

## ABSTRACT

TYAGI, PRIYANKA. The Structure of Genetic Diversity in Upland Cotton (*Gossypium hirsutum* L.) Cultivars and Dissecting the Components of Hybrid Cotton Yield. (Under the direction of Vasu Kuraparthy and Daryl Bowman.)

Upland cotton (*Gossypium hirsutum* L.) is the major source of natural fiber. Domesticated in Yucatan peninsula 5,000 years ago, Upland cotton has narrow genetic base. With a heightened risk of genetic vulnerability to disease epidemics and climate change, elite breeding programs could benefit from the unexploited standing genetic variation of obsolete cultivars without the yield drag typically associated with wild accessions. Previous studies to estimate genetic diversity in Upland cotton have either been specific to germplasm from a single breeding program or a specific geographical area and many used limited samples of markers for genetic analysis. A thorough understanding of genetic diversity and population structure of Upland cotton germplasm with improved genotyping platforms would allow efficient utilization of these resources in genetic and breeding efforts. In this study, 378 Upland cotton accessions and 3 cultivars of *G. barbadense* were evaluated for genetic diversity and population structure using 120 SSR markers. These accessions represent different public and private breeding programs in the U.S. Average genetic distance of 0.195 indicated low level of overall genetic diversity. STRUCTURE analysis identified four groups of Upland cotton accessions that correspond to western, southwestern, mid-south and eastern regions of the cotton growing area in the US. Significant admixture was detected between these groups. Genetic diversity was highest for the group of eastern cultivars. In addition, core sets of Upland cotton accessions that represent most of the genetic diversity within the germplasm pool were identified.

We also studied three inbreds and three hybrids across five locations to identify the relationship of lint yield components with lint yield heterosis in different environment conditions. Yield was significantly by both boll number and lint per boll. However, boll number was the major contributor to lint yield for both parents and hybrids. There was a significant interaction between environment and percent heterosis for lint yield. Different yield components contributed to change in percent heterosis in low to high yielding environments in different hybrids. Increase in boll number due to higher boll retention and increase in lint per boll due to more lint per seed were found to be the main yield components affecting relationship of lint yield heterosis with environment.

© Copyright 2013 by Priyanka Tyagi

All Rights Reserved

The Structure of Genetic Diversity in Upland Cotton (*Gossypium hirsutum* L.) Cultivars and  
Dissecting the Components of Hybrid Cotton Yield

by  
Priyanka Tyagi

A dissertation submitted to the Graduate Faculty of  
North Carolina State University  
in partial fulfillment of the  
requirements for the degree of  
Doctor of Philosophy

Crop Science

Raleigh, North Carolina

2013

APPROVED BY:

---

Vasu Kuraparthi  
Committee Co-Chair

---

Daryl Bowman  
Committee Co-chair

---

James Holland

---

Keith Edmisten

## **DEDICATION**

This work is dedicated to my family for their love, support and encouragement.

## **BIOGRAPHY**

Priyanka Tyagi was born on June 11, 1982 in the small industrial town of Modinagar, UP, India to Janardan Tyagi and Sunita Tyagi. She completed her high school education in Modinagar. She completed her BS as well as MS from CCS University, Meerut, India. During her MS, she joined Microbiology Department at Indian Agricultural Research Institute in New Delhi, India as an intern for a research project. Upon graduation she joined the corn genetics program at Indian Agricultural Research Institute, New Delhi as a research fellow. In 2008 she joined cotton breeding program at Crop Science department, NC State University for a PhD.

## **ACKNOWLEDGMENTS**

First and foremost, I would like to thank Dr. Vasu Kuraparthi for giving me the opportunity of joining North Carolina State University and for his guidance and patience during my PhD program. I would also like to thank my committee members: Dr. Daryl Bowman, Dr. Jim Holland and Dr. Keith Edmisten for their guidance, support and suggestions. Thanks to Dr. Consuelo Arellano, Dr. Susana Milla-Lewis and Dr. Michael Gore for their suggestions and expertise. I am thankful to Dr. Gina Brown-Guedira for allowing me access to their lab and to Jared, Kim and Sharon for providing technical guidance. I thank Cotton Incorporated and Monsanto for funding my graduate studies. I gratefully acknowledge the assistance of Dwight Parrish, Johnny Denton, Phil Johnson, Cathy Herring and Blake Bowen for field experiments. Thanks to the amazing team of undergraduate students Debbi, Megan, Molly, Evan, Wesley and James who brought fun and cheers to field days. Heartfelt thanks to my fellow graduate students Ryan Andres, Linglong Zhu, and Baljinder Kaur for their support and friendship. They are the most wonderful people to work with. Finally, I would like to thank my family, friends and my husband Devarshi Selote, for giving me the strength and encouragement to face all challenges in life.

