Mississippi State University
Scholars Junction

Theses and Dissertations Theses and Dissertations

5-3-2008

SNP characterizaiton and genetic and molecular analysis of mutants affecting fiber development in cotton

Chuanfu An

Follow this and additional works at: https://scholarsjunction.msstate.edu/td

Recommended Citation
An, Chuanfu, "SNP characterizaiton and genetic and molecular analysis of mutants affecting fiber development in cotton" (2008). *Theses and Dissertations*. 3971.
https://scholarsjunction.msstate.edu/td/3971

This Dissertation - Open Access is brought to you for free and open access by the Theses and Dissertations at Scholars Junction. It has been accepted for inclusion in Theses and Dissertations by an authorized administrator of Scholars Junction. For more information, please contact scholcomm@msstate.libanswers.com.

SNP CHARACTERIZAITON AND GENETIC AND MOLECULAR ANALYSIS OF MUTANTS AFFECTING FIBER DEVELOPMENT IN COTTON

By

Chuanfu An

A Dissertation
Submitted to the Faculty of
Mississippi State University
in Partial Fulfillment of the Requirements
for the Degree of Doctor of Philosophy
in Agronomy
in the Department of Plant and Soil Sciences

Mississippi State, Mississippi

May 2008

SNP CHARACTERIZATION AND GENETIC AND MOLECULAR ANALYSIS OF

MUTANTS AFFECTING FIBER DEVELOPMENT IN COTTON

By

Chuanfu An

Approved:

Johnie N. Jenkins
Adjunct Professor of Agronomy
and Research Geneticist, USDA, ARS
(Co-Major Professor)

Din-Pow Ma
Professor of Biochemistry and
Molecular biology
(Minor Professor)

Sukumar Saha
Adjunct Professor of Agronomy
and Research Geneticist, USDA, ARS
(Co-Major Professor)

Jack C. McCarty, Jr.
Adjunct Professor of Agronomy and
Research Agronomist, USDA, ARS
(Committee member)

Frank B. Matta
Professor of Horticulture
(Committee member)

Teddy P. Wallace
Associate Professor of Agronomy
(Committee member)

John A. Boyle
Professor Emeritus of Biochemistry
and Molecular Biology
(Committee member)

William L. Kingery
Graduate Coordinator, Department of
Plant and Soil Sciences

Vance H. Watson
Dean of the College of
Agriculture and Life Sciences

Name: Chuanfu An

Date of Degree: May 2, 2008

Institute: Mississippi State University

Major Field: Agronomy

Major Professor: Jonnie N. Jenkins and Sukumar Saha

Title of study: SNP CHARACTERIZATION AND GENETIC AND MOLECULAR ANALYSIS OF MUTANTS AFFECTING FIBER DEVELOPMENT IN COTTON

Pages in Study: 155

Candidate for Degree for Doctor of Philosophy

Cotton (*Gossypium* spp.) is the world's leading textile fiber crop, and an important source of oil and protein. Insufficient candidate gene derived-markers suitable for genetic mapping and limited information on genes that control economically important traits are the major impediments to the genetic improvement of Upland cotton (*G. hirsutum* L.). The objectives of this study were to develop a SNP marker discovery strategy in tetraploid cotton species, SNP characterization and marker development from fiber initiation and elongation related genes, chromosomal assignment of these genes by SNP marker-based deletion analysis or linkage mapping, and genetic and molecular analysis of mutants affecting cotton fiber development. Phylogenetic grouping and comparision to At- and Dt-genome putative ancestral diploid species of allotetraploid cotton facilitated differentiation between genome specific polymorphisms (GSPs) and marker-suitable locus-specific polymorphisms (LSPs). By employing this strategy, a total of 222 and 108 SNPs were identified and the average frequency of SNP was 2.35%

and 1.30% in six *EXPANSIN A* genes and six *MYB* genes, respectively. Both gene families showed independent and incongruent evolution in the two subgenomes and a faster evolution rate in Dt-genome than that in At-genome. SNPs were concordantly mapped to different chromsomes, which confirmed their value as candidate gene marker and indicated the reliability of SNP discovery stragey. QTL mapping by two F_2 populations developed from fiber mutants detected major QTL which explain 62.8-87.1% of the phenotypic variation for lint percentage or lint index in the vicinity of BNL3482-138 on chromosome 26. Single marker regression analyses indicated STV79-108, which was located to the long arm of chromosome 12 (the known location of N_1 and perhaps n_2 loci), also had significant association (R^2 % value 15.4-30.6) with lint percentage, lint index, embryo protein percentage and micronaire. Additional QTL and significant markers associated with other seed and fiber traits were detected on different chromosomes. Inheritance analysis indicated that both genetic models $N_1N_1n_2n_2$ and $n_2n_2li_3li_3$ could lead to the fiberless phenotype. The observation of fuzzless-short lint phenotype indicated fiber initiation and elongation were controlled by different mechanisms. The penetrance of Li_2 gene expression was observed in this study.

DEDICATION

To those who need food and clothes in the world. I wish I could do something for them in my life.

ACKNOWLEDGEMENTS

Acknowledgement of dissertation is not only expressing the gratefulness to people who helped me but also a good opportunity to summary the past several years and look forward the road ahead.

In August 2004, I quitted my first Ph. D study, which started one year ago, in pharmacognosy at Nanjing University (Nanjing, China) and came to Mississippi State, a typical small southern America town, with love and dreams. Here, I got another opportunity to work on plant genetics and genomics after switching from sciences of silkworm and mulberry to botany, and then to pharmacognosy. I was wondering is this my final major and can I realize my dream in this not good "enough" university? Four years passed, the answers turn out to be positive. I find interest in molecular genetics. I think my education here is worthwhile, if not completely satisfactory to me. All these are due to wonderful people I met here.

Grateful acknowledge goes to my major professor Dr. Johnie N. Jenkins. I am so proud of being one of his students. He treated me as student instead of worker or assistant. He gave me opportunity and enough freedom to read references, design experiments, and draft manuscripts. He told me Ph. D degree is not enough and the most important is what you can bring to table. He set me a model of how to be a good scientist, a good person as well. My sincere thankfulness is extended to Dr. Jack C. McCarty for his selflessness

help on this dissertation. Advice and encouragement from him always make me feel confident and give me reasons to continue. Special acknowledge goes to Dr. Sukumar Saha for the time we worked together.

I would like to thank my minor professor Dr. Din-Pow Ma for his guidance and friendship. My sincere gratitude also goes to Drs. John A. Boyle, Frank B. Matta, and Teddy P. Wallace for their severing in the committee. Their kindness help will never be forgotten. I thank Dr. Russell W. Hayes, Mr. Douglas A. Dollar, and Mr. Kimber E. Gourley for their technical assistant. I appreciate the communication and help from Drs. Brian E. Scheffler, Rickie B. Turley, Franklin E. Callahan, Martin J. Wubben, and Jixiang Wu.

Last, for sure is not least, I would like to acknowledge my family members, parents, parents-in law, and especially my wife, Yufang Guo. They are the driving force for me to stride ahead.

Getting Ph. D degree is only a stop sign in the one-way traffic life. There is only three seconds to pause. Even the road ahead will be struggles and frustration, but the temptation of sunshine and flowers can not be resisted. The reason is that the author believes the final success and peace will be defined as how much time, energy and hope we save to realize our self-improvement and to build up our personalities.

TABLE OF CONTENTS

DEDICATION ii

ACKNOWLEDGEMENTS iii

LIST OF TABLES viii

LIST OF FIGURES xi

CHAPTER

I. INTRODUCTION 1

Challenges for current genetic improvement of cotton 2
Current need in cotton genetics and genomics research 2
Objectives of this research 4
Justification and usefulness of this research 4

II. SNP CHARACTERIZATION AND SNP MARKER-BASED CHROMOSOMAL ASSIGNMENT OF SIX *EXPANSIN A* GENES IN COTTON 5

Abstract 5
Introduction 6
Materials and Methods 9
- Plant materials and DNAs isolation 9
- PCR primer design, amplification, cloning, and sequencing 11
- Sequence analysis and SNP primer design 13
- SNP genotyping 14
- Chromosomal assignment of *EXPANSIN A* genes 14

Results 15
- Sequence characterization of six *EXPANSIN A* genes 15
- Nucleotide diversity 20
- Chromosomal assignment 23

Discussion 25
- SNP marker discovery in tetraploid cotton 25

SNP features in cotton 27
Unequal evolutionary rate of the At- and Dt-genome in cotton 28
Chromosomal locations of six *EXPANSIN A* genes 29

III. R2R3-MYB TRANSCRIPTION FACTORS SNP IDENTIFICATION, PHYLOGENOMIC CHARACTERIZATION, CHROMOSOME LOCALIZATION, AND LINKAGE MAPPING IN COTTON 33

Abstract 33
Introduction 34
Materials and Methods 36
Plant materials... 36
PCR amplification, cloning, and sequencing... 37
SNP characterization and phylogenetic analysis 40
Chromosomal assignment and linkage mapping 40
Results 42
SNP characterization and halotype analysis of six *MYB* genes 42
Phylogenomic sequence characterization 45
Chromosome localization of six *MYB* genes 49
Linkage mapping of *MYB* genes by SNP markers.................................... 52
Discussion 54
SNP in cotton 54
MYB gene phylogenomic features 55
Chromosomal locations of *MYB* genes 57

IV. DETECTING QTL FOR YIELD COMPONENTS, SEED, AND FIBER TRAITS OF UPLAND COTTON UNDER MULTIPLE FUZZLESS LOCI GENETIC BACKGROUNDS 58

Abstract 58
Introduction 59
Materials and Methods 61
Mapping populations and phenotypic data collection 61
DNA isolation and molecular marker analysis 62
Linkage mapping and QTL analysis 63
Chromosomal assignment of linkage groups and trait-associated markers 64
Statistical analysis.. 64
Results and Discussion 65
Phenotypic variation and trait correlation... 65
Linkage maps 70
QTL associated with different traits 71
Two-locus epistatic interactions 78

V. INSIGHTS INTO THE INHERITANCE OF LINT AND FUZZ FIBER INITIATION IN UPLAND COTTON (*Gossypium hirsutum* L.) 83

Abstract 83
Introduction 84
Materials and Methods 86
Plant materials 86
Data collection, phenotypic classification, and data analysis 89
Results and Discussion 90
Genetic analysis on FM966, MD17, 177, and 181 derived populations ... 95
Observations and primary analysis on crosses related to Li_1 and Li_2 101
The abnormal phenotypes observed in Li_2 related crosses and generations 106

VI. CONCLUSIONS 111

SNP marker development strategy 111
SNP characterization on *EXPANSIN A* and *MYB* gene families 111
EXPANSIN A and *MYB* gene families' chromosomal locations in tetraploid cotton 113
QTL for yield components, seed, and fiber traits 114
Inheritance of lint and fuzz fiber development 115

REFERENCES 116

APPENDIX 133

LIST OF TABLES

2.1 PCR Primers Designed from Six *EXPANSIN A* Genes 12

2.2 Distribution and Types of SNPs in PCR Amplicons of Six Cotton *EXPANSIN A* Genes 18

2.3 The Nucleotide Diversity (π values, $\times10^{-2}$) among Six PCR-amplified Cotton *EXPANSIN A* Gene Fragments in A- and D-genomes 22

2.4 Chromosomal Locations of Six *EXPANSIN A* Genes in Cotton....................... 24

3.1 SNP Makers Derived from Six *MYB* Genes for Genotyping in Four Allotetraploid Cotton Species 39

3.2 SNP Characterization of Six *MYB* Genes in Selected Cotton Genotypes 44

3.3 Pairwise Comparison Matrix of the Eight Cotton Lines Showing the Nucleotide Diversity (π value, $\times10^{-2}$) in A- and D-genomes 46

3.4 Pairwise Comparison Matrix of Molecular Evolutionary Rates for Six *MYB* Genes in Cotton.................................. 48

3.5 Chromosomal Locations (*Lo*, long arm; *sh*, short arm) of Six *MYB* Genes with Previously Reported QTL 51

4.1 Phenotype Values for Yield Components, Seed, and Fiber Traits of F_1, F_2 and their Parents in Two Mapping Populations 66

4.2 Spearman Correlation Coefficients between Yield Components, Seed, and Fiber Traits in Two Mapping Populations 69

4.3 QTL and Single Non-mapped Markers Associated with Yield Components, Seed, and Fiber Traits Identified by Interval Mapping with MapQTL 5.0 and Single Marker Regression Analysis, Respectively 72

4.4 Two-locus Epistatic Interactions Identified in Two Mapping Populations 79

5.1 Phenotype, Origin, and Seed Source of Six Parental Lines 87

5.2 Segregation of Fiberless (fls), Fuzzless (N), and Fuzzy-linted (F) Phenotypes in Four of the Six Parental Lines 97

5.3 Observed Phenotypic Segregation of Crosses Related to Li_1 and Li_2 and Six Parental Lines 103

5.4 Abnormal Phenotypes Observed in Li_2 Associated Crosses and Generations 108

5.5 The Next Generation Segregation from the Abnormal Li_2 Plant of 2006 Field 110

A.1 Fragment EXPA1 A Genome Amplicon Haplotypes 134

A.2 Fragment EXPA1 D Genome Amplicon Haplotypes 134

A.3 Fragment EXPA2-1 D Genome Amplicon Haplotypes 134

A.4 Fragment EXPA2-2 A Genome Amplicon Haplotypes 135

A.5 Fragment EXPA2-2 D Genome Amplicon Haplotypes 135

A.6 Fragment EXPA3-1 A Genome Amplicon Haplotypes 135

A.7 Fragment EXPA3-2 A Genome Amplicon Haplotypes 135

A.8 Fragment EXPA3-2 D Genome Amplicon Haplotypes 136

A.9 Fragment EXPA4-1 A Genome Amplicon Haplotypes 136

A.10 Fragment EXPA4-2 A Genome Amplicon Haplotypes 136

A.11 Fragment EXPA4-2 D Genome Amplicon Haplotypes 136

A.12 Fragment EXPA4-3 D Genome Amplicon Haplotypes 137

A.13 Fragment EXPA5-1 A Genome Amplicon Haplotypes 137

A.14 Fragment EXPA5-1 D Genome Amplicon Haplotypes 137

A.15 Fragment EXPA5-2 A Genome Amplicon Haplotypes 137
A.16 Fragment EXPA5-2 D Genome Amplicon Haplotypes 138
A.17 Fragment EXPA6-1 A Genome Amplicon Haplotypes 138
A.18 Fragment EXPA6-1 D Genome Amplicon Haplotypes 138
A.19 Fragment EXPA6-2 A Genome Amplicon Haplotypes 138
A.20 Fragment EXPA6-2 D Genome Amplicon Haplotypes 139
A.21 Fragment EXPA6-3 A Genome Amplicon Haplotypes 139
A.22 Fragment EXPA6-3 D Genome Amplicon Haplotypes 139
A.23 A-genome Haplotypes of Gene *MYB1* 139
A.24 D-genome Haplotypes of Gene *MYB1* 140
A.25 A-genome Haplotypes of Gene *MYB2* 140
A.26 D-genome Haplotypes of Gene *MYB2* 140
A.27 A-genome Haplotypes of Gene *MYB3* 140
A.28 D-genome Haplotypes of Gene *MYB3* 141
A.29 A-genome Haplotypes of Gene *MYB4* 141
A.30 D-genome Haplotypes of Gene *MYB4* 141
A.31 A-genome Haplotypes of Gene *MYB5* 141
A.32 D-genome Haplotypes of Gene *MYB5* 142
A.33 A-genome Haplotypes of Gene *MYB6* 143
A.34 D-genome Haplotypes of Gene *MYB6* 143

LIST OF FIGURES

3.1 Linkage Maps of Selected SNP Markers Derived from *MYB* Genes with SSR Markers 53

5.1 Phenotype of Four Parental Lines and Their Half Diallele Crosses Derived F_1 92

5.2 Phenotype of Six Parental Lines and Some of the Crosses Derived F_1 93

5.3 Abnormal Phenotypes Observed in Li_2 and its Related Crosses with the other Parental Lines 94

A.1 Phylogenetic Analysis and Genome Specific Sequences Distinguishing of Fragment EXPA1 144

A.2 Phylogenetic Analysis and Genome Specific Sequences Distinguishing of Fragment EXPA2-1 144

A.3 Phylogenetic Analysis and Genome Specific Sequences Distinguishing of Fragment EXPA2-2 145

A.4 Phylogenetic Analysis and Genome Specific Sequences Distinguishing of Fragment EXPA3-1 145

A.5 Phylogenetic Analysis and Genome Specific Sequences Distinguishing of Fragment EXPA3-2 146

A.6 Phylogenetic Analysis and Genome Specific Sequences Distinguishing of Fragment EXPA4-1 146

A.7 Phylogenetic Analysis and Genome Specific Sequences Distinguishing of Fragment EXPA4-2 147

A.8 Phylogenetic Analysis and Genome Specific Sequences Distinguishing of Fragment EXPA4-3 147

A.9 Phylogenetic Analysis and Genome Specific Sequences Distinguishing of Fragment EXPA5-1 148

A.10 Phylogenetic Analysis and Genome Specific Sequences Distinguishing of Fragment EXPA5-2 148

A.11 Phylogenetic Analysis and Genome Specific Sequences Distinguishing of Fragment EXPA6-1 149

A.12 Phylogenetic Analysis and Genome Specific Sequences Distinguishing of Fragment EXPA6-2 149

A.13 Phylogenetic Analysis and Genome Specific Sequences Distinguishing of Fragment EXPA6-3 150

A.14 Phylogenetic Grouping Based Duplicated Loci Differentiation in *MYB1* Gene 151

A.15 Phylogenetic Grouping Based Duplicated Loci Differentiation in *MYB2* Gene 152

A.16 Phylogenetic Grouping Based Duplicated Loci Differentiation in *MYB3* Gene 152

A.17 Phylogenetic Grouping Based Duplicated Loci Differentiation in *MYB4* Gene 153

A.18 Phylogenetic Grouping Based Duplicated Loci Differentiation in *MYB5* Gene 153

A.19 Phylogenetic Grouping Based Duplicated Loci Differentiation in *MYB6* Gene 154

A.20 Phylogenetic Analysis of the Overall Six *MYB* Genes 155

CHAPTER I

INTRODUCTION

Cotton (*Gossypium* spp.) is the most important natural textile fiber and a significant oilseed crop in the world. The genus *Gossypium* is composed of approximately 50 species split across two ploidy levels, diploid (2n=2x=26) and tetraploid (2n=4x=52). Among them, two diploid species from Africa-Asia (*G. herbaceum* L., A_1; *G. arboreum* L., A_2) and two tetraploid species from Americas [*G. hirsutum* L., $(AD)_1$, Upland cotton; *G. barbadense* L., $(AD)_2$, extra long staple cotton] have been independently domesticated for the specialized single-celled trichomes, or fibers, that occur on the epidermis of the seeds (Wendel and Cronn 2003). However, more than 95% of the annual cotton crop worldwide is Upland cotton and extra long staple cotton accounts for less than 2% (National Cotton Council, http://www.cotton.org, 2006). Diploid species are divided into eight genomic groups (A to G, and K). The five extant allotetraploid species originated in the New World from interspecific hybridization between an A genome-like ancestral African species and a D genome-like American species (Wendel and Cronn 2003). The closet extant relatives of the original tetraploid progenitors are the A genome species *G. herbaceum* (A_1) or *G. arboreum* (A_2) and the D genome species *G. raimondii* Ulbrich (D_5) (Brubaker et al. 1999).

In addition to their economic importance, cottons are an excellent model system for several important biological studies, including crop domestication, evolutionary genomics, plant polyploidization (Wendel and Cronn 2003), and single cell elongation and cell wall and cellulose biosynthesis (Kim and Triplett 2001). Decoding cotton genomes through genetics, genomics, and molecular biology approaches will not only facilitate understanding the above important biological issues but also eventually aid in cotton improvement.

Challenges for current genetic improvement of cotton

Two of the major impediments to the genetic improvement of cotton fiber are: (1) insufficient information about genes that control important fiber traits and (2) lack of suitable markers useful for marker assisted breeding program. A logical prelude to a breeding strategy for fiber improvement is the establishment of research programs that identify and characterize important candidate genes and markers associated with fiber development.

Current need in cotton genetics and genomics research

Molecular markers used in cotton (*Gossypium* spp.) genome mapping and genetic diversity analysis have evolved from hybridization-based RFLPs to PCR-based markers such as RAPDs, AFLPs, and microsatellites. However, the low level of polymorphism associated with candidate genes limits their use in cotton genetic diversity analysis, integrated genetic mapping, and important candidate gene mapping (Liu et al. 2000). Single-nucleotide polymorphisms (SNP), including single-base changes or indels

(insertion or deletion) at specific nucleotide position, have been shown to be the most abundant source of DNA polymorphisms in many organisms and the ideal molecular marker associated with functional genes (Brookes 1999). The research on SNP analysis in cotton lags behind other crops due to its large genome size, tetraploid nature, the presence of high repetitive DNA content, and paucity of information on genomic sequences. The presence of homoeologous and paralogous sequences in polyploidy complicates SNP identification (Mochida et al. 2003; Somers et al. 2003). Discovery of SNP markers from candidate genes especially from fiber development related genes would increase the efficiency and available marker number in cotton molecular mapping.

The seed coat of cultivated cotton is covered with lint and fuzz fiber. To date, three types of spontaneous fiber mutants (fiberless, fuzzless-linted, and fuzzy-short lint) have been identified or developed and the inheritance modle has been reported. Inheritance of fuzzlessness indicated that at least two loci are involved, which designated *N1* and n_2, in addition to the presence of epistatic and modifier gene effect in some cases (Nadarajan and Rangasamy 1988; Du et al. 2001; Turley and Kloth, 2002). Inheritance of the short lint phenotype in cotton was reported to be controlled by two independent and completely dominant genes designaged Li_1 and Li_2, respectively (Narbuth and Kohel 1990; Kohel et al. 1992). To date, six genotypic models have been reported for the five fiberless lines collected from different parts of world (Musaev and Abzalov 1972; Nandarajan and Rangasamy 1988; Zhang and Pan 1991; Du et al. 2001; Turley 2002; Turley and Kloth 2002 and 2008; Turley 2008). Therefore, a general inheritance model integrating information from both conventional and molecular genetics associated with

fiber development is not available (Rong et al 2005 and 2007; Zhang et al. 2005; Guo et al. 2006; Abdurakhmonov et al. 2007; Wan et al. 2007).

Objectives of this research

The objectives of this research include: 1) develop a suitable SNP maker discovery strategy in tetraploid cotton; 2) SNP characterization and marker discovery from selected fiber development related genes; 3) chromosomal assignment of SNP markers; 4) molecular mapping of yield component, fiber and seed traits with selected fiber mutants; and 5) genetic analysis of fuzz and lint fiber development in the selected fiber mutants.

Justification and usefulness of this research

A suitable and applicable strategy was developed for discovering SNP markers in a polyploidy species like cotton. This strategy overcomes the complications of duplicated genes, which are common to many polyploidy crop species. SNP markers derived from genes associated with fiber initiation and elongation will not only be useful for chromosome localization and molecular mapping but also will facilitate genetic dissection of economically important fiber and yield traits. Molecular mapping of mutants is a valuable step toward understanding the role of functional genes in complex quantitative traits. Genetic analysis of the tested fiber mutants in this study will enhance understanding and clarification of genetic control models of cotton fiber development.

CHAPTER II

SNP CHARACTERIZATION AND SNP MARKER-BASED CHROMOSOMAL ASSIGNMENT OF SIX *EXPANSIN A* GENES IN COTTON

Abstract

Knowledge of biological significance associated with DNA markers is very limited in cotton. SNPs are potential target functional marker loci to tag genes of biological importance. Plant expansins are a group of extracellular proteins that directly modify the mechanical properties of cell walls, enable turgor-driven cell extension, and likely affect length and quality of cotton fibers. Objectives were development of SNP markers, assess SNP characteristics, and chromosomally localize six *EXPANSIN A* genes. Ancestral and homoeologous relationships of six *EXPANSIN A* genes were revealed by phylogentic grouping and comparison to extant A- and D-genome relatives of contemporary AD-genome cottons. The average rate of SNP per nucleotide was 2.35% (one SNP per 43 bp), with 1.74% and 3.99% occurring in coding and noncoding regions, respectively, in the selected genotypes. An unequal evolutionary rate of the *EXPANSIN A* genes at the subgenomes level of tetraploid cotton was recorded. Chromosomal locations for each of six *EXPANSIN A* genes were established by gene-specific SNP markers. Results revealed a strategy of discovering SNP markers in a polyploidy species like

cotton. These markers could be useful in associating candidate genes with complex fiber traits in MAS.

Introduction

Cotton (*Gossypium* spp.) is the leading natural fiber crop of the world. Approximately 90% of cotton's value resides in the fiber (lint). Although lint production has increased recently, fiber quality has been declining over the last decade. Fiber quality, determined by micronaire, length, strength, elongation, and uniformity, is the most important factor in modern spinning technology and profitability. Botanically, the fiber is a single-celled trichome developing from individual epidermal cells on the outer integument of cotton ovules. The development of fiber cells undergoes four discrete, yet overlapping stages: differentiation, expansion/primary cell wall (PCW) synthesis, secondary cell wall (SCW) synthesis, and maturation (Wilkins and Jernstedt 1999; Wilkins and Arpat 2005). So far, many genes involved in cotton fiber development have been isolated and characterized (Arpat et al. 2004). There is strong interest from both a biological and economical standpoint to identify genes for which expression closely parallels the rate of cotton fiber expansion and elongation.These genes will likely impart major influences on cell wall development, fiber quality and many other plant attributes. Most of the economically important fiber traits are controlled by quantitative trait loci (QTLs). Knowledge of functional genes underlying fiber quality QTLs is very limited in cotton.

Expansins are a large family of extracellular proteins that loosen the components of rigid plant cell walls and thereby allow cell expansion (Darley et al. 2001; Li et al.

2002; Sampedro and Cosgrove 2005). Regulation of cell wall extensibility during cell expansion was reported to be controlled, in part, by different expression of *EXPANSIN* genes in tomato (Vogler et al. 2003) and cotton (Arpat et al. 2004). These functions suggest expansins could significantly affect economic fiber properties such as length and elongation. Following the recommendation of an ad hoc working group for nomenclature of the *EXPASIN* gene family, '*α-expansin*' were be referred to *EXPANSIN A* (abbreviated as *EXPA*) (Kende et al. 2004). A key research model for cell biogenesis research is the cotton fiber, which arises from a single epidermal cell of the ovule integument that commences extensive unipolar expansion on the day of anthesis (0 days post-anthesis [dpa]) and lasts to approximately 20 dpa (Smart et al. 1998; Wilkins and Jernstedt 1999). Elongation of cotton fibers is highly polar and rapid; growth rates in cultivated species exceed 2mm/day during peak growth (Wilkins and Jernstedt 1999). Expression of several *EXPANSIN* genes parallels fiber elongation (Shimizu et al. 1997; Orford and Timmis 1998; Ruan et al. 2001; Harmer et al. 2002; Ji et al. 2003). Comprehensive analyses of the cotton fiber transcriptome showed that *GhEXPA1* (AF043284) is one of the top 15% expressed genes in *G. arboreum* L. cv. AKA8401 (Arpat et al. 2004).

Molecular markers used in cotton genome mapping and genetic diversity analysis have evolved from hybridization-based RFLPs (Reinisch et al. 1994; Shappley et al. 1998; Rong et al. 2004) to PCR-based markers such as RAPDs (Kohel et al. 2001), AFLPs (Abdalla et al. 2001; Mei et al. 2004), and microsatellites (Zhang et al. 2002; Han et al. 2004, 2006; Park et al. 2005; Frelichowski et al. 2006). For these markers, relatively low levels of intraspecific polymorphism and limited association with candidate genes have hampered integrated genetic mapping and important candidate gene mapping. DNA

markers specific to these candidate genes will help to associate biologically important genes with complex fiber QTLs. Single-nucleotide polymorphism (SNP), including single-base changes or indels (insertion or deletion) at specific nucleotide positions, has been shown to be the most abundant class of DNA polymorphisms in many organisms (Kwok et al. 1996; Wang et al. 1998; Brookes 1999; Cho et al. 1999). SNP variation analysis and SNP marker development from candidate genes could provide valuable information regarding their evolution and effects on complex traits. The anticipated value of SNPs for analysis of candidate gene evolution and their effects on complex traits have stimulated large scale SNP characterization and marker mapping in rice (Feltus et al. 2004), wheat (Mochida et al. 2003; Somers et al. 2003; Zhang et al. 2003; Caldwell et al. 2004), maize (Ching et al. 2002; Batley et al. 2003), soybean (Zhu et al. 2003; Kim et al. 2005), and barley (Kanazin et al. 2002; Bundock et al. 2003; Bundock and Henry 2004). Most cotton sequence variation analyses have been confined to single gene or DNA fragments for phylogenetic analysis (Small et al. 1998; Small and Wendel 2000; Cronn et al. 2002; Alvarez et al. 2005). Candidate gene-based association mapping using SNP markers has emerged as a powerful tool to determine the role of genes in complex traits (Glazier et al. 2002). Despite the use of SNPs in studies on human diseases, few SNP analyses have been carried out in plants, especially polyploid crops compared to other types of markers (Kanazin et al. 2002; Batley et al. 2003; Neale and Savolainen 2004). The research on SNP analysis in cotton is almost nil due to the large genome size (Grover et al. 2004), tetraploid nature, the presence of high repetitive DNA content (Zhao et al. 1998), and paucity of information on genomic sequences (Chee et al. 2004). Sequence-tagged sites (STS) sequencing results indicated that the rate of variation per nucleotide

was 0.35% between *G. hirsutum* and *G. barbadense*, and the variation per nucleotide were 0.14% and 0.37% within these two species, respectively (Rong et al. 2004).

Here, SNP analysis and marker-based gene chromosomal assignment to six well characterized cotton *EXPANSIN A* genes (Harmer et al. 2002) were studied. It provides a useful homoeologous sequences distinguishing strategy for allotetraploid cotton SNP discovery. The chromosomal locations and SNP markers derived from six *EXPANSIN A* genes will be useful for integrated genetic mapping and fiber quality related QTLs analysis in cotton.

Materials and Methods

Plant materials and DNAs isolation

Four tetraploid species, TM-1, HS46 and MARCABUCAG8US-1-88 (MAR) (*G. hirsutum* L., AD_1), 3-79 (*G. barbadense* L., AD_2), *G. tomentosum* Nuttall ex Seemann (AD_3), *G. mustelinum* Miers ex Watt (AD_4), and two diploid genome species, *G. arboreum* L. (A_2) and *G. raimondii* Ulbrich (D_5), were used for PCR amplification and SNP marker identification. TM-1 is a genetic standard for *G. hirsutum*. 3-79 is a double haploid line of *G. barbadense*. HS46 and MAR are the two parents of recombinant inbred lines developed for mapping projects (Shappley et al. 1998; Ulloa et al. 2005). *G. tomentosum* and *G. mustelinum* are two wild cotton species. *G. arboreum* and *G. raimondii* are extant relatives of species that donated the A and D genomes of the original AD allotetraploid that gave rise to modern 52-chromosome *Gossypium* species (Brubaker et al. 1999; Wendel and Cronn 2003). Two kinds of genetic stocks were used for

chromosomal assignment of *EXPANSIN A* genes by deletion analysis: (1) Quasi-isogenic hypoaneuploid interspecific F_1 hybrid chromosome substitution stocks, each involving chromosomally identified primary monosomy, monotelodisomy or tertiary monosomy (Liu et al. 2000; Ulloa et al. 2005), and (2) quasi-isogenic euploid CS-B lines BC_5S_1-derived interspecific backcrossed chromosome substitution lines of 3-79 in TM-1 (Stelly et al. 2005; Jenkins et al. 2006; Saha et al. 2006a, b). The primary monosomic plants (2n=51) lacked an entire chromosome of the normal *G. hirsutum* complement, whereas monotelodisomic plants (2n=52) lacked most or all of just one *G. hirsutum* chromosome arm. Tertiary monosomics were deficient for one of the two reciprocally translocated *G. hirsutum* chromosomes described in detail by Brown et al. (1981) and Menzel et al. (1985). Thus, each tertiary monosomic plant lacked the centric segment from one *G. hirsutum* chromosome and the distal acentric segment from the other chromosome involved in the translocation. Two sets of monosomic and monotelodisomic F_1 interspecific hybrids, from *G. hirsutum* aneuploids crossed with 3-79 (*G. barbadense*) versus *G. tomentosum*; and one set of tertiary monosomic hybrids, from crosses with *G. tomentosum*, only were used for chromosomal assignment. DNA samples of the diploid species (*G. arboreum* and *G. raimondii*) were kindly provided by Dr. John Yu (USDA-ARS, Crop Germplasm Research Unit, College Station, TX). Genomic DNA of the different genotypes and species were isolated from young leaves of individual plants using a DNeasy Plant Maxi Kit (Qiagen Inc, Valencia, CA).

PCR primer design, amplification, cloning, and sequencing

A total of 13 primer-pairs were designed for six cotton *EXPANSIN A* genes (Table 2.1) using Primer 3 software (http://frodo.wi.mit.edu/). For each gene, there was at least one primer-pair designed to amplify 400-800 bp for sequencing. Gene-specificity of each primer was tested using BLASTN against cotton genomic sequences in GenBank. *Pfu* polymerase (Stratagene, La Jolla, CA) was used for PCR amplification according to the manufacturer's protocol on a PTC-225 Peltier Thermal Cycler (MJ Research Inc, Waltham, MA). The PCR products were excised from agarose gels following electrophoresis, purified using QIAEX II gel extraction kit (Qiagen Inc, Valencia, CA), and cloned into the pCR4-TOPO vector (Invitrogen, Carlsbad, CA) after adding the 3'-A to the purified DNA fragment according to the manufacturer's protocol. Plasmid DNA isolated from kanamycin-resistant colonies was bi-directionally sequenced with ABI Prism BigDye Terminator Cycle Sequencing Kit v3.1 (Applied Biosystems, Foster City, CA) on an ABI automated sequencer. PCR products were initially cloned because allotetraploid cotton contains two homoeologous copies for each gene derived from At- and Dt-genomes if duplications and heterozygous loci were not considered. In this study, 12 recombinant colonies were sequenced for each *EXPANSIN A* gene generated from each genotype to avoid possible complications due to PCR recombination (Cronn et al. 2002). Forward and reverse matched sequences from at least three clones were used to determine the sequence for each duplicated copy (Cedroni et al. 2003; Rong et al. 2004).

Table 2.1 PCR Primers Designed from Six *EXPANSIN A* Genes

Gene	NCBI accession No.[1]	Primer	Primer sequence (5'→3')
GhEXPA1	AF512539	GhEXPA1	F: CCCCACGAGAACACTTTGAT R: CTAATGGCACTTGCTTGCCT
GhEXPA2	AF512540	GhEXPA2-1	F: GTCAGCCAATTGTTTGAGCTA R: TAGATAAAGCATAGTTAGGGG
		GhEXPA2-2	F: ACAGCCACCAACTTTTGTCCC R: AGTTTGTCCGAATTGCCAAC
GhEXPA3	AF512541	GhEXPA3-1	F: GTATGCTTTTTGGTATGCAG R: GTGTGTCGGTGGAAAATG
		GhEXPA3-2	F: TTTGACAATGGCTTGAGCTG R: TCCGTTACTGGTTGTGACGA
GhEXPA4	AF512542	GhEXPA4-1	F: TACGCCCGATATTCAACACA R: CGTTTGCGCACTTAATCTCA
		GhEXPA4-2	F: TGAGATTAAGTGCGCAAACG R: GCCAGTTTTGACCCCAGTTA
		GhEXPA4-3	F: TGAAGGTGAAGGGAACCAAC R: CCCAACCCCCATTTTTACTT
GhEXPA5	AF512543	GhEXPA5-1	F: GTGCCCACCCAATAATTAAA R: ATGTTAATCGTACCTCCGAT
		GhEXPA5-2	F: TGATTTTCGAAGGGTGCCAT R: TATTGCATGCTCCCAAACAC
GhEXPA6	AF512544	GhEXPA6-1	F: CTGTTGTTTGTTCGCAGGAA R: GAAGCAGCAAAAGGCAAAAC
		GhEXPA6-2	F: TGGCCACTCCTACTTCAACC R: GACCCGAAAGTCCCACTACA
		GhEXPA6-3	F: GTTTTGCCTTTTGCTGCTTC R: GAGAGGCTTTGTCCGTTGAG

[1]The NCBI accession numbers represent the original genes isolated from *G. hirsutum* L. cv Siokra 1-4 (Harmer et al. 2002); In this report, the standard nomenclature of the *EXPANSIN* genes was used (Kende et al. 2004).

