregation and recombination. Genetics 62**:** 753.

Rieseberg, L. H., C. R. Linder and G. J. Seiler, 1995 Chromosomal and genic barriers to introgression in Helianthus. Genetics 141**:** 1163-1171.

Rieseberg, L. H., and J. F. Wendel, 1993 Introgression and its consequences in plants. pp. 70-109 in Hybrid zones and the evolutionary process, edited by Richard G. Harrison. Oxford University Press, New York.

Robinson, A., A. Bell, N. Dighe, M. Menz, R. Nichols *et al.*, 2007 Introgression of Resistance to Nematode Rotylenchulus reniformis into Upland Cotton (*Gossypium hirsutum*) from *Gossypium longicalyx*. Crop Science 47**:** 1865-1877.

Rooney, W. L., and D. M. Stelly, 1991 Preferential Transmission and Somatic Elimination of a *Gossypium-Sturtianum* Chromosome in *G-Hirsutum*. Journal of Heredity 82**:** 151-155.

Sachs, E. S., J. H. Benedict, D. M. Stelly, J. F. Taylor, D. W. Altman *et al.*, 1998 Expression and segregation of genes encoding CryI, a insecticidal proteins in cotton. Crop Science 38**:** 1-11.

Saha, S., D. A. Raska and D. M. Stelly, 2006 Upland cotton (*Gossypium hirsutum L.*) x Hawaiian cotton (*G. tomentosum* Nutt. ex. Seem) F1 hybrid hypoaneuploid chromosome substitution series. J Cotton Sci 10**:** 146-154.

Saha, S., and A. Zipf, 1997 Genetic diversity and phylogenetic relationships in cotton based on isozyme markers. Journal of crop production 1**:** 79-93.

Sandler, L., and E. Novitski, 1957 Meiotic Drive as an Evolutionary Force. American Naturalist 91**:** 105-110.

Sari-Gorla, M., M. E. Pe, D. L. Mulcahy and E. Ottaviano, 1992 Genetic dissection of pollen competitive ability in maize. Heredity 69**:** 423-430.

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

Vos, P., R. Hogers, M. Bleeker, M. Reijans, T. Van de Lee *et al.*, 1995 AFLP: a new technique for DNA fingerprinting. Nucleic acids research 23**:** 4407-4414.

Wang, D. G., J.-B. Fan, C.-J. Siao, A. Berno, P. Young *et al.*, 1998 Large-scale identification, mapping, and genotyping of single-nucleotide polymorphisms in the human genome. Science 280**:** 1077-1082.

Welsh, J., and M. McClelland, 1990 Fingerprinting genomes using PCR with arbitrary primers. Nucleic acids research 18**:** 7213-7218.

Wendel, J., C. Brubaker, I. Alvarez, R. Cronn and J. Stewart, 2009 Evolution and Natural History of the Cotton Genus, pp. 3-22 in Genetics and Genomics of Cotton, edited by A. Paterson. Springer US.

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

Willett, C. S., and R. S. Burton, 2003 Environmental influences on epistatic interactions: viabilities of cytochrome c genotypes in interpopulation crosses. Evolution 57**:** 2286-2292.

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

Zhang, Z., J. Rong, V. N. Waghmare, P. W. Chee, O. L. May *et al.*, 2011 QTL alleles for improved fiber quality from a wild Hawaiian cotton, *Gossypium tomentosum*. Theoretical and applied genetics 123**:** 1075-1088.