## TABLE OF CONTENTS

LIST OF TABLES .....	vii
LIST OF FIGURES .....	ix
INTRODUCTION .....	1
Origin and domestication of Upland cotton ( <i>Gossypium hirsutum</i> L.).....	2
Molecular marker resources in cotton.....	6
Breeding for improved yield.....	9
References.....	12
CHAPTER I. Genetic diversity and population structure in Upland cotton ( <i>Gossypium</i> <i>hirsutum</i> L.).....	19
Abstract.....	20
Introduction.....	21
Material and Methods .....	25
Plant material .....	25
SSR genotyping .....	26
Preliminary analysis of genotypic data and genetic diversity.....	27
Analysis of genetic structure.....	28
Construction of core sets.....	29
Results.....	30
SSR analysis.....	30
Unique alleles.....	31
Population Structure.....	31
Genetic diversity .....	33
Core sets of Upland cotton accessions.....	34
Discussion.....	35
References.....	41

CHAPTER II. Components of hybrid cotton yield and their relationship with environment ...	68
.....	68
Abstract .....	69
Introduction.....	70
Materials and Methods.....	73
Results.....	75
Environment effect on entry means .....	75
Entry means across locations .....	76
Correlations between means of yield components.....	76
Correlations between environmental mean lint yield and percent heterosis for lint yield and yield components .....	77
Discussion.....	80
References.....	85

## APPENDICES

Appendix A. List of 381 cotton accessions with identification number (PI number)	100
.....	100
Appendix B. List of SSR primers used to genotype a panel of 378 cotton accessions	112
.....	112
Appendix C. Proportional Membership of cotton accessions to clusters as determined by model based analysis using STRUCTURE. Lines were assigned to a group based on membership probability higher than 0.70. The identified clusters roughly correspond to following geographical areas of cotton belt: Cluster 1-Western, cluster 2-Eastern, cluster 3-southwest, cluster 4-midsouth and cluster 5- <i>G. barbadense</i> ...	117

## LIST OF TABLES

CHAPTER I. Genetic diversity and population structure in Upland cotton (*Gossypium hirsutum* L.)

Table 1.1 Summary statistics for SSR markers used to genotype <i>G. hirsutum</i> accessions ...	48
Table 1.2 Summary of unique and rare alleles observed in a combined panel of <i>G. hirsutum</i> and <i>G. barbadense</i> accessions vs. only the Upland cotton accessions .....	52
Table 1.3 List of <i>G. hirsutum</i> accessions with unique alleles with frequency less than 0.5% in the diversity panel .....	53
Table 1.4 Summary of rare alleles found in Upland cotton accessions grouped in clusters based on STRUCTURE analysis. The color name in bracket shows the color assigned to the group in STRUCTURE Q-plot. The geographic region corresponding to the color codes are as follows: Red- western region, green- southeastern region, blue-southwest, and yellow-midsouth.....	55
Table 1.5 Genetic distance estimates calculated using Nei's 1983 distance within and between <i>G. hirsutum</i> groups identified by model based population structure analysis .....	56
Table 1.6 Analysis of molecular variance for upland cotton accessions between and within four groups corresponding to four major regions of cotton production in United States as identified by STRUCTURE.....	57
Table 1.7 Pairwise $F_{ST}$ estimates for the four groups corresponding to four major regions of cotton production in United States as identified by STRUCTURE.....	58
Table 1.8 Core sets of Upland cotton accessions identified by simulated annealing algorithm using Powermarker software. Agronomically unfavorable accessions were removed from the complete panel to yield a reduced panel of 324 accessions. Acala Maxxa,	

Wannamaker Cleveland, Dixie King, Deltapine 14, DES 56 and Paymaster HS200 are used as positive constrains in core sets of all sizes .....	59
CHAPTER II. Components of hybrid cotton yield and their relationship with environment	
Table 2.1 ANOVA table showing source of variation, degrees of freedom, mean squares and significance for main effects and interaction for lint yield tested in five locations ...	89
Table 2.2 Mean values for yield components for the five locations averaged across six entries .....	90
Table 2.3 Mean values for entries for lint yield and its components across locations .....	91
Table 2.4 Correlation of lint yield with yield components over entries .....	92
Table 2.5 Correlation between environment mean lint yield and percent heterosis values for lint yield and yield components for three hybrids .....	93

## LIST OF FIGURES

CHAPTER I. Genetic diversity and population structure in Upland cotton (*Gossypium hirsutum* L.)