Sequence analysis and SNP primer design

Alignment of diverse sequences was conducted using ClustalX (Thompson et al. 1997), and phylogenetic relationships were analyzed by the neighbor joining (NJ) method using MEGA version 3.1 (Kumar et al. 2004). The putative assignment of a sequence to a particular locus or subgenome was based on phylogenetic analysis and the relationship to diploid ancestral species (A_2 and D_5) of the tetraploid cotton. DnaSP 4.0 software (Rozas et al. 2003) was used to identify SNPs based on a comparative alignment of sequences at a putative locus, and estimate of nucleotide and haplotype diversity. Final average nucleotide diversities (π) were calculated from all pairwise comparisons (Tajima 1983; Nei 1987). Interspecies SNP primers were designed based on single nucleotide differences in the sequences at a putative locus between TM-1 and 3-79 or TM-1 and *G. tomentosum* for chromosomal assignment of SNP markers. Primers were selected to anneal immediately upstream or downstream of the SNP site as the forward or reverse primer, respectively, so that the polymorphism could be detected by one-base extension technology with the ABI Prism SNaPshot^{TM} multiplex kit. Selected primers were evaluated by Primer 3 (http://frodo.wi.mit.edu/) using criteria of a primer length of 18-26 nt (20 nt as the optimum), an optimum annealing temperature of 50 °C, and a 40-60% GC content.

SNP genotyping

The ABI Prism SNaPshot™ multiplex kit and an ABI 3100 capillary electrophoresis system were used for screening SNP markers following a slight modification of the manufacturer's protocol. Templates for SNP genotyping were amplified using *Pfu* polymerase and then purified by incubation with Shrimp Alkaline Phospatase (SAP) and *Exo* I (2 units of SAP and 4 units of *Exo* I in a 20 µl PCR reaction volume) at 37°C for 1 hr and then at 75°C for 15 min. The thermal cycle reaction mixture (7 µl reaction system in 384-well plates) contained 1.5 µl of SNaPshot Multiplex Ready Reaction Mix, 0.5 µl of purified PCR product, 0.2 µl of SNP primer (10 µM), and 4.8 µl of distilled water. The thermal cycle reaction was carried out for 25 cycles at 96°C for 10 sec, 50°C for 5 sec, and 60°C for 30 sec. After treatment with SAP (1 unit) at 37°C for 1 hr, and followed by incubating the reactions at 75°C for 15 min, 1 µl of 10-fold diluted SNaPshot product was mixed with 0.2 µl of GeneScan-120 LIZ size standard, and 8.8 µl of Hi-Di formamide, denatured at 95°C for 5 min, and then loaded onto the ABI 3100 Genetic Analyzer (Applied Biosystems, Foster City, CA) in SNaShot mode for SNP marker analysis.

Chromosomal assignment of EXPANSIN A genes

The deletion analysis method was used for identifying chromosomal locations of six *EXPANSIN A* genes (Liu et al. 2000; Ulloa et al. 2005). All of the aneuploid chromosome substitution F_1 lines, except the particular aneuploid line missing a specific chromosome or chromosome arm, would expectedly show two alleles originating from both parents, similar to an F_1 heterozygous locus. The absence of the TM-1 (*G. hirsutum*)

allele in any one of the aneuploid F_1 plant indicated the missing chromosome or chromosome arm was the most likely location of the gene of interest. In euploid CS-B stocks, the absence of the TM-1 allele but presence of the 3-79 allele suggested that the gene was probably located in the substituted chromosome or chromosome arm.

Results

Sequence characterization of six EXPANSIN A genes

Sequence data were obtained from PCR fragments amplified from *EXPANSIN A* genes in selected cotton genotypes using gene-specific primer-pairs (Table 2.1). First, the identities of all the sequences in these results were confirmed to the respective original gene by BLASTN. Phylogenetic analyses were then used to discriminate homoeologous gene sequences based on the relationship with the ancestrally related extant genomes A_2 and D_5. Within each genome cluster, sequences from the available genotypes were classified to one putative locus based on the clustering results of the phylogram. No duplicated loci were found within either of the At- or Dt-genomes for any of six *EXPANSIN A* genes. Any locus-specific sequence difference identified between the selected genotypes was considered as a SNP. Nucleotide sequences from six *EXPANSIN A* genes were deposited to GenBank database with the accession numbers EF644199 to EF644335. Genome assignments and phylogenetic analyses of all PCR-amplified gene fragments are summarized in Figure A.1-A.13.

In total, 52.9 kb of DNA sequences (33.7 kb and 19.2kb from coding and noncoding regions, respectively) were acquired from both genomes after eliminating overlapping sequences (Table 2.2). Available sequence from A- and D-genomes was 31.7

kb and 21.2 kb, respectively. BLASTN analysis showed that amplified gene segments, ranging in size from 340 bp to 607 bp, had 98%-100% nucleotide identity to the target gene. A total of 222 SNPs, including 120 single-base changes and 102 indels, were identified in 134 amplicons (Table 2.2). Transitions accounted for 69 (57.5%) and transversions for 51 (42.5%) of the total 120 single-base changes. The ratio of 'A/G' to 'T/C' transitions was 1.23:1, with no significant difference between the four types of transversions. Analysis of indel sequences indicated a bias toward 'A' and 'T' nucleotides, which is similar to the 'A' nucleotide bias in maize (Batley et al. 2003). The average rate of SNPs per nucleotide was 2.35%, with 1.74% and 3.99% occurring in coding regions and noncoding regions, respectively, in selected genotypes. Based on the average rate of SNP per nucleotide, a higher nucleotide change rate was discovered in the D-genome (2.90%) than that of the A-genome (1.98%). Furthermore, an uneven distribution of SNP was observed among the six genes, suggesting that the occurrence of SNP varies among six *EXPANSIN A* genes. Amplicons of EXPA2-2 and EXPA5-1 in *G. tomentosum* and *G. raimondii* had a 46- and 35-nucleotide deletion, respectively, thereby contributing to a higher rate of observed nucleotide polymorphism in the Dt-genome of tetraploid species. One putative triallelic SNP site ('A/C/G') was discovered in the A-genome amplicon of EXPA5-1, similar to triallelic SNP loci observed in soybean ESTs (Van et al. 2005). The simple sequence repeat ('TA') motif is present in the intron of the EXPA5-1 amplicon, as previously reported (Kumar et al. 2006). The 119 cSNPs included 83 single-base changes and 36 indels in the 33.7 kb of coding sequence. Among the single-base changes, 26, 31 and 26 were detected in the first, second, and third codon positions, respectively. Results revealed that a total of 56 out of the 119 cSNPs (47%)

were nonsynonymous changes, provided that the aforementioned fragment-deletions of EXPA2-2 and EXPA5-1 in *G. tomentosum* and *G. raimondii* were not considered.

Detail results of haplotype organization are included in the Tables A.1-A.22. The number of SNP haplotypes present in each group was determined when the available sequence number exceeded two and more than one SNP variable site presented in the aligned sequences. Results showed that the haplotype number ranged from two to six out of the maximum seven available genotype sequences, with an estimated haplotype diversity that varied between 0.3 and 1 (Table 2.2). The relatively high haplotype number indicated the distinct and diverse SNP characters among the selected cotton species, especially at the inter-species level, which represents valuable sources of germplasm for Upland cotton improvement.

Table 2.2 Distribution and Types of SNPs in PCR Amplicons of Six Cotton *EXPANSIN A* Genes

Fragment	Geno.	Seq. No.	Haplotypes[1]		Transitions		Transversions				Indels				Coding region		Total	
			H[2]	H_d[3]	A/G	T/C	A/T	G/C	A/C	G/T	A	C	G	T	Length (bp)	SNPs No. (non-synonymous changes)	Length (bp)	SNPs No.
EXPA1	A	4	3	0.83±0.22	3	1	0	0	0	1	1	1	0	1	392	4(3)	607	8
	D	7	5	0.86±0.14	4	4	0	2	0	2	0	0	0	1	392	7(6)	607	13
EXPA2-1	A	1	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
	D	7	3	0.67±0.16	1	0	0	0	1	0	0	0	0	0	358	2(1)	474	2
EXPA2-2	A	4	4	1.00±0.18	4	1	0	1	0	1	1	1	0	0	435	6(3)	511	9
	D	7	5	0.86±0.14	4	2	0	2	0	2	1	0	0	0	435	10(5)[4]	511	11
EXPA3-1	A	7	3	0.52±0.21	1	0	0	1	1	0	1	2	0	0	416	2(1)	595	6
	D	1	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
EXPA3-2	A	6	3	0.60±0.21	1	0	0	0	1	0	0	0	0	0	444	2(2)	531	2
	D	5	3	0.70±0.22	1	1	0	0	1	0	0	0	0	0	444	2(1)	527	3
EXPA4-1	A	7	4	0.81±0.13	0	2	0	0	1	0	0	0	0	1	293	3(1)	552	4
	D	1	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
EXPA4-2	A	6	6	1.00±0.10	1	2	2	2	0	0	0	0	0	0	383	4(3)	538	7
	D	6	5	0.93±0.12	0	3	2	0	1	1	0	0	0	0	383	2(0)	538	7
EXPA4-3	A	3	-	-	0	0	1	0	0	0	0	0	0	0	204	0(0)	340	1
	D	7	5	0.91±0.10	1	1	3	1	0	1	0	0	0	0	204	3(2)	340	7
EXPA5-1	A	7	4	0.71±0.18	3[5]	0	0	1	1	1	1	0	0	1	310	6(5)	420	8
	D	5	3	0.70±0.22	1	2	0	0	0	0	11	12	7	10	310	37(1)[6]	447	43
EXPA5-2	A	6	4	0.87±0.13	0	0	1	0	2	0	1	0	0	0	313	2(2)	509	4
	D	3	2	0.67±0.31	0	1	1	0	0	0	1	0	0	0	313	2(2)	510	3
EXPA6-1	A	5	4	0.90±0.16	3	4	0	2	0	1	0	0	1	0	440	8(4)	572	11
	D	7	5	0.86±0.14	4	2	1	0	0	1	0	0	0	0	440	7(5)	572	8
EXPA6-2	A	5	3	0.70±0.22	1	0	0	0	0	1	0	0	0	0	265	2(2)	417	2
	D	6	2	0.33±.22	1	1	0	0	0	0	0	2	0	1	265	2(2)	420	5
EXPA6-3	A	7	4	0.71±0.18	2	3	0	0	1	2	12	2	8	19	204	2(2)	591	49
	D	7	4	0.81±0.13	2	1	2	0	0	2	1	0	0	1	204	4(3)	550	9

Table 2.2 (continued)
[1]When the number of SNPs in the amplified fragment is less than two, the haplotype analysis had not been conducted. Sites with alignment gaps were not considered for haplotypte analysis.
[2]H means haplotype number.
[3]H_d refers to haplotype diversity.
[4]*G. tomentosum* and *G. raimondii* have one gap in the coding region, which was not considered in the amino acid change analysis.
[5]There was one triallelic SNP (A/C/G) site. Here, A/G transition was recorded at this position.
[6]Thirty-five gaps appeared between *G. tomentosum* and *G. raimondii* were not considered in the amino acid change analysis.

Nucleotide diversity

Nucleotide diversity (π) across all possible comparisons among the eight genotypes was measured using available sequences from A- or D-genomes. Because only two sequences were compared at a time, and no segregating populations were considered, only π value for the nucleotide diversity assay which measures the average number of nucleotide differences per site between two sequences (Nei 1987) were reported. The mean value of π in the A- and D-genome is summarized in Table 2.3. Results showed that nucleotide diversity was lowest among the three *G. hirsutum* lines (TM-1, HS46, and MAR), revealing that polymorphisms among six *EXPANSIN A* genes were much higher at the interspecific level than intraspecific level in *Gossypium* species. The lowest nucleotide diversity was found between HS46 and MAR in both At- and Dt-genomes, which was observed as 0.02 and 0.10, respectively. Nucleotide diversities (π values) among the three *G. hirsutum* lines in the Dt-genome were significantly higher than that in the At-genome, indicating that loci in the Dt-genome exhibit a faster evolutionary rate among six *EXPANSIN A* genes compared to the At-genome in tetraploid cotton.

The independent and incongruent evolution of the two subgenomes (At and Dt) was also revealed by the different phylogenetic topologies detected in polyploid duplicated genes or fragments of the selected tetraploid species. The highest nucleotide diversity in the A-genome was observed between the ancestral A-genome species (*G. arboreum*) and *G. tomentosum* (π=0.67), while in the D-genome, the highest nucleotide diversity was obtained from *G. raimondii* (ancestral D-genome donor) and *G. mustelinum* (π=0.77) (Table 2.3). When considering all other genotypes, the different phylogenetic patterns of At- and Dt-genomes became even more evident. Analyses of six *EXPANSIN A*

genes suggest that polyploid speciation in cotton was accompanied by a diverse subgenome molecular evolution.

Table 2.3 The Nucleotide Diversity (π values, $\times 10^{-2}$) among Six PCR-amplified Cotton *EXPANSIN A* Gene Fragments in A- and D-genomes[1]

Genotype[2]	*G. arboreum*	TM-1	HS46	MAR	3-79	*G. mustelinum*	*G. tomentosum*
G. raimondii	-	0.56	0.55	0.71	0.62	0.77	0.68
TM-1	0.51	-	0.12	0.22	0.59	0.64	0.47
HS46	0.48	0.03	-	0.10	0.51	0.55	0.41
MAR	0.52	0.07	0.02	-	0.60	0.67	0.42
3-79	0.64	0.30	0.41	0.38	-	0.37	0.37
G. mustelinum	0.55	0.31	0.24	0.30	0.38	-	0.46
G. tomentosum	0.67	0.35	0.30	0.30	0.32	0.39	-

[1]π value, a measure of the average number of nucleotide differences per site between two sequences; The numbers below and above the diagonal represent A- and D-genome sequences pairwise comparisons, respectively.

[2]*G. arboreum* and *G. raimondii* were considered related to the diploid ancestral A- and D-genome progenitors of tetraploid species; TM-1 was considered as genetic standard of *G. hirsutum* and 3-79 was includes as a representative sample of *G. barbadense*; HS46 and MAR (abbreviation of MARCABUCAG8US-1-88) were used as two diverse *G. hirsutum* lines. One accession of *G. mustelinum* and *G. tomentosum* were included to represent another two allotetraploid species.

Chromosomal assignment

Single-nucleotide extensions with primers shown in Table 2.4 were used to genotype 25 gene-specific SNP markers in six *EXPANSIN A* genes. The chromosomal location of each SNP locus was delimited by deletion analysis with one or more hypoaneuploid or euploid stocks, and the results were compared for SNPs within and across six *EXPANSIN A* genes. The individual SNP loci were assigned to the long arms of chromosomes 20, 10, 9, 1, and 3 (Table 2.4). For the four genes (*GhEXPA3*, *GhEXPA4*, *GhEXPA5* and *GhEXPA6*) and the At-genome locus of gene *GhEXPA2* represented here by multiple SNPs, chromosomal assignments were concordant among SNPs within a gene. SNP markers Exp1-1_Gbmt_193F and Exp2-1_Gbmt_378R from gene *GhEXPA1* and Dt-genome locus of gene *GhEXPA2*, respectively, were both localized to chromosome arm 20Lo, which indicated chromosome locations of the two genes. Chromosomes 10 and 20 are homeologous, so it is possible that At- and Dt-homoeologous loci of *GhEXPA2* on the long arms of chromosomes 10 and 20 were detected. In a few cases where aneuploid analysis could not be confirmed by euploid CS-B lines or visa versa, chromatin losses during backcrossing or other types of cytological abnormalities in the development of these cytogenetic stocks could explain the results, suggesting that these particular cytogenetic stocks warrant further characterization.

Table 2.4 Chromosomal Locations of Six *EXPANSIN A* Genes in Cotton

Gene	SNP marker[1]	SNP primer sequence (5'→3')	Chromosome location[2]		
			Aneuploid *G. barbadense*	Aneuploid *G. tomentosum*	Euploid CS-B
GhEXPA1	Exp1-1_Gbmt_193F	GTCCGAATTGCCAACCAGC	20*Lo*	20*Lo*	N/A
GhEXPA2	Exp2-1_Gbmt_378R	CACTTTTCTTCTTTTTGTTCAGT	20*Lo*	20*Lo*	N/A
	Exp2-2_Gbt_58F	CAGCAGGCACTACATTGTAG	10*Lo*	10*Lo*	10
	Exp2-2_Gbt_59R	AGCGATGGCAGGACTATCACA	10*Lo*	10*Lo*	10
	Exp2-2_Gbt_93F	CATCGCTGGCAGTCACTTT	10*Lo*	10*Lo*	10
	Exp2-2_Gbt_108R	AGAGCAATGCTTACCTTAACGG	10*Lo*	10*Lo*	10
	Exp2-2_Gbt_175F	CTGGACATAGGTAGCCATCCTGTT	10*Lo*	10*Lo*	10
	Exp2-2_Gb_182R	GATATAACGTCAGTGTCCATCAAG	N/A	N/A	10
	Exp2-2_Gbt_312F	TGACACCCTGCAAAAGGT	10*Lo*	10*Lo*	10
	Exp2-2_Gbmt_345R	AAACTCAATTCAAATCATCAC	10*Lo*	10*Lo*	10
	Exp2-2_Gbt_415F	ACGATTCCAGCTCGATATTC	10*Lo*	10*Lo*	10
	Exp2-2_Gbmt_422R	CCGAACCGGCATTCTTGC	10*Lo*	10*Lo*	10
	Exp2-2_Gbt_508R	ACAGCCACCAACTTTTGTCC	N/A	10*Lo*	10
GhEXPA3	Exp3-1_Gb_489F	TGATCTCTCTCAGCCTATTTTT	10*Lo*	N/A	10
	Exp3-2_Gb_372F	TCTGTATTGGGCAATGTGTT	N/A	N/A	10
GhEXPA4	Exp4-1_Gbmt_65F	GCCATTATTGAAAAGTGCAG	9*Lo*	9*Lo*	9
	Exp4-1_Gbmt_147R	ATGGTGTTTGAATTTTTTT	9*Lo*	N/A	9
	Exp4-1_Gbm_412F	GCATTCACCACAGCCAAAAA	9*Lo*	N/A	9
	Exp4-2_Gbmt_244F	CAAAATCTTTCCCCTTTTACT	9*Lo*	9*Lo*	9
GhEXPA5	Exp5-1_Gmt_205R	CTACCAACTCAGGTGCGATTAC	N/A	1	N/A
	Exp5-2_Gbt_444F	ATACGTCATTAAATTTTCCC	1*Lo*	N/A	1
GhEXPA6	Exp6-1_Gb_77F	CTCGCGATGGTCTCTGGT	3*Lo*	N/A	N/A
	Exp6-1_Gbm_89F	GGTCTCTGGTGTTCAGGGATAT	3*Lo*	N/A	N/A
	Exp6-1_Gbmt_96R	TTGCATGTGCATTAGTCCAA	3*Lo*	3*Lo*	N/A
	Exp6-1_Gb_156R	CTAAAAGATGGCTTCATTTGAAGC	3*Lo*	N/A	N/A

[1]The nomenclature of the SNP markers followed the order: amplified fragment, polymorphic character, SNP site position in the amplified fragment, and forward or reverse primer. e.g. Exp2-2_Gbmt_422R means: this marker is located at position 422 of amplified fragment EXPA2-2; this SNP site is polymorphic between TM-1 and 3-79, *G. mustelinum* or *G. tomentosum*; it is a reverse amplification primer.[2]*Lo* means on the long arm of the chromosome.

Discussion

Cotton fiber elongation is the net result of the complex interplay between cell turgor and cell wall extensibility. The elongation is coupled with the expression of many genes, among which *EXPANSIN* is one of most highly expressed (Arpat et al. 2004). To more fully explore and differentiate among *EXPANSIN* genes, SNPs were assessed, SNP markers were developed in six well-characterized cotton *EXPANSIN A* genes, and each of them were assigned to chromosome. Success of the SNP marker development approach used here indicates a workable strategy to distinguish homoeologous sequences and thus SNP marker development for genes within multigene family. The chromosome localization results will expectedly contribute to the comparative map of cotton chromosomes and facilitate research that specifically tests whether any of six *EXPANSIN A* genes impact fiber quality, e.g., by enabling fiber quality SNP-QTL association analysis.

SNP marker discovery in tetraploid cotton

Polyploidy complicates the identification of SNPs by the fact that highly similar sequences exist in different subgenomes (homoeologous sequences) or duplicate copies within a genome (paralogous sequences) (Mochida et al. 2003; Somers et al. 2003). So far, two common methodologies have been employed to overcome this barrier. One is locus specific-PCR amplification and the other is an *in silico* approach based on clustering that distinguishes paralogs (Richert et al. 2002; Mochida et al. 2003; Somers et al. 2003; Caldwell et al. 2004). However, the high sequence conservation of homoeologous loci that occurs in both intergenic and genic regions of tetraploid cotton

hampered the design of genome-specific PCR primers for SNP identification (Grover et al. 2004).

Phylogenetic analyses of duplicated low- and single-copy sequences in cotton showed that homoeologs exhibit independent evolution. Cronn et al. (1999) proposed that most genes duplicated during polyploidization in *Gossypium* were expected to exhibit independent evolution in the allopolyploid nucleus. The evolution analysis on six cotton *MYB* transcriptional factors supported this conclusion (Cedroni et al. 2003). Results in this study, which were obtained in 13 amplified fragments from six *EXPANSIN A* genes, also indicated an independent evolution pattern. Here, SNPs were identified based on the theory of independent gene evolution in the At- and Dt-genomes of tetraploid species, and were combined with an integrated phylogenetic approach that incorporated orthologous sequences of ancestral diploid species represented in the closest living descendants (extant species) of the donor species as a base reference. Putative SNP identification was validated by deletion analysis for chromosomal assignment of six *EXPANSIN A* genes. The success of this approach was facilitated by the inbred nature of the plant materials (highly inbred lines or doubled haploid line) and use of gene-specific PCR primers, which avoided problems inherent to heterozygous alleles and orthologous sequences. Sequence assembly of *G. arboreum* fiber ESTs generated around 14,000 candidate genes, which offer vast potential for identifying genes that affect fiber quality (Arpat et al. 2004). The identification of candidate genes is relatively difficult in tetraploid species. Only a very small percent (~5%) of cotton fiber genes have been genetically mapped (Rong et al. 2004) due to the low polymorphism nature. Even more disappointing is the fact that <1% of the ~3000 fiber genes identified by transcriptome

profiling as being developmentally regulated and therefore representing important candidate genes show a lower than average polymorphism when using traditional RFLP and other DNA markers (Alabady and Wilkins unpublished data). The SNP discovery strategy presented here holds great promise for developing SNP markers from important fiber candidate genes, even though discovery is complicated by polyploidy. Success of this pilot study also indicates that SNP is applicable to large diverse gene super-families and eliminates the need to focus only on single- or low-copy genes in cultivated cotton species.

SNP features in cotton

Analysis of DNA sequence diversity among a subset of *EXPANSIN A* genes in diploid and tetraploid cotton revealed a mean SNP frequency of 2.35% (1 SNP per 43 bp of sequence), with 1.74% and 3.99% occurring in coding regions and noncoding regions, respectively, in selected cotton lines. Depending on the species and the genomic region under investigation, recent studies indicated that SNP frequencies range from as high as one SNP in 10 to 15 bp in some noncoding regions of the human genome (Brookes 1999) to a much lower one SNP per 4 kb in some highly conserved regions (Nickerson et al. 1998). In plants, SNP frequency also varies among species and is distributed unevenly across genomes. The average SNP frequency was 2.8 SNPs/kb in selected regions of rice and sorghum genomes (Feltus et al. 2004), one SNP per 70 bp in maize (Ching et al. 2002), and one SNP in 78bp, 189bp, and 9 bp in a particular gene(s) of grape (*Vitis vinifera*), barley (*Hordeum vulgare*), and wheat (*Triticum aestivum*), respectively (Kanazin et al. 2002; Caldwell et al. 2004; Salmaso et al. 2004). The nucleotide variation

generated from six *EXPANSIN A* genes in interspecifiec cotton lines was similar to the SNP frequency of selected genes in other crops. However, Rong et al. (2004) showed that the nucleotide variation of intraspecific DNA sequence (*G. hirsutum* and *G. barbadense*) within amplicons derived from genetically mapped sequence-tagged sites (STS) was 0.35%.

Six *EXPANSIN A* gene sequences showed greater nucleotide variation in introns than in exons. Due to the large and direct effect on the phenotype, exons exhibit the least amount of nucleotide variation compared to introns (Holland et al. 2001). Polymorphisms occur at a frequency two to three times higher in introns than in exons in *Arabidopsis* and soybean, respectively (The Arabidopsis Genome Initiative 2000; Zhu et al. 2003). In cotton species, Wilkins et al. (1994) showed that intron sequences were less conserved than exons of the V-ATPase catalytic subunit superfamily, which was recently supported by comparative sequences analysis (Chee et al. 2004; Kumar et al. 2006).

Unequal evolutionary rate of the At- and Dt-genome in cotton

Results suggested a different evolutionary pattern of six *EXPANSIN A* genes in the two subgenomes, such that the At-genome preferentially exhibits point mutations resulting in amino acid substitutions, while the Dt-genome preferentially displays short fragment deletions or insertions. The higher rate of SNP occurrence and nonsynonymous change in coding regions of the Dt-genome, and the nucleotide diversity comparison demonstrated that At- and Dt-genome of the tetraploid cotton evolve at unequal rates. These results were also supported by previous studies on phylogenetic analysis of low-copy gene sequences and molecular maker-based QTLs analysis. Phylogenetic analysis

of *Adh* (Small et al. 1998) and *FAD2-1* (Liu et al. 2001) showed a faster evolutionary rate in the Dt-genome than in At-genome of tetraploid cotton. Furthermore, a study on the molecular evolutionary rate variation of 48 nuclear genes in cotton showed that D-Dt comparisons displayed higher sequence divergence than in A-At comparisons (Senchina et al. 2003). Reinisch et al. (1994) reported that Dt-genome RFLP marker polymorphism levels were 10% higher than those of the At-genome. Moreover, inferences of the location of QTLs such as fiber-related traits (Jiang et al. 1998; Lacape et al. 2005), disease resistance (Wright et al. 1998), and leaf morphology (Jiang et al. 2000) repeatedly implied a higher evolutionary rate in the Dt-genome than the At-genome of Upland cotton. However, independent evolution of duplicated low- and single-copy sequences after polyploid formation in cotton has also been found in several independent phylogenetic analyses (Cronn et al. 1999; Small and Wendel 2000; Cedroni et al. 2003). These observations collectively indicate that evolutionary forces on the two genomes may have been fundamentally different, which results in independent evolution between the homoeologous loci in tetraploid cotton.

Chromosomal locations of six EXPANSIN A genes

Interspecific SNP markers were utilized by single nucleotide extension technology and deletion analysis (Liu et al. 2000) for chromosomal assignment of six *EXPANSIN A* genes. In cotton, developing candidate gene markers based on conventional PCR methods has been limited by the paucity of genomic sequence data and relatively high levels of monomorphism for PCR amplicon mobility (Chee et al. 2004; Kumar et al. 2006). SNP markers derived from candidate genes would facilitate not only chromosomal

assignment of these genes, but also in QTLs mapping and discovery of the roles of candidate genes in complex traits. In this study, the chromosomal locations of duplicated homoeologous copies of *GhEXPA1*, *GhEXPA3*, *GhEXPA4*, *GhEXPA5*, and *GhEXPA6* genes were not detected. This could be due to the incomplete genome coverage in cytogenetic stocks, lacking homoeologous duplicated sequences, and relatively low interspecific nucleotide diversity. Results revealed that all six *EXPANSIN A* genes may share a polyploid duplication event like *GhEXPA2* genes on the long arm of two homoeologous chromosomes 10 and 20 in tetraploid cotton. However, segmental duplication, an independent duplication event after polyploidization, might also play a role in gene evolution in cotton. For example, *GhEXPA2* and *GhEXPA3* both localize to the long arm of chromosome 10. Pfeil et al. (2004) suggested that the fate of the duplicated loci could be one of the several forms: (1) long-term maintenance of the same or similar function; (2) divergence in function; (3) loss of one copy; or (4) intralocus or interlocus gene conversion. Results showing chromosomal locations of some *EXPANSIN A* genes in non-homoeologous chromosomes, supports earlier studies that polyploidy event in cotton created some unique avenues for response to selection (Wright et al. 1998; Rong et al. 2004). However for *rpb2* gene evolution in *Gossypium* species Pfeil et al. (2004) reported that single gene or segmental duplication was more likely the cause than ancient polyploidy. Rong et al. (2004) observed several duplication events within each subgenome, in addition to homoeologus duplication in cotton. They suggested that this could be due to retrotransposition or present-day cotton may be derived from a putative ancestor containing six or seven chromosomes. Segmental duplications, as a part of polyploidization events, account for 12 out of 21 *EXPANSIN* genes in *Arabidopsis* and 16

out of 44 in rice (Sampedro et al. 2005). Identification of chromosomal locations of SNP markers using deletion analysis further confirmed the true allelic nature of the SNP markers and supported the merit of the strategy for discovery of SNP based on the phylogenetic and comparative analysis of sequences from tetraploid and closely related ancestral diploid species sequences for other candidate genes in allotetraploid cotton.

Fiber length is critical to fiber quality in the global cotton market. Identifying the genetic variants, diagnostic markers and chromosomal locations of candidate genes that are associated with fiber cell elongation is likely to accelerate cotton improvement. Identification of SNPs for six *EXPANSIN A* genes may also help in functional genomics analysis because any changes in the nucleotides of the coding or regulatory regions of the gene may also have functional consequences. Paterson et al. (2003) reported one fiber length-related QTL on chromosome 20 with an LOD threshold of 3.75. They also found another fiber length QTL, significant in water limited treatment, located on chromosome 9. One SNP marker specific to *GhEXPA1* and *GhEXPA2*, respectively, were located on the long arm of chromosome 20 and four different SNP markers derived from *GhEXPA4* on the long arm of chromosome 9 (Table 2.4). A QTL associated with fiber elongation was detected on chromosome 9 (Mei et al. 2004). Another QTL for fiber elongation was mapped on chromosome 3 (Ulloa et al. 2005), the same chromosome where four SNP markers generated from different regions of gene *GhEXPA6* were assigned. Furthermore, QTLs affecting fiber elongation were found on chromosome 1 and 20 where *GhEXPA1*, *GhEXPA2*, and *GhEXPA5* genes were also localized (Chee et al. 2005a). Lacape et al. (2005) reported the association of fiber length QTLs with chromosome 3 and 10, and fiber elongation with chromosome 9, 10 and 20. QTL analysis by newly developed

microsatellite markers suggested that loci on chromosome 3, 9 and 10 affect the fiber elongation, and 2.5% and 50% fiber span length, respectively (Frelichowski et al. 2006). The coincidence of chromosomal locations of six *EXPANSIN A* genes and fiber length and elongation QTLs leaves open the possibility that the *EXPANSIN A* genes affect the important QTLs, especially considering their functional role in fiber cell expansion and elongation (Smart et al. 1998; Arpat et al. 2004). SNP markers derived from candidate genes may be useful for exploring the roles of candidate genes in the complex traits.

In conclusion, a SNP marker discovery strategy in tetraploid cotton was developed, SNPs in six cotton *EXPANSIN A* genes were characterized, and each of them was localized to chromosome. The overlap of the *EXPANSIN A* genes chromosomal locations and previously reported fiber quality QTLs for fiber length and elongation may reflect the putative roles of some *EXPANSIN A* genes in these QTLs. These markers will have potential as candidate gene markers in marker-assisted selection program of fiber traits.

CHAPTER III

R2R3-MYB TANSCRIPTION FACTORS SNP IDENTIFICATION, PHYLOGENOMIC CHARACTERIZATION, CHROMOSOME LOCALIZATION, AND LINKAGE MAPPING IN COTTON

Abstract

R2R3-MYB transcription factors of plants are involved in the regulation of trichome length and density. Several are differentially expressed during initiation and elongation of cotton fibers. Objectives were sequence phylogenomic characterization of six *MYB* genes, their chromosomal localization, and linkage mapping via SNP marker in AD-genome cotton (2n=52). Phylogenetic grouping and comparison to At- and Dt-genome putative ancestral diploid species of allotetraploid cotton facilitated differentiation between genome-specific polymorphisms (GSPs) and marker-suitable locus-specific polymorphisms (LSPs). The SNP frequency was an average one per 77 bases overall, and one per 106 and 30 bases in coding and noncoding regions, respectively. SNP-based multivariate relationships conformed to independent evolution of six *MYB* homoeologs in the four tetraploid species. Nucleotide diversity analysis indicated that six *MYB* loci evolved more quickly in the Dt- than At-genome. The greater variation in the Dt-D genome comparisons than that in At-A genome comparisons showed no significant bias among synonymous substitution, nonsynonymous substitution,

and nucleotide change in noncoding regions. SNPs were concordantly mapped by deletion analysis and linkage mapping, which confirmed their value as candidate gene markers and indicated the reliability of the SNP discovery strategy. These SNPs may be useful for genetic dissection of economically important fiber and yield traits because of the role of these genes in fiber development.

Introduction

Cotton (*Gossypium* spp.) is the world's most important natural textile fiber warranting increased exploration of fiber-related traits through various molecular genetic approaches. Currently, two types of molecular markers are primarily used in molecular mapping of cotton genome. One is genomic markers which primarily target noncoding regions of the cotton genome such as RFLP (Reinisch et al. 1994), RAPD (Kohel et al. 2001), AFLP (Mei et al. 2004), STS (Rong et al. 2004), and SSR (Zhang et al. 2002; Frelichowski et al. 2006). The other is candidate gene markers represented by EST-SSR (Chee et al. 2004; Park et al. 2005; Guo et al. 2007), cDNA probe-based STS or RFLP markers (Rong et al. 2004), and SNP i.e. single nucleotide polymorphism (An et al. 2007). Development of candidate gene markers has received much attention in recent years because of the possible association of functional genes with complex traits. However, the low polymorphism level of cDNA probe-based STS or RFLP markers hampered candidate gene mapping (Rong et al. 2004). SNPs have recently been used as the choice for candidate gene marker in many plant species and are reported to be the most abundant molecular markers (Cho et al. 1999; Ching et al. 2002; Zhang et al. 2003; Zhu et al. 2003). However, SNP study in cotton lags behind the other major crops due to its

allotetraploid nature, high repetitive DNA content, and the lack of genome sequence information.

The candidate gene approach is widely accepted as a strategy for identification of loci influencing complex and economically important traits (Faris et al. 1999; Giroux et al. 2000; Pflieger et al. 2001; Beecher et al. 2002). Candidate gene markers derived from resistance genes or deference response genes were placed on regions containing major resistance QTL in wheat (Faris et al. 1999), pepper (Pflieger et al. 1999), and rice (Wang et al. 2001). The storage protein genes for puroindoline in wheat (Giroux et al. 2000) and hordoindolines in barley (Beecher et al. 2002) were both implicated to play role in grain hardiness and texture by QTL analysis. Markers developed from genes related to carbohydrate and nitrogen metabolism were found to be associated with sugar content and yield in sugar beet (Schneider et al. 2002). Wilson et al. (2004) detected significant association between candidate genes involved in kernel starch biosynthesis and traits for maize kernel composition and starch quality. In cotton, Rong et al. (2007) also found evidence of a general association between concentrations of candidate genes and cotton fiber-related QTL.

R2R3-MYB transcription factors, characterized by two imperfect repeats (R2 and R3) in the DNA-binding domain, are one of the largest regulatory gene families in plants (Riechmann et al. 2000). Some of them were shown to control trichome initiation, expansion, branching, and maturation in *Arabidopsis* (Oppenheimer et al. 1991; Glover et al. 1998; Szymanski et al. 2000; Schiefelbein 2003). Cotton fibers are elongated trichomes derived from the ovule epidermis. Previous reports suggested a similarity in genetic control of MYB transcription factors in *Arabidopsis* trichomes and cotton fibers

(Suo et al. 2003; Wang et al. 2004; Humphries et al. 2005; Perez-Rodriguez et al. 2005; Wu et al. 2006). Expression analysis demonstrated six R2R3-MYB transcription factors were expressed in fiber cells but differentially regulated during fiber initiation and expansion (Loguercio et al. 1999; Cedroni et al. 2003). In addition, several other *MYB* genes had been indicated to play important role in cotton fiber initiation (Suo et al. 2003; Hsu et al. 2005; Lee et al. 2006; Yang et al. 2006).

The objectives of this study were to characterize the sequence phylogenomic characterization of six *MYB* genes in selected tetraploid and diploid cotton species, determine their chromosomal locations, and develop molecular linkage map using candidate gene derived SNP markers. The chromosomal locations and genetic linkage mapping of SNP markers with framework SSR markers will improve the resolution of the cotton comparative map. SNP markers derived from *MYB* genes in this study will be useful as diagnostic markers for exploration of the roles of these candidate genes in complex fiber traits.

Materials and Methods

Plant materials

HS46 and MARCABUCAG8US-1-88 (MAR), two *G. hirsutum* (AD_1) lines with diverse agronomic and fiber properties, and three lines of other tetraploid species including *G. barbadense* L. (AD_2, accession 3-79), *G. tomentosum* Nuttall ex Seemann (AD_3), and *G. mustelinum* Miers ex Watt (AD_4) were used for PCR amplification, cloning, and sequencing of six *MYB* genes. Chromosomal assignment of SNP markers

were accomplished using three different sets of hypoaneuploid F_1 stocks developed from an interspecific cross between TM-1 (genetic standard for *G. hirsutum*, AD_1) and one of the three species, 3-79, *G. tomentosum* or *G. mustelinum*, together with one set of euploid interspecific backcrossed chromosome substitution lines (CS-B, BC_5S_1) of 3-79 in TM-1. Hypoaneuploid F_1 cytogenetic stocks between TM-1 and 3-79 consisted of 10 primary monosomic and 28 monotelodisomic lines; whereas, hypoaneuploid F_1 lines between TM-1 and *G. tomentosum* included 11 primary monosomic and 27 monotelodisomic lines (Liu et al. 2000; Saha et al. 2006b). The new hypoaneuploid F_1 chromosome substitution stocks between TM-1 and *G. mustelinum* (unpublished information) were also used for deletion analysis. Euploid CS-B stocks contain 12 different chromosome and 8 chromosome arm substitutions by 3-79 in TM-1 background (Stelly et al. 2005). Fresh leaves were collected from individual plant, frozen in liquid nitrogen, and then subjected to genomic DNAs extraction by Qiagen DNeasy plant maxi kit (Qiagen Inc., Valencia, CA). A set of 186 recombinant inbred lines (RILs) generated from an interspecific cross between TM-1 and 3-79 were used as a mapping population for constructing molecular linkage map of SNP markers specific to the *MYB* genes and the selected framework SSR markers in cotton (Park et al. 2005; Frelichowski et al. 2006).

PCR amplification, cloning, and sequencing

Gene-specific PCR primers of *MYB1* (COT105 and COT106), *MYB2* (Myb2F and COT108), *MYB3* (Myb3F and COT110), *MYB4* D-genome locus (COT111 and COT112), and *MYB6* (Myb6F and COT116) were adopted from Loguercio et al. (1999). Gene-specific PCR primers of *MYB4* A-genome locus (Myb4A_F and Myb4A_R) and *MYB5*

(Myb5_F and Myb5_R) were designed based on GenBank deposited sequences generated from the previous works by Loguercio et al. (1999) and Cedroni et al. (2003) (Table 3.1). Pfu polymerase (Stratagene, La Jolla, CA) was used for PCR amplification following the protocol described elsewhere (An et al. 2007). The PCR products were separated on a 1% (w/v) agarose gel and purified using QIAEX II gel extraction kit (Qiagen Inc, Valencia, CA). The purified products were ligated into TOPO TA cloning vector and transformed into TOPO10 competent *E. coli* cells (Invitrogen, Carlsbad, CA). Both strands of the recombinant plasmid were sequenced using an ABI 3730XL automated sequencer with ABI Prism BigDye Terminator Cycle Sequencing Kit v3.1 (Applied Biosystems, Foster City, CA). In order to avoid possible complications from PCR recombination (Cronn et al. 2002) and identify the duplicated copies in the genome, the protocols employed by other studies in cotton were followed (Cedroni et al. 2003; Rong et al. 2004; An et al. 2007). It consisted of picking up multiple clones (12 clones) for sequencing of each amplicon and considering one identical sequence from at least three clones.

Table 3.1 SNP Makers Derived from Six *MYB* Genes for Genotyping in Four Allotetraploid Cotton Species

Gene	Genome	PCR primer (5'-3')	SNP primer (5'-3')	Genotyping result for SNP			
				TM-1	3-79	*Gm*[1]	*Gt*[2]
MYB1	A	Myb1A_F: GAGTTTCCTCGTATTATTTTCAGC COT106: ACCCTATGAATCCAAGGGTC	Myb1Gbmt_238_R: CTTGTGCAGCTGTTCTTGTTG Myb1Gt_272_R: CTTGCCCATTGCTCGGAC	C C	T C	T C	T A
	D	Myb1D_F: GGAAACAAGTGAGTTTCCTCCTTG COT106: ACCCTATGAATCCAAGGGTC	Myb1Gb_500_F: CTGTCAAAATTCAAATTCAAAT	G	T	G	G
MYB2	A	Myb2A_F: CTGTTCTGACATCATGGTTTATT Myb2_R: GGCTAATAATGGCTCCAAAGAAGG	Myb2Gb_204_R: TAGTTCAACCATCTCAACC Myb2Gb_334_F: TGATTGCTGGGAGACTTC	T C	A T	T C	T C
	D	Myb2D_F: GTTCTGTTCTGACATCATGGTTTATC Myb2_R: GGCTAATAATGGCTCCAAAGAAGG	N/A[3]	N/A	N/A	N/A	N/A
MYB3	A/D[4]	Myb3_F: GGGCCACTAAAGAATGGAGCA COT110: GCTACAGTTCACTATGTCGG	Myb3Gt_730_F: GAGCTTTTTGTTGTTTGTTTATCTTCT Myb3Gb_742_R: CATGGATTGCATCAAATTCAT	A G	A A, G[5]	A G	G G
MYB4	A	Myb4A_F: TAACTCAAAGCATGCCAGTC Myb4A_R: CACTAAAGCAGAAGGAGCATTAG	Myb4Gbmt_428_F: CTGAGAAAGGGACTGGTAT	C	T	T	T
	D	COT111: CCATTAACTCAAAGCATGCC COT112: CAGGAGGAACAAGGAGGAGC	Myb4Gbmt_73_R: CCTTCATCGACCCTTCGCTA Myb4Gbmt_105_R: TCCAGTATCTTTCTGGGTACTGC	G A	T G	T G	T G
MYB5	A/D[4]	Myb5_F: GACCTTTTGAAACTGTGAAGGT Myb5_R: CTGCTGTAAAAACCAAACATTC	Myb5Gbmt_511_R: ACCAACCAACAAACCATTGC	C, T[6]	T	T	T
MYB6	A	Myb6A_F: CGTTCCCTCCCCAAAGCT COT116: ACAATGGCGTGCATGTTGCC	Myb6Gm_341_F: CATTTCCAAAACATAACGTTTAC	G	G	A	G
	D	Myb6A_R: CGTTCCCTCCCCAAAGCC COT116: ACAATGGCGTGCATGTTGCC	Myb6Gb_856_F: GTAACAAATGATGAAGAAGAAC	A	G	A	A

[1]Abbreviation of *G. mustelinum*.
[2]Abbreviation of *G. tomentosum*.
[3]No suitable SNP marker could be designed in D-genome sequence of gene *MYB2*
[4]Not applicable for genome-specific PCR primers design.
[5]3-79 gave two alleles for this SNP marker, one from each subgenome. CS-B lines were used for allele 'A' (in Dt-genome) chromosomal assignment.
[6]TM-1 gave two alleles for this SNP marker, one from each subgenome. Allele 'C' was chromosomally assigned by deletion analysis with hypoaneuploid F_1 cytogenetic stocks.

SNP characterization and phylogenetic analysis

Six *MYB* gene sequences from the five allotetraploid cotton lines together with GenBank deposited sequences from TM-1 and living models of two allotetraploid ancestral genomes: *G. herbaceum* L. (A-genome; accession A1-73) and *G. raimondii* Ulbrich (D-genome; 'Galau's') were used for SNP characterization. The GenBank sequences of *Gossypoides kirkii* (Masters) J.B. Hutchinson were used as outgroup to cotton genus (Malvaceae) in phylogenetic analyses (Cedroni et al. 2003). DNASTAR (DNASTAR Inc., Madison, Wisconsin) and Clustalx (Thompson et al. 1997) were used for vector-trimming and sequence alignment. Before SNP characterization, differentiation between paralogous and homoeologous loci was performed by phylogenetic grouping and comparison of sequences from the two diploids (An et al. 2007). Phylogenetic analyses were performed by maximum parsimony (MP) method using MEGA 3.1 (Kumar et al. 2004). To determine the confidence levels for each tree, an MP bootstrap analysis with 100 replicates was conducted. DnaSP 4.0 software was used to identify SNP by comparative analysis of aligned sequences from different genotypes at a putative locus (Rozas et al. 2003). Nucleotide diversities (π), haplotype number (H) and diversity (H_d), rate of silent (K_{sil}) and nonsynonymous (K_a) substitutions of pairwise comparisons were also calculated by DnaSP 4.0 software (Tajima 1983; Nei 1987; Rozas et al. 2003).

Chromosomal assignment and linkage mapping

In order to minimize the potential problems associated with homoeologous sequences in SNP genotyping, genome-specific (or locus-specific) PCR primers were designed according to sequences differences between two subgenomes in tetraploid

cotton if applicable (Table 3.1). Interspecies SNP primers were designed based on a single nucleotide difference among sequences at a putative locus (each clade or group in the phylogram of individual *MYB* gene) between the genotypes of TM-1 and 3-79, *G. tomentosum* or *G. mustelinum*. The primer was selected, based on the sequence information, to anneal just upstream or downstream of the SNP site as the forward or reverse primer, respectively, so that the polymorphism could be detected by one base extension technology with an ABI Prism SNaPshot^{TM} multiplex kit (Applied Biosystems, Foster City, CA). All primers used for genotyping are summarized in Table 3.1. The deletion analysis method frequently used for molecular marker chromosomal assignment in cotton (Liu et al. 2000; An et al. 2007) was employed to assign chromosomal locations for six *MYB* genes using the four sets of cytogenetic stocks mentioned in the plant materials. Ninety SSR markers which are polymorphic between TM-1 and 3-79 and span the cotton genome were selected based on the information available in cotton microsatellite database (CMD, http://www.cottonmarker.org/; Blenda et al. 2006), and used as anchored markers for linkage mapping of sections of selected chromosomes with SNP markers. Chromosomal assignment of the constructed linkage groups was achieved by deletion analysis, comparison to the allele size with CMD panel, published integrated molecular maps (Lacape et al. 2005; Park et al. 2005; Frelichowski et al. 2006; Guo et al. 2007), and the assignment of cotton linkage maps to chromosomes (Wang et al. 2006b). The SSR markers used in this study were fluorescent labeled by Sigma Genosys (The Woodlands, TX) or Applied Biosystems (Foster City, CA). PCR reaction and thermal cycle protocol for genotyping the RILs population were conducted according to the method of Gutierrez et al. (2002). One polymorphic SNP marker between TM-1 and 3-79

were selected, if available, from each gene for linkage mapping. The procedures of SNP marker genotyping described in An et al. (2007) were employed for cytogenetic stocks and RILs population genotyping. An automated capillary electrophoresis system ABI3100 Genetic Analyzer with GeneMapper software 4.0 (Applied Biosystems, Foster City, CA) was used to analyze both PCR-amplified DNA fragments of SSR markers and the single nucleotide extension of SNP markers. The genotyping output data of both SNP and SSR markers were coded for linkage analysis using JoinMap® 4 (Van Ooijen 2006). Chi-square test was used to check if the marker segregation was compatible with the 1:1 ratio. Recombination frequencies were converted into map distances (centiMorgan, cM) using the Kosambi mapping function (Kosambi 1944) and linkage groups were determined at LOD scores ≥6.

Results

SNP characterization and haplotype analysis of six MYB genes

In vitro SNP discovery through amplicon cloning and sequencing has been accomplished by homoeologous differentiation and gene specific fragment amplification in cotton (Figirue A.14-19; Table 3.2). In this study, no duplicated or heterogeneous loci were found within each subgenome. SNPs including indels were detected from 8,301 bp of aligned sequences (7,084 bp and 1,217 bp of coding and noncoding regions, respectively). Nuclotide sequences generated from six *MYB* genes were deposited to the GenBank database under the accession numbers EU249397 to EU249456. From the eight cotton genotypes, 108 SNPs were detected from both A- and D-genomes (Table 3.2),

giving an average SNP frequency of one SNP every 77 bases. Results showed the presence of one SNP per 106 bp in the coding regions and one SNP per 30 bp in the noncoding regions (Table 3.2). The SNP distribution varied among the six gene examined. The highest rate of SNP occurrence was observed in *MYB6* (one SNP every 34 bp) and the lowest rate of SNP frequency was present in *MYB3* (one SNP every 260 bp). Transitions ('A/G' or 'C/T') were the most common cause of sequence variation in the selected cotton genotypes (49%) compared to transversions ('A/T', 'G/C', 'A/C' or 'G/T', 26%) and indels (25%). In *MYB6*, two nucleotide ('C' and 'T') substitutions were observed in three indel positions (A-genome sites 101 and 111, D-genome site 99). A significant bias to 'T' insertion/deletion was detected in the overall sequences (59.30%). In coding regions of six *MYB* genes, 41 out of 67 cSNPs (SNPs in coding region) sites were predicted to result in amino acid changes (Table 3.2). Sequence polymorphism defined haplotypes number ranged from two to seven among the seven selected cotton genotypes and haplotype diversity varied from 0.29±0.20 to 1.00±0.08 among six *MYB* genes (Table 3.2 and Table A.23-A.34).

Table 3.2 SNP Characterization of Six *MYB* Genes in Selected Cotton Genotypes

Gene	Genome	Haplotypes[1]		Transitions		Transversions				Indels				Coding region		Total	
		H	H_d	A/G	T/C	A/T	G/C	A/C	G/T	A	C	G	T	Length (bp)	SNP No. (nonsynonymous changes)	Length (bp)	SNP No.
MYB1	A	5	0.86±0.14	2	2	1	0	0	1	0	0	0	0	492	6(4)	492	6
	D	4	0.71±0.18	2	2	0	0	0	0	0	0	0	0	492	4(3)	492	4
MYB2	A	4	0.71±0.18	1	2	1	1	1	1	0	0	0	2	473	5(3)	551	9
	D	4	0.71±0.18	1	3	1	0	0	0	0	0	0	1	473	5(4)	550	6
MYB3	A	2	0.29±0.20	0	0	0	1	0	0	0	0	0	0	537	1(0)	778	1
	D	4	0.71±0.18	3	1	0	1	0	0	0	0	0	0	537	2(1)	779	5
MYB4	A	5	0.86±0.14	1	3	2	2	2	2	0	0	0	0	842	12(8)	853	12
	D	5	0.86±0.14	2	3	2	0	1	0	0	0	0	0	842	8(5)	853	8
MYB5	A	3	0.71±0.13	1	0	1	0	0	0	0	0	0	0	571	2(1)	591	2
	D	2	0.29±0.20	1	2	0	0	0	0	0	0	0	0	571	3(0)	591	3
MYB6	A	5	0.86±0.14	2	6	1	2	0	0	0	4[2]	1	5	627	7(4)	882	21
	D	7	1.00±0.08	6	7	2	0	2	0	1	4[3]	1	8	627	12(8)	889	31

[1]Sites with alignment gaps were not considered for haplotypte analysis; H means haplotype number; H_d means haplotype diversity.
[2]At positions 101 and 111 of the aligned sequences, two kinds of nucleotide substitution (‘C’ and ‘T’) occur at the indel positions. Here, it was considered as ‘C’ nucleotide indel in data analysis.
[3]At position 99 of the aligned sequences, two kinds of nucleotide substitution (‘C’ and ‘T’) occurs at the indel position. Here, it was considered as ‘C’ nucleotide indel in data analysis.

Phylogenomic sequence characterization

SNP-based multivariate relationships suggested independent evolution of six *MYB* homoeologs in the four tetraploid species. Parsimony analyses revealed that sequences (Figure A.14-A.19) fell into two clades, each containing one of the two homoeologs from the allotetraploid cotton lines and the corresponding copy from the progenitor diploid genomes. Pairwise comparisons of the nucleotide diversity (π) of six *MYB* genes in both A- and D-genomes are summarized in Table 3.3. The π value measures the average number of nucleotide differences per site between two sequences (Nei 1987). The lowest nucleotide diversities occurred among the three *G. hirsutum* lines in both A- and D-genomes. Results from both A- and D-genomes showed the highest nucleotide diversities were between *G. mustelinum* and the extant species of the ancestral genome donors. Nucleotide diversities of *MYB* genes were higher in the D-Dt comparisons than for the A-At comparison of the allotetraploid cotton species, indicating that *G. herbaceum* may be a closer ancestor of the At-genome donor than *G. raimondii* is of the Dt-genome donor.

Table 3.3 Pairwise Comparison Matrix of the Eight Cotton Lines Showing the Nucleotide Diversity (π value, $\times 10^{-2}$) in A- and D-genomes[1]

Genotype[2]	A_1	TM-1	HS46	MAR	3-79	*G. mustelinum*	*G.tomentosum*
D_5	-	0.53	0.58	0.56	0.52	0.68	0.57
TM-1	0.37	-	0.11	0.09	0.15	0.29	0.18
HS46	0.37	0	-	0.05	0.20	0.32	0.21
MAR	0.41	0.04	0.04	-	0.18	0.29	0.19
3-79	0.42	0.21	0.21	0.25	-	0.30	0.19
G. mustelinum	0.56	0.33	0.33	0.29	0.38	-	0.27
G. tomentosum	0.50	0.29	0.27	0.33	0.34	0.46	-

[1]The numbers above and below the diagonal line represent sequence comparisons within D- and A-genomes, respectively.
[2]Taxa *G. herbaceum* and *G. raimondii* are designated by their genome designations, A_1 and D_5, respectively. TM-1 is genetic standard of *G. hirsutum*. 3-79 is a a double haploid line of G. *barbadense*. HS46 and MAR are another two diverse elite *G. hirsutum* lines.

To further explore the nature of substitutions contributing to overall divergence in cotton, pairwise comparisons among orthologous copies for six *MYB* genes of both A- and D-genomes are tabulated separately for nonsynonymous substitution (K_a), silent substitution (K_{sil}), and the K_a:K_{sil} ratio (Table 3.4). Results indicated that K_a and K_{sil} values in the D-Dt comparisons were higher than the corresponding values in the A-At comparisons except for the comparison between MAR (*G. hirsutum*) and its two genome living models of K_{sil} value. This higher variation in the Dt-genomes (when compared to a progenitor genome) came not only from greater amino acid substitutions, but also from nucleotide changes in noncoding regions and synonymous changes in the coding regions. Although these predictions were based on the genomic sequence, they may allow speculation of evolutionary constraints placed on amino acid substitutions without knowing the exact effect of the SNPs on predicted codons. Nucleotide diversities among the three *G. hirsutum* lines in the Dt-genome were higher than that in the At-genome, indicating six *MYB* genes loci in Upland cotton Dt-genome exhibited a faster evolutionary rate than the At-genome (Table 3.3). Most of the substitution ratios (K_a:K_{sil}) of pairwise comparisons were less than 1, indicating a high level of evolutionary constraint placed on amino acid substitution in the six *MYB* genes (Table 3.4).

Table 3.4 Pairwise Comparison Matrix of Molecular Evolutionary Rates for Six *MYB* Genes in Cotton [1]

Genotype[2]	A_1	TM-1	HS46	MAR	3-79	*G. m*	*G. t*
D_5	-	0.005[3] 0.007 0.696	0.005 0.009 0.541	0.005 0.007 0.710	0.005 0.008 0.601	0.005 0.011 0.416	0.004 0.009 0.428
TM-1	0.003 0.007 0.404	-	0.001 0.002 0.705	0.002 0 -	0.002 0.001 4.069	0.002 0.004 0.519	0.001 0.002 0.637
HS46	0.003 0.007 0.404	0 0 -	-	0.001 0.002 0.221	0.002 0.002 0.831	0.002 0.005 0.312	0.001 0.004 0.309
MAR	0.003 0.008 0.416	0 0.001 0	0 0.001 0	-	0.002 0.001 4.241	0.002 0.004 0.546	0.002 0.002 0.682
3-79	0.004 0.005 0.765	0.002 0.003 0.584	0.002 0.003 0.584	0.002 0.004 0.465	-	0.002 0.004 0.365	0.001 0.003 0.380
G. m	0.004 0.008 0.448	0.002 0.007 0.276	0.002 0.007 0.276	0.002 0.006 0.318	0.003 0.005 0.601	-	0.001 0.005 0.187
G. t	0.004 0.008 0.499	0.002 0.006 0.245	0.002 0.006 0.245	0.002 0.007 0.214	0.003 0.005 0.602	0.003 0.008 0.318	-

[1]The numbers above and below the diagonal line represent D- and A-genome sequences comparisons, respectively.
[2]Taxa *G. herbaceum* and *G. raimondii* are designated by their genome designation, A_1 and D_5, respectively. TM-1 is genetic standard of *G. hirsutum*. 3-79 is a double haploid line of G. *barbadense*. HS46 and MAR are another two diverse *G. hirsutum* lines. *G. m* and *G. t* are abbreviation of *G. mustelinum* and *G. tomentosum*, respecitively.
[3]Three parameters were used to characterize each comparison: nonsynonymous substitution per nonsynonymous sites in coding sequences (K_a, top), substitutions per site including intron and synonymous sites (K_{sil}, middle), and K_a:K_{sil} ratio (bottom).

Chromosome localization of six MYB genes

Hypoaneuploid stocks, developed from three interspecific crosses between TM-1 (*G. hirsutum*) and 3-79 (*G. barbadense*), *G. tomentosum* or *G. mustelinum*, and one set of euploid interspecific backcrossed chromosome substitution lines (CS-B, BC_5S_1) of 3-79 in TM-1 were used for chromosomal assignment of SNP markers by deletion analysis (An et al. 2007). Thirteen different SNP sites between the common parent TM-1 and 3-79, *G. mustelinum* or *G. tomentosum*, respectively, were selected for SNP primer design in six *MYB* genes (Table 3.1). Chromosomal locations were confirmed using deletion lines from different sources. Due to the conserved character of the homoeologous sequences in gene *MYB3* and *MYB5*, no suitable genome-specific PCR primers could be designed. However, chromosomal assignment of genome-specific alleles was still possible by euploid CS-B or hypoaneuploid F_1 stocks (Table 3.1). Moreover, no SNP marker could be designed from the Dt-genome of gene *MYB2*, therefore only At-genome location was considered for chromosomal assignment by either deletion analysis or linkage mapping. SNP markers used for chromosomal assignment and the according genotyping results are listed in Table 3.1. Deletion analyses of the six genes were performed using all the available cytogenetic stocks and the results are summarized in Table 3.5. Chromosomal locations of the gene *MYB4* were detected on the long arm of two homoeologous chromosomes: 7 and 16. Only one subgenomic location of genes *MYB1*, *MYB2*, *MYB5*, and *MYB6* was found by deletion analysis using SNP markers, which were on the long arm of chromosomes 18, short arm of chromosome 8, short arm of chromosome 11, and short arm of chromosome 11, respectively. Complete coverage for all the chromosomes are not aviabile in the cytogenetic stocks. The putative chromosome location of gene

MYB3 in Dt-genome could not be determined due to lack of complete coverage of Dt-genome; however, it is probably on one of the chromosomes for which there is no aneuploid stock (long arm of chromosome 14, 15, or chromosome 19, 21, 23, and 24).

Table 3.5 Chromosomal Locations (*Lo*, long arm; *sh*, short arm) of Six *MYB* Genes with Previously Reported QTL

Gene	Subgenome	Chromosomal location		Previously reported QTL[1]
		Deletion analysis	Linkage mapping	
MYB1	A	-	13	EL[2, 13, 14]; FF[14]; FL[3, 11]; FS[11, 14];
	D	18*Lo*	18	FS[5, 13]; EL[2, 5]; FF[4, 11]; FL[10, 21]; 2.5% SL[5]; 50% SL[5]; FU[21]; LP[21]
MYB2	A	8*sh*	8	EL[9, 14, 17]; FL[3, 8, 14]; FU[21]; LI[8]; LY[7, 8]; SCY[9]; MIC[12]; FS[8, 12, 14, 17, 21]
	D	-	-	-
MYB3	A	-	-	-
	D	Possibly 14*Lo*/15*Lo*/19/21/23/24	-	EL[2, 5, 9, 11, 13, 14, 17, 20]; FF[4, 11, 14]; FL[3, 8, 9, 11, 12, 14, 17, 18, 19, 20]; LP[19]; FU[11, 21]; LY[7, 8, 18, 19]; MIC[8, 12, 17, 18, 19, 20]; SCY[8, 16, 18, 19]; FS[5, 8, 9,10, 11, 13, 12, 14, 17, 19, 20, 21];
MYB4	A	7*Lo*	-	EL[17]; FL[3, 17]; FU[16]; SCY[7, 8]
	D	16*Lo*	16	EL[15]; FF[10, 11]; FL[15, 17]; LI[15]; LP[15]; MIC[17]; FS[11, 17, 18]
MYB5	A	11*sh*	-	EL[2, 18, 19]; FF[4]; FL[3, 14]; FU[14]; LP[6]; LY[7, 8]; MIC[12, 17]; FS[14, 17]
	D	-	-	-
MYB6	A	11*sh*	-	EL[2, 18, 19]; FF[4]; FL[3, 14]; FU[14]; LP[6]; LY[7, 8]; MIC[12, 17]; FS[14, 17]
	D	-	21	EL[2]; FF[4, 14]; FL[3, 17, 18]; FS[9, 14, 16]

[1]EL-elongation; FF-fineness; FL-length; FS-strength; FU-uniformity; LI-lint index; LP-lint percentage; LY-lint yield; MIC-micronaire; SCY-seed cotton yield; 2.5% SL-length at 2.5%; 50% SL-length at 50% [2]Chee et al. (2005a) [3]Chee et al. (2005b) [4]Draye et al. (2005) [5]Frelichowski et al. (2006) [6]Guo et al. (2006) [7]He et al. (2005) [8]He et al. (2007) [9]Jiang et al. (1998) [10]Kohel et al. (2001) [11]Lacape et al. (2005) [12]Lin et al. (2005) [13]Park et al. (2005) [14]Paterson et al. (2003) [15]Ren et al. (2002) [16]Saranga et al. (2001) [17]Shen et al. (2005) [18]Shen et al. (2006b) [19]Shen et al. (2007) [20]Wang et al. (2006a) [21]Zhang et al. (2005).

Linkage mapping of MYB genes by SNP markers

Framework SSR markers were utilized to construct linkage maps with SNP markers. One hundred eighty-six RILs, from the cross of TM-1 and 3-79, were used for genotyping by 90 SSR markers and five polymorphic SNP markers specific to gene *MYB1*, *MYB2*, *MYB4*, and *MYB6*. Genetic linkage mapping results confirmed the deletion analysis for the chromosomal locations of *MYB1*, *MYB2*, and *MYB4*. Linkage mapping also revealed chromosomal locations of two genes' homoeologous loci (At-genome of gene *MYB1* and Dt-genome of gene *MYB6*), which were on chromosome 13 and 21, respectively (Table 3.5). Moreover, it also showed the linkage relationship between 15 SSR marker and five SNP markers (Figure 3.1). Three SNP markers showed distorted segregation in the mapping population. The segregation of SNP markers Myb1Gbmt_238_R and Myb4Gbmt_105_R was skewed toward TM-1 and the segregation of SNP marker Myb2Gb_204_R was skewed toward 3-79.

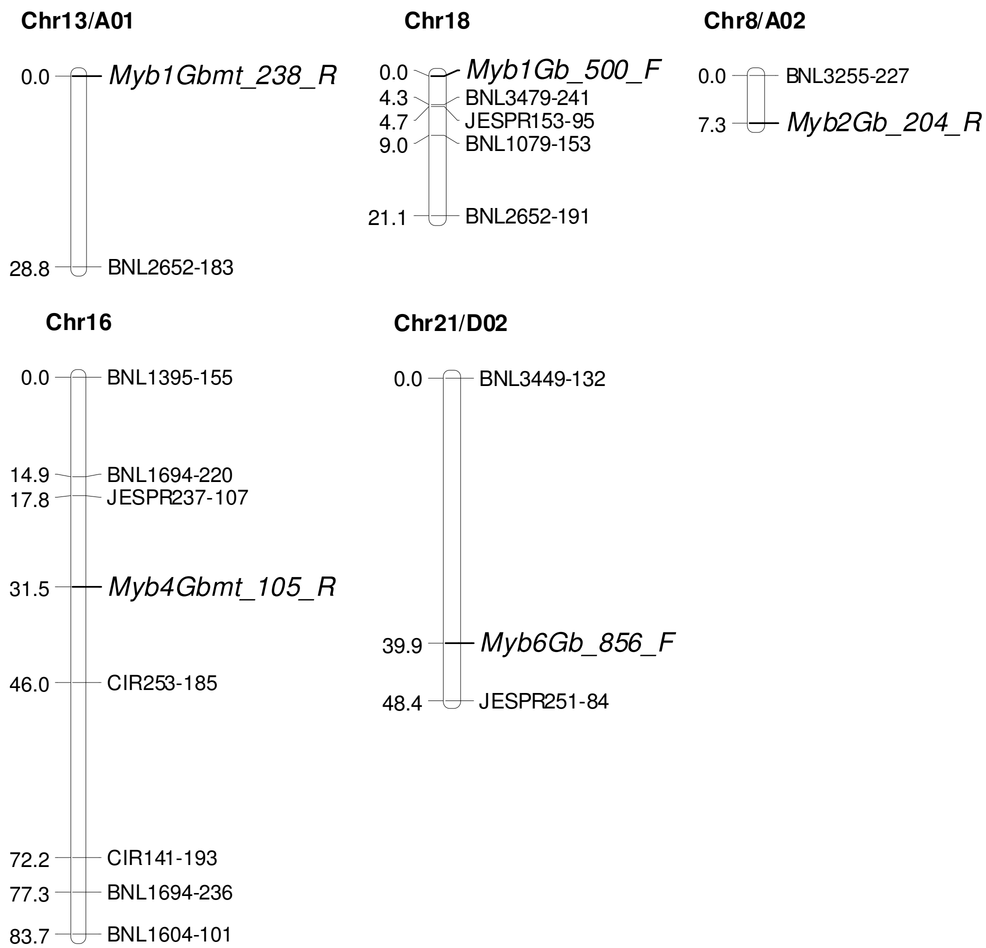

Figure 3.1 Linkage Maps of Selected SNP Markers Derived from *MYB* Genes with SSR Markers.

Chromosomal assignments of linkage group A01, A02, and D02 to chromosome 13, 8, and 21, respectively, were based on Wang et al. (2006).

Discussion

SNP in cotton

Efficient SNP discovery in polyploids, such as cotton, must address the problem of appropriate methods that can distinguish between genome-specific polymorphisms (GSPs) and locus-specific polymorphisms (LSPs). In this study, the possibility of identifying false SNP were reduced by applying the following approaches: 1) PCR primers were designed from well characterized genes to generate an amplicon pool from each genotype; 2) multiple clones were sequenced to avoid the random error of sequencing and to ensure getting the duplicated loci of the gene; 3) putative locus was identified by phylogenetic clustering and comparing to the two progenitor diploid genome species of allotetraploid cottons; and 4) locus-specific PCR and SNP primer were designed for SNP marker genotyping to confirm the reliability of the procedures (An et al. 2007). Thus, a total of 108 putative SNPs were identified among selected genotypes at the same locus. The average frequency of SNP was one SNP per 77 bp (1.30%), with one SNP per 106 bp (0.94%) and one SNP per 30 bp (3.33%) in coding and noncoding regions, respectively. In *Arabidopsis thaliana*, the rate of variation per nucleotide were detected as 1.09% and 0.27% in *GL1* gene (a member of the *MYB* gene family) of 26 accessions (Hauser et al. 2001) and *Atmyb2* gene of 20 ecotypes (Kamiya et al. 2002), respectively. In cotton, the average rates of SNP per nucleotide were observed as 2.35% in six *EXPANSIN A* genes (An et al. 2007). Another pilot SNP study revealed the rate of variation per nucleotide was 0.35% between *G. hirsutum* and *G. barbadense* (one SNP

every 286 bp), and the variation per nucleotide were 0.14% and 0.37% within these two species, respectively (Rong et al. 2004).

In other crops, Ching et al. (2002) reported the presence of one SNP per 31 bp in noncoding regions and one per 124 bp in coding regions when analyzing 18 maize genes in 36 inbred lines. One SNP in every 273 bp was present in soybean (Zhu et al. 2003). Genome-wide sequence alignment between rice subspecies *Indica* and *Japonica* revealed a polymorphism rate of 1.70 SNP/kb and 0.11 indel/kb (Feltus et al. 2004). In wheat, SNP frequency was one SNP per 540 bp (Somers et al. 2003). The incidence of SNP in barley was reported as one SNP per 27 bases in the intronless *Isa* gene (Bundock et al. 2003), and approximately one SNP per 131 bases in the exonic region of the *P450* gene family members (Bundock and Henry 2004). Although varying frequencies of SNP per length of DNA sequence have been reported, it is highly dependent upon what kind of sequence data and genotypes are used to generate SNP in each species. As expected a higher number of SNPs were observed at the interspecific level compared to the intraspecific level of six cotton *MYB* genes in this study.

MYB gene phylogenomic features

The cotton genus contains about 50 species with a basic chromosome number of 13. The five tetraploid cotton species (AADD, 2n=4x=52) are a monophyletic assemblage putatively derived from a single allopolyploidization event that occurred 1.5 million years ago (MYA) after divergence of the diploid progenitors about 6.7 MYA (Senchina et al. 2003). The two diploid species that gave rise to the allotetraplods were from the A- and D-genome groups which are best represented by the extant species *G.*

herbaceum L. and *G. raimondii* Ulbr., respectively (Wendel and Cronn 2003). Results showed that the tetraploid *MYB* genes could be broadly separated into two origins representing the putative A- and D-genomes based on their similarity with the sequences of the diploid ancestral species (Figure A.14-A.20). SNP-based multivariate relationships conformed to independent evolution of six *MYB* homoeologs in the four tetraploid species (Cronn et al. 1999; Cedroni et al. 2003). The nucleotide diversity observed was higher in the Dt-genome compared to the At-genome of the three *G. hirsutum* species. Previous studies with *Adh* (Small et al. 1998, 1999; Small and Wendel 2002) and *FAD2-1* (Liu et al. 2001) showed a faster evolutionary rate in the Dt-genome than in the At-genome of cotton. Reinisch et al. (1994) reported that the RFLP marker polymorphism levels of the Dt-genome were 10% higher than the At-genome. The Dt-genome, from an ancestor that does not produce spinnable fiber, contributes substantially to fiber quality of tetraploid cottons (Jiang et al. 1998; Saranga et al. 2001; Paterson et al. 2003; Lacape et al. 2005; Rong et al. 2007). Many QTL that positively affect fiber quality have been detected on the Dt-genome (Table 3.5). In addition, many EST loci associated with fiber development have also been mapped to the Dt-genome (Park et al. 2005). However, some QTL influencing fiber quality and yield have been identified in the At-genome as well (Mei et al. 2004; Frelichowski et al. 2006). Whether the spreading of the At-genome repetitive DNA elements to the Dt-genome (Zhao et al. 1998) or different evolutionary pressures operating on the two genomes (Small and Wendel 2002) caused the different evolutionary dynamics is still obscure. But, all these facts collectively indicated the importance of further investigations of the Dt-genome for fiber improvement in the tetraploid cottons.

Chromosomal locations of MYB genes

The chromosomal locations of six *MYB* genes were identified via deletion analysis or linkage mapping (Table 3.5, Figure 3.1). The low level of polymorphism of other kinds of molecular markers derived from functional genes such as EST-SSR (Park et al. 2005; Guo et al. 2007) or cDNA probe-based STS or RFLP (Rong et al. 2004) among mapping parents has hindered their use in candidate gene mapping. Results presented here show the great potential for using SNP markers to tag functional genes and improve the comparative maps in cotton.

Previous studies have led to the discoveries of important QTL on different chromosomes in cotton. A comprehensive summary of the previously reported cotton fiber quality and yield component traits related QTL on the same chromosomes as six *MYB* genes are summarized in Table 3.5. Analyses on the effects of chromosome-specific introgression in Upland cotton indicated that substitutions for chromosomes 16 and 18 from 3-79 had additive effects related to reduced yield (Saha et al. 2006a). These chromosomes are the locations of genes *MYB1* and *MYB4*. Further studies using topcrosses of 13 CS-B lines with five commercial cultivars showed that chromosomes 7 and 18 (locations of gene *MYB4* and *MYB1*, respectively) had additive effects for fiber strength (Jenkins et al. 2007). Given the role of MYB transcription factors in fiber cell initiation and expansion, the agreement of the chromosomal locations between *MYB* genes and previously reported fiber yield and quality QTL suggested these SNP markers may be useful in studying the association between important fiber development genes and economically important QTL in cotton.

CHAPTER IV

DETECTING QTL FOR YIELD COMPONENTS, SEED, AND FIBER TRAITS OF UPLAND COTTON UNDER MULTIPLE FUZZLESS LOCI GENETIC BACKGROUNDS

Abstract

Cotton (*Gossypium* spp.) is the world's leading fiber crop and an important source of protein and oil. The purpose of this research was to detect QTL or molecular markers associated with yield components, fiber, and seed traits under multiple fuzzless loci genetic backgrounds. Two F_2 populations developed from crossing MD17, a fuzzless-lintless genetic stock containing three fuzzless loci, N_1, n_2 and a postulated n_3, with line 181 (fuzzless-linted) and with FM966, a fuzzy-linted cultivar, were used for molecular mapping by polymorphic SSR markers. Major QTL which explain 68.3 (population with FM966) to 87.1% (population with 181) of the phenotypic variation for lint percentage and 62.8% (population with 181) for lint index were detected in the vicinity of BNL3482-138 on chromosome 26. Single marker regression analyses indicated STV79-108, which was located to the long arm of chromosome 12 (the known location of N_1 and perhaps n_2 loci), also had significant association with lint percentage (R^2 % 26.7), lint index (R^2 % 30.6), embryo protein percentage (R^2 % 15.4) and micronaire R^2 % 20.0). Additional QTL and significant markers associated with other seed and fiber traits were detected on

different chromosome locations and explained large percentages of the phenotypic variation. Two-locus epistatic interactions were also observed. Results from this research will facilitate further understanding the complex network of cotton fiber development and seeds traits.