- Figure 1.1 (a) Histogram of allele frequencies for complete panel of *G. hirsutum* and *G. barbadense* accessions (b) Histogram of allele frequencies for alleles amplified in *G. hirsutum* accessions ..... 62
- Figure 1.2 (a) Ln (probability of data) for K ranging from 2 to 12 (b) Estimating number of subpopulations using delta K values for K ranging from 2 to 12 using method proposed by Evanno et al. (2005)..... 63
- Figure 1.3 Q-plot showing clustering of 381 *Gossypium* accessions based on analysis of genotypic data using STRUCTURE. Each accession is represented by a vertical bar. The colored subsections within each vertical bar indicate membership coefficient (Q) of the accession to different clusters. Colors correspond to the following groups: Group 1 (red), Group 2 (green), Group 3 (blue), Group 4 (yellow) and Group 5 (pink) ..... 64
- Figure 1.4 Dendrogram of 381 *G. hirsutum* accessions by NJ analysis. Colors in the dendrogram correspond to population structure as identified in Structure analysis. Membership threshold of 70% was used to assign accessions to different cluster. Accessions with less than 70% membership to any cluster were considered as mixed and are indicated in black in the dendrogram..... 65
- Figure 1.5 Three dimensional Principal Coordinate analysis (PCA) of Upland cotton diversity panel genotyped with SSR markers..... 66
- Figure 1.6 Plot showing percentage of alleles included in core sets of different sizes ranging from 8 to 53 accessions of *G. hirsutum*..... 67

CHAPTER II. Components of hybrid cotton yield and their relationship with environment

Figure 2.1 Relationship of percent heterosis for lint yield with mean environmental yield.. 95

Figure 2.2 Lint yield and major yield components over low to high yielding environments for  
DP51 x STV474 and its parents ..... 96

Figure 2.3 Lint yield and major yield components over low to high yielding environments for  
STV474 x LA887 and its parents ..... 97

Figure 2.4 Lint yield and major yield components over low to high yielding environments for  
LA887 x DP51 and its parents..... 98

## **INTRODUCTION**

### **Origin and domestication of Upland cotton (*Gossypium hirsutum* L.)**

Upland cotton is a tetraploid ( $2n=4x=52$ ) member of the genus *Gossypium* in the family Malvaceae. Almost 90% of the total cotton fiber produced globally comes from Upland cotton. Only three other species of this genus are used for fiber production; *G. barbadense* (Pima cotton), which contributes around 8%, and two diploid species (Old world cotton,  $2n=2x=26$ ): *G. arboreum* (Tree cotton) and *G. herbaceum* (Levant cotton), which contribute to roughly 2% of the global cotton fiber production (Wendel 1992). Cotton fiber is the leading natural fiber in textile industry. Apart from textiles, cotton finds uses in industries ranging from automobiles to insulations as well as in the food and feed industries as a source of cottonseed oil and protein rich meal for livestock. The United States is the third largest producer and the largest exporter of cotton with an industry worth more than 25 billion dollars ([www.icac.org](http://www.icac.org)). In 2011, approximately 15 million bales of cotton were produced from 14.7 million planted acres in the United States (<http://www.cotton.org/econ/world/detail.cfm>). The cotton growing area in the United States constitutes of 17 states in the south collectively referred to as the US cotton belt area.

Cotton is not indigenous to the United States. Two centers of diversity have been identified for *G. hirsutum*, one in southern Mexico-Guatemala, considered the primary center of diversity, and the other in the Caribbean region (Wendel et al. 1992; Brubaker et al. 1999). Cotton cultivation in the United States started in seventeenth century from a mixture of green seed and black seed biotypes along-with the Old world cotton introduced by British colonists (Smith et al. 1999). In late 1700s, green seeded *G. hirsutum* was the principal form of cotton being grown in the United States. Day neutral Mexican highland stocks were introduced in

early 1800s and were later hybridized with green and black seeded cottons to improve their agronomic characters (Smith et al. 1999). These Mexican highland stocks with favorable agronomic characters eventually replaced most green and black seed lines to form the base of Upland cotton cultivars grown today (Smith et al. 1999). Locally adapted cultivars were developed from these Mexican highland stocks during 1800s primarily through selections as well as crossing to other introductions. Between 1892 and 1904, cotton production in the US was highly affected by a boll weevil (*Anthonomus grandis*) outbreak (Niles and Feaster 1984). The breeding efforts during this era were focused on developing early maturing cultivars to evade boll weevil damage and led to a change in the germplasm pool with short staple early maturing cultivars becoming more popular than late maturing long staple cultivars.