Introduction

Cotton (*Gossypium* spp.) is the most important natural textile fiber and the second-important oilseed source in the world. Lint fiber of cotton is the major commercial product for textile industries. Cottonseed protein cake and oil are secondary products and have been drawing a large interest as animal feed (Arieli 1998), human consumption (O'Brien and Wakelyn 2005), and biodiesel feedstock (Royon et al. 2007). With the advent of molecular marker technology, it became possible to construct high density molecular genetic maps (Rong et al. 2004; Guo et al. 2007) and to locate QTL for lint yield (Shen et al. 2006b; He et al. 2007), fiber quality (Paterson et al. 2003; Lacape et al. 2005; Park et al. 2005), seed traits (Song and Zhang 2007), and response to biotic or abiotic stress (Wright et al. 1998; Saranga et al. 2004; Shen et al. 2006a), to linkage groups or chromosomes in cotton. However, limited information is known about the direct association or interaction among specific genes and the variations of traits of interest such as cotton fiber (Ruan et al. 2003; Arpat et al. 2004; Lee et al. 2007).

Genetic mapping of mutants with discrete phenotypes has been regarded as a primary step toward their isolation, and could provide clues to their organization and function (Jander et al. 2002). Identification of discrete mutations associated with QTL also represents a rapid and efficient way to dissect quantitative traits (Paterson 1995).

Mutant mapping offers the potential to explore the nonlinear interactions among mutant genes and QTL in the complex development network (Rong et al. 2007). Robertson (1985) suggested that qualitative mutant alleles and wild type alleles at loci affecting quantitative traits were the extremes of a possible range of effects and QTL resulting from the segregation of naturally available wild type alleles represented the milder effects. QTL mapping using populations segregating for mutants might also have the potential to increase the sensitivity due to the enlarged quantitative variation scale by the expression of qualitative mutant genes. In cotton, several studies were conducted to detect QTL related to yield components or fiber traits using mapping populations derived from different fiber mutants (Rong et al. 2005, 2007; Abdurakhmonov et al. 2007) or a multiple dominant marker line T586 (including a fuzzless locus N_1) (Zhang et al. 2005; Guo et al. 2006; Wan et al. 2007).

Cotton fibers are single-celled trichomes developed from the ovule epidermis. There are two distinct types of fiber. The spinnable long fibers are called lint (25-35 mm) and initiate around fertilization and elongate rapidly afterwards; whereas the short fibers are called fuzz (~5 mm) which develop at a later stage (Stewart 1975). Several qualitative mutants in fiber development were identified and their genetics for controlling lint length and fuzz fiber production have been studied (Kohel 1972; Narbuth and Kohel 1990; Zhang and Pan 1991; Du et al. 2001; Karaca et al. 2002; Turley and Kloth 2002; Rong et al. 2005). Among these, two loci (N_1 and n_2) were reported to inhibit fuzz fiber development (Kearney and Harrison 1927; Ware et al. 1947) and they also show considerable negative effect on lint production (Ware 1940; Ware et al. 1947; Rong et al. 2005).

The objective of this study was to detect QTL or molecular markers associated with yield components, fiber, and seed traits under multiple fuzzless loci genetic backgrounds.

Materials and Methods

Mapping populations and phenotypic data collection

Two fiber mutant lines, MD17 (PI 616493), 181, and one commercial cultivar FiberMax 966 (FM966) of Upland cotton (*Gossypium hirsutum* L.) were used as parents for this study. MD17 is a fuzzless-lintless germplasm line that combined two fuzzless loci N_1, n_2, and a postulated n_3 locus and the proposed genotype is $N_1N_1n_2n_2n_3n_3$ (Turley 2002; Turley and Kloth 2002). Fiber mutant 181 is a fuzzless-linted line from China. FM966 is an Upland commercial cultivar with normal phenotype (fuzzy-linted) from Bayer CropScience. The F_2 populations were developed by crossing MD17 as common female parent with FM966 and 181, respectively. Population with FM966 (MD17×FM966 F_2) consisted of 100 plants and population with 181 (MD17×181 F_2) consisted of 164 plants. Three parental lines along with their resulting F_1 and F_2 progenies were grown at the Plant Science Research Center of Mississippi State University in 2006. Standard cultural, insect and weed control practices were followed. To avoid possible shattering and loss of lintless seeds, individual plants were labeled and hand harvested every other day. All the open bolls on each individual F_2 plant were harvested and the multiple harvestings were combined for each plant at the end of the season. Randomly selected samples of parental and F_1 plants were also hand harvested for all measurements.

Seed cotton samples were weighed and ginned on a laboratory 10-saw gin to determine lint percentage (LP), lint index (LI g fiber 100^{-1}), and seed weight (SW g wt of 100 acid delinted seed). Lint samples of population with FM966 were sent to STARLAB, Inc. in Knoxville, TN, for determination of micronaire (MIC), elongation (E1), fiber strength (T1), 50% span length (SL50), and 2.5% span length (SL2.5) by single instruments. All measurements were made on individual plants in the F_2 generations. Fiber parameters in population with 181 were not measured because many plants did not produce lint fibers. Acid delinted seed samples of population FM966 were sent to the Mississippi State Chemical Laboratory for seed crude oil and whole protein percentage measurement by petroleum ether extraction gravimetric method and Leco Nitrogen Combustion analyzer, respectively. Using 5 to 10 grams acid delinted seed samples, each seed was cut by scissors and the embryo was removed by needle to measure the seed hull percentage (HP). Embryo crude oil or protein percentage (EOP or EPP) was determined by dividing delinted seed crude oil or protein percentage by embryo percentage (i.e. 100%-hull percentage). F_2 mean values represent an average of individual plants.

DNA isolation and molecular marker analysis

Leaf tissues were collected from individual F_2 plant and bulks of parents and F_1 plants in the field. Genomic DNA was isolated from frozen dried leaf samples using DNeasy Mini Kit (Qiagen Inc, Valencia, CA, USA) according to the manufacture protocol. A total of 1407 fluorescent-labeled SSR primer pairs and 31 SNP primers generated from previously two chapters studies (An et al. 2007, 2008; Hsu et al. 2008) were used to screen the parents. These SSR primers included 379 BNL, 205 CIR, 47 CM,

310 JESPR, 84 MGHES, 17 MUCS, 8 MUSB, 23 MUSS, 48 NAU, 94 STV, and 192 TMB series (Blenda et al. 2006; http://www.cottonmarker.org/). SSR markers PCR reaction, amplification, and capillary electrophoresis analysis were conducted with an ABI 3100 genetic analyzer (Applied Biosystems, Foster City, CA, USA) followed the protocol of Gutierrez et al. (2002). The procedures for genotyping SNP marker by the single nucleotide extension method was described in detail by An et al. (2007). All polymorphic markers were genotyped across the entire F_2 populations to construct the linkage maps for QTL analysis.

Linkage mapping and QTL analysis

Linkage mapping and QTL analyses were performed on each population separately. Linkage maps were constructed by Joinmap 4.0 (Van Ooijen 2006). A minimum LOD score of 5 and a maximum recombination fraction of 0.40 were set as thresholds to establish linkage groups. Recombination frequencies were converted into map distance (cM) using Kosambi mapping function (Kosambi 1944). Segregation of alleles at each marker locus was tested against the expected ratios (1:2:1 and 3:1 for codominant and dominant markers, respectively) using a chi-square goodness of fit test function in Joinmap 4.0.

QTL was indentified using MapQTL 5.0 (Van Ooijen 2004) with interval mapping and a non-parametric genome scan based on the Kurskal-Wallis (K-W) analysis. Major QTL are considered those with LOD score $\geq$ 3.0 and putative QTL with LOD score between 2.0 and 3.0. The percentage of the phenotypic variance explained by a QTL (R^2) was estimated at the highest probability peak.

Chromosomal assignment of linkage groups and trait-associated markers

Chromosomal assignment of linkage groups and trait-associated markers to specific chromosome was achieved by the following strategies: (1) CMD inquiry (http://www.cottonmarker.org); (2) comparison to the published integrated molecular maps (Lacape et al. 2005; Guo et al. 2007); (3) the complete assignment of cotton linkage maps to chromosomes (Wang et al. 2006); (4) published information on SSR makers, chromosomal assignment by deletion analysis (Guo et al. 2007b); and (5) additional deletion analysis results performed in this study.

Statistical analysis

The distribution of phenotypic traits and correlation analysis were performed by 'PROC MEANS' and 'PROC CORR SPEARMAN' command of SAS 9.1 (SAS Institute Inc., NC, USA), respectively. Genetic effects associated with single markers were conducted by regression analysis and stepwise multiple regression analyses using 'PROC GLM' command of SAS 9.1 (SAS Institute Inc., NC, USA). Single marker regression was performed on all polymorphic markers of each population. In order to detect the combined effect of significant trait-associated markers ($P \leq 0.05$), multiple markers regression analyses were conducted. The significance of these markers was confirmed by both single marker regression analysis and Kurskal-Wallis (K-W) analysis in MapQTL 5.0 (Van Ooijen 2004). Two-locus epistatic interaction among QTL-related markers and significant markers outside mapped QTL were also analyzed by 'PROC GLM' command of SAS 9.1 (SAS Institute Inc., NC, USA).

Results and Discussion

Phenotypic variation and trait correlation

The phenotypic values for yield components, fiber, and seed traits of the two F_2 populations, F_1s, and their parents are summarized in Table 4.1. F_2 mean values are means of all plants in an F_2 population that could be used to measure the trait. Measurements of fiber quality traits were only applied to plants that produced lint. There were 12 plants that did not produce lint and two plants that only produced limited lint fiber in population with FM966. These plants were considered as missing values for data analyses and thus did not affect population mean values. The F_1 plants in both populations had a fuzzless-linted phenotype, but had a smaller lint percentage (LP) and lint index (LI) value than the linted parent line (Table 4.1). Values in Table 4.1 for parents and F_1 were measured from random boll samples and are not replicated. In both populations, LP and LI varied within each qualitative phenotype category, which revealed the quantitative character of these traits as previously reports on fiber mutants QTL mapping (Rong et al. 2005, 2007; Zhang et al. 2005; Guo et al. 2006; Abdurakhmonov et al. 2007; Wan et al. 2007). MD17 did not produce any lint fiber and accordingly fiber quality traits could not be measured.

Table 4.1 Phenotype Values for Yield Components, Seed, and Fiber Traits of F_1, F_2 and their Parents in Two Mapping Populations

Population	Trait[1]	P_1[2] value	P_2[3] value	F_1 value	F_2				
					Mean±SD	Min	Max	Skewness	Kurtosis
MD17×FM966	LP (%)	0	39.69	32.00	24±15	0.00	40.76	-0.60	-1.29
	LI (g)	0	6.31	4.97	3.78±2.53	0.00	8.03	-0.29	-1.38
	SW (g)	10.36	8.63	10.57	10.29±0.99	7.14	12.82	-0.14	0.53
	HP (%)	35.96	39.36	37.02	38±2	33.72	46.40	1.07	2.34
	EPP (%)	29.57	33.03	30.17	32±3	26.43	39.41	0.24	0.16
	EOP (%)	31.91	29.46	32.71	30±2	25.04	37.08	0.20	0.05
	MIC	N/A	4.20	5.50	4.98±0.77	2.90	6.55	-0.22	-0.36
	E1 (%)	N/A	4.50	4.75	5.84±0.80	4.25	9.00	0.98	2.37
	T1 (k N m kg^{-1})	N/A	269	230	227±22	176	271	-0.01	-0.46
	SL50 (mm)	N/A	14.35	14.86	14.46±0.72	12.32	16.00	-0.31	0.03
	SL2.5 (mm)	N/A	29.21	29.08	28.03±1.50	24.00	31.12	-0.20	-0.11
MD17×181	LP (%)	0	22.04	8.32	9.44±7.59	0.00	31.26	0.60	-0.35
	LI (g)	0	2.28	0.94	1.06±0.96	0.00	4.58	1.10	1.18
	SW (g)	10.36	8.08	10.36	9.55±1.21	6.93	13.02	0.16	-0.09

[1]LP-lint percentage; LI-lint index; SW-seed weight; HP-hull percentage; EPP-embryo protein percentage; EOP-embryo crude oil percentage; MIC-micronaire; E1-elongation; T1-fiber strength; SL50- fiber span length 50%; SL2.5-fiber span length 2.5%.
[2]P_1 represents the female parent MD17 in both populations.
[3]P_2 represents the male parent FM966 and 181 in population MD17×FM966 and MD17×181, respectively.

To determine the distribution pattern of each trait, skewness and kurtosis values were calculated (Table 4.1). All tested traits in both populations had absolute skewness value around or less than 1, but the absolute values of kurtosis for LP, LI, HP, and elongation (E1) of population FM966 and for LI of population 181 were greater than 1. A common method to treat this kind of abnormally distributed data for QTL analysis is to conduct log transformation. However, Mutschler et al. (1996) was concerned that normalizing the data could misrepresent the differences of trait among individuals by pulling the skewed tails of the distribution toward the center, thus reducing QTL detection ability. In addition, non-transformed abnormal distribution data for disease resistance (Wright et al. 1998), pubescence (Wright et al. 1999), and flowering time (Guo et al. 2007a) have been repeatedly used for QTL analysis in cotton. Similar QTL mapping result were found when transformed and non-transformed nematode resistance data were compared (Shen et al. 2006a). Thus, the original non-transformed data were used for QTL analysis in this study.

Spearman correlations were calculated to detect the relationship among the non-transformed data of yield components, fiber, and seeds traits (Table 4.2). Significant ($P \leq 0.001$) positive correlations were present between two yield components traits, LP and LI, with the coefficient value as 0.96 and 0.99 in population with FM966 and 181, respectively. Among measured seed traits (SW, HP, EPP, and EOP), significant ($P \leq 0.05$) correlation were only observed between SW and HP with the coefficient as -0.23. Within the five fiber quality traits, fiber strength (T1), fiber span length 50% (SL50) and 2.5% (SL2.5) positively correlated with each other and T1 also had a negative correlation with micronaire (MIC). Both yield components traits were positively correlated with EPP and

T1, but negatively correlated with HP, EOP, and MIC. Phenotypic correlation coefficients of SW with T1, SL50, and SL2.5 were 0.31, 0.37, and 0.33, respectively. EPP correlated with MIC and T1 in a negative and positive manner, respectively; while EOP had the reversed correlation with these two traits to that of EPP.

Table 4.2 Spearman Correlation Coefficients between Yield Components, Seed, and Fiber Traits in Two Mapping Populations[1, 2]

Trait[3]	LP	LI	SW	HP	EPP	EOP	MIC	E1	T1	SL50
LP	1	0.99***	-0.13							
LI	0.96***	1	0.02							
SW	-0.15	0.06	1							
HP	-0.35***	-0.42***	-0.23*	1						
EPP	0.49***	0.49***	0.10	0.04	1					
EOP	-0.30**	-0.26**	-0.03	-0.08	-0.19	1				
MIC	-0.63***	-0.61***	0.04	0.06	-0.53***	0.25*	1			
E1	-0.15	-0.20	0.04	0.15	0.01	-0.11	-0.00	1		
T1	0.27*	0.39***	0.31**	-0.28**	0.35**	-0.32***	-0.22*	-0.07	1	
SL50	-0.07	0.07	0.37***	-0.06	0.12	-0.12	0.07	-0.14	0.42***	1
SL2.5	0.02	0.16	0.33**	-0.12	0.18	-0.11	-0.12	-0.19	0.46***	0.88***

[1]The number below and above the diagonal represent correlation analysis results from population MD17×FM966 and MD17×181, respectively.
[2]*, **, and *** denote significance at ≤0.05, ≤0.01, and ≤0.001 level, respectively.
[3]LP-lint percentage; LI-lint index; SW-seed weight; HP-hull percentage; EPP-embryo protein percentage; EOP-embryo crude oil percentage; MIC-micronaire; E1-elongation; T1-fiber strength; SL50- fiber span length 50%; SL2.5-fiber span length 2.5%.

Linkage maps

One hundred and forty-four SSR primer pairs out of the 1407 SSR and 31 SNP primer pairs (10.0%) amplified polymorphisms between MD17 and FM966 at 159 loci across the genome. Of these, 115 were assigned to 29 linkage groups with a total map distance of 779 cM and covered approximately 17.3% of the total recombination length of the cotton genome (Rong et al. 2004). One hundred and forty-six SSR loci, which were obtained from 132 SSR polymorphic primer pairs selected from the total of 1438 primer pairs (1407 SSR and 31 SNP) between MD17 and 181 (9.2%), were used for linkage analysis. Of these, 132 markers were assigned to 33 linkage groups which covered 1003 cM or approximately 22.3% of the total recombination length of the cotton genome (Rong et al. 2004).

No polymorphic SNP marker was found between parents for either population; however there were only 31 SNP markers aviabile. There were 11 (6.9%) and 16 (9.8%) segregation distorted ($P \leq 0.05$) markers detected in population FM966 and 181, respectively. In population FM966, linkage map did not cover chromosomes 1, 4, 5, 7, 12, 22, 23, 24, and 25; whereas in population 181 it did not cover chromosomes 1, 5, 13, 21, 22, 23, and 24. Thus, these chromosomes were not used for QTL mapping, but the single and multiple marker regression analyses were still applicable using the AD model. Some of these markers could be assigned to a chromosome or chromosome arm, but not to a specific linkage location on the chromosome.

QTL associated with different traits

Analysis of both F_2 populations for yield components, seed, and fiber traits resolved 26 QTL that could be assigned to a map, (Table 4.3). These QTL were designated as '*q*' followed orderly by an abbreviation of a trait name, the chromosome location, and then the number of the detected QTL related to the trait on that chromosome (McCouch et al. 1997). Additional markers significantly associated with QTL but not mapped were detected by both single marker regression analysis and K-W analysis.

Lint Percentage

One major QTL *qLP-c26-1* was mapped for LP in both populations, which could explain 87.1 and 68.3% of the phenotypic variation, respectively. In each population the QTL shared the same marker BNL3482-138 on chromosome 26. In addition, two QTL, *qLP-c11-1* and *qLP-c11-2*, mapped on chromosome 11 and one QTL mapped on chromosomes of 10 (*qLP-c10-1*) and 12 (*qLP-c12-1*) were only observed in population 181. The percentage of the phenotypic variation explained by these QTL in population 181 ranged from 6.1% for *qLP-c11-2* to 28.4% for *qLP-c11-1*. However, two makers on the long arm of chromosome 12, STV79-108 and BNL2621-198, were also significantly associated with LP in populations FM966. They contributed 26.7 and 10.4% of the phenotypic variation, respectively. When three other markers which explained less than 10% each of the phenotypic variation were combined with these two markers, the five non-mapped markers accounted for 47.2% of the phenotypic variation for LP in population FM966.

Table 4.3 QTL and Single Non-mapped Markers Associated with Yield Components, Seed, and Fiber Traits Identified by Interval Mapping with MapQTL 5.0 and Single Marker Regression Analysis, Respectively

Trait[1]	QTL	Marker	Position (cM)	LOD	R^2 (%)[2]	A[2,3]	D[2]	Chro.[4]	Reference
MD17×FM966									
LP	*qLP-c26-1*	BNL3482-138 to BNL598-119	19.00	5.79	87.1	-14.93	12.27	26	
		STV79-108			26.7*	-5.48	-11.83*	12*Lo*	Deletion analysis
		BNL2621-198			10.4*	3.56	-2.34	12*Lo*	Deletion analysis
		BNL3282-158			7.9*	6.80*	3.33	UL	
		BNL2631-185			6.4*	2.98	-6.46*	UL	
		BNL3590-188			6.1*	-1.57	-6.84*	17*Lo*	Deletion analysis
LI		STV79-108			30.6**	-1.26	-1.96*	12*Lo*	Deletion analysis
		BNL2621-198			9.8*	0.58	-0.57	12*Lo*	Deletion analysis
		BNL3282-158			9.1*	1.37	0.50	UL	
		BNL3482-138			8.4*	-0.96**	0.14	26	Deletion analysis
		NAU1369-253			8.2*	-2.06	0.84	UL	
		BNL3971-181			6.3*	1.03	0.25	2*Lo*	Deletion analysis
SW	*qSW-c9-1*	BNL1317-179	0.00	2.50	11.0	0.39	0.40	9	
	qSW-c16-1	TMB561-245	2.57	5.50	22.6	0.65	0.15	16	
		BNL3545-184			9.4*	0.29	0.43	2*Lo*	Deletion analysis
		BNL2650-201			8.9*	0.29*	-0.40*	UL	
		CIR166-113			8.4*	0.52	-0.14	10	Guo et al., 2007
		MGHES31-202			6.1*	0.47	-0.03	12	Lacape et al., 2005
HP	*qHP-c3-1*	MGHES8-178 to MUCS400-219	13.00	2.75	16.1	-0.88	-1.10	3	
	qHP-c14-1	BNL3034-153 to JESPR6-208	15.00	2.92	15.3	0.89	-1.10	14	
		BNL2650-201			9.6*	-0.46	1.07**	UL	
		NAU1369-253			8.9*	1.71	-1.87	8	Guo et al., 2007
		BNL3400-173			8.5*	-1.14	2.35*	UL	
		BNL3347-142			8.4*	-0.19	1.17	19	Guo et al., 2007
		BNL1395-156			8.0*	-1.88	0.53	16	Guo et al., 2007
		BNL673-127			7.9*	0.45	0.58	UL	
		CM42-138			7.5*	-0.45	-0.89*	9*sh*	Deletion analysis

Table 4.3 (continued)

EPP	*qEPP-c6-1*	NAU1151-215	22.62	2.28	10.9	-1.06	-1.16	6	
	qEPP-c15-1	CIR234-283 to CIR270-182	24.28	2.09	11.0	-0.44	1.73	15	
		STV79-108			15.4**	-3.03**	1.14	12*Lo*	Deletion analysis
		MUSS298-167			8.6*	-1.67	-0.14	9	Park et al., 2005
		BNL3590-181			6.2*	-0.30	1.31*	2	Guo et al., 2007
EOP		STV164-173			10.3*	-0.22	2.18*	9*sh*	Deletion analysis
		CIR393-200			7.9*	1.77	-0.46	7*Lo*	Deletion analysis
		BNL3400-173			7.9*	-0.92	-0.89	UL	
		BNL3261-203			6.1*	-0.64	0.78	12	Guo et al., 2007
		BNL119-224			4.7*	-0.58*	N/A	20	Guo et al., 2007
MIC	*qMIC-c2-1*	JESPR101-117	1.20	2.34	12.4	-0.24	0.41	2	
	qMIC-c26-1	BNL3482-138	0.00	2.30	12.1	0.29	0.27	26	
		STV79-108			20.0**	0.31	0.43	12*Lo*	Deletion analysis
		BNL-3099-172			13.2*	0.06	0.47*	UL	
		MUCS6-155			9.9*	-0.25	-0.28	7	Park et al., 2005
		MUSS298-165			9.8*	0.77*	-0.30	9	Park et al., 2005
		BNL2705-159			9.6*	0.07	-0.47**	10	Park et al., 2005
		CIR99-81			9.0*	-0.01	-0.45**	18	Lacape et al., 2005
		BNL1162-243			7.5*	-0.27*	0.13	9	Guo et al., 2007
E1		JESPR211-226			10.4*	-1.15**	-1.12*	UL	
T1	*qT1-c16-1*	BNL1395-16	0.00	2.08	11.1	9.95	-3.16	16	
		BNL3854-146			11.0*	-12.09*	5.65	UL	
		BNL3255-216			7.6*	7.78*	2.41	8	Guo et al., 2007
SL50	*qSL50-c10-1*	BNL3790-165 to CIR166-113	6.00	3.51	21.2	-0.14	0.65	10	
	qSL50-c21-1	BNL3442-110 to BNL3171-228	20.00	2.57	28.5	0.20	-0.75	21	
		BNL4028-166			11.9*	0.55	-0.98*	UL	
SL2.5	*qSL2.5-c10-1*	TMB317-191 to CIR166-113	12.96	3.02	15.7	1.08	0.97	10	
	qSL2.5-c21-1	BNL3442-110 to BNL3171-228	24.00	2.42	22.3	0.23	-1.41	21	
		BNL3482-138			9.8*	-0.59**	0.26	26	Guo et al., 2007
		BNL1066-128			8.9*	-0.62	-0.27	26*Lo*	Deletion analysis

Table 4.3 (continued)

		BNL3031-160			7.5*	0.06	-0.82*	9*Lo*	Deletion analysis
		BNL3902-196			5.2*	0.38*	N/A	15	Guo et al., 2007
MD17×181									
LP	*qLP-c10-1*	TMB317-191	0.68	2.47	6.8	-1.73	2.83	10	
	qLP-c11-1	BNL1231-193 to BNL2650-210	20.00	5.20	28.4	-5.30	-3.30	11	
	qLP-c11-2	CIR196-190	1.32	2.19	6.1	1.96	-2.61	11	
	qLP-c12-1	BNL1227-185	0.00	2.69	7.3	-3.20	-0.42	12	
	qLP-c26-1	BNL3482-138	0.00	31.17	68.3	-9.31	-1.49	26	
LI	*qLI-c10-1*	TMB317-191	0.68	2.30	6.3	-0.21	0.34	10	
	qLI-c11-1	BNL1231-193 to BNL2650-210	19.00	6.56	34.7	-0.72	-0.57	11	
	qLI-c11-2	CIR196-190	1.32	2.19	6.1	0.23	-0.36	11	
	qLI-c12-1	BNL1227-185	0.00	2.15	5.9	-0.36	-0.07	12	
	qLI-c26-1	BNL3482-138	0.00	26.65	62.8	-1.11	-0.29	26	
		TMB1437-153			4.8*	0.25*	-0.21	UL	
		TMB1473-173			4.5*	0.10	-0.38*	UL	
		CM76-196			4.3*	-0.24*	-0.20	15*sh*	Deletion analysis
		TMB353-103			4.2*	0.05	-0.38*	UL	
SW	*qSW-c7-1*	BNL1122-175 to CIR393-200	1.00	13.08	32.2	0.98	0.17	7	
	qSW-c11-1	BNL1231-193 to BNL2650-210	16.00	2.04	13.4	-0.01	-0.89	11	
		MGHES6-192			5.3*	-0.33*	0.34	UL	
		BNL4064-142			2.4*	-0.23*	N/A	UL	

[1]LP-lint percentage; LI-lint index; SW-100 seed weight; HP-hull percentage; EPP-embryo protein percentage; MIC-micronaire; T1-fiber strength; SL50- fiber span length 50%; SL2.5-fiber span length 2.5%.
[2]*, **, and *** denote significance at ≤0.05, ≤0.01, and ≤0.001 level, respectively.
[3]Positive additive value indicates the allele from parent MD17 increases the trait value; negative additive value indicates the allele from another parent (FM966 or 181) increases trait value.
[4]Chromosome assignment by deletion analysis was performed by chromosome substitution stocks or aneuploid stocks; UL-unlocated marker; QTL chromosome assignment accomplished by strategies stated in the material and methods part of the text; *Lo* and *sh* mean on the long and short arm of the chromosome, respectively.

Lint Index

QTL for LI could only be mapped in population 181 (Table 4.3). Five QTL shared the same chromosome (10, 11, 12, and 26) locations and intervals as QTL for LP in this population. The phenotypic variation explained by these QTL varied from 6.1 to 62.8%. In this population, *qLI-c11-1* (34.7%) and *qLI-c26-1* (62.8%) were the two major QTL for LI with similar values to QTL for LP. There were four non-mapped markers significantly associated with LI in population 181, but their effects were less than 5% each (Table 4.3). Among all the polymorphic markers tested in population FM966, STV79-108 had the largest effect (30.6%) on LI. Combined with the other five significant markers on chromosome 2, 12, 26, and unknown locations, the six markers explained 55.1% of the phenotypic variation.

Seed Weight

QTL for SW were different in the two mapping populations (Table 4.3). In population FM966, QTL were mapped on chromosome 9 (*qSW-c9-1*, 11.0%) and 16 (*qSW-c16-1*, 22.6%); while in population 181, mapped QTL on chromosome 7 and 11 explained 32.2 and 13.4% of the phenotypic variation, respectively. In both mapping populations, additional significant non-mapped markers were found, Table 4.3, but none explained a large amount of the phenotypic variation.

Hull Percentage

Two suggestive QTL, designated as *qHP-c3-1* and *qHP-c14-1*, were detected explaining 16.1 and 15.3% of the phenotypic variation, respectively (Table 4.3). The 181 allele increased additive HP at *qHP-c3-1* and the MD17 allele increased additive HP at *qHP-c14-1*. Seven non-mapped markers were also significantly associated with HP at

different chromosome locations and they explained 38.5% of the phenotypic variation (Table 4.3).

Embryo Protein Percentage

Two suggestive QTL, *qEPP-c6-1* and *qEPP-c15-1*, for EPP were identified, which explained 10.9 and 11.0% of the phenotypic variation, respectively (Table 4.3). The alleles from MD17 decreased the additive EPP. Three other significant QTL markers on chromosome 2 (BNL3590-181), 9 (MUSS298-167), and 12 (STV79-108) were associated with EPP, explaining 6.2, 8.6, and 15.4% of the phenotypic variation, respectively (Table 4.3). Together they contributed 27.7% of the phenotypic variation.

Embryo Crude Oil Percentage

QTL for EOP was not be mapped; however, five markers significantly associated with QTL for EOP were identified (Table 4.3). Among them, STV164-173 on the short arm of chromosome 9 explained the largest phenotypic variation (10.3%). The phenotypic variations explained by the other four markers were in the range of 4.7 to 7.9%. The combined effect of these five markers was 32.7%.

Micronaire

Two suggestive QTL for MIC, *qMIC-c2-1* and *qMIC-c26-1*, explaining 12.4 and 12.1% of the phenotypic variation, respectively, were identified (Table 4.3). *qMIC-c2-1* was in the vicinity of JESPR101-117 located on chromosome 2, and *qMIC-c26-1* was in the vicinity of BNL3482-138 located on chromosome 26. The MD17 additive allele decreased MIC at *qMIC-c2-1* and increased it at *qMIC-c26-1*. Two additional markers (STV79-108 and BNL3099-172) explained 20.0 and 13.1% of the phenotypic variation,

respectively (Table 4.3). Together with the other small effect markers, they explained 52.4% of the phenotypic variation.

Elongation

Only one marker (JESPR211-226) with unknown chromosomal location were found significantly associated with E1. It explained 10.4% of the phenotypic variation.

Fiber Strength

A suggestive QTL, *qT1-c16-1*, and two markers, BNL3854-146 and BNL3255-216, had significant association with T1. The QTL was in the vicinity of BNL1395-166 located on chromosome 16, which explained 11.1% of the phenotypic variation. FM966 allele increased additive fiber strength at this locus. The two non-mapped QTL contributed 17.1% of the phenotypic variation.

Fiber Span Length 50%

Two QTL, *qSL50-c10-1* and *qSL50-c21-1*, were detected for SL50, which were on chromosome 10 and 21, accounting for 21.2 and 28.5% of the phenotypic variation, respectively (Table 4.3). Interestingly, the allele from MD17 increased additive effects on SL50 at *qSL50-c21-1*, which revealed that the fuzzless-lintless line harbors some good alleles for fiber quality. A non-mapped marker BNL4028-166 also had significant association with SL50 and explained 11.9% of the phenotypic variation.

Fiber Span Length 2.5%

Similar marker intervals of QTL for SL50 on chromosome 10 and 21 and two QTL for SL2.5 (*qSL2.5-c10-1* and *qSL2.5-c21-1*) were mapped, which explained 15.7 and 22.3% of the phenotypic variation, respectively (Table 4.3). However, at locus *qSL2.5-c10-1*, unlike that in SL50, it was the MD-17 allele that increased the additive

effect for SL2.5. Four other significant non-mapped markers together explained 22.2% of the phenotypic variation (Table 4.3).

Two-locus epistatic interactions

Marker defined QTL intervals and significant markers for non-mapped QTL were used in the two-locus epistatic interaction analysis. A number of two-locus epistatic interaction significant at $P \leq 0.05$ level were identified for HP, EPP, MIC, and SL2.5 in population FM966 and for LP, LI, and SW in population 181 (Table 4.4). Among six pairs of markers showing epistatic interaction for HP of population FM966, two pairs (JESPR6-208 and MGHES8-178; JESPR6-208 and MUCS400-219) suggested epistatic interactions might be present between *qHP-c3-1* and *qHP-c14-1*. In the same population, molecular marker BNL3482-138 associated with *qMIC-c26-1* showed significant epistastic interactions with marker STV79-108 on chromosome 12. For SL2.5 of population FM966, two markers (CIR166-113 and TMB317-191) of QTL *qSL2.5-c10-1* both interacted with markers on chromosome 15 and 26. The results of the two-locus epistatic analysis detected in population 181 revealed interactions between QTL on chromosome 11 and 26 for both LP and LI. Furthermore, QTL for LI and SW on chromosome 11 interacted with markers on chromosome 26 and non-mapped markers in population 181.

Table 4.4 Two-locus Epistatic Interactions Identified in Two Mapping Populations

Population	Trait[1]	Locus 1/Chromosome	Locus 2/Chromosome	R^2 (%)[2]
MD17×FM966	HP	JESPR6-208/14	BNL3347-142/19	31.1*
		JESPR6-208/14	NAU1369-253/8	34.3**
		JESPR6-208/14	MGHES8-178/3	32.2**
		JESPR6-208/14	MUCS400-219/3	37.3**
		NAU1369-253/8	BNL1395-156/16	31.8***
		NAU1369-253/8	BNL3400-173/unknown	26.4*
	EPP	MUSS298-167/9	BNL3590-181/2	22.4*
	MIC	BNL3482-138/26	STV79-108/12	40.9*
	SL2.5	BNL3031-160/9	BNL3902-196/15	20.1*
		CIR166-113/10	BNL3482-138/26	36.4*
		CIR166-113/10	BNL3902-196/15	29.6**
		TMB317-191/10	BNL3482-138/26	34.7*
		TMB317-191/10	BNL3902-196/15	27.7*
MD17×181	LP	BNL1231-193/11	BNL3482-138/26	38.8*
	LI	CM76-196/15	TMB353-103/unknown	14.1*
		TMB1473-173/unknown	TMB353-103/unknown	18.7**
		TMB1437-153/unknown	CM76-196/15	16.2*
		CIR196-190/11	CM76-196/15	20.0*
		BNL2650-210/11	TMB1437-153/unknown	19.9*
		BNL1231-193/11	BNL3482-138/26	38.8*
	SW	BNL1231-193/11	MGHES6-192/unknown	16.7*

[1]LP-lint percentage; LI-lint index; SW-seed weight; HP-hull percentage; EPP-embryo protein percentage; MIC-micronaire; SL2.5-fiber span length 2.5%.
[2]*, **, and *** denote significance at ≤0.05, ≤0.01, and ≤0.001 level, respectively.

Major population types used for cotton QTL mapping consists of F_2, $F_{2:3}$, BC_x, and RILs. Due to the fast development and the ability to measure both additive and dominant effect, F_2 populations have been considered as the desirable population type to detect QTL for various traits of cotton including yield components (Rong et al. 2005, 2007; Guo et al. 2006), fiber quality (Kohel et al. 2001; Rong et al. 2007), biotic resistance (Wright et al. 1998; Shen et al. 2006a), photoperiod response (Guo et al. 2007a), and leaf morphology (Wright et al. 1999; Waghmare et al. 2005).

In addition to conferring the fuzzless trait, the mutant qualitative genes (N_1 and n_2) affects lint production in a quantitative manner (Ware 1940; Ware et al. 1947). Using mapping populations developed from a *G. barbadense* line (fuzzy-linted) and single fuzzless locus mutant (*G. hirsutum*, N_1 and n_2), Rong et al. (2005) observed considerable quantitative variation in lint fiber production superimposed on the discrete effects of fuzz mutants and mapped both loci within the intervals of QTL for lint percentage and lint index. These results were confirmed by their following study on QTL mapping for lint percentage and fiber quality traits (micronarie, uniformity, and fiber elongation) using n_2 mutant derived population (Rong et al. 2007). However, all these results were based on their conclusion that both fuzzless loci (N_1 and n_2) were on chromosome 12 of the tetraploid cotton genome; however, they were originally regarded as a pair of homoeologous loci on the long arm of chromosomes 12 and 26 (Endrizzi and Ramsay 1980; Endrizzi et al. 1984; Samora et al. 1994). Instead of using the single fuzzless loci mutant developed mapping populations (Rong et al. 2005, 2007; Zhang et al. 2005; Guo et al. 2006; Wan et al. 2007), the common female parent line of the two F_2 populations combined at least two fuzzless loci (N_1 and n_2, and probably n_3) (Turley and Kloth 2002).