Upland cotton growing area in the United States can be broadly categorized into four areas: Western zone, Southwestern plains, Mid-south or delta region, and the Southeast. Acala cottons grown in the western zone of Cotton Belt were developed from Mexican stocks introduced during early 1900s with inputs from *G. barbadense* and Triple hybrid (*G. thurberi* Todaro x *G. arboreum* L. x *G. hirsutum* L.) (Staten 1970; Smith et al. 1999; Bowman et al. 2006). Selections were made from original Acala that was introduced in 1907 from the village of Acala in Mexico and led to development of Acala 1517 series of lines (Staten 1970). In the San Joaquin Valley, CA, SJV germplasm was developed using Acala 8 which was a selection from original Acala (Smith et al. 1999). Breeding efforts for Acala cottons were driven by efforts to improve fiber quality and resistance to Verticillium wilt which posed a major challenge in the western region (Staten 1970). The SJ series of Acala cottons

was developed from a combination of California acalas with New Mexico acalas and Triple Hybrid-Early Fluff (THEF) (Staten 1970, Smith et al. 1999).

Due to high fiber quality of Acala cultivars, these lines have been frequently used in other breeding programs across the cotton belt to develop high yield and high fiber quality cultivars. Bowman et al. (1996) studied the pedigrees of 260 lines released between 1970 and 1990 and found that the cultivar 'Acala' was present in pedigrees of 73% of the modern Upland cotton cultivars. From 1980 to 2000, Acala lines developed by New Mexico State University were found to be the major source of fiber strength improvement in US cotton cultivars (Bowman and Gutierrez 2003).

Cotton breeding in southwestern plains of US focused on cultivars that could avoid yield damages caused by summer storms and boll weevil (Smith et al. 1999). These cultivars were developed from Stormproof stocks collected from Mexico in 1850s. Bohemian, which was selected from these early stocks in 1860, became the major source of stormproof material. Mebane, Lone Star and Rowden were developed from Bohemian and formed the core of southwestern cultivars. Paymaster cultivars were developed following a complex series of crosses using lines from Stoneville breeding program, Acala, Macha as well as Kekchi, which was introduced in 1904 from Guatemala (Smith et al. 1999). Another major breeding effort in southwestern plains has been the development of Multiple Adversity Resistance (MAR) germplasm. MAR germplasm development was based on breeding for resistance to biotic and abiotic stresses, like bacterial blight, Verticillium and Fusarium wilts, Phymatotrichum root rot, leaf spots, bolls rots, root-knot and reinform nematodes, drought stresses, along with improved yield and fiber quality (Bird 1982; Thaxton and El-Zik 2003).

In the Mississippi delta region, Delta and Pine Land Company and Stoneville Pedigreed Seed Company have been the major private breeding programs. Deltapine material was developed from Express, Mebane Triumph, Polk, Foster and Wannamaker Cleveland (Smith et al. 1999; Bowman et al. 2006). The ancestry of these line trace back to Bohemian, Belle creole and early Mexican highland stocks (Smith et al. 1999). Late 1900s marked the rise in use of transgenic technology and transgenic varieties soon replaced conventional varieties for commercial cotton production. Selections within segregating material generated from artificial hybridizations were frequently used during the development of Deltapine cultivars unlike programs like Acala cotton which were mostly developed through reselection from existing germplasm (Smith et al. 1999). Stoneville breeding program traces its ancestry to Lone star which itself was developed from Bohemian (Bowman et al. 2006). Stoneville LA887 is a high fiber strength cultivar also resistant to fusarium wilt, root-knot and reniform nematodes (Bowman 1999; Bowman et al. 2006; Bowman and Gutierrez 2003).

Pee Dee germplasm program is one of the most important breeding programs from southeastern cotton belt. This program was initiated in 1935 to develop early maturing long staple cultivars but shifted its focus in 1957 to develop high fiber quality medium staple cultivars (Culp and Harrell 1973). In the early years Earlystaple 7 and Pee Dee 4381-54 were the only two successful cultivars that were developed as selections within the germplasm pool. However, hybridization and selection became the main method of development of material since 1950s. Culp and Harrell (1973) emphasized the importance of hybridization and selection to create new recombinants for development of Pee Dee germplasm.

Development of Triple hybrid strains (*G. arboreum* L. x *G. thurberi* Todaro x *G. hirsutum*

L.) by Beasley in 1940 was a major event and this hybrid was used as a source of high fiber strength (Beasley 1940; Kerr 1960). Interspecific introgressions using *G. barbadense* and Triple hybrid have significantly contributed to the fiber quality improvements in Acala as well as Pee Dee germplasm (Smith et al. 1999; May 2001; Zhang et al. 2005a; Zhang et al. 2005b). The complex breeding history of Pee Dee cottons involves crosses between diverse germplasm including Triple hybrid, Sealand, Earlistaple, C 6-5, AHA, Dixie King, Auburn 56 and Coker 421 (Culp and Harrell 1973; Bowman et al. 2006). Other breeding programs from southeastern region include Coker Pedigreed Seed Company and McNair Seed Company. Continuous germplasm improvement and breeding efforts combined with improvements in management practices has led to great improvement in cotton lint yield through the cotton breeding history in United States.

### **Molecular marker resources in cotton**

Historically, germplasm characterization was based on morphological markers. But availability of a limited amount of morphological variations and variability in these traits caused by environment limited the level of resolution that could be achieved. The advent of molecular marker technology was a major breakthrough since these markers detected variation at DNA level without being affected by environment. A wide array of molecular markers is available in cotton such as: Restriction fragment length polymorphism (RFLP), Random amplified polymorphic DNA (RAPD), Amplified fragment length polymorphism (AFLP), Sequence related amplified polymorphism (SRAP), Resistance gene analogs (RGA), Simple sequence repeats (SSR) and Single nucleotide polymorphism (SNP) (Zhang et al. 2008).

The use of molecular markers for genetic studies in cotton started with RFLP markers (Meredith 1992). RFLP is a hybridization-based marker system based on presence or absence of a restriction site in the target sequence. RFLP markers have been used for estimation of genetic diversity (Becelaere et al. 2005) as well as linkage map construction in cotton (Rong et al. 2004; Ulloa et al. 2002; Waghmare et al. 2005). Even though these are robust markers, RFLP markers are highly labor intensive and require handling of radioactive material. Moreover, these markers need a large amount of DNA, making it difficult to use a large number of loci in any study. These challenges faced by early researchers were overcome by the advent of PCR (Polymerase Chain Reaction) based markers which were easier to use and required less DNA making it easier to scan a large number of loci throughout the genome with increased efficiency.

RAPDs are PCR based multi-locus markers which use short primer of 8-10 bp to amplify random DNA fragments (Williams et al. 1990; Powell et al. 1996). Polymorphism detected by these markers is based on variation at primer binding site as well as the length of DNA between two flanking primers. RAPD markers have been used to characterize genetic variation in elite cotton germplasm (Tatineni et al. 1996; Iqbal et al. 1997) and genetic mapping (Khan et al. 1999). Although RAPD markers are easy to use, they are not very reproducible across different studies.

AFLP are widely used dominant, multi-locus molecular markers which combines both hybridization and PCR based approach to detect polymorphism (Vos et al. 1995). A major advantage of AFLP is that they generate a large number of fragments per primer pair and therefore are very informative markers (Abdalla et al. 2001).

Among PCR based molecular markers SSRs have become highly popular for genetic studies in cotton. SSRs are codominant, easy to use, highly reproducible, highly polymorphic, and ubiquitous molecular markers. SSR primers are 18-25 bp long. These primers are complementary to flanking regions of microsatellites, which are tandem repeats of small core nucleotide units of 2-6nt (Saghai-Marooft et al. 1994). The variation in number of these tandem repeats gives rise to the polymorphism detected by SSR markers. The only disadvantage associated with these markers is the initial cost of developing primers, since designing SSR primers requires prior knowledge of DNA sequence. SSR markers can be developed by analyzing SSR enriched small insert genomic DNA libraries, from BAC libraries or through mining Expressed Sequence Tags (ESTs) (Zhang et al. 2008). Currently, a large number of cotton SSRs are publically available. SSR markers are highly automatable and can be labeled with fluorescent tags to allow multiplexing of multiple markers in a single reaction (Yu et al 2012). SSR markers are most commonly used molecular markers for genetic studies in cotton (Zhang et al. 2005 ab; Bertini et al. 2006; Fang et al. 2013; Song et al. 2005; Guo et al. 2007).