QTL mapping in this study revealed that QTL for LP or LI which explained a high percentage of the phenotypic variation (62.8 to 87.1%) were probably located near the same location of chromosome 26 because they share the common marker BNL3482-138 in both populations in this research. The MD17 BNL-3482 allele at this locus decreased both LP and LI. In addition, a putative QTL for MIC (LOD=2.30, R^2=12.1%) was in the vicinity of the same marker. Abdurakhmonov et al. (2007) also found a close association between SSR markers on chromosome 26 and lint percentage using fiber mutant derived RIL population. An EST-SSR maker, STV79-108, assigned to the long arm of chromosome 12 (the location of N_1) showed significant association with LP, LI, EPP, and MIC (Table 4.3). But there was no significant hit of this primer original sequence in the GenBank database (Taliercio et al. 2006). The allele from MD17 at this locus reduced LP and LI similar to that of the QTL on chromosome 26. On chromosome 12, where N_1 is located, suggestive and small effect (R^2=7.30 and 5.90%, respectively) QTL for LP and LI in the vicinity of BNL1227-185 were detected in population 181. Even though STV79-108 showed polymorphism between the two parents of population 181, no significant association was detected between this marker and any trait. this might indicate that the two fuzzless parental lines share the common fuzzless locus (N_1 or n_2). Additional QTL for yield components and fiber quality traits not on chromosome 12 or 26 had a general chromosomal agreement with previous reports (Guo et al. 2006; Rong et al. 2007). Because the common parent line used in this study combined multiple fuzzless loci, the fiber mutant trait could not be used as a morphological marker and be placed on the linkage map as Rong et al. (2005). Previous studies on fuzzless mutant mapping used different molecular marker type (Rong et al. 2005, 2007) or there were very few SSR

markers on chromosome 12 and 26 (Zhang et al. 2005). Advanced experiments incorporating widely used SSR marker into the integrated STS markers framework map (Rong et al. 2004, 2005) would be helpful to further verify the association of these fuzzless loci with lint production traits in tetraploid cotton.

Cotton seed protein and oil percentage are controlled by multiple genes (Dani and Kohel 1989; Ye et al. 2003). A pilot study on molecular mapping of cotton seed physical and nutrient traits detected QTL for kernel protein and oil percentage on chromosome 23 and 24, which explained above 20% of the phenotypic variations (Song and Zhang 2007). In this study, additional QTL on different chromosomes were associated with seed traits (SW, HP, EPP, and EOP) on different chromosomes. Phenotypic variation explained by each individual QTL ranged from 10.9 to 32.2% (Table 4.3). EST-SSR marker STV79-108, which was associated with LP, LI, and, MIC, was also tightly associated ($P \leq 0.001$, $R^2=15.4\%$) with EPP. Marker STV164-173, which had high similarity with a gene (*Pho1*) coding for α-1,4-glucan phosphorylase (Taliercio et al. 2006), showed a significant association with EOP ($P \leq 0.05$, $R^2=10.3\%$). The gene *Pho1* is related to starch metabolism (Buchner et al. 1996). It might be reasonable to find the association between EOP and starch metabolism related genes when considering the proposed model of carbon partitioning among fiber, seed coat, storage protein, and oil in developing cotton seed (Ruan et al. 1997).

In summary, QTL associated with yield components, seed, and fiber traits of Upland cotton were detected under the multiple fuzzless loci genetic backgrounds. These results should help accelerate genetic dissection of the complex cotton fiber development network and optimize both fiber and seed traits through breeding.

CHAPTER V

INSIGHTS INTO THE INHERITANCE OF LINT AND FUZZ FIBER INITIATION IN UPLAND COTTON (*Gossypium hirsutum* L.)

Abstract

Upland cotton (*Gossypium hirsutum* L.) fibers are single-celled trichomes derived from the ovule epidermis. Based on initiation time and final length, fiber can be distinguished as adherent fuzz and spinnable lint. Segregating populations (F_2 and BC_1) derived from a half diallele cross among two fiberless (MD17 and 177), two fuzzy-short lint (Li_1 and Li_2), one fuzzless-linted (181), and one normal phenotype commercial cultivar (FM966) lines were employed to evaluate the inheritance pattern of fuzz and lint fiber development. These lines were identified or developed from US (MD17, Li_1, Li_2, and FM966) and China (177 and 181) germplasm. Inheritance analysis indicated that both genetic models of $N_1N_1n_2n_2$ and $n_2n_2li_3li_3$ could lead to the fiberless phenotype. The observation of fuzzless-short lint phenotype in crosses between Li_1 or Li_2 and three other fiber mutant lines provided evidences that the original phenotype of Li_1 and Li_2 was fuzzy-short lint and cotton fiber initiation and elongation were controlled by different mechanisms. The three categories of abnormal phenotypes (typical phenotype, two-phenotype on same plant, and fuzzy/fuzzless-long lint) detected in Li_2 related generations represented the report on the penetrance of fiber mutant gene expression.

Introduction

Upland cotton (*Gossypium hirsutum* L.) fiber is a differentiated single epidermal cell of the ovule. The coat of normal cottonseed is covered with spinnable lint and adherent fuzz, which are separated by ginning. These two types of fibers differ in initiation timing and final length. Lint fiber initiates between anthesis and two days post-anthesis (dpa) and can elongate 2.5-3.5 cm, whereas the fuzz fiber starts initiation approximately a week later and the elongation stops at around 0.5 cm (Stewart 1975). Three different mutant lines have been identified in nature or derived by crossing mutants, which are fuzzless-linted (naked seed, N), fuzzy-short lint, and fiberless (fuzzless-lintless, fls). Two loci, the dominant naked seed gene N_1 and the recessive naked seed gene n_2, have been reported to inhibit fuzz fiber initiation on the cottonseed coat (Kearney and Harrison 1927; Ware 1940; Ware et al. 1947; Kohel 1973; Endrizzi et al. 1984; Percy and Kohel 1999). Lint fiber elongation could be completely arrested by two nonallelic genes Li_1 and Li_2 (Kohel 1972; Narbuth and Kohel 1990; Kohel et al. 1992). Both Li_1 and Li_2 were characterized by fuzzy and uniform layer of very short lint fiber, while only Li_1 had stunted plant and deformed vegetative morphology. To date, six genotype models have been reported for the five fiberless lines collected from different parts of world: L40 ($Ft_1Ft_1Ft_2Ft_2Ft_cFt_cN_2N_2$, Musaev and Abzalov 1972), MCU5 (2 to 4 undetermined recessive genes, Peter et al. 1984; Nandarajan and Rangasamy 1988), two different models for XZ142w ($n_2n_2li_3li_3$, Zhang and Pan 1991; $n_1n_1n_2n_2li_3li_3li_4li_4$, Du et al. 2001), MD17 ($N_1N_1n_2n_2n_3n_3$ or $N_1N_1n_2n_2$, Turley 2002; Turley and Kloth 2002 and 2008), and

SL1-7-1 ($N_1N_1fl_1fl_1n_3n_3$, Turley and Kloth 2008). Both fiberless lines MD17 and SL1-7-1 were developed in US, while fiberless lines L40, MCU5, and XZ142w were identified in Uzbekistan, India, and China, respectively. Except for the genetic analysis on fiberless line L40, which conferred to the inheritance of fuzz distribution instead of presence or absence, the five fiberless inheritance models for the other four fiberless lines postulated four new loci: n_3, li_3, li_4, and fl_1. However, the phenotypes solely determined by these loci have not been reported.

Therefore, initiation and elongation of an epidermal cell into fiber requires a complex network of gene interaction. Whether the interaction is dominant epistasis (Zhang and Pan 1991; Ding et al. 2007), recessive epistasis (Du et al. 2001), or several genotype models determining the fiberless phenotype (Turley and Kloth 2002, 2008), it is still contentious. The disparity of interpretations among previous reports warrants further clarification of the genetic control of cotton fiber development. With the attempt to further understand the genetic mechanisms and the curiosity to look for a common or general genetic interaction model, three kinds of fiber mutant lines together with one normal phenotype commercial cultivar were used to derive segregating populations and evaluate the inheritance of fuzz and lint fiber development. Parental lines 177 (fiberless) and 181 (fuzzless-linted) originated from the cross of fiberless XZ142w with fuzzless-linted GZNn (Du et al. 2002). Both XZ142w and GZNn are China fiber mutant germplasm. So, genetic analysis of these parental lines derived segregating populations will shed light on comparing different interpretations given by China and US scientists.

Materials and Methods

Plant materials

Six inbreed lines including five fiber mutants, MD17 (PI 616493), 177, 181, Li_1, and Li_2, along with one commercial cultivar FiberMax 966 (FM966, Bayer CropScience) of Upland cotton (*Gossypium hirsutum* L.) were used as parental lines for deriving segregating populations. MD17 and 177 are fiberless i.e. fuzzless-lintless (fls) lines with different origin. MD17 derived from the cross between Mexican fuzzless seed UA3-3 (MOVC accession 143, recessive n_2) and Ballard fuzzless seed (MOVC accession 243, dominant N_1), whereas 177 originated from the cross between two China germplasm lines XZ142w (fiberless) and GZNn (fuzzless-linted/naked seed, N) (Zhang and Pan 1991; Du et al. 2001; Turley 2002; Turley and Kloth 2002; Turley and Kloth 2008). Li_1 and Li_2 are two nonallelic dominant fiber mutants (Narbuth and Kohel 1990; Kohel et al. 2002). In addition to control short and thick (fuzzy-short lint) fibers development by both mutants, Li_1 is characterized by stunted and deformed stem and leaves contrasting to the normal vegetative morphology of Li_2 (Kohel 1972; Narbuth and Kohel 1990). Line 181 has the same pedigree origin as 177, but it was selected as fuzzless-linted i.e. naked seed (N). The detail description of phenotypes, origins, and seed sources of the six lines are described in Table 5.1.

Table 5.1 Phenotype, Origin, and Seed Source of Six Parental Lines

Parent	Phenotype	Origin	Seed source
FM966	Normal i.e. fuzzy-linted (F)	-	Bayer CropScience Co.
MD17	Fiberless i.e. fuzzless-lintless (fls)	Cross between MOVC accessions 143 (recessive n_2) and 243 (dominant N_1)[1]	Dr. R.B. Turley, USDA-ARS, Stoneville, MS
177	Fiberless i.e. fuzzless-lintless (fls)	Cross between XZ142w and GZNn[2]	Dr. X.M. Du, CRI-CAAS, China
181	Fuzzless i.e. fuzzless-linted (N)	Cross between XZ142w and GZNn[2]	Dr. X.M. Du, CRI-CAAS, China
Li_1	Fuzzy-short lint (Stunted and deformed vegetative morphology)	Completely dominant single-gene spontaneous mutation and an isogenic line of TM-1[3]	Dr. R.J. Kohel, USDA-ARS, College Station, TX
Li_2	Fuzzy-short lint (normal vegetative morphology)	Completely dominant single-gene spontaneous mutation[4]	Dr. R.J. Kohel, USDA-ARS, College Station, TX

[1]Turley and Kloth 2002
[2]XZ142w is a spontaneous fiberless (fuzzless-lintless) mutant from Upland cotton cultivar Xuzhou 142 of China. Two different fiberless genotype models $n_2n_2li_3li_3$ and $n_1n_1n_2n_2li_3li_3li_4li_4$ have been proposed for XZ142w by Zhang and Pan 1991 and Du et al. 2001, respectively. GZNn is a fuzzless (fuzzless-linted) line characterized by expression of N_1N_1 fuzzless gene (Du et al. 2001)
[3]Kohel 1972; Kohel et al. 1992
[4]Narbuth and Kohel 1990; Kohel et al. 1992

In summer of 2005, the six lines were grown in the greenhouse at Mississippi State (MS) for deriving F_1 seeds by half diallele cross. All plants were grown in nursery containers (20 cm diameter by 30 cm depth) filled with a commercially available potting mix. Nutrition and water was supplied regularly. No supplemental lighting was employed. The F_1 seeds were harvested and grown in the same conditions as their parents in November of 2005. Flowers were self-pollinated to generate F_2 seeds. The six parental lines, 15 half diallele cross derived F_1s and F_2s were grown in the field in 2006 as two-row, two-row, and six-row plots (12 m in length), respectively. In summer of 2006, backcrosses were made in the field. Fourteen crosses (not including cross $Li_2 \times Li_1$) derived F_1s were crossed with one of their parent lines to derive BC_1 seeds. For both greenhouse and field experiments, plants for selfing and crossing were sequentially numbered. Plant numbers were recorded on the tags attached to each self and cross boll for later verification of the phenotype of the parent plant at harvest. The six parental lines, 15 cross derived F_2s, and 14 BC_1s produced from the 2006 summer field crosses were grown in the field in 2007 as two-row, three-row, and six-row plots (12 m in length), respectively. F_2 data were combined across years after a test of heterogeneity between years demonstrated the results were homogeneous. Field experiments were conducted at the Plant Science Research Center of Mississippi State University. In all field trials, where applicable, plots were over-seeded and after the plants reached the first true leaf stage, seedling were thinned to three plants per meter. Standard cultural practices, weed and insect control were followed.

At the end of the 2006 season, it was discovered that parental line Li_2 and it related F_1s and F_2s showed abnormal phenotypes. The greenhouse verified typical

parental Li_2 plants (fuzzy-short lint) derived progenies showed three different phenotypes in S_1 (selfed parental line) and F_1, which were typical Li_2 phenotype, two-phenotype on same plant, and fuzzy/fuzzless-long lint. This phenomenon happed again in the 2007 field trials of parental line Li_2 and its related F_2s and BC_1s. However, only two-phenotype on same plant could be detected among segregating populations (F_2s and BC_1s). In an attempt to further detect this abnormal phenomenon, not previously reported (Narbuth and Kohel 1990), several abnormal plants were transplanted to the greenhouse after trimming branches and grown in nursery containers mentioned before to observe plant performance in the greenhouse. Before transplanting, open-pollinated seeds from different phenotype branches/bolls or fuzzy-long lint plants of Li_2 were harvested separately. These seeds were grown in 2007 in the field as one-row plot (12 m in length) per source.

Data collection, phenotype classification, and data analysis

In order to avoid the possible bias seed phenotypic classification by examining seeds at certain stages or specific positions on the cotton plant, the following two-step method were employed. First, individual plants from the F_2 and BC_1 populations were tagged and two open bolls at the first branch node between main stem nodes 7-9 were collected and placed in labeled paper bags and brought to lab for classification; second, at the end of the season, all plants in the field were examined and the phenotype was recorded. The fuzzy/fuzzless phenotypes were scored according to the standards set up by Ware (1940), with the fuzzy seed corresponding to classes 1-9 and fuzzless seed corresponding to classes 13-19. The grouping grades of fuzzy and fuzzless seed were

employed to discover recessive fuzzless seed phenotype (Ware et al. 1947) and genetic analyses of different fiberless lines (Turley and Kloth 2002, 2008). Lint/short-lint classification was performed by comparing to the two fuzzy-short lint lines Li_1 and Li_2. The final phenotype classification was confirmed by the two-step process. All population sizes were tested to ensure they exceeded the suggested sizes for a probability level of 0.05 (Hansen 1959). Chi-squares were calculated to test the goodness-of-fit for genetic models proposed.

Results and Discussion

Fiber phenotypes conferred by the six parental lines and their half diallele cross derived F_1s are shown in Figure 5.1 and 5.2. Abnormal phenotypes were observed in Li_2 related crosses and generations during both years of the field trials (Figure 5.3), however, images shown in Figure 5.2 are typical Li_2 and its related F_1s phenotypes obtained in the greenhouse experiment. Due to the deformed leaves, stems and stunted plants, the Li_1 phenotype did not survive well under field conditions, therefore, segregation in these populations tended to be biased. Considering the biased segregation of Li_1 associated populations and the abnormal phenotypes detected in Li_2 related populations, data were split into two parts for analysis purpose, which were (1) crosses among FM966, MD17, 177, and 181, and (2) the populations related to Li_1 and Li_2. Subsequently, the performance of Li_2 associated abnormal phenotypes will be reported. To simplify the terminology, fls, N, and F will be used as abbreviations for fiberless i.e. fuzzless-lintless, fuzzless-linted (naked seed), and fuzzy-linted, respectively (Turley and Kloth 2002).

When crosses not related to Li_1 or Li_2, ‘linted’ indicates the seed has the long lint fiber like normal phenotype.

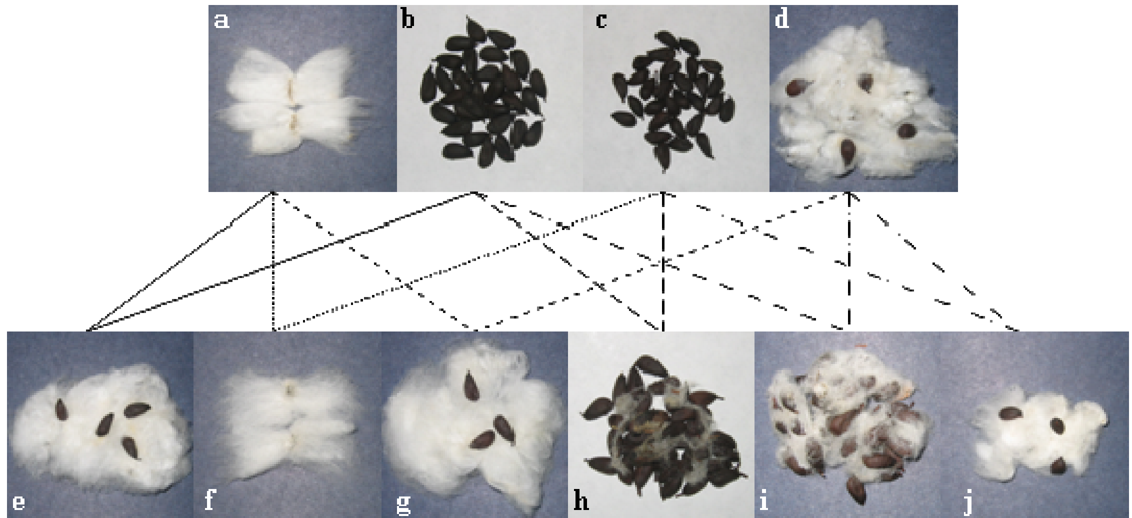

Figure 5.1 Phenotype of Four Parental Lines and Their Half Diallele Crosses Derived F_1.

Seed coat and fiber are maternal tissues, so F_2 seeds showed here (e-j) representing the genotype and phenotype of F_1 plant. (a) fuzzy-linted (F) normal phenotype FM966, (b) fiberless (fls) MD17, (c) fiberless (fls) 177, (d) fuzzless-linted (N) 181, (e) F_1 phenotype of cross MD17 × FM966, (f) F_1 phenotype of cross FM966 × 177, (g) F_1 phenotype of cross FM966 × 181, (h) F_1 phenotype of cross MD17 × 177, (i) F_1 phenotype of cross MD17 × 181, (j) F_1 phenotype of cross 181 × 177.

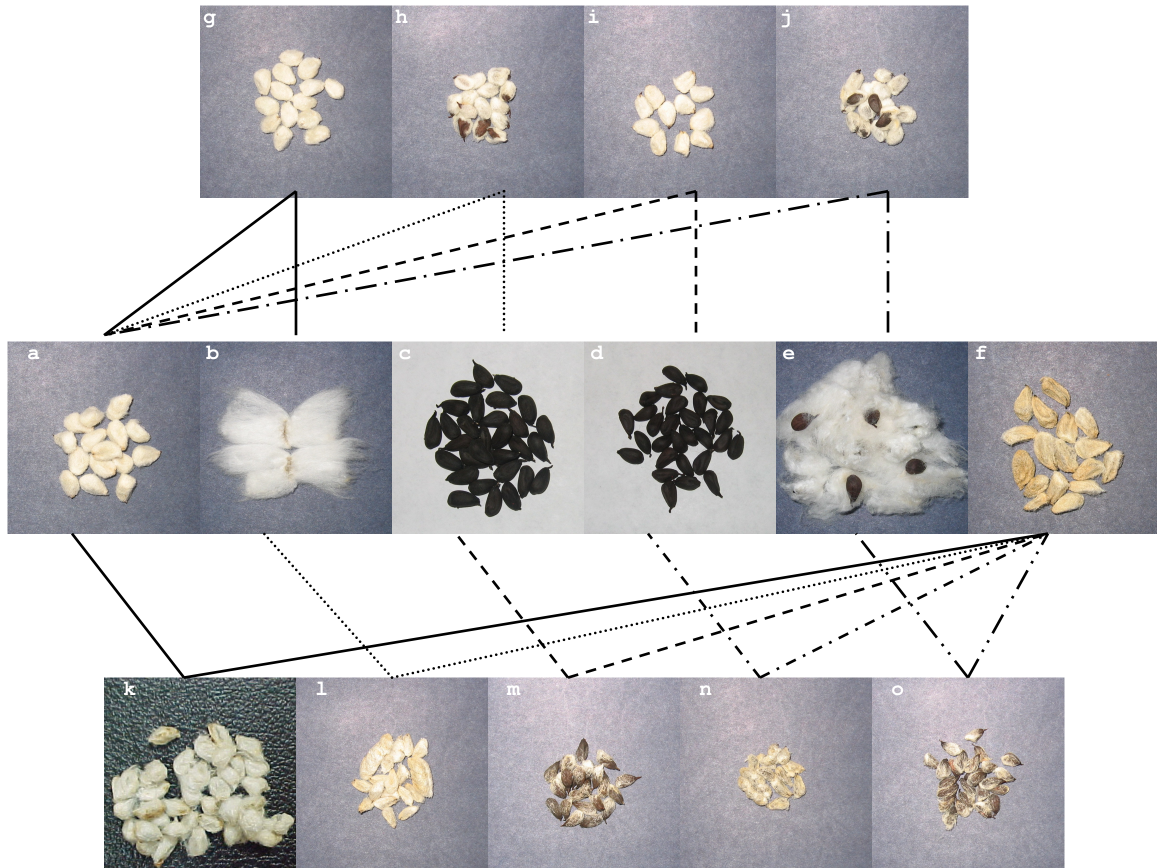

Figure 5.2 Phenotype of Six Parental Lines and Some of the Crosses Derived F_1.

Because seed coat and fiber are maternal tissues, F_2 seeds show here (e-o) represent the genotype and phenotype of F_1 plant. (a) fuzzy-short lint Li-1, (b) fuzzy-linted (F) normal phenotype FM966, (c) fiberless (fls) MD17, (d) fiberless (fls) 177, (e) fuzzless-linted (N) 181, (f) fuzzy-short lint Li-2, (g) F_1 phenotype of cross FM966 × Li-1, (h) F_1 phenotype of cross MD17 × Li-1, (i) F_1 phenotype of cross 177 × Li-1, (j) F_1 phenotype of cross 181 × Li-1, (k) F_1 phenotype of cross Li-2 × Li-1, (l) F_1 phenotype of cross FM966 × Li-2, (m) F_1 phenotype of cross MD17 × Li-2, (n) F_1 phenotype of cross 177 × Li-2, (o) F_1 phenotype of cross 181 × Li-2.

Figure 5.3 Abnormal Phenotypes Observed in *Li_2* and its Related Crosses with the other Parental Lines.

(a-d) the two phenotypes (fuzzy-short lint and fuzzy-long lint) observed in field trials of 2006 and 2007, (e) the two phenotypes F_1 plant of FM966 × *Li_2* still characterized by two phenotypes after transported into greenhouse, (f) the fuzzy-long lint phenotype of *Li_2* observed in field of 2007.

Genetic analysis on FM966, MD17, 177, and 181 derived populations

MD17 is a fiberless germplasm selected from F_2 progeny of a cross between Mexican fuzzless seed UA3-3 (MOVC accession 143) and Ballard fuzzless seed (MOVC accession 243) (Turley 2002). Because accession 243 and accession 143 expresses the dominant N_1N_1 and recessive n_2n_2 naked seed allele, respectively, genotype of MD17 was presumably and originally considered as $N_1N_1n_2n_2$ (Turley 2002). Turley and Kloth (2002) analyzed the fuzzless seed phenotypes using accessions 243, 143, MD17, and commercial cultivar (DP 5690) and found that fuzzless seeds were obtained in n_2n_2 plants only when a second recessive locus was present, which was designated as n_3. There were three alleles responsible for the phenotype of fiberless MD17 instead of two. Therefore, the genotype of MD17 was modified as $N_1N_1n_2n_2n_3n_3$ (Turley and Kloth 2002). However, the genotype of MD17 was challenged again when Turley and Kloth (2008) used MD17 as a check line to verify similarity with another fiberless line SL1-7-1. The genotype of MD17 was revised back to $N_1N_1n_2n_2$ (Turley and Kloth 2008). Both $N_1N_1n_2n_2n_3n_3$ and $N_1N_1n_2n_2$ genotypes were possible for MD17. However, only two homogenous loci (N_1 and n_2) were needed to give the fiberless phenotype (personal communication with Dr. R.B. Turley, USDA-ARS). In other words, the two known loci (N_1 and n_2) were homogenously present in this fiberless line.

MD17 was used as a check line for inheritance designation of other fiber mutant lines together with their pedigree information. Two China germplasm fiber mutant lines, 177 (fls) and 181 (N), used herein for genetic analysis were selected from the cross between XZ142w (fls) and GZNn (N). Even though the final conclusion for the genotype of XZ142w and GZNn has not been reached, it was clear that they harbored recessive

n_2n_2 and dominant N_1N_1 fuzzless allele, respectively (Zhang and Pan 1991; Du et al. 2001; Ding et al. 2007). So, fuzzless allele(s) in 177 and 181 could be either N_1 or n_2. Based on the above background information and segregation data analyses, the observed phenotypic segregation ratios and proposed inheritance models for both F_2 and BC_1 data are given in Table 5.2.

Table 5.2 Segregation of Fiberless (fls), Fuzzless (N), and Fuzzy-linted (F) Phenotypes in Four of the Six Parental Lines

Parents, cross, and generation	Genotype	No. of plant observed			Expected ratio (fls : N : F)	χ^2	*P*
		fls	N	F			
FM966	$n_1n_1N_2N_2N_3N_3Li_3Li_3$	0	0	169	-	-	-
MD17	$N_1N_1n_2n_2n_3n_3Li_3Li_3$	147	0	0	-	-	-
177	$n_1n_1n_2n_2n_3n_3li_3li_3$	183	0	0	-	-	-
181	$N_1N_1N_2N_2n_3n_3Li_3Li_3$	0	193	0	-	-	-
MD17 × FM966							
F_1	$N_1n_1N_2n_2N_3n_3Li_3Li_3$	0	31	0	-	-	-
F_2	$N_1n_1N_2n_2N_3n_3Li_3Li_3 \otimes$	26	185	67	4: 45: 15	4.894	0.087
BC_1 (F_1× FM966)	$N_1n_1N_2n_2N_3n_3Li_3Li_3 \times n_1n_1N_2N_2N_3N_3Li_3Li_3$	0	163	179	0: 1: 1	0.749	0.387
FM966 × 177							
F_1	$n_1n_1N_2n_2N_3n_3Li_3li_3$	0	0	44	-	-	-
F_2	$n_1n_1N_2n_2N_3n_3Li_3li_3 \otimes$	26	38	284	4: 3: 57	31.835	<0.001
BC_1 (F_1 × 177)	$n_1n_1N_2n_2N_3n_3Li_3li_3 \times n_1n_1n_2n_2n_3n_3li_3li_3$	40	30	150	2: 1: 5	5.455	0.065
FM966 × 181							
F_1	$N_1n_1N_2N_2N_3n_3Li_3Li_3$	0	69	0	-	-	-
F_2	$N_1n_1N_2N_2N_3n_3Li_3Li_3 \otimes$	0	244	65	0: 3: 1	2.590	0.108
BC_1 (F_1× FM966)	$N_1n_1N_2N_2N_3n_3Li_3Li_3 \times n_1n_1N_2N_2N_3N_3Li_3Li_3$	0	152	166	0: 1: 1	0.616	0.432
MD17 × 177							
F_1	$N_1n_1n_2n_2n_3n_3Li_3li_3$	0	31	0	-	-	-
F_2	$N_1n_1n_2n_2n_3n_3Li_3li_3$	144	202	0	7: 9: 0	0.639	0.424
BC_1 (F_1 × MD17)	$N_1n_1n_2n_2n_3n_3Li_3li_3 \times N_1N_1n_2n_2n_3n_3Li_3Li_3$	119	123	0	1: 1: 0	0.066	0.797
MD17 × 181							
F_1	$N_1N_1N_2n_2n_3n_3Li_3Li_3$	0	25	0	-	-	-
F_2	$N_1N_1N_2n_2n_3n_3Li_3Li_3 \otimes$	98	238	0	1: 3: 0	3.111	0.078
BC_1 (F_1× MD17)	$N_1N_1N_2n_2n_3n_3Li_3Li_3 \times N_1N_1n_2n_2n_3n_3Li_3Li_3$	142	154	0	1: 1: 0	0.486	0.486
181 × 177							
F_1	$N_1n_1N_2n_2n_3n_3Li_3li_3$	0	82	0	-	-	-
F_2	$N_1n_1N_2n_2n_3n_3Li_3li_3 \otimes$	69	232	66	7: 45: 12	23.494	<0.001
BC_1 (F_1 × 177)	$N_1n_1N_2n_2n_3n_3Li_3li_3 \times n_1n_1n_2n_2n_3n_3li_3li_3$	65	123	78	1: 2: 1	2.774	0.250

The crosses MD17 × FM966, FM966 × 177, and FM966 × 181 were used to initiate the inheritance study of the three fiber mutants, which were then confirmed by backcrossing to one of their parental lines. The F_1 plants from MD17 × FM966 were all fuzzless-linted (N) (Figure 5.1) and pooled segregation comprised 26 fls, 185 N, and 67 F plants (Table 5.2). The χ^2 value for a 4: 45: 15 modified trihybrid ratio was 4.894 with a probability value of 0.087. The fuzzless-linted phenotype of all F_1 plants confirmed the inheritance of the N_1 locus of MD17. Except for the presence of N_1 and n_2 loci, the observed phenotype ratio implied a third locus. Whether it was n_3 (Turley and Kloth 2002), li_3 (Zhang and Pan 1991), or li_4 (Du et al. 2001) could not be determined at this point. Forty-four F_1 plants of cross FM966 × 177 produced fuzzy-linted (F) phenotype seeds and 69 F_1 plants were scored as fuzzless-linted (N) in cross FM966 × 181 (Table 5.2; Figure 5.1). With the knowledge of pedigree information, the inheritance of N_1 and n_2 in 181 and 177, respectively, could be deduced. Due to the same F_1 phenotype when crossing with normal phenotype lines between the two fiberless line (177 and XZ142w) and the pedigree information, it was reasonable to test the previously reported model of XZ142w for 177 (Zhang and Pan 1991; Du et al. 2001). The F_2 population of FM966 × 177 segregated into 26fls: 38N: 284F. This observed ratio fit neither a 1fls: 3N: 12F proposed by Zhang and Pan (1991) nor the more complex model (4fls: 15N: 45F) proposed by Du et al. (2001). The F_2 plants from the cross FM966 × 181 showed a phenotypic segregation of 3N: 1F and the backcross population with FM966 gave a phenotypic segregation of 1N: 1F. Therefore, there was one locus (N_1) or more loci difference between FM966 and 181 and to determine the genotype of the other one must consider other crosses.

To further explore the inheritance of the three fiber mutant lines, the segregation data derived from crosses among them were analyzed. In the cross between MD17 and 181, all F_1 progenies were fuzzless-linted (Figure 5.1). Evidence for the recessive n_2 in MD17 could be obtained from the segregation in F_2 and BC_1 populations, for which the segregation fit 1N: 3F and 1N: 1F, respectively (Table 5.2). It further supported that there was one locus difference between MD17 and 181, which was n_2 in MD17 and N_2 in 181.

The F_1 plants generated from two fiberless lines MD17 and 177 were fuzzless-linted (N) and pooled F_2 segregation comprised 144fls: 202N, which gave a good fit to 7fls: 9N ratio (χ^2=0.639 and P=0.424). The observed phenotype ratio suggested the segregation of two pairs of genes. From the above analyses, it was concluded that both MD17 and 177 had the recessive fuzzless gene n_2 and they were different at N_1 locus (N_1 for MD17 and n_1 for 177).

So, an additional locus between MD17 and 177 was needed to explain the inheritance model. In addition, the presence of fuzzless-linted plants by crossing the two fiberless lines indicated that inheritance models determining the fiberless phenotype of MD17 and 177 was different. Considering the fiberless inheritance model $N_1N_1n_2n_2$ for MD17 (Turley and Kloth 2008), it was necessary to select a fiberless model for 177 in order to find the common connections linking the genotype of different fiberless lines (Zhang and Pan 1991; Du et al. 2001; Turley and Kloth 2008). After study of the F_2 and BC_1 segregation patterns of crosses between 177 (a similar line as XZ142w) and other lines, the $n_2n_2li_3li_3$ model proposed by Zhang and Pan (1991) was selected and used for inheritance analysis (Turley and Kloth 2008). Du et al. (2001) invoked four genes for fiberless XZ142w genetic analysis and proposed the fiberless model as $n_1n_1n_2n_2li_3li_3li_4li_4$.

However, this change ignored the nomenclature and functions of fuzzless seed alleles N_1 and n_2, which had been studied over decades (Kohel 1973; Endrizzi et al. 1984; Samora et al. 1994; Rong et al. 2005). Based on the chosen fiberless model ($n_2n_2li_3li_3$), the second locus difference between MD17 and 177 could be li_3, which defined the fiberless phenotype of 177 together with n_2. Therefore, to this point, the genotype of MD17 and 177 would be $N_1N_1n_2n_2Li_3Li_3$ and $n_1n_1n_2n_2li_3li_3$, respectively, and $n_1n_1N_2N_2Li_3Li_3$ and $N_1N_1N_2N_2Li_3Li_3$ could be designated to FM966 and 181, respectively.

So far, two problems still exist: (1) how many different alleles are related to the inheritance model of fiberless 177 in the cross FM966 × 177, and (2) what exactly was the third different locus between MD17 and FM966, n_3 or li_3? Chi-square tests were performed to assess the segregation results obtained both in F_2 and BC_1 populations between FM966 and 177. It was noted that the segregation among F_2 progenies deviated significantly from the expected 4fls: 3N: 57F ratio, however, the segregation among progenies from backcrosses to 177 fit a 2fls: 1N: 5F ratio (Table 5.2). A similar situation occurred in the crosses between 181 and 177, in which the segregation of the three loci only fit in the backcross population with 177 (Table 5.2). The expression of n_2n_2 had been shown to vary with its background genotype, producing either completely or partially naked seed (Percy and Kohel 1999; Rong et al. 2005). The n_2 locus also might have multiple alleles and might be influenced by modifier genes (Endrizzi and Ray 1991). So, the biased grouping of fuzzy and fuzzless plants led to the deviation from the expected segregating ratios. But, the segregation still indicated that there should be another locus difference between FM966 and 177 after considering n_2 and li_3. Considering the genetic background of MD17, it was reasonable to deduce that this locus

was n_3 in two fiberless lines (MD17 and 177) and N_3 in FM966, which also solved the problem of the third different locus between MD17 and FM966. Consequently, the n_3 allele should also be in the fuzzless-linted line 181. Therefore, the final genotypes would be $n_1n_1N_2N_2N_3N_3Li_3Li_3$, $N_1N_1n_2n_2n_3n_3Li_3Li_3$, $n_1n_1n_2n_2n_3n_3li_3li_3$, and $N_1N_1N_2N_2n_3n_3Li_3Li_3$ for FM966, MD17, 177, and 181, respectively. Detail genetic models are summarized in Table 5.2.

Observations and primary analysis on crosses related to Li_1 and Li_2

In the crosses with normal phenotype commercial cultivar FM966, both F_2 segregations of Li_1 and Li_2 deviated significantly from the single dominant-gene model and P values were calculated as 0.2×10^{-3} and 0.3×10^{-3}, respectively (Table 5.3). Segregation at the Li_1 locus in the backcross population with FM966 also favored the normal phenotype, although the observed ratio did not significantly (P=0.159) deviate from expected 1 fuzzy-short lint: 1 fuzzy-long lint ratio. The main reason for this deviation might be the lower vigor of many homozygous mutants (which had deformed stems and leaves with stunted plants) especially in the drought field condition of 2006 and 2007. This phenomenon was observed by Rong et al. (2005). They reported that a F_2 population of Pima S-7 by Li_1 had 66 plants out of 151 that were characterized as normal phenotype. In the first genetic study on Li_1, Kohel (1972) also reported a variation in goodness-of-fit to the single gene model which was related in part as whether the plants were classified in the greenhouse or in the field. The Li_2 locus at the backcross population with FM966 significantly deviated from the expected 1 fuzzy-short lint: 1 fuzzy-long lint ratio (P=0.0034) (Table 5.3), which probably resulted from the disturbing of the

abnormal phenotypes observed in *Li_2* related generations and populations. Some of plants in the segregating populations might be genetically fuzzy-short lint, but appeared as fuzzy-long lint. Fuzzy-short lint fiber mutants *Li_1* and *Li_2* had similar fiber phenotypes but the two single dominant genes were not allelic (Narbuth and Kohel 1990) and are located on different chromosomes (Karaca et al. 2002; Kohel et al. 2002; Rong et al. 2005). However, the segregation in the F_2 population from crosses between *Li_1* and *Li_2* deviated significantly from the expected 15 fuzzy-short lint: 1 fuzzy-long lint ratio ($P<0.0001$) (Table 5.3), which might have been affected together by lower vigor of *Li_1* plant and the abnormal phenotypes related to *Li_2*. The F_1 plants from crossing *Li_2* with *Li_1* gave the fuzzy-short lint phenotype, which was the same as the two parental lines and difficult to distinguish (Figure 5.2).

Table 5.3 Observed Phenotypic Segregation of Crosses Related to Li_1 and Li_2 and Six Parental Lines

Parents, cross, and Generation	Total plant No.	No. of plant observed				
		Fuzzless-lintless i.e. fiberless (fls)	Fuzzless-short lint	Fuzzless-long lint (N)	Fuzzy-short lint	Fuzzy-long lint (F)
FM966	169	0	0	0	0	169
MD17	147	147	0	0	0	0
177	183	183	0	0	0	0
181	193	0	0	193	0	0
Li_1	138	0	0	0	138	0
Li_2	8[1]	0	0	0	8	0
FM966 × Li_1						
F_1	29	0	0	0	29	0
F_2	158	0	0	0	98	60
BC_1 (F_1 × FM966)	244	0	0	0	111	133
FM966 × Li_2[2]						
F_1	4	0	0	0	4	0
F_2	316	0	0	0	209	107
BC_1 (F_1 × FM966)	127	0	0	0	47	80
MD17 × Li_1						
F_1	17	0	17	0	0	0
F_2	226	14	84	72	31	25
BC_1 (F_1 × MD17)	191	56	53	82	0	0
MD17 × Li_2[2]						
F_1	5	0	5	0	0	0
F_2	231	60	60	62	30	19
BC_1 (F_1 × MD17)	274	111	35	128	0	0
177 × Li_1						
F_1	20	0	0	0	20	0

Table 5.3 (continued)

F_2	300	11	72	17	128	72
BC_1 ($F_1 \times 177$)	251	66	42	48	30	65
177 × Li_2[2]						
F_1	7	0	0	0	7	0
F_2	289	11	83	10	110	75
BC_1 ($F_1 \times 177$)	169	23	44	46	13	43
181 × Li_1						
F_1	54	0	54	0	0	0
F_2	196	0	108	47	25	16
BC_1 ($F_1 \times 181$)	228	0	78	150	0	0
181 × Li_2[2]						
F_1	6	0	6	0	0	0
F_2	275	0	128	66	47	34
BC_1 ($F_1 \times 181$)	200	0	63	137	0	0
$Li_2 \times Li_1$[2,3]						
F_1	4	0	0	0	4	0
F_2	211	0	0	0	180	31

[1]The number of Li_2 plant is the original plant individuals for crossing with the other lines to derive F_1 seeds in the greenhouse in the summer of 2005. The other lines' plant number is the combined number of field trials in 2006 and 2007.
[2]It is the original F_1 plant number to derive F_2 seeds by selfing in greenhouse of 2005 winter. F_1 plants from the field trial of 2006 are not included because of the observed abnormal phenotype of the F_1s. Though backcrosses were made in field of 2006, only seed produced from typical F_1 plants were selected and grown for the BC_1 populations. The typical Li_2 plants were also selfed in greenhouse of 2005 summer to produce seeds for field trials of 2006 and 2007. So, the original Li_2 and F_1 plants for generating selfing seed, F_2 and BC_1 populations are typical ones i.e. fuzzy-short lint. A Few of two-phenotype plants were also observed in F_2 and BC_1 populations. However, phenotypic classification herein was based on the phenotype of the majority of open bolls on the plant at the end of the season. The two-phenotype observations are summarized in Table 5.4.
[3]No backcross population available.

Three other fiber mutant lines (MD17, 177, and 181) were also crossed with Li_1 or Li_2. Because of harboring the dominant fuzzless gene N_1 in MD17 and 181, F_1 plants from crossing MD17 or 181 with Li_1 or Li_2 were characterized with fuzzless-short lint (Figure 5.2). However, the F_1 phenotype of both crosses 177 × Li_1 and 177 × Li_2 was fuzzy-short lint (Table 5.3), which further indicated the recessive fuzzless character of 177. The short lint could be very easily twisted from the fuzzless-short lint seed by two fingers, whereas twisting short lint from the fuzzy-short lint seed was very difficult and this could only be accomplished with the aid of a knife. Although the final inheritance model of these mutants could not be reached due to biased phenotype segregation caused by Li_1, Li_2, and 177, segregating data in F_2 and BC_1 populations (Table 5.3) might still provide useful information: (1) Dominant genes Li_1 and Li_2 were factors that activate during fiber elongation instead of initiation. In all the F_2 and BC_1 populations from crosses of MD17, 177, and 181 with Li_1 or Li_2, fuzzless-short lint phenotype could be identified. These plants had lint fiber initiation but the late elongation was arrested by the Li_1 or Li_2 locus. In the original reports on the identification, characterization, and genetic analysis of these two fiber mutants, even the '*Li*' was the abbreviation of '*Ligon lintless'*, they were regarded as short lint instead of lintless (Kohel 1972, 1992; Narbuth and Kohel 1990). The short lint feature of these two mutants was also confirmed by microscopy analysis on Li_1 (Karaca et al. 2002) and fiber development studies on both of them (Kohel et al. 1992). However, Li_1 and Li_2 sometimes were misrepresented as lintless instead of short lint in communications. (2) Fuzzless genes (N_1, n_2, or n_3) combining with other unidentified loci such as li_3 and fl_1 would lead to different fiberless genetic models. However, Li_1 or Li_2 mutated loci had effect only on lint fiber elongation and Li_1 also

influenced vegetative development. So, there could be two different mechanisms for cotton fiber initiation and elongation even though there might be some overlapping of gene function in both processes. Li_1 and Li_2 were independent inherited from fiber initiation genes. Segregating populations (F_2 and BC_1) derived from two fiberless lines (MD17 and 177) crossed with Li_1 or Li_2 had all five possible phenotype categories: fiberless (fls), fuzzless-short lint, fuzzless-long lint (N), fuzzy-short lint, and fuzzy-long lint (F) (Table 5.3). Comparing the two phenotypic categories (N and F) in segregating F_2 and BC_1 of cross FM966 × 181, the cross between 181 and Li_1 or Li_2 had four types, which resulted from splitting N and F phenotypes into short and long lint fiber, respectively (Table 5.3).

The abnormal phenotypes observed in Li_2 related crosses and generations

The original seed sources for S_1 and F_1s were increased in the greenhouse. The typical phenotype of these seeds was verified through several generations or different plants. However, the S_1 and F_1s phenotype in the field 'segregated' abnormally into typical phenotype, two-phenotype on same plant, and fuzzy/fuzzless-long lint in two years' field trials (Figure 5.3). Specifically, the typical phenotype for S_1 of Li_2 and F_1s of crosses FM966 × Li_2, 177 × Li_2, and Li_2 × Li_1 was fuzzy-short lint, but fuzzy-long lint individual plants were found. More surprising, some plant had two-phenotype i.e. some branches/bolls were fuzzy-short lint and others were fuzzy-long lint. Similar situation happened in F_1s of crosses MD17 × Li_2 and 181 × Li_2: the typical phenotype was fuzzless-short lint, but there were also fuzzless-long lint plants and two-phenotype (some branches/bolls were fuzzless-short lint and others were fuzzless-long lint) plants detected

in the populations (Table 5.4). For the segregation populations (BC_1s and F_2s), a few of two-phenotype plants were also identified (Table 5.4). Because this phenomenon happened in four generations (S_1, F_1, F_2, and BC_1) of all Li_2 related entries, it could not be the activation of some transposed elements by crossing. Otherwise, the S_1 generation should be stable. In addition, the transposed elements hypothesis also could not explain the two-phenotype on same plant. Similarly, fiberless mutant line SL1-7-1 identified and repeatedly characterized as the fiberless phenoytpe in US (Turley and Ferguson 1996; Ruan and Choury 1998; Turley and Kloth 2008) produced little lint at maturity in Australia (Wu et al. 2006). Another fiberless line 177 used in this study was characterized as producing a small amount of lint on some seeds under greenhouse condition of Mississippi State, though in its origin place (China) and the field condition of Mississippi State it was a profoundly fiberless line as showed in Figure 5.1 and 5.2. Personal communication with Drs. R.B Turley (USDA-ARS) and X.M. Du (CRI-CAAS) also showed their finding of different phenotypes on one fiber mutant plant in the field. But this finding does not appear in print. This evidence indicates the fiber mutants' phenotype expression is affected by factors such as environment conditions in addition to genetics. Analysis on the pooled two years data on S_1s of Li_2 and five F_1s grown in 2006 indicated a varied percentage of penetrance, which was from 50.4% for Li_2 to 100% for the F_1s of $Li_2 \times Li_1$ (Table 5.4).

Table 5.4 Abnormal Phenotypes Observed in *Li*$_2$ Associated Crosses and Generations[1]

Parents, cross, and generation	Total plant No.	Typical Phenotype[2]	Two-phenotype	Fuzzy/fuzzless-long lint[3]
Li$_2$	125	63	28	34
FM966 × *Li*$_2$				
F_1	19	11	8	0
F_2	316	-	4	-
BC_1 (F_1 × FM966)	127	-	6	-
MD17 × *Li*$_2$				
F_1	13	8	4	1
F_2	231	-	4	-
BC_1 (F_1 × MD17)	274	-	1	-
177 × *Li*$_2$				
F_1	30	20	8	2
F_2	289	-	5	-
BC_1 (F_1 × 177)	169	-	3	-
181 × *Li*$_2$				
F_1	48	29	8	11
F_2	275	-	3	-
BC_1 (F_1 × 181)	200	-	5	-
Li$_2$ × *Li*$_1$[4]				
F_1	8	8	0	0
F_2	211	-	13	-

[1]In segregating populations (F_2 and BC_1), total plant number means the individual plant number in the entire population. Here, only two-phenotype could be identified and considered as abnormal phenotype. So, only this category is summarized.
[2]The typical phenotype is fuzzy-short lint for crosses FM966 × *Li*$_2$, 177 × *Li*$_2$, and *Li*$_2$ × *Li*$_1$; and fuzzless-short lint for MD17 × *Li*$_2$ and 181 × *Li*$_2$.
[3]Fuzzy-long lint for crosses FM966 × *Li*$_2$, 177 × *Li*$_2$, and *Li*$_2$ × *Li*$_1$; and fuzzless-long lint for MD17 × *Li*$_2$ and 181 × *Li*$_2$.
[4]No backcross population was available.

Detection of the abnormal phenotype plants' performance after transplanting to the greenhouse showed that one survived F_1 plant of FM966 × Li_2 still produced two phenotypes on different branches (Figure 5.3). It showed the similar events of the two fiber phenotype under two different environments (field/greenhouse). However, it does not provide enough evidence to claim that the two-phenotype was stable. The next generation field confirmation experiment on the fuzzy-long lint and two-phenotype Li_2 showed contrasting results. Both fuzzy-short lint and fuzzy-long lint seeds from the three two-phenotype Li_2 plants produced three types of seeds (Table 5.5). Seeds from one fuzzy-long lint Li_2 plant produced different number of all three types of abnormal phenotypes (Table 5.5). Therefore, the fuzzy-long lint character could switch back to the typical fuzzy-short lint phenotype in the next generation, which also indicated the instability of the expression of the fiber mutant.

Table 5.5 The Next Generation Segregation from the Abnormal Li_2 Plant of 2006 Field

Seed source	Total plant No.	Fuzzy-short lint	Fuzzy-long lint	Two-pheno type
Fuzzy-short lint branch of plant 1	39	11	13	15
Fuzzy-long lint branch of plant 1	34	13	14	7
Fuzzy-short lint branch of plant 2	42	21	14	7
Fuzzy-long lint branch of plant 2	11	6	3	2
Fuzzy-short lint branch of plant 3	18	4	9	5
Fuzzy-long lint branch of plant 3	17	5	7	5
Fuzzy-long lint Li_2	50	21	19	10

In summary, genetic models proposed by Zhang and Pan (1991) and Turley and Kloth (2002, 2008) were combined to analyze the inheritance pattern of fuzz and lint fiber initiation using mutant lines indentified from both US and China germplasm. Result indicated that both genetic models ($N_1N_1n_2n_2$ and $n_2n_2li_3li_3$) could lead to the fiberless phenotype. The observation of fuzzless-short lint phenotype in crosses between Li_1 or Li_2 and three other fiber mutant lines not only provided evidence that the original phenotype of Li_1 and Li_2 were fuzzy-short lint but also indicated different mechanisms for cotton fiber initiation and elongation at the conventional genetics level. The abnormal phenotypes detected in Li_2 related generations represented the report on the penetrance of fiber mutant gene. Future investigation will help to detect the true nature of the penetrance in cotton.

CHAPTER VI

CONCLUSIONS

SNP marker development strategy

Efficient SNP discovery in polyploids, such as cotton, must address the problem with appropriate methods that can distinguish between genome-specific polymorphisms (GSPs) and locus-specific polymorphisms (LSPs). In this study, the possibility of identifying false SNP were reduced by applying the following approaches: 1) PCR primers were designed from well characterized genes to generate an amplicon pool from each genotype; 2) multiple clones were sequenced to avoid random error of sequencing and to ensure getting the duplicated loci of the gene; 3) putative locus wasidentificatied by phylogenetic clustering and compared to the two progenitor diploid genome species of allotetraploid cottons; 4) locus/gene-specific PCR and SNP primer were designed for SNP marker genotyping of *EXPANSIN A* and *MYB* gene families based on sequence difference within one clad; and 5) confirmed the true nature of the SNP markers with genetic test by detecting its chromosomal location using deletion analysis.

SNP characterication on *EXPANSIN A* and *MYB* gene families

A total of 222 SNPs, including 120 single-base changes and 102 indels, were identified from six *EXPANSIN A* genes. Transitions accounted for 69 (57.5%) and

transversions for 51 (42.5%) of the total 120 single-base changes. Analysis of indel sequences indicated a bias toward ‘A’ and ‘T’ nucleotides. The average rate of SNPs per nucleotide was 2.35%, with 1.74% and 3.99% occurring in coding regions and noncoding regions, respectively. Based on the average rate of SNP per nucleotide, a higher nucleotide change rate was discovered in the D-genome (2.90%) than that of the A-genome (1.98%). Results revealed that a total of 56 out of the 119 cSNPs (47%) were nonsynonymous changes. Haplotype number ranged from two to six out of the maximum seven available genotype sequences, with an estimated haplotype diversity that varied between 0.33 and 1. The independent and incongruent evolution of the two subgenomes (At and Dt) was also revealed by the different phylogenetic topologies detected in polyploid duplicated *EXPANSIN A* genes or fragments of the selected tetraploid species.

From 8,301 bp aligned sequences of six *MYB* genes, 108 SNPs were detected from both A- and D-genomes, giving an average SNP frequency of one SNP for every 77 bases (1.30%). Results showed the presence of one SNP per 106 bp (0.94%) in the coding regions and one SNP per 30 bp (3.33%) in the noncoding regions. The SNP distribution varied among the six examined genes. The highest rate of SNP occurrence was observed in *MYB6* (one SNP every 34 bp) and the lowest rate of SNP frequency was present in *MYB3* (one SNP every 260 bp). Transitions (‘A/G’ or ‘C/T’) were the most common cause of sequences variation in the selected cotton genotypes (49%) compared to transversions (‘A/T’, ‘G/C’, ‘A/C’ or ‘G/T’, 26%) and indels (25%). A significant bias to ‘T’ insertion/deletion was detected in the overall sequences (59.30%). In coding regions of six *MYB* genes, 41 out of 67 cSNPs (SNPs in coding region) sites were predicted to result in amino acid changes. Sequence polymorphism defined haplotypes number ranged

from two to seven among the seven selected cotton genotypes and haplotype diversity varied from 0.286±0.196 to 1.000±0.076 among six *MYB* genes. The lowest nucleotide diversities occurred among the three *G. hirsutum* lines in both A- and D-genomes. Results from both A- and D-genomes showed the highest nucleotide diversities were between *G. mustelinum* and the extant species of the ancestral genome donors. Nucleotide diversities of *MYB* genes were higher in the D-Dt comparisons than for the A-At comparison of the allotetraploid cotton species. Nucleotide diversities among the three *G. hirsutum* lines in the Dt-genome were higher than that in the At-genome, indicating six *MYB* genes loci in Upland cotton Dt-genome exhibited a faster evolutionary rate than the At-genome. Most of the substitution ratios (K_a:K_{sil}) of pairwise comparisons were less than 1, indicating a high level of evolutionary constraint placed on amino acid substitution in six *MYB* genes.

EXPANSIN A and *MYB* genes chromosomal locations in tetraploid cotton

Gene chromosome localization was accomplished by SNP marker-based deletion analysis or linkage mapping with framework SSR marker. Six *EXPANSIN A* genes were assigned to the long arms of chromosome 20 (*EXPANSIN A1* and *EXPANSIN A2*), 10 (*EXPANSIN A2* and *EXPANSIN A3*), 9 (*EXPANSIN A4*), 1 (*EXPANSIN A5*), and 3 (*EXPANSIN A6*). Among them, only homoeologus loci (chromosome 10 and 20) for gene *EXPANSIN A2* were identified based on available SNP markers and sequence information. Six *MYB* genes chromosomal assignment results showed that homoeologous loci for *MYB1*, *MYB4*, and *MYB6* were on three pairs of homoeologous chromosomes: 13 and 18, 7 and 16, and 11 and 21, respectively. A-genome loci for gene *MYB2* and *MYB5*

were assigned on the short arm of chromosome 8 and 11, respectively. The possible chromosome location for *MYB3* is on chromosome 14 (long arm), 15 (long arm), 19, 21, 23, or 24. The general agreement between these fiber development-related genes and QTL for fiber yield and quality traits might also indicate the potential of these SNP markers for tagging important traits in cotton.

QTL for yield components, seed, and fiber traits

Two F_2 populations developed from crossing MD17, a fuzzless-lintless genetic stock containing three fuzzless loci, N_1, n_2 and a postulated n_3, with line 181 (fuzzless-linted) and with FM966, a fuzzy-linted cultivar, were used for molecular mapping by polymorphic SSR markers. Major QTL which explain 68.3 (population with FM 966) to 87.1% (population with 181) of the phenotypic variation for lint percentage and 62.8% (Population 181) for lint index were detected in the vicinity of BNL3482-138 on chromosome 26. Single marker regression analyses indicated STV79-108, which was located to the long arm of chromosome 12 (the known location of N_1 and perhaps n_2 loci), also had significant association with lint percentage (R^2 26.7), lint index (R^2 30.6), embryo protein percentage (R^2 15.4) and micronaire (R^2 20.0). Additional QTL and significant markers associated with other seed and fiber traits were detected on different chromosome locations and explained large percentages of the phenotypic variation. Two-locus epistatic interactions were also observed. Results from this research will facilitate further understanding the complex network of cotton fiber development and seeds traits.

Inheritance of lint and fuzz fiber development

Segregating populations (F_2 and BC_1) derived from half diallele crosses among two fiberless (MD17 and 177), two fuzzy-short lint (Li_1 and Li_2), one fuzzless-linted (181), and one commercial cultivar (FM966) were employed to evaluate the inheritance pattern of fuzz and lint fiber development. Inheritance analysis indicated that both genetic models of $N_1N_1n_2n_2$ and $n_2n_2li_3li_3$ could lead to the fiberless phenotype. The observation of fuzzless-short lint phenotype in crosses between Li_1 or Li_2 and the other three fiber mutant lines provided evidences that the original phenotype of Li_1 and Li_2 was fuzzy-short lint and cotton fiber initiation and elongation were controlled by different mechanisms. The three categories of abnormal phenotypes (typical penotype, two-phenotype on one plant, and fuzzy/fuzzless-long lint) detected in Li_2 related generations represented the report on the penetrance of fiber mutant gene expression.

REFERENCES

Abdalla AM, Reddy OUK, Pepper AE (2001) Genetic diversity and relationships of diploid and tetraploid cottons using AFLP. Theor Appl Genet 102:222-229

Abdurakhmonov IY, Buriev ZT, Saha S, Pepper AE, Musaev JA, Almatov A, Shermatov SE, Kushanov FN, Mavlonov GT, Reddy UK, Yu JZ, Jenkins JN, Kohel RJ, Abdukarimov A (2007) Microsatellite markers associated with lint percentage trait in cotton, *Gossypium hirsutum*. Euphytica 156:141-156

Alvarez I, Cronn R, Wendel JF (2005) Phylogeny of the New World diploid cottons (*Gossypium* L., Malvaceae) based on sequences of three low-copy nuclear genes. Plant Syst Evol 252:199-214

An C, Saha S, Jenkins JN, Ma DP, Scheffler BE, Kohell RJ, Yu JZ, Stelly DM (2008) Cotton (*Gossypium* spp.) R2R3-MYB transcription factors SNP identification, phylogenomic characterization, chromosome localization, and linkage mapping. Theor Appl Genet DOI 10.1007/s00122-008-0732-4

An C, Saha S, Jenkins JN, Scheffler BE, Wilkins TA, Stelly DM (2007) Transcriptome profiling, sequence characterization, and SNP-based chromosomal assignment of the *EXPANSIN* genes in cotton. Mol Gen Genomics 278:539-553

Arieli A (1998) Whole cottonseed in dairy cattle feeding: a review. Anim Feed Sci Tech 72:97-110

Arpat AB, Waugh M, Sullivan JP, Gonzales M, Frisch D, Main D, Wood T, Leslie A, Wing RA, Wilkins TA (2004) Functional genomics of cell elongation in developing cotton fibers. Plant Mol Biol 54:911-929

Batley J, Barker G, O'Sullivan H, Edwards KJ, Edwards D (2003) Mining for single nucleotide polymorphisms and insertions/deletions in maize expressed sequence tag data. Plant Physiol 132:84-91

Beecher B, Bowman J, Martin JM, Bettge AD, Morris CF, Blake TK, Giroux MJ (2002) Hordoindolines are associated with a major endosperm-texture QTL in barley (*Hordeum vulgare*). Genome 45:584-591

Blenda A, Scheffler J, Scheffler B, Palmer M, Lacape JM, Yu JZ, Jesudurai C, Jung S, Muthukumar S, Yullambalase P, Ficklin S, Staton M, Eshelman R, Ulloa M, Saha S, Burr B, Liu S, Zhang T, Fang D, Pepper A, Kumpatla S, Jacobs J, Tomkins J, Cantrell R, Main D (2006) CMD: a cotton microsatellite database resource for *Gossypium* genomics. BMC Genomics 7:132

Brookes AJ (1999) The essence of SNPs. Gene 234:177-186

Brown MS, Menzel MY, Hasenkampf CA, Naqi S (1981) Chromosome configurations and orientations in 58 heterozygous translocations in *Gossypium hirsutum*. J Hered 72:161-168

Brubaker CL, Paterson AH, Wendel JF (1999) Comparative genetic mapping of allotetraploid cotton and its diploid progenitors. Genome 42:184-203

Buchner P, Borisjuk L, Wobus U (1996) Glucan phosphorylases in *Vicia faba* L.: cloning, structural analysis and expression patterns of cytosolic and plastidic forms in relation to starch. Planta 199:64-73

Bundock PC, Christopher JT, Eggler P, Ablett G, Henry RJ, Holton TA (2003) Single nucleotide polymorphisms in cytochrome P450 genes from barley. Theor Appl Genet 106:676-682

Bundock PC, Henry RJ (2004) Single nucleotide polymorphism, haplotype diversity and recombination in the *Isa* gene of barley. Theor Appl Genet 109:543-551

Caderas D, Muster M, Vogler H, Mandel T, Rose JKC, McQueen-Mason S, Kuhlemeier C (2000) Limited correlation between *EXPANSIN* gene expression and elongation growth rate. Plant Physiol 123:1399-1413

Caldwell KS, Dvorak J, Lagudah ES, Akhunov E, Luo MC, Wolters P, Powell W (2004) Sequence polymorphism in polyploid wheat and their D-genome diploid ancestor. Genetics 167:941-947

Cedroni ML, Cronn RC, Adams KL, Wilkins TA, Wendel JF (2003) Evolution and expression of *MYB* genes in diploid and polyploid cotton. Plant Mol Biol 51:313-325

Chee PW, Draye X, Jiang CX, Decanini L, Delmonte TA, Bredhauer R, Smith CW, Paterson AH (2005a) Molecular dissection of interspecific variation between *Gossypium hirsutum* and *Gossypium barbadense* (cotton) by a backcross-self approach: I. fiber elongation. Theor Appl Genet 111:757-763

Chee PW, Draye X, Jiang CX, Decanini L, Delmonte TA, Bredhauer R, Smith CW, Paterson AH (2005b) Molecular dissection of phenotypic variation between

Gossypium hirsutum and *Gossypium barbadense* (cotton) by a backcross-self approach: III. Fiber length. Theor Appl Genet 111:772-781

Chee PW, Rong J, Williams-Coplin D, Schulze SR, Paterson AH (2004) EST derived PCR-based markers for functional gene homologues in cotton. Genome 47:449-462

Ching A, Caldwell KS, Jung M, Dolan M, Smith OS, Tingey S, Morgante M, Rafalski AJ (2002) SNP frequency, haplotype structure and linkage disequilibrium in elite maize inbred lines. BMC Genet 3:19

Cho RJ, Mindrinos M, Richards DR, Sapolsky RJ, Anderson M, Drenkard E, Dewdney J, Reuber TL, Stammers M, Federspiel N, Theologis A, Yang WH, Hubbell E, Au M, Chung EY, Lashkari D, Lemieux B, Dean C, Lipshutz RJ, Ausubel FM, Davis RW, Oefner PJ (1999) Genome-wide mapping with biallelic markers in *Arabidopsis thaliana*. Nat Genet 23:203 – 207

Cronn R, Cedroni M, Haselkorn T, Grover C, Wendel JF (2002) PCR-mediated recombination in amplification products derived from polyploid cotton. Theor Appl Genet 104:482-489

Cronn R, Small RL, Wendel JF (1999) Duplicated genes evolve independently after polyploid formation in cotton. Proc Natl Acad Sci USA 96:14406-14411

Dani RG, Kohel RJ (1989) Maternal effects and generation mean analysis of seed-oil content in cotton (*Gossypium hirsutum* L.). Theor Appl Genet 77:569-575

Darley CP, Forrester AM, McQueen-Mason SJ (2001) The molecular basis of plant cell wall expansion. Plant Mol Biol 47:179-195

Ding YZ, Guo WZ, Zhang TZ (2007) Genetic analysis of two mutants of fiber in *Gossypium hirsutum* L. Cotton Sci 19:179-182

Draye X, Chee PW, Jiang CX, Decanini L, Delmonte TA, Bredhauer R, Smith CW, Paterson AH (2005) Molecular dissection of interspecific variation between *Gossypium hirsutum* and *G. barbadense* (cotton) by a backcross-self approach: II. Fiber fineness. Theor Appl Genet 111:764-771

Du XM, Pan JJ, Wang RH, Zhang TZ, Shi YZ (2001) Genetic analysis of presence and absence of lint and fuzz in cotton. Plant Breed 120:519-522

Endrizzi JE, Ramsay G (1980) Identification of ten chromosome deficiencies in cotton. J Hered 71:45-48

Endrizzi JE, Ray DT (1991) Monosomic and monotelodisomic analysis of 34 mutant loci in cotton. J Hered 82:53-57

Endrizzi JE, Turcotte EL, Kohel RJ (1984) Qualitative genetics, cytology, and cytogenetics. In: Kohel R.J. and Lewis D.F. (eds) Cotton. Agronomy Soc. of America, Madison, WI, p 81-129

Faris JD, Li WL, Liu DJ, Chen PD, Gill BS (1999) Candidate gene analysis of quantitative disease resistance in wheat. Theor Appl Genet 98:219-225

Feltus FA, Wan J, Schulze SR, Estill JC, Jiang N, Paterson AH (2004) An SNP resource for rice genetics and breeding based on subspecies India and Japonica genome alignments. Genome Res 14:1812-1819

Frelichowski JE, Palmer MB, Main D, Tomkins JP, Cantrell RG, Stelly DM, Yu J, Kohel RJ, Ulloa M (2006) Cotton genome mapping with new microsatellites from Acala 'Maxxa' BAC-ends. Mol Gen Genomics 275:479-491

Giroux MJ, Talbert L, Habernicht DK, Lanning S, Hemphill A, Martin JM (2000) Association of puroindoline sequence type and grain hardiness in hard red spring wheat. Crop Sci 40:370-380

Glazier AM, Nadeau JH, Aitman TJ (2002) Finding genes that underlie complex traits. Science 298:2345-2349

Glonek G, Solomon PJ (2004) Factorial and time course designs for cDNA microarray experiments. Biostatistics 5:89-111

Glover BJ, Perez-Rodriguez M, Martin C (1998) Development of several epidermal cell types can be specified by the same MYB-related plant transcription factor. Development 125:3497-3508

Grover CE, Kim H, Wing RA, Paterson AH, Wendel JF (2004) Incongruent patterns of local and global genome size evolution in cotton. Genome Res 14:1474-1482

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

Guo WZ, Ma GJ, Zhu YC, Yi CX, Zhang TZ (2006) Molecular tagging and mapping of quantitative trait loci for lint percentage and morphological marker genes in Upland cotton. J Integr Plant Biol 48:320-326

Guo Y, McCarty JC, Jenkins JN, Saha S (2007a) QTL for node of first fruiting branch in a cross of an upland cotton, *Gossypium hirsutum* L., cultivar with primitive accession Texas 701 (in press). Euphytica DOI 10.1007/s10681-007-9613-1

Guo Y, Saha S, Yu JZ, Jenkins JN, Kohel RJ, Scheffler BE, Stelly DM (2007b) BAC-derived SSR markers chromosome locations in cotton (in press). Euphytica DOI 10.1007/s10681-007-9585-1

Gutierrez OA, Basu S, Saha S, Jenkins JN, Shoemaker DB, Cheatham CL, McCarty JC (2002) Genetics distance among selected cotton genotypes and its relationship with F_2 performance. Crop Sci 42:1841-1847

Han Z, Wang C, Song X, Guo W, Gou J, Li C, Chen X, Zhang T (2006) Characteristics, development and mapping of *Gossypium hirsutum* derived EST-SSR in allotetraploid cotton. Theor Appl Genet 112:430-439

Han ZG, Guo WZ, Song XL, Zhang TZ (2004) Genetic mapping of EST-derived microsatellites from the diploid G*ossypium arboreum* in allotetraploid cotton. Mol Gen Genomics 272:308-327

Hanson WD (1959) Minimum family size for planning genetic experiments. Agronomy J 51:711-715

Harmer SE, Orford SJ, Timmis JN (2002) Characterisation of six *α-expansin* genes in *Gossypium hirsutum* (Upland cotton). Mol Genet Genomics 268:1-9

Hauser MT, Harr B, Schlotterer C (2001) Trichome distribution in *Arabidopsis thaliana* and its close relative *Arabidopsis lyrata*: Molecular analysis of the candidate gene *GLABROUS1*. Mol Biol Evol 18:1754-1763

He DH, Lin ZX, Zhang XL, Nie YC, Guo XP, Feng CD, Stewart JMcD (2005) Mapping QTLs of traits contributing to yield and analysis of genetic effects in tetraploid cotton. Euphytica 144:141-149

He DH, Lin ZX, Zhang XL, Nie YC, Guo XP, Zhang YX, Li W (2007) QTL mapping for economic traits based on a dense genetic map of cotton with PCR-based markers using the interspecific cross of *Gossypium hirsutum* × *Gossypium barbadense*. Euphytica 153:181-197

Holland JB, Helland SJ, Sharopova N, Rhyne DC (2001) Polymorphism of PCR-based markers targeting exons, introns, promoter regions, and SSRs in maize and intron and repeat sequences in oat. Genome 44:1065-1076

Hsu CY, An C, Saha S, Ma DP, Jenkins JN, Scheffler BE, Stelly DM (2008) Molecular and SNP characterization of two genome specific transcription factor genes *GhMyb8* and *GhMyb10* in cotton species. Euphytica 159:259-273

Hsu CY, Jenkins JN, Saha S, Ma DP (2005) Transcriptional regulation of the lipid transfer protein gene *LTP3* cotton fiber by a novel MYB protein. Plant Sci 168:167-181

Humphries JA, Walker AR, Timmis JN, Orford SJ (2005) Two WD-repeat genes from cotton are functional homologues of the *Arabidopsis thaliana TRANSPARENT TESTA GLABRA1* (*TTG1*) gene. Plant Mol Biol 57:67-81

Jander G, Norris SR, Rounsley SD, Bush DF, Levin IM, Last RL (2002) Arabidopsis map-based cloning in the post-genome era. Plant Physiol 129:440-450

Jenkins JN, McCarty JC, Wu J, Saha S, Gutierrez O, Hayes R, Stelly DM (2007) Genetic effect of thirteen *Gossypium barbadense* L. chromosome substitution line in topcrosses with Upland cotton cultivars: II. Fiber quality traits. Crop Sci 47:561-570

Ji SJ, Lu YC, Feng JX, Wei G, Li J, Shi YH, Fu Q, Liu D, Luo JC, Zhu YX (2003) Isolation and analysis of genes preferentially expressed during early cotton fiber development by subtractive PCR and cDNA array. Nucleic Acids Res 31:2534-543

Jiang C, Wright RJ, Woo SS, DelMonte TA, Paterson AH (2000) QTL analysis of leaf morphology in tetraploid *Gossypium* (cotton). Theor Appl Genet 100:409-418

Jiang CX, Wright RJ, El-Zik KM, Paterson AH (1998) Polyploid formation created unique avenues for response to selection in *Gossypium* (cotton). Proc Natl Acad Sci USA 95:449-4424

Kamiya T, Kawabe A, Miyashita NT (2002) Nucleotide polymorphism at *Atmyb2* locus of the wild plant *Arabidopsis thaliana*. Genet Res Camb 80:89-98

Kanazin V, Talbert H, See D, DeCamp P, Nevo E, Blake T (2002) Discovery and assay of single-nucleotide polymorphisms in barley (*Hordeum vulgare*). Plant Mol Biol 48:529-537

Karaca M, Saha S, Jenkins JN, Zipf A, Kohel RJ, Stelly DM (2002) Simple sequence repeat (SSR) markers linked to the *Ligon Lintless* (Li_1) mutant in cotton. J Hered 93:221-224

Kearney TH, Harrison GJ (1927) Inheritance of smooth seeds in cotton. J Agri Res 35:193-217

Kende H, Bradford KJ, Brummell DA, Cho H-T, Cosgrove DJ, Fleming AJ, Gehring C, Lee Y, McQueen-Mason SJ, Rose JKC, Voesenek LACJ (2004) Nomenclature for members of the expansin superfamily of genes and proteins. Plant Mol Biol 55:1105-1110

Kerr MK, Churchill GA (2001) Bootstrapping cluster analysis: assessing the reliability of conclusions from microarray experiments. Proc Natl Acad Sci USA 98:8961-8965

Kim HJ, Triplett BA (2001) Cotton fiber growth in planta and in vitro: models for plant cell elongation and cell wall biogenensis. Plant Physiol 127:1361-1366

Kim MY, Van K, Lestart P, Moon JK, Lee SH (2005) SNP identification and SNAP marker development for a *GmNARK* gene controlling supernodulation in soybean. Theor Appl Genet 110:1003-1010

Kohel RJ (1972) Linkage tests in Upland cotton, *Gossypium hirsutum* L. II.. Crop Sci 12:66-69

Kohel RJ (1973) Genetic nomenclature in cotton. J Hered 64:291-295

Kohel RJ, Narbuth EV, Benedict CR (1992) Fiber development of ligon lintless-2 mutant of cotton. Crop Sci 32:733-735

Kohel RJ, Stelly DM, Yu J (2002) Tests of six cotton (*Gossypium hirsutum* L.) mutants for association with aneuploids. J Hered 93:130-132

Kohel RJ, Yu J, Park YH, Lazo GR (2001) Molecular mapping and characterization of traits controlling fiber quality in cotton. Euphytica 121:163-172

König R, Baldessari D, Pollet N, Niehrs C, Eils R (2004) Reliability of gene expression ratios for cDNA microarrays in multiconditional experiments with a reference design. Nucleic Acids Res 32:e29

Kosambi DD (1944) The estimation of map distances from recombination values. Ann Eugen 12:172-175

Kumar P, Paterson AH, Chee PW (2006) Predicting intron sites by aligning cotton ESTs with *Arabidopsis* genomic DNA. J Cotton Sci 10:29-38

Kumar S, Tamura K, Nei M (2004) MEGA3: integrated software for molecular evolutionary genetics analysis and sequence alignment. Briefings in Bioinformatics 5:150-163

Kwok PY, Deng Q, Zakeri H, Nickerson DA (1996) Increasing the information content of STS-based genome maps: identifying polymorphisms in mapped STSs. Genomics 31:123-126

Lacape JM, Nguyen TB, Courtois B, Belot JL, Giband M, Gourlot JP, Gawryziak G, Roques S, Hau B (2005) QTL analysis of cotton fiber quality using multiple

Gossypium hirsutum × *Gossypium barbadense* backcross generations. Crop Sci 45:123-140

Lee JJ, Hassan OSS, Gao W, Wei NE, Kohel RJ, Chen XY, Payton P, Sze SH, Stelly DM, Chen ZJ (2006) Developmental and gene expression analysis of a cotton naked seed mutant. Planta 223:418-432

Lee JJ, Woodward AW, Chen ZJ (2007) Gene expression changes and early events in cotton fibre development. Ann Bot 100:1391-1401

Li Y, Darley CP, Ongaro V, Fleming A, Schipper O, Baldauf SL, McQueen-Mason SJ (2002) Plant expansins are a complex multigene family with an ancient evolutionary origin. Plant Physiol 128:854-864

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

Liu Q, Brubaker CL, Green AG, Marshall DR, Sharp PJ, Singh SP (2001) Evolution of the *FAD2-1* fatty acid desaturase 5' UTR intron and the molecular systematics of *Gossypium* (Malvaceae). Am J Bot 88:92-102

Liu S, Saha S, Stelly DM, Burr B, Cantrell RG (2000) Chromosomal assignment of microsatellite loci in cotton. J Hered 91:326-332

Loguercio LL, Zhang JQ, Wilkins TA (1999) Differential regulation of six novel *MYB*-domain genes defines two distinct expression patterns in allotetraploid cotton (*Gossypium hirsutum* L.) Mol Gen Genet 261:660-671

McCouch SR, Cho YG, Yano PE, Blinstrub M, Morishima H, Kinoshita T (1997) Report on QTL nomenclature. Rice Genet Newslett 14:11-13

Mei M, Syed NH, Gao W, Thaxton PM, Smith CW, Stelly DM, Chen ZJ (2004) Genetic mapping and QTL analysis of fiber-related traits in cotton (*Gossypium*). Theor Appl Genet 108:280-291

Menzel MY, Richmond KL, Dougherty BJ (1985) A chromosome translocation breakpoint map of the *Gossypium hirsutum* genome. J Hered 76:406-414

Mochida K, Yamazaki Y, Ogihara Y (2003) Discrimination of homoeologous gene expression in hexaploid wheat by SNP analysis of contigs grouped from a large number of expressed sequence tags. Mol Gen Genomics 270:371-377

Musaev DA, Abzalov MM (1972) Some questions concerning the inheritance of fuzzy in cotton seeds (*G. hirsutum* L.). Genetika 8:7-16

Mutschler MA, Doerge RW, Liu SC, Kuai JP, Liedl BE, Shapiro FA (1996) Analysis of pest resistance in the wild tomato Lycopersicon pennellii: QTL controlling acylsugar level and composition. Theor Appl Genet 92:709-718

Nadarajan N, Sree Rangasamy SR (1988) Inheritance of the fuzzless-lintless character in cotton (*Gossypium hirsutum*). Theor Appl Genet 75:728-730

Narbuth EV, Kohle RJ (1990) Inheritance and linkage analysis of a new fiber mutant in cotton. J Hered 81:131-133

Neale DB, Savolainen O (2004) Association genetics of complex traits in conifers. Trends Plant Sci 9:325-330

Nei M (1987) Molecular Evolutionary Genetics. Columbia University Press, New York

Nickerson DA, Taylor SL, Weiss KM, Clark AG, Hutchinson RG, Stengard J, Salomaa V, Vartiainen E, Boerwinkle E, Sing CF (1998) DNA sequence diversity in a 9.7-kb region of the human lipoprotein lipase gene. Nat Genet19:233-240

O'Brien RD, Wakelyn PJ (2005) Cottonseed oil: An oil for *trans*-free options. Food Tech 16:677-679

Oppenheimer OG, Hermn PL, Sivakumaran S, Esch J, Marks DM (1991) A *myb* gene required for leaf trichome differentiation in *Arabidopsis* is expressed in stipules. Cell 67:483-493

Orford SJ, Timmis JN (1998) Specific expressioin of an *expansin* gene during elongation of cotton fibers. BBA-Gene Struct Expr 1398:342-346

Park YH, Alabady MS, Ulloa M, Sickler B, Wilkins TA, Yu J, Stelly DM, Kohel RJ, El-Shihy OM, Cantrell RG (2005) Genetic mapping of new cotton fiber loci using EST-derived microsatellite in an interspecific recombinant inbred line cotton population. Mol Gen Genomics 274:428-441

Paterson AH (1995) Molecular dissection of quantitative traits: Progress and prospects. Genome Res 5:321-333

Paterson AH, Saranga Y, Menz M, Jiang CX, Wright RJ (2003) QTL analysis of genotype × environment interactions affecting cotton fiber quality. Theor Appl Genet 106:384-396

Percy RG, Kohel RJ (1999) Qualitative genetics. In: Smith C.W. and Cothren J.T. (eds) Cotton: origin, history, technology, and production. John Wiley & Sons, Inc, p 316-360

Perez-Rodriguez M, Jaffe FW, Butelli E, Glover BJ, Martin C (2005) Development of three different cell types is associated with the activity of a specific MYB transcription factor in the ventral petal of *Antirrhinum majus* flowers. Development 132:359-370

Peter SD, Muralidharan V, Vaman Bhat M (1984) Fuzzless-lintless mutant of MCU 5 cotton. Cotton Dev 14:43

Pfeil BE, Brubaker CL, Craven LA, Crisp MD (2004) Paralogy and orthology in the Malvaceae *rpb2* gene family: investigation of gene duplication in *Hibiscus*. Mol Biol Evol 21:1428-1437

Pflieger S, Lefebvre V, Caranta C, Blattes A, Goffinet B, Palloix A (1999) Disease resistance gene analogs as candidates for QTLs involved in pepper-pathogen interactions. Genome 42:1100-1110

Pflieger S, Lefebvre V, Causse M (2001) The candidate gene approach in plant genetics: a review. Mol Breed 7:275-291

Reinisch AJ, Dong J, Brubaker CL, Stelly DM, Wendel JF, Paterson AH (1994) A detailed RFLP map of cotton, *Gossypium hirsutum* × *Gossypium barbadense*: chromosome organization and evolution in a disomic polyploid genome. Genetics 138:829-847

Ren LH, Guo WZ, Zhang TZ (2002) Identification of quantitative trait loci (QTLs) affecting yield and fiber propreties in chromosome 16 in cotton using substitution line. Acta Botanica Sinica 44:815-820

Richert AM, Premstaller A, Gebhardt C, Oefner PJ (2002) Genotyping of SNPs in a polyploid genome by pyrosequencing. Biotechniques 32:592-593,596-598,600

Riechmann JL, Heard J, Martin G, Reuber L, Jiang CZ, Keddie J, Adam L, Pineda O, Ratcliffe OJ, Samaha RR, Creelman R, Pilgrim M, Broun P, Zhang JZ, Ghandehari D, Sherman BK, Yu GL (2000) Arabidopsis transcription factors: Genome-wide comparative analysis among eukaryotes. Science 290:2105-2110

Robertson DS (1985) A possible technique for isolating genic DNA for quantitative traits in plants. J Theor Biol 117:1-10

Rong J, Abbey C, Bowers JE, Brubaker CL, Chang C, Chee PW, Delmonte TA, Ding X, Garza JJ, Marler BS, Park C, Pierce GJ, Rainey KM, Rastogi VK, Schulze SR, Trolinder NL, Wendel JF, Wilkins TA, Williams-Coplin TD, Wing RA, Wright RJ, Zhao X, Zhu L, Pateson AH (2004) A 3347-locus genetic recombination map of

sequence-tagged sites reveals features of genome organization, transmission and evolution of cotton (*Gossypium*). Genetics 166:389-417

Rong J, Feltus FA, Waghmare VN, Pierce GJ, Chee PW, Draye X, Saranga Y, Wright RJ, Wilkins TA, May OL, Wayne S, Gannaway JR, Wendel JF, Paterson AH (2007) Meta-analysis of polyploid cotton QTLs shows unequal contributions of subgenomes to a complex network of genes and gene clusters implicated in lint fiber development. Genetics 176:2577-2588

Rong J, Pierce GJ, Waghmare VN, Rogers CJ, Desai A, Chee PW, May OL, Gannaway JR, Wendel JF, Wilkins TA, Paterson AH (2005) Genetic mapping and comparative analysis of seven mutants related to seed fiber development in cotton. Theor Appl Genet 111:1137-1146

Royon D, Daz M, Ellenrieder G, Locatelli S. (2007) Enzymatic production of biodiesel from cotton seed oil using *t*-butanol as solvent. Bioresource Technol 98:648-653

Rozas J, Sánchez-DelBarrio JC, Messeguer X, Rozas R (2003) DnaSP, DNA polymorphism analysis by the coalescent and other methods. Bioinformatics 19:2496-2497

Ruan YL, Chourey PS, Delmer DP, Perez-Grau L (1997) The differential expression of sucrose synthase in relation to diverse patterns of carbon partitioning in developing cotton seed. Plant Physiol. 115:375-385

Ruan YL, Llewellyn DJ, Furbank RT (2001) The control of single-celled cotton fiber elongation by developmentally reversible gating of plasmodesmata and coordinated expression of sucrose and K^+ transporters and expansion. Plant Cell 13:47-60

Ruan YL, Llewellyn DJ, Furbank RT (2003) Suppression of sucrose synthase gene expression repression cotton fiber cell initiation, elongation, and seed development. Plant Cell 15:952-964

Runa YL, Choury PS (1998) A fiberless seed mutation in cotton is associated with lack of fiber cell initiation in ovule epidermis and alterations in sucrose synthase expression and carbon partitioning in developing seeds. Plant Physiol 118:399-406

Saha S, Jenkins JN, Wu J, McCarty JC, Gutierrez OA, Percy RG, Cantrell RG, Stelly DM (2006a) Effects of chromosome-specific introgression in Upland cotton on fiber and agronomic traits. Genetics 172:1927-1938

Saha S, Raska DA, Stelly DM (2006b) Upland cotton (*Gossypium hirsutum* L.) × Hawaiian cotton (*G. tomentosum* Nutt. ex Seem.) F_1 hybrid hypoaneuploid chromosome substitution series. J Cotton Sci 12:263-272

Salmaso M, Faes G, Segala C, Stefanini M, Salakhutdinov I, Zyprian E, Toepfer R, Grando MS, Velasco R (2004) Genome diversity and gene haplotypes in the grapevine (*Vitis vinifera* L.), as revealed by single nucleotide polymorphisms. Mol Breeding 14:385-395

Samora PJ, Stelly DM, Kohel RJ (1994) Localization and mapping of Le_1 and Gl_2 loci of cotton (*Gossypium hirsutum* L.). J Hered 85:152-157

Sampedro J, Cosgrove DJ (2005) The expansin superfamily. Genome Biol 6:242

Sampedro J, Lee Y, Carey RE, dePanphilis C, Cosgrove DJ (2005) Use of genomic history to improve phylogeny and understanding of births and deaths in a gene family. Plant J 44:409-419

Saranga Y, Jiang CX, Wright RJ, Yakir D, Paterson AH (2004) Genetic dissection of cotton physiological responses to arid conditions and their interrelationships with productivity. Plant Cell Environ 27:263-277

Saranga Y, Menz M, Jiang CX, Wright RJ, Yakir D, Paterson AH (2001) Genomic dissection of genotype x environment interactions conferring adaptation of cotton to arid conditions. Genome Res 11:1988-1995

SAS Institute (2003) SAS proprietary software version 9.1. SAS Inst., Cary, NC

Schiefelbein J (2003) Cell-fate specification in the epidermis: a common patterning mechanism in the root and shoot. Curr Opin Plant Biol 6:74-78

Schneider K, Schafer-Pregl R, Borchardt DC, Salamini F (2002) Mapping of QTLs for sucrose content, yield and quality in a sugar beet population fingerprinted by EST-related markers. Theor Appl Genet 104:1107-1113

Senchina DS, Alwarez I, Cronn RC, Liu B, Rong J, Noyes RD, Paterson AH, Wing RA, Wilkins TA, Wendel JF (2003) Rate variation among nuclear genes and the age of polyploidy in *Gossypium*. Mol Biol Evol 20:633-643

Shappley ZW, Jenkins JN, Meredith WR, McCarty JC (1998) An RFLP linkage map of Upland cotton, *Gossypium hirsutum* L.. Theor Appl Genet 97:756-761

Shen X, Becelaere GV, Kumar P, Davis RF, May OL, Chee P (2006a) QTL mapping for resistance to root-knot nematodes in the M-120 RNR Upland cotton line (*Gossypium hirsutum* L.) of the Auburn 623 RNR source. Theor Appl Genet 113:1539-1549

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

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

Shen X, Zhang T, Guo W, Zhu X, Zhang X (2006b) Mapping fiber and yield QLTs with main, epistatic, and QTL × environment interaction effects in recombinant inbred lines of Upland cotton. Crop Sci 46:61-66

Shimizu Y, Aotsuka S, Hasegawa O, Kawada T, Sakuno TF, Hayashi T (1997) Changes in the level of mRNAs for cell wall-related enzymes in growing cotton fiber cells. Plant Cell Physiol 38:375-378

Small RL, Ryburn JA, Cronn RC, Seelanan T, Wendel JF (1998) The tortoise and the hare: choosing between noncoding plastome and nuclear *Adh* sequences for phylogeny reconstruction in a recently diverged plant group. Am J Bot 85:1301-1315

Small RL, Ryburn JA, Wendel JF (1999) Low levels of nucleotide diversity at homoeologous *Adh* loci in allotetraploid cotton (*Gossypium* L.). Mol Biol Evol 16:491-501

Small RL, Wendel JF (2000) Copy number lability and evolutionary dynamics of the *Adh* gene family in diploid and tetraploid cotton (*Gossypium*). Genetics 155:1913-1925

Small RL, Wendel JF (2002) Differential evolutionary dynamics of duplicated paralogous *Adh* loci in alloteraploid cotton (*Gossypium*). Mol Biol Evol 19:597-607

Smart LB, Vojdani F, Maeshima M, Wilkins TA (1998) Genes involved in osmoregulation during turgor-driven cell expansion of developing cotton fibers are differentially regulated. Plant Physiol 116:1539-1549

Somers DJ, Kirkpatrick R, Moniwa M, Walsh A (2003) Mining single-nucleotide polymorphisms from hexaploid wheat ESTs. Genome 49:431-437

Song XL, Zhang TZ (2007) Identification of quantitative trait loci controlling seed physical and nutrient traits in cotton. Seed Sci Res 17:243-251

Stelly DM, Saha S, Raska DA, Jenkins JN, McCarty JC, Gutierrez OA (2005) Registration of 17 Upland (*Gossypium hirsutum*) cotton germplasm lines disomic for different *G. barbadense* chromosome or arm substitutions. Crop Sci 45:2663-2665

Stewart JM (1975) Fiber initiation on the cotton ovule (*Gossypium hirsutum*). Amer J Bot 62:723-730

Suo J, Liang X, Pu L, Zhang Y, Xue Y (2003) Identification of *GhMYB109* encoding a R2R3 MYB transcription factor that expressed specifically in fiber initials and elongating fibers of cotton (*Gossypium hirsutum* L.) BBA-Gene Struct Expr 1630:25-34

Szymanski DB, Lloyd AM, Marks MD (2000) Progress in the molecular genetic analysis of trichome initiation and morphogenesis in Arabidopsis. Trends Plant Sci 5:214-219

Tajima F (1983) Evolutionary relationship of DNA sequences in finite populations. Genetics 105:437-460

Taliercio E, Allen RD, Essenberg M, Klueva N, Nguyen H, Patil MA, Paxton P, Millena ACM, Phillips AL, Pierce ML, Scheffler B, Turley R, Wang J, Zhang D, Scheffler J (2006) Analysis of ESTs from multiple *Gossypium hirsutum* tissues and identification of SSRs. Genome 49:306-319

The Arabidopsis Genome Initiative (2000) Analysis of the genome sequence of flowering plant *Arabidopsis thaliana*. Nature 408:796-815

Thompson JD, Gibson TJ, Plewniak F, Jeanmougin F, Higgins DG (1997) The ClustalX windows interface: flexible strategies for multiple sequence alignment aided by quality analysis tools. Nucleic Acids Res 24:4876-4882

Turley RB (2002) Registration of MD17 fiberless Upland cotton as a genetic stock. Crop Sci 42:994-995

Turley RB, Ferguson DL (1996) Changes in ovule proteins during early fiber development in a normal and a fiberless line of cotton (*Gossypium hirsutum* L.). J Plant Physiol 149:695-702

Turley RB, Kloth RH (2002) Identification of a third fuzzless seed locus in upland cotton (*Gossypium hirsutum* L.). J Hered 93:359-364

Turley RB, Kloth RH (2008) The inheritance model for the fiberless trait in upland cotton (*Gossypium hirsutum* L.) line SL1-7-1: variation on a theme (in press). Euphytica DOI 10.1007/S10681-008-9670-0

Ulloa M, Saha S, Jenkins JN, Meredith WR, McCarty JC, Stelly DM (2005) Chromosomal assignment of RFLP linkage groups harboring important QTLs on an intraspecific cotton (*Gossypium hirsutum* L.) joinmap. J Hered 96:132-144

Van K, Hwang EY, Kim MY, Park HJ, Lee SH, Cregan PB (2005) Discovery of SNPs in soybean genotypes frequently used as the parents of mapping populations in the United States and Korea. J Hered 96:529-535

Van Ooijen JW (2006) JoinMap® 4, Software for the calculation of genetic linkage maps in experimental populations. Kyazma B.V., Wageningen, Netherlands

Van Ooijen, JW (2004) MapQTL® 5.0, Software for the mapping of quantitative trait loci in experimental populations. Kyazma B.V., Wageningen, Netherlands

Vogler H, Caderas D, Mandel T, Kuhlemeier C (2003) Domains of *expansin* gene expression define growth regions in the shoot apex of tomato. Plant Mol Biol 53:267-272

Waghmare VN, Rong J, Rogers CJ, Pierce GJ, Wendel JF, Paterson AH (2005) Genetic mapping of a cross between *Gossypium hirsutum* (cotton) and the Hawaiian endemic, *Gossypium tomentosum*. Theor Appl Genet 111:665-676

Wan Q, Zhang Z, Hu M, Chen L, Liu D, Chen X, Wang W, and Zheng J (2007) T_1 locus in cotton is the candidate gene affecting lint percentage, fiber quality and spiny bollworm (*Earias* spp.) resistance. Euphytica 158:241-247

Wang B, Guo W, Zhu X, Wu Y, Huang N, Zhang T (2006) QTL mapping of fiber quality in an elite hybrid derived-RIL population of upland cotton. Euphytica 152:367-378

Wang B, Wu Y, Guo W, Zhu X, Huang N, Zhang T (2007) QTL analysis and epistasis effects dissection of fiber qualities in an elite cotton hybrid grown in second generation. Crop Sci 47:1384-1392

Wang DG, Fan JB, Siao CJ, Berno A, Young P, Sapolsky R, Ghandour G, Perkins N, Winchester E, Spencer J, Kruglyak L, Stein L, Hsie L, Topaloglou T, Hubbell E, Robinson E, Mittmann M, Morris MS, Shen N, Kilburn D, Rioux J, Nusbaum C, Rozen S, Hudson TJ, Lipshutz R, Chee M, Lander ES (1998) Large-scale identification, mapping, and genotyping of single-nucleotide polymorphisms in the human genome. Science 280:1077-1082

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

Wang S, Wang JW, Yu N, Li CH, Luo B, Gou JY, Wang LJ, Chen XY (2004) Control of plant trichome development by a cotton fiber MYB gene. Plant Cell 16:2323-2334

Wang Z, Taramino G, Yang D, Liu G, Tingey SV, Miao GH, Wang GL (2001) Rice ESTs with disease-resistance gene- or defense-response gene-like sequences mapped to regions containing major resistance genes or QTLs. Mol Genet Genomics 265:302-310

Ware JO (1940) Relation of fuzz pattern to lint in an upland cotton cross. J Hered 31:489-498

Ware JO, Benedict LI, Rolfe WH (1947) A recessive naked-seed character in upland cotton. J Hered 38:313-320

Wendel JF, Cronn C (2003) Polyploidy and the evolutionary history of cotton. Adv Agron 78:139-186

Wilkins TA, Arpat AB (2005) The cotton fiber transcriptome. Physiologia Plantarum 124:295-300

Wilkins TA, Jernstedt JA (1999) Molecular genetics of developing cotton fiber. In: Basra A.S. (eds) Cotton fiber: development biology, quality improvement, and textile processing. Binghamton, NY, pp 231-269

Wilkins TA, Wan CY, Lu CC (1994) Ancient origin of the vacuolar H^+-ATPase 69-kilodalton catalytic subunit superfamily. Theor Appl Genet 89:514-524

Wilson LM, Whitt SR, Ibanez AM, Rocheford TR, Goodman MM, Buckler ES (2004) Dissection of maize kernel composition and starch production by candidate gene association. Plant Cell 16:2719-2733

Wright RJ, Thaxton PM, El-Zik KM, Paterson AH (1998) D-subgenome bias of *Xcm* resistance genes in tetraploid *Gossypium* (cotton) suggests that polyploid formation has created novel avenues for evolution. Genetics 149:1987-1996

Wright RJ, Thaxton PM, El-Zik KM, Paterson AH (1999) Molecular mapping of genes affecting pubescence of cotton. J Hered 90:215-219

Wu Y, Machado AC, White RG, Llewellyn DJ, Dennis ES (2006) Expression profiling identifies genes expression early during lint fibre initiation in cotton. Plant Cell Physiol 47:107-127

Yang SS, Cheung F, Lee JJ, Ha M, Wei NE, Sze SH, Stelly DM, Thaxton P, Triplett B, Town CD, Chen ZJ (2006) Accumulation of genome-specific transcripts, transcription factors and phytohormonal regulators during early stages of fiber cell development in allotetraploid cotton. Plant J 47:761-775

Ye ZH, Lu ZZ, Zhu J (2003) Genetic analysis for developmental behavior of some seed quality traits in Upland cotton (*Gossypium hirsutum* L.). Euphytica 129:183-191

Zhang J, Guo W, Zhang T (2002) Molecular linkage mapping of allotetraploid cotton (*Gossypium hirsutum* L. × *Gossypium barbadense* L.) with a haploid population. Theor Appl Genet 105:1166-1174

Zhang TZ, Pan JJ (1991) Genetic analysis of a fuzzless-lintless mutant in *Gossypium hirsutum* L.. Jiangsu J Agri Sci 7:13-16

Zhang W, Gianibelli MC, Ma W, Rampling L, Gale KR (2003) Identification of SNPs and development of allele-specific PCR markers for γ–gliadin alleles in *Triticum aestivum*. Theor Appl Genet 107:130-138

Zhang ZS, Xiao YH, Luo M, Li XB, Luo XY, Hou L, Li DM, Pei Y (2005) Construction of a genetic linkage map and QTL analysis of fiber-related traits in upland cotton (*Gossypium hirsutum* L.) Euphytica 144:91-99

Zhao XP, Si Y, Hanson RE, Crane CF, Price HJ, Stelly DM, Wendel JF, Paterson AH (1998) Dispersed repetitive DNA has spread to new genomes since polyploid formation in cotton. Genome Res 8:479-492

Zhu YL, Song QJ, Hyten DL, Tassell CPV, Matukumalli LK, Grimm DR, Hyatt SM, Fickus EW, Young ND, Cregan PB (2003) Single-nucleotide polymorphisms in soybean. Genetics 163:1123-1134

APPENDIX

Table A.1 Fragment EXPA1 A Genome Amplicon Haplotypes

Haplotype	Position					Genotype[1]
	43	153	226	230	252	
I	G	G	A	T	G	HS46 and Mar
II	A	A	A	C	T	A_2
III	A	G	G	T	T	Gm

[1]Mar, HS46, Gt, Gm, A_2, and D_5 represented *G. hirsutum* line MARCABUCAG8US-1-88 and HS46, *G. tomentosum*, *G. mustelinum*, *G. arboreum* and *G. raimondii*, respectively. TM-1 is genetic standards of *G. hirsutum*. 3-79 is a double haploid line of *G. barbadense*. The followings Table A.1-34 are same as this one.

Table A.2 Fragment EXPA1 D Genome Amplicon Haplotypes

Haplotype	Position												Genotype
	65	134	198	229	234	283	391	402	423	424	486	555	
I	A	G	C	G	G	G	A	G	C	C	C	C	D_5
II	G	G	C	G	G	G	A	C	C	G	C	C	TM-1, HS46 and Mar
III	A	A	T	G	G	T	G	C	T	G	T	C	Gm
IV	A	G	C	G	G	T	G	C	T	G	T	T	3-79
V	A	G	C	T	A	G	G	C	T	G	T	C	Gt

Table A.3 Fragment EXPA2-1 D Genome Amplicon Haplotypes

Haplotype	Position		Genotype
	95	97	
I	G	A	3-79 and Gm
II	A	C	D_5, TM-1, HS46 and MAR
III	G	C	Gt

Table A.4 Fragment EXPA2-2 A Genome Amplicon Haplotypes

Haplotype	Position							Genotype
	69	107	217	290	294	304	316	
I	G	A	A	A	C	G	T	A_2
II	A	G	G	A	T	G	G	Mar
III	A	G	G	A	T	C	G	TM-1
IV	A	A	G	G	T	G	T	Gm

Table A.5 Fragment EXPA2-2 D Genome Amplicon Haplotypes

Haplotype	Position										Genotype
	129	198	262	293	298	347	455	466	487	488	
I	A	G	C	G	G	G	A	G	C	C	D_5
II	G	G	C	G	G	G	A	C	C	G	HS46, Mar and TM-1
III	A	A	T	G	G	T	G	C	T	G	Gm
IV	A	G	C	T	A	G	G	C	T	G	Gt
V	A	G	C	G	G	T	G	C	T	G	3-79

Table A.6 Fragment EXPA3-1 A Genome Amplicon Haplotypes

Haplotype	Position			Genotype
	155	367	483	
I	C	G	C	HS46, Mar, Gm, Gt and TM-1
II	C	G	A	3-79
III	G	A	C	A_2

Table A.7 Fragment EXPA3-2 A Genome Amplicon Haplotypes

Haplotype	Position		Genotype
	50	166	
I	G	A	3-79
II	G	C	TM-1, HS46 and MAR
III	A	C	A_2

Table A.8 Fragment EXPA3-2 D Genome Amplicon Haplotypes

Haplotype	Position			Genotype
	96	184	225	
I	C	A	T	TM-1
II	C	A	C	HS46, Gm and D_5
III	A	G	C	3-79

Table A.9 Fragment EXPA4-1 A Genome Amplicon Haplotypes

Haplotype	Position			Genotype
	142	492	506	
I	C	C	T	TM1, HS46 and Mar
II	C	A	C	A_2
III	T	A	T	Gt
IV	C	A	T	Gm and 3-79

Table A.10 Fragment EXPA4-2 A Genome Amplicon Haplotypes

Haplotype	Position							Genotype
	24	112	247	264	337	473	503	
I	C	C	T	C	G	T	G	3-79
II	C	C	T	G	C	A	G	A_2
III	T	C	T	G	G	A	G	Gm
IV	C	T	A	G	G	A	G	TM-1
V	C	C	A	G	G	A	A	Mar
VI	C	C	A	G	G	A	G	HS46

Table A.11 Fragment EXPA4-2 D Genome Amplicon Haplotypes

Haplotype	Position							Genotype
	112	244	247	290	312	327	492	
I	C	T	A	A	A	C	T	Gt
II	T	T	A	A	A	T	T	HS46 and Mar
III	C	T	A	A	A	T	C	D_5
IV	C	T	A	A	A	T	T	3-79
V	C	G	T	T	C	T	T	TM-1

Table A.12 Fragment EXPA4-3 D Genome Amplicon Haplotypes

Haplotype	Position							Genotype
	28	30	137	238	239	273	279	
I	G	T	G	C	T	T	T	D_5
II	A	A	C	T	A	A	T	Gt and Mar
III	G	A	C	T	A	A	G	TM-1 and HS46
IV	G	A	C	T	A	A	T	Gm
V	G	A	C	T	A	T	T	3-79

Table A.13 Fragment EXPA5-1 A Genome Amplicon Haplotypes

Haplotype	Position						Genotype
	223	224	256	306	337	345	
I	C	C	G	G	A	G	HS46, Mar, 3-79 and TM-1
II	G	C	G	G	A	T	Gt
III	A	C	G	G	A	G	Gm
IV	C	G	A	A	C	G	A_2

Table A.14 Fragment EXPA5-1 D Genome Amplicon Haplotypes

Haplotype	Position			Genotype
	82	249	283	
I	C	T	A	Gt, Gm and HS46
II	C	T	G	Mar
III	T	C	A	D_5

Table A.15 Fragment EXPA5-2 A Genome Amplicon Haplotypes

Haplotype	Position			Genotype
	158	237	444	
I	A	T	C	Gt and 3-79
II	C	T	A	A_2
III	A	A	A	Gm
IV	A	T	A	HS46 and Mar

Table A.16 Fragment EXPA5-2 D Genome Amplicon Haplotypes

Haplotype	Position		Genotype
	145	309	
I	T	C	TM-1 and Gt
II	A	T	D_5

Table A.17 Fragment EXPA6-1 A Genome Amplicon Haplotypes

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

Table A.18 Fragment EXPA6-1 D Genome Amplicon Haplotypes

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

Table A.19 Fragment EXPA6-2 A Genome Amplicon Haplotypes

Haplotype	Position		Genotype
	189	209	
I	T	G	Gm
II	G	A	A_2
III	G	G	TM-1, 3-79 and MAR

Table A.20 Fragment EXPA6-2 D Genome Amplicon Haplotypes

Haplotype	Position		Genotype
	25	204	
I	C	A	3-79, TM-1, HS46, Gt and D_5
II	T	G	Gm

Table A.21 Fragment EXPA6-3 A Genome Amplicon Haplotypes

Haplotype	Position								Genotype
	67	70	99	200	211	346	454	538	
I	C	A	T	T	C	T	T	G	TM-1, HS46, Mar and Gm
II	A	A	G	T	C	C	T	G	A_2
III	C	G	G	A	T	T	T	A	Gt
IV	C	G	G	A	T	T	C	A	3-79

Table A.22 Fragment EXPA6-3 D Genome Amplicon Haplotypes

Haplotype	Position							Genotype
	70	91	180	348	368	413	497	
I	G	T	A	G	A	C	A	HS46, Mar and TM-1
II	G	T	T	G	A	T	G	Gm
III	G	T	T	G	A	C	A	3-79 and Gt
IV	A	G	T	T	T	C	A	D_5

Table A.23 A-genome Haplotypes of Gene *MYB1*

Haplotype	Position						Genotype[1]
	100	105	145	238	272	348	
I	C	C	T	G	G	A	TM-1, HS46, and Mar
II	C	C	A	A	G	A	Gm
III	C	C	T	A	G	A	3-79
IV	T	T	T	A	T	A	Gt
V	C	C	T	A	G	G	A_1

Table A.24 D-genome Haplotypes of Gene *MYB1*

Haplotype	Position				Genotype
	96	100	150	209	
I	C	T	A	A	TM-1, 3-79, Gm, and Gt
II	C	C	G	A	HS46
III	C	T	G	A	MAR
IV	T	T	A	G	D_5

Table A.25 A-genome Haplotypes of Gene *MYB2*

Haplotype	Position							Genotype
	120	153	197	204	285	334	425	
I	G	C	G	A	C	C	C	TM-1, HS46, MAR, and Gt
II	G	A	G	A	C	C	C	A_1
III	G	C	G	T	T	T	C	3-79
IV	T	C	A	A	C	C	C	Gm

Table A.26 D-genome Haplotypes of Gene *MYB2*

Haplotype	Position					Genotype
	14	196	284	429	446	
I	A	G	C	C	C	TM-1, HS46, MAR, and Gt
II	A	G	C	T	T	D_5
III	A	G	T	C	C	3-79
IV	T	A	C	C	C	Gm

Table A.27 A-genome Haplotypes of Gene *MYB3*

Haplotype	Position	Genotype
	512	
I	C	TM-1, HS46, MAR, 3-79, Gm, and Gt
II	G	A_1

Table A.28 D-genome Haplotypes of Gene *MYB3*

Haplotype	Position					Genotype
	145	452	480	674	686	
I	A	A	C	A	T	3-79
II	A	A	C	G	C	Gt
III	A	A	C	A	C	TM-1, HS46, MAR, and Gm
IV	G	G	G	A	C	D_5

Table A.29 A-genome Haplotypes of Gene *MYB4*

Haplotype	Position												Genotype
	72	83	114	115	132	133	194	405	428	463	581	834	
I	A	A	G	C	T	G	A	A	T	T	T	G	A_1
II	C	G	G	C	T	G	A	A	T	C	T	C	Gt
III	A	G	G	C	T	G	T	T	T	C	T	C	3-79
IV	A	G	G	C	T	G	A	A	C	C	T	C	TM-1, HS46, and MAR
V	A	G	C	A	G	T	A	A	T	C	C	C	Gm

Table A.30 D-genome Haplotypes of Gene *MYB4*

Haplotype	Position								Genotype
	30	59	73	78	105	224	279	332	
I	C	G	A	G	C	A	A	T	3-79
II	C	A	A	G	C	A	T	T	Gt
III	T	G	A	G	C	A	A	T	Gm
IV	C	G	C	G	T	A	A	T	TM-1, HS46, and MAR
V	C	G	A	A	C	T	A	C	D_5

Table A.31 A-genome Haplotypes of Gene *MYB5*

Haplotype	Position		Genotype
	423	511	
I	A	G	TM-1, MAR
II	A	A	HS46, 3-79, Gm, and A_1
III	T	A	Gt

Table A.32 D-genome Haplotypes of Gene *MYB5*

Haplotype	Position			Genotype
	117	219	363	
I	C	A	T	TM-1, HS46, MAR, 3-79, Gm, and Gt
II	T	G	C	D_5

Table A.33 A-genome Haplotypes of Gene *MYB6*

Haplotype	Position															Genotype
	40	76	98	123	323	360	646	685	815	847	876	40	76	98	123	
I	T	T	C	C	C	G	T	A	C	C	C	T	T	C	C	TM-1, HS46, and 3-79
II	T	T	C	G	C	A	T	A	C	C	C	T	T	C	G	Gt
III	A	C	C	C	C	G	T	A	C	C	C	A	C	C	C	MAR
IV	A	C	T	C	C	G	T	A	C	C	C	A	C	T	C	Gm
V	T	T	C	C	T	G	C	G	T	G	T	T	T	C	C	A_1

Table A.34 D-genome Haplotypes of Gene *MYB6*

Haplotype	Position																	Genotype
	40	76	109	135	324	380	382	621	627	652	740	776	815	842	845	855	883	
I	A	C	C	T	T	G	T	A	C	A	G	C	C	G	G	G	T	3-79
II	A	C	C	C	T	A	T	A	C	A	A	C	A	A	A	A	T	D_5
III	A	C	C	T	T	G	T	G	T	A	G	C	C	G	G	A	T	TM-1
IV	A	C	C	T	T	G	T	A	C	A	G	C	C	G	G	A	T	HS46
V	A	C	C	T	T	G	T	A	C	A	G	A	C	G	G	A	T	MAR
VI	T	T	T	T	C	G	C	A	C	T	G	C	C	G	G	A	T	Gm
VII	A	C	C	T	T	G	C	A	C	A	G	C	C	G	G	A	C	Gt

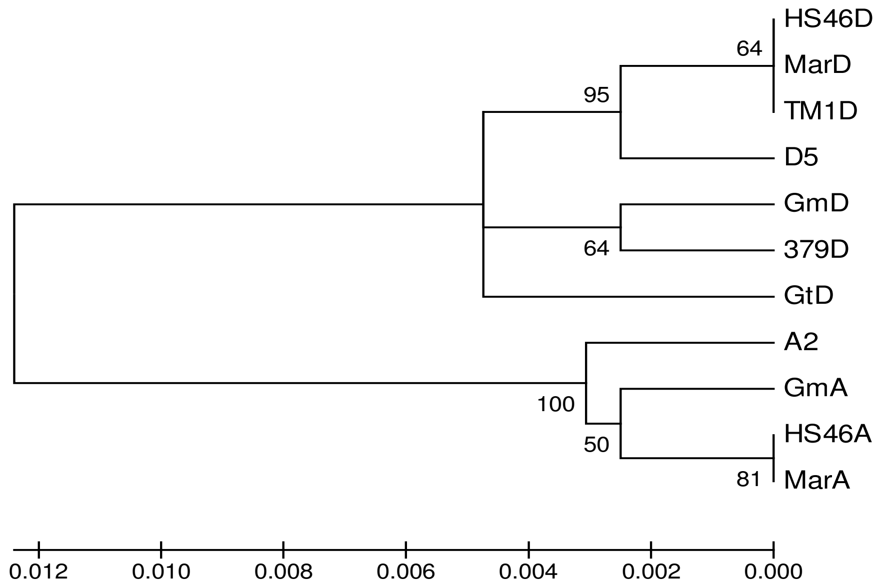

Figure A.1 Phylogenetic Analysis and Genome Specific Sequences Distinguishing of Fragment EXPA1.

Clades supporting brootstrap values were shown adjacent the branches. Mar, Gt, Gm, A_2, and D_5 represented *G. hirsutum* line MARCABUCAG8US-1-88, *G. tomentosum*, *G. mustelinum*, *G. arboreum* and *G. raimondii*. The suffix of A or D at the simplified name indicated A_T or D_T genome sequence. For example, GtD was the *G. tomentosum* D genome sequence. Following Figures A.2 to A.13 have the same legend notes as Figure A. 1.

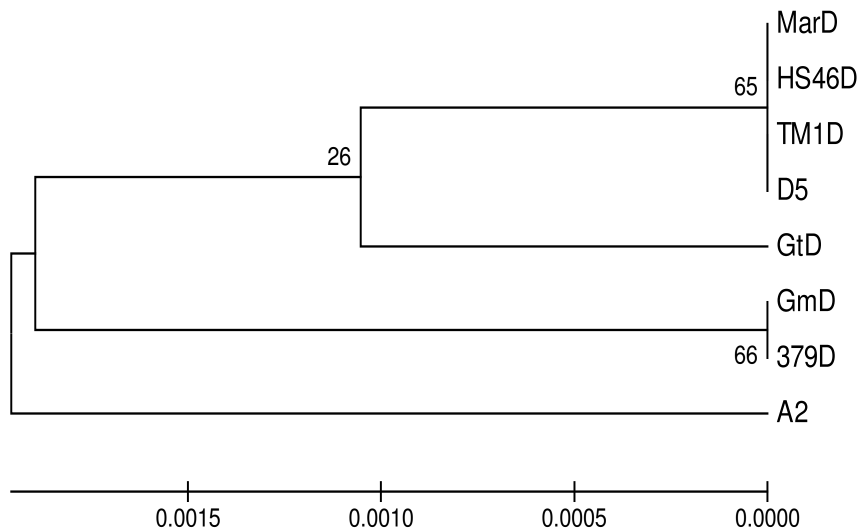

Figure A.2 Phylogenetic Analysis and Genome Specific Sequences Distinguishing of Fragment EXPA2-1.

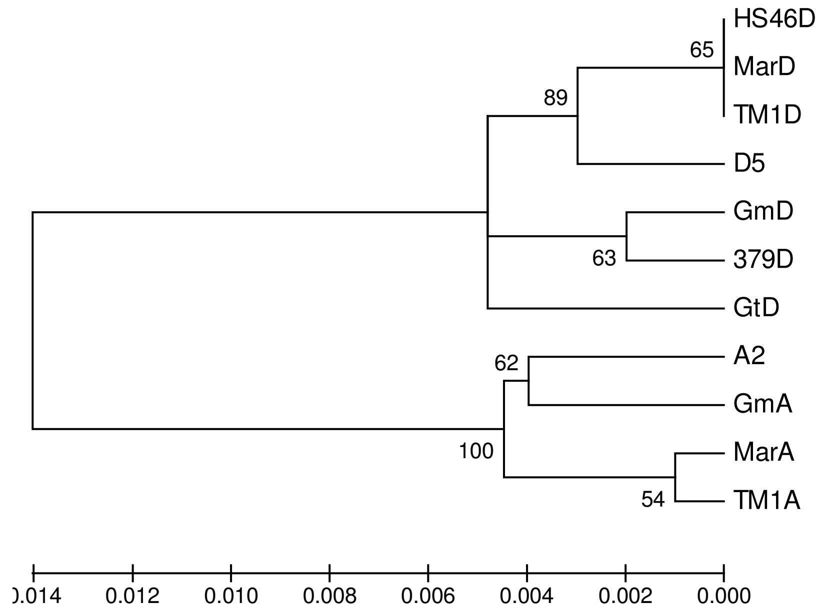

Figure A.3 Phylogenetic Analysis and Genome Specific Sequences Distinguishing of Fragment EXPA2-2.

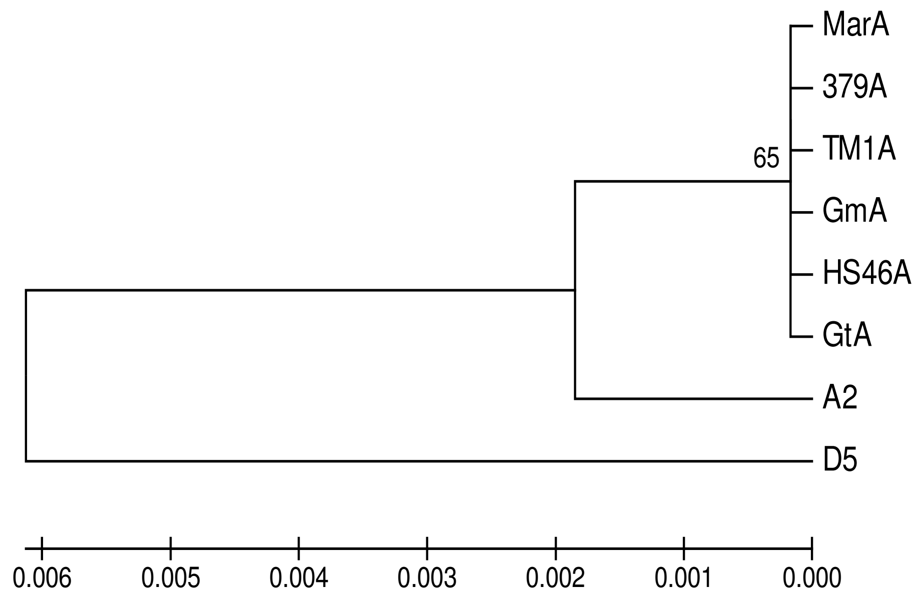

Figure A.4 Phylogenetic Analysis and Genome Specific Sequences Distinguishing of Fragment EXPA3-1.

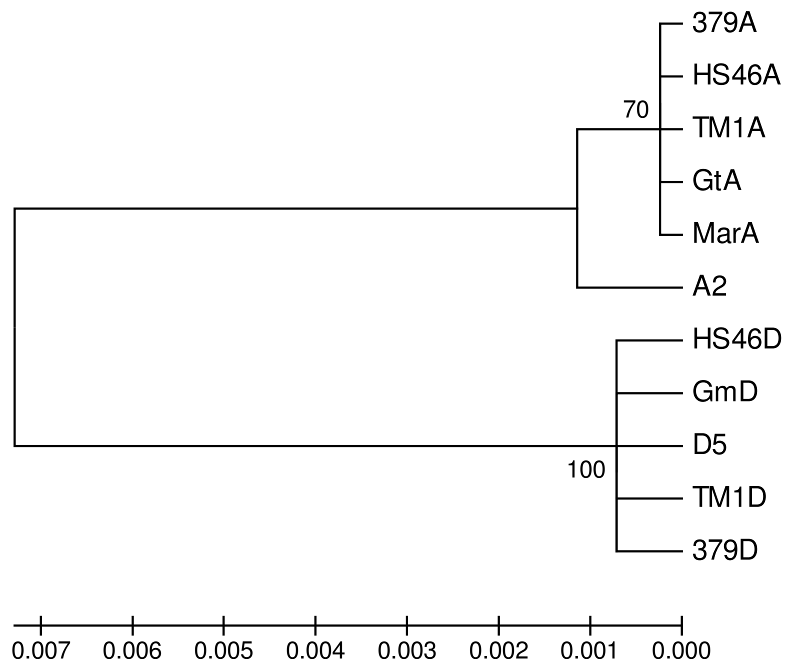

Figure A.5 Phylogenetic Analysis and Genome Specific Sequences Distinguishing of Fragment EXPA3-2.

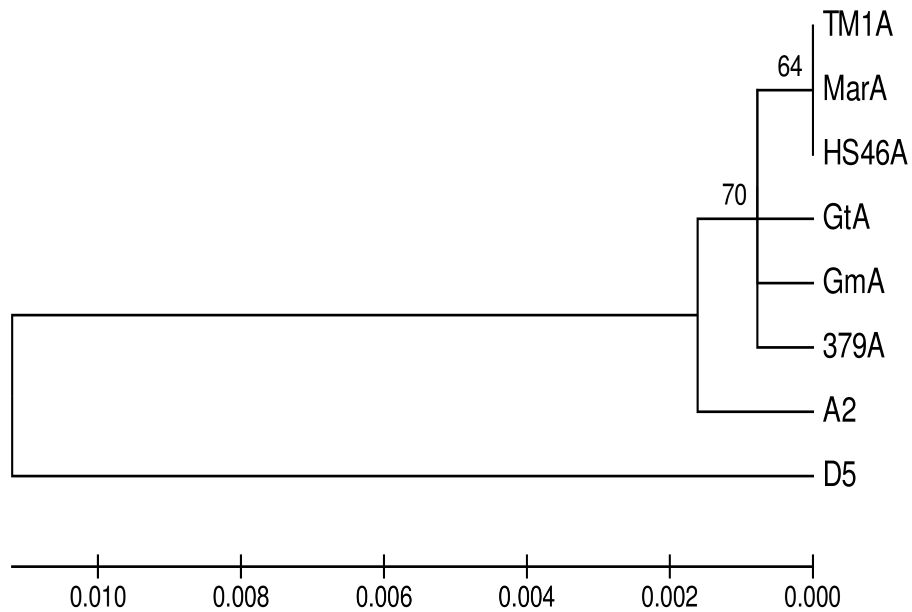

Figure A.6 Phylogenetic Analysis and Genome Specific Sequences Distinguishing of Fragment EXPA4-1.

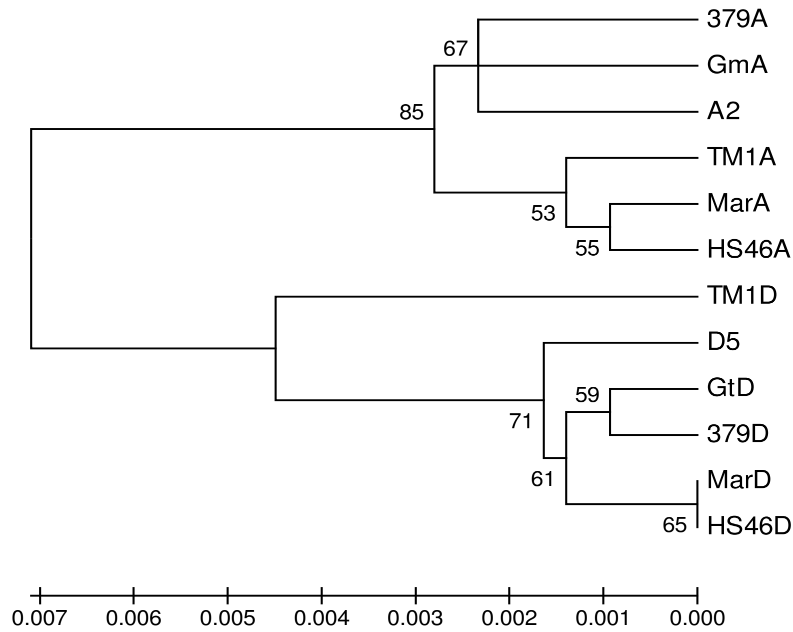

Figure A.7 Phylogenetic Analysis and Genome Specific Sequences Distinguishing of Fragment EXPA4-2.

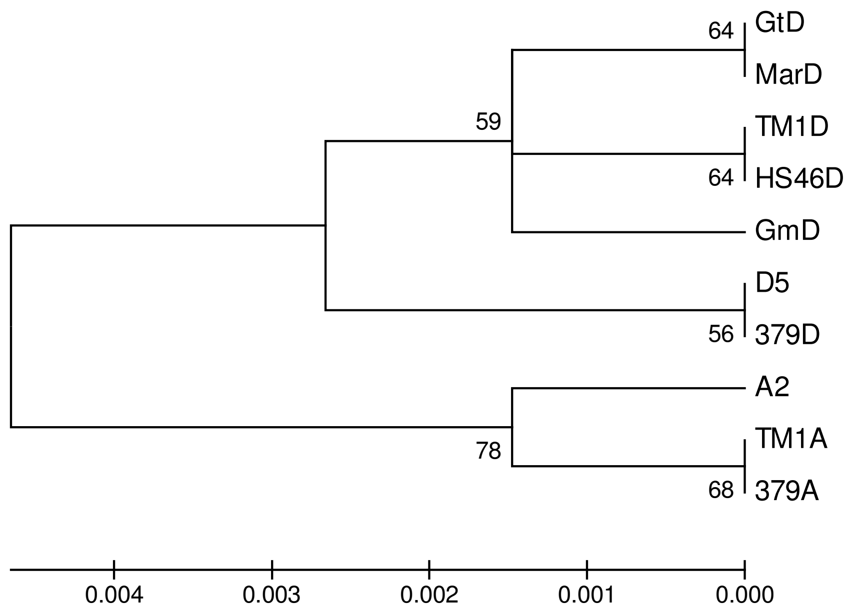

Figure A.8 Phylogenetic Analysis and Genome Specific Sequences Distinguishing of Fragment EXPA4-3.

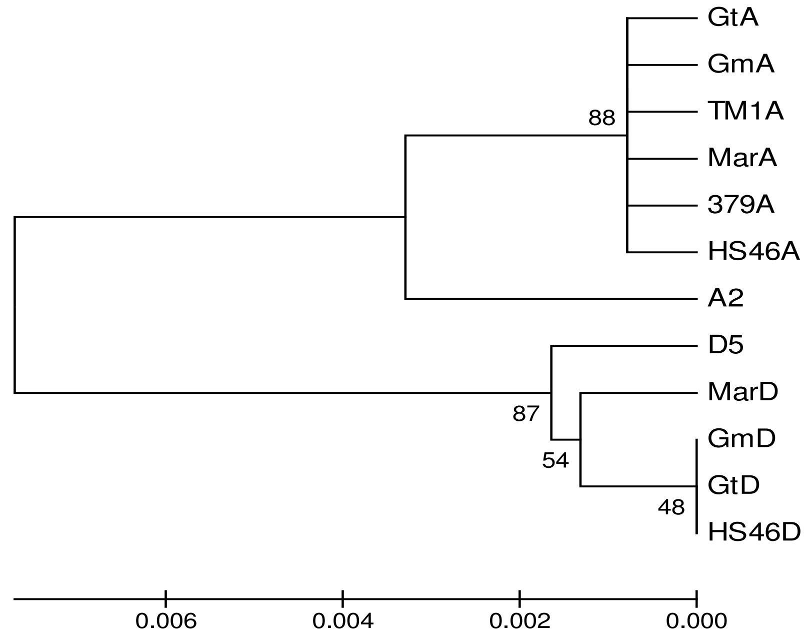

Figure A.9 Phylogenetic Analysis and Genome Specific Sequences Distinguishing of Fragment EXPA5-1.

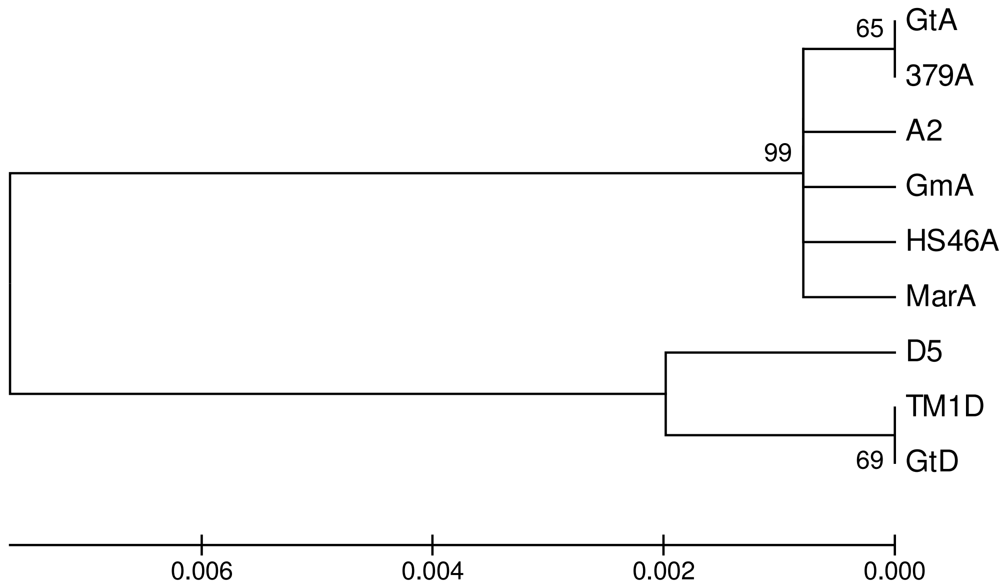

Figure A.10 Phylogenetic Analysis and Genome Specific Sequences Distinguishing of Fragment EXPA5-2.

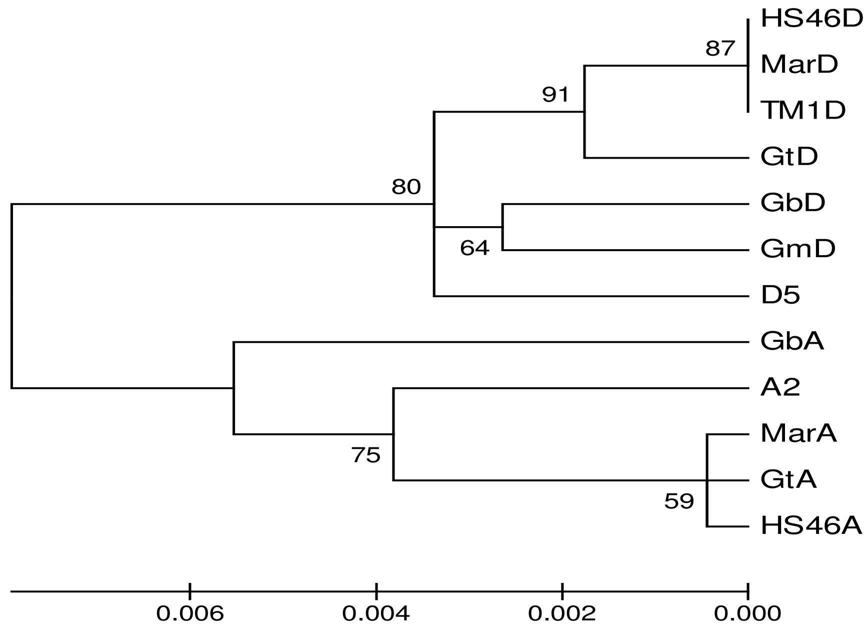

Figure A.11 Phylogenetic Analysis and Genome Specific Sequences Distinguishing of Fragment EXPA6-1.

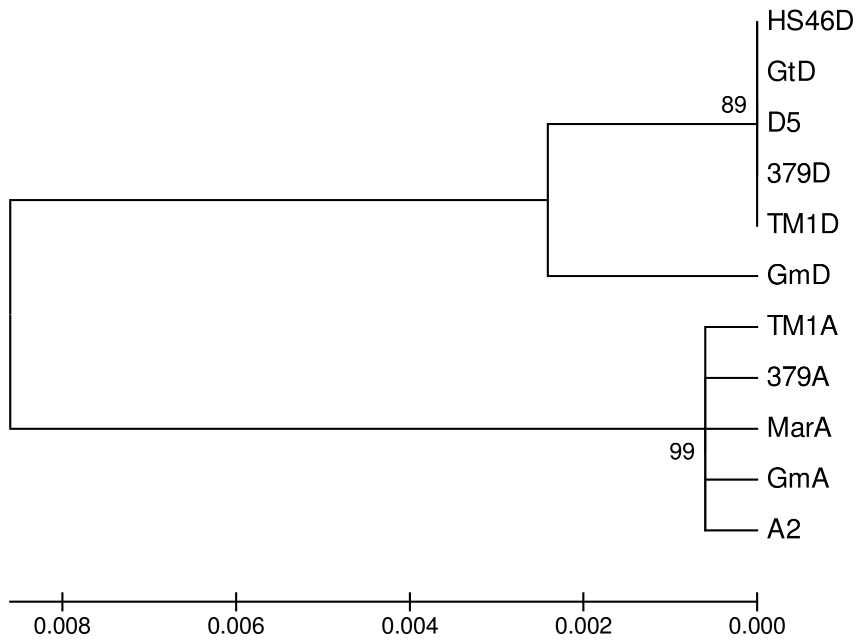

Figure A.12 Phylogenetic Analysis and Genome Specific Sequences Distinguishing of Fragment EXPA6-2.

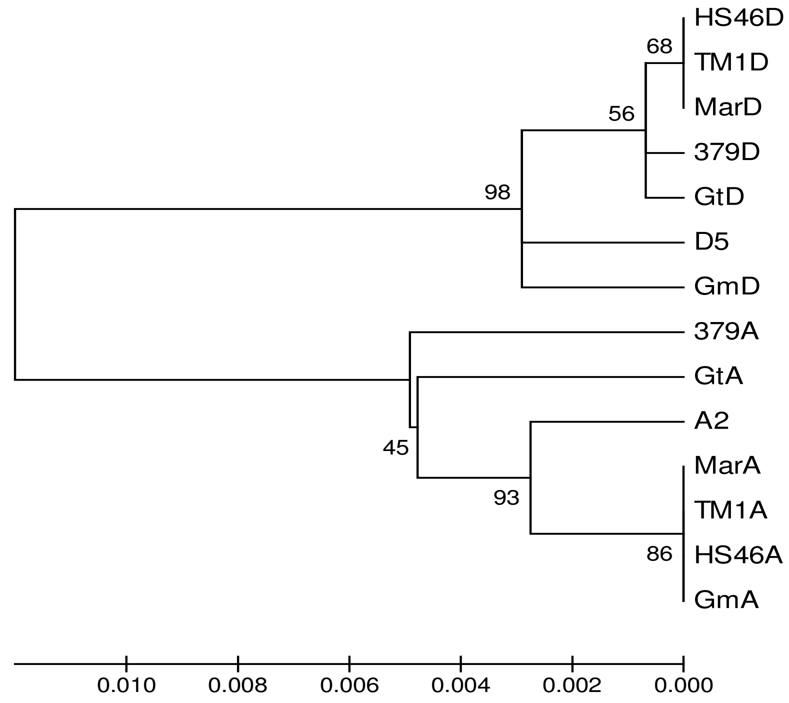

Figure A.13 Phylogenetic Analysis and Genome Specific Sequences Distinguishing of Fragment EXPA6-3.

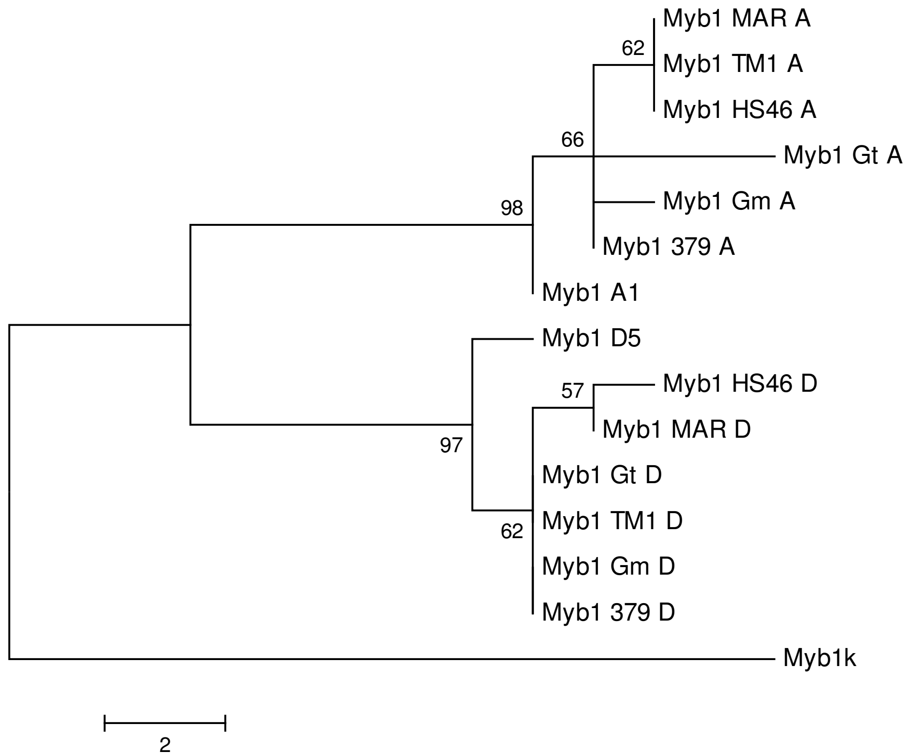

Figure A.14 Phylogenetic Grouping Based Duplicated Loci Differentiation in *MYB1* Gene.

Taxa *G. herbaceum* and *G. raimondii* are designated by their genome designation, A_1 and D_5, respectively. Gm and Gt are the abbreviations of *G. mustelinum* and *G. tomentosum*. HS46 and MAR are two *G. hirsutum* lines. TM-1 is genetic standard of *G. hirsutum* and 3-79 is a double haploid line of *G. barbadense*. 'A' and 'D' followed each taxa name are the subgenome differentiation of each sequence. Numbers on nodes indicate bootstrap values. Following Figures A.15 to A.18 have the same legend notes as Figure A. 1.

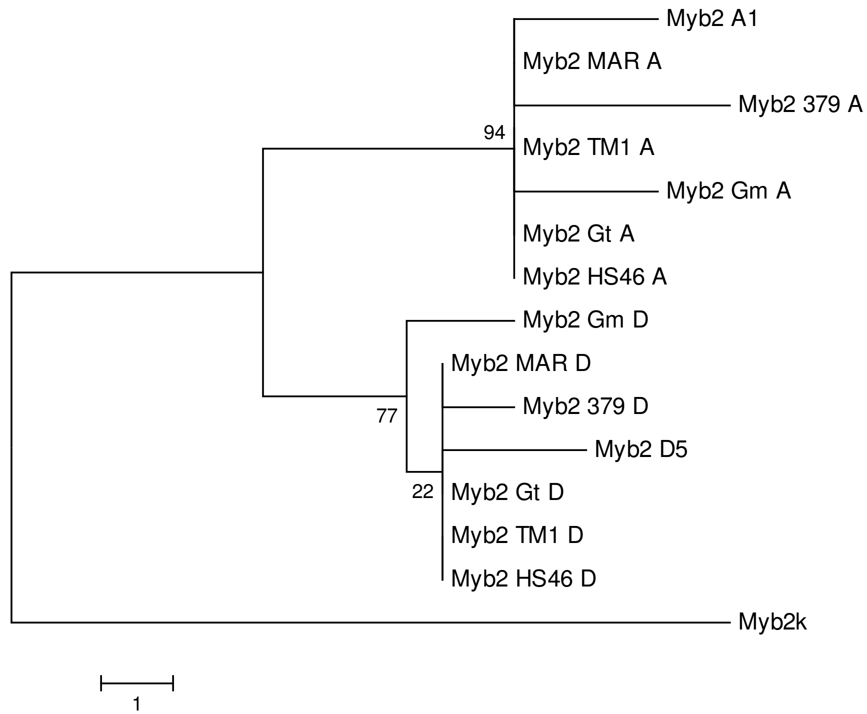

Figure A.15 Phylogenetic Grouping Based Duplicated Loci Differentiation in *MYB2* Gene.

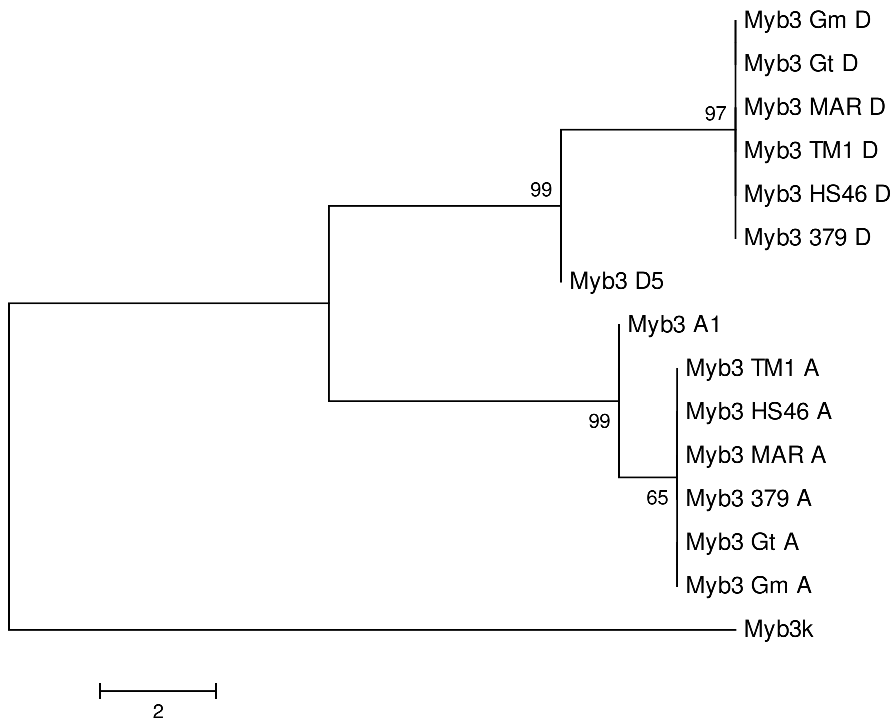

Figure A.16 Phylogenetic Grouping Based Duplicated Loci Differentiation in *MYB3* Gene.

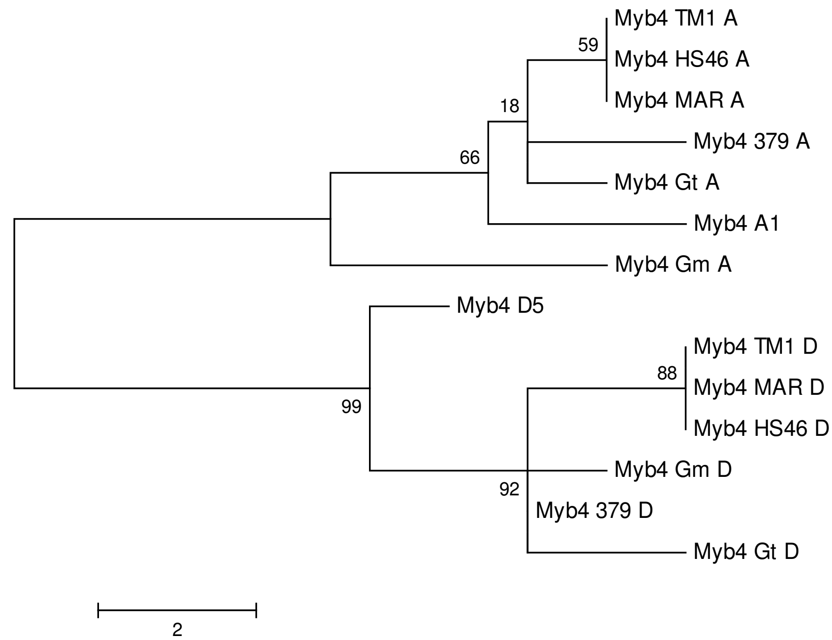

Figure A.17 Phylogenetic Grouping Based Duplicated Loci Differentiation in *MYB4* Gene.

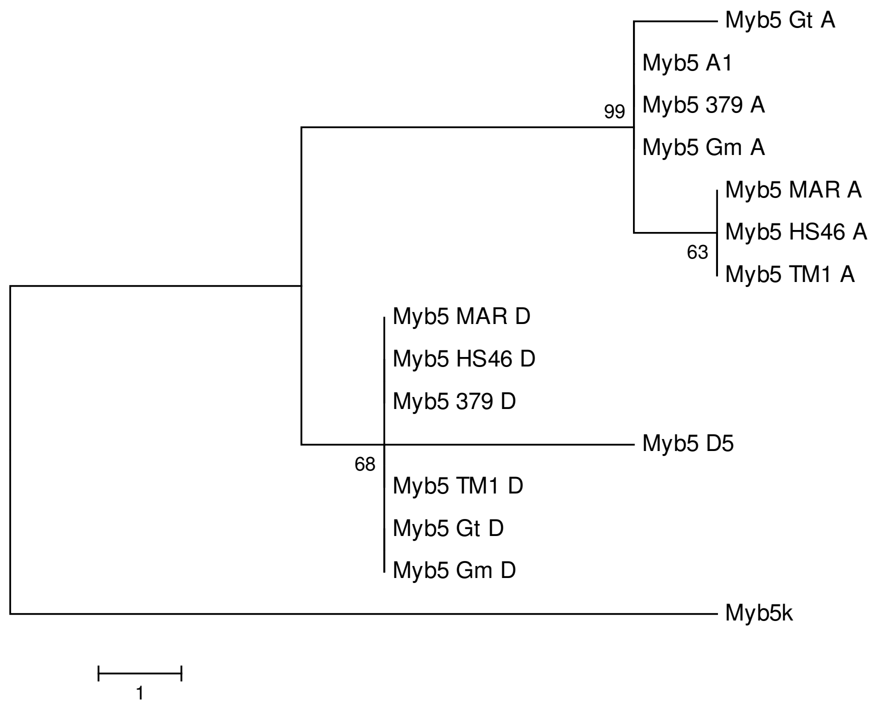

Figure A.18 Phylogenetic Grouping Based Duplicated Loci Differentiation in *MYB5* Gene.

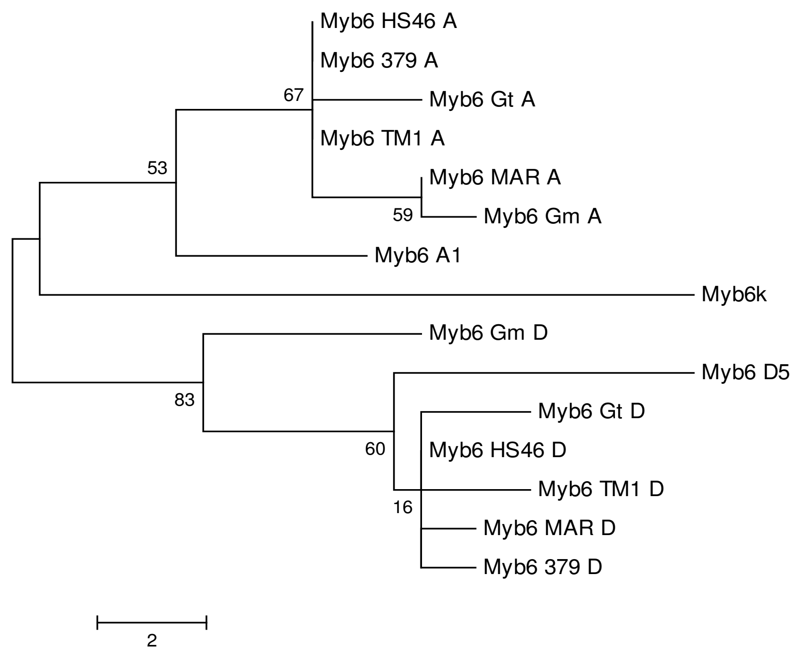

Figure A.19 Phylogenetic Grouping Based Duplicated Loci Differentiation in *MYB6* Gene.

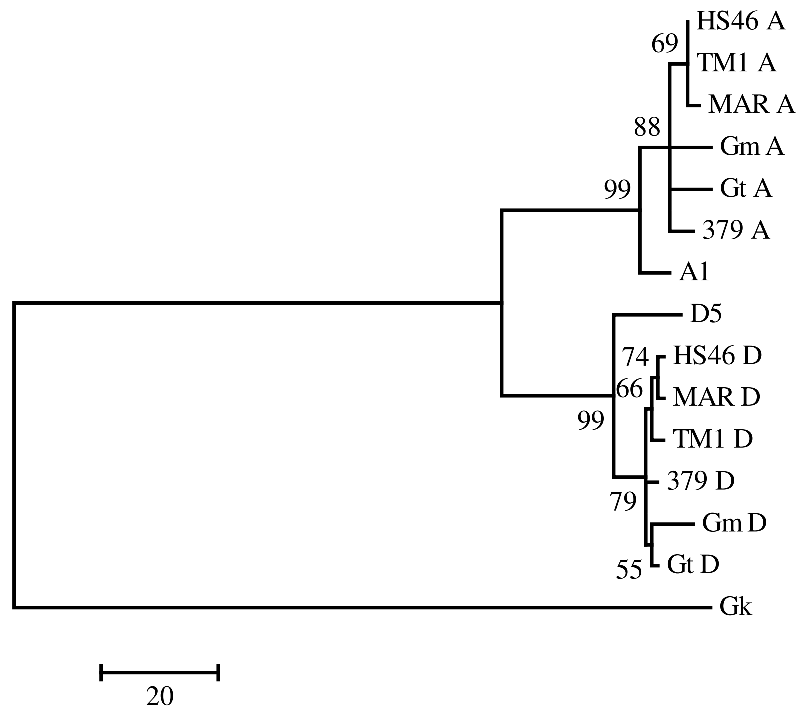

Figure A.20 Phylogenetic Analysis of the Overall Six *MYB* Genes.

Taxa *G. herbaceum* and *G. raimondii* are designated by their genome designation, A_1 and D_5, respectively. Gm and Gt are the abbreviation of *G. mustelinum* and *G. tomentosum*. HS46 and MAR are two *G. hirsutum* lines. TM-1 is genetic standard of *G. hirsutum*. 3-79 is a double haploid line of *G. barbadense*, respectively. 'A' or 'D' followed each taxa name is the subgenome differentiation of each sequence; Numbers on nodes indicate bootstrap values.