Recently, SNPs have become a popular marker system. They are the most abundant molecular markers that detect polymorphism at a single nucleotide level (Van Deynze et al. 2009). SNPs are co-dominant, robust, and highly automatable molecular markers. But, unlike multi-allelic SSR markers, SNPs have a disadvantage of being predominantly biallelic. The low allelic variability of SNP markers can however be compensated by their abundance in the genome. Previous studies have found that SSR markers perform better than SNP markers for genetic diversity studies (Hamblin et al. 2007; Yang et al 2011).

## **Breeding for increased yield**

Breeding for higher yield is one of the primary objectives of any breeding program. Cotton lint yield showed continuous improvement until the 1980s (Bridge et al. 1971; Bridge and Meredith 1983; Bassett and Heyer 1985; Culp and Green 1992). However since the 1980s, there has been a decline in rate of genetic gain in cotton lint yield (Meredith et al. 1997; Bayles et al. 2005; Meredith 2002). This decrease in breeding progress has been attributed to extensive use of backcross breeding strategy to create transgenic cultivars using very few recurrent parents as well as low level of genetic diversity present in Upland cotton germplasm pool (Meredith et al. 1997; Meredith 2006). However, the rate of genetic gain has improved during the early 21<sup>st</sup> century (Kerby and Hugie 2006; Kuraparthi and Bowman 2013).

In order to increase genetic gain and end the yield plateau associated with Upland cotton yield, more efficient management technologies should be developed along with continuous improvement of Upland cotton germplasm (Meredith et al. 1997). Another approach of increasing yield is to exploit hybrid vigor. Cotton hybrids show yield heterosis at both inter- and intraspecific levels (Marani 1967). The conventional method of producing hybrid seed is by manual emasculation and pollination of flowers. Due to high labor costs associated with conventional method, several studies have been conducted to identify male sterility systems in cotton (Lu et al. 1994; Wang and Li 2002). Attempts have also been made to exploit insect pollination for production of hybrid seed. China and India are two major cotton producers who are using cotton hybrids for increased yields (Paroda and Basu 1993;

Dong et al. 2006). There have been attempts to identify male sterility and fertility restorer systems but none have been successful so far (Wu et al. 2011).

Previous studies have found that an increase in yield is associated with boll number and high yielding lines have smaller bolls (Culp and Harrell 1975; Cole et al. 2009; Jenkins 2009). Lint yield in cotton is a multiplicative product of several yield components which share complex relationships among them. The lint yield model for cotton has been described as:  $\text{Yield} = \text{Boll number} \times \text{seeds boll}^{-1} \times \text{lint seed}^{-1}$  (Kerr 1966; Worley et al. 1974). Apart from genetics, environment has a significant effect on these yield components which affects yield (Guinn and Mauney 1984; Reddy et al. 1992; Cole et al. 2009). The influence of environment on hybrid yield has been found to be variable (Young and Murray 1966). Cole et al. (2009) showed that mid-parent heterosis is higher in low yielding environments than high yielding environments. Boll number was identified as the major yield component affecting this relationship (Cole et al. 2009). However, it was not identified if this change was due to more fruiting sites or higher retention of bolls by hybrids in low yielding environments (Cole et al. 2009).

Presence of sufficient genetic diversity is critical for genetic improvement in a breeding program. Genetic diversity provides a buffer against any biotic or abiotic stresses and is a source for new genes that could be used to overcome such challenges. Introgressions from wild or exotic germplasm often lead to unfavorable agronomic effects which limits the practical utility of wild resources (VanEsbroeck and Bowman 1998). Most of the genetic diversity studies using molecular markers have revealed a low level of diversity in Upland

cotton germplasm (Becelaere et al. 2005; Multani and Lyon, 1995; Iqbal et al. 1997; Campbell et al. 2009; Bertini et al. 2006; Tatineni et al. 1996; Abdurakhmonov et al. 2008).

Campbell et al. (2009) used 41 SSR to characterize genetic diversity among Pee Dee germplasm. The genetic similarity in Pee Dee germplasm was found to range from 0.64-0.96 (Campbell et al. 2009). The average genetic distance among New Mexico Acala germplasm was revealed as 0.19 ranging from 0.06-0.38 (Zhang et al. 2005a). Among the Acala germplasm from California, genetic diversity study using 88 SSR primer pair revealed average genetic similarity of 0.77 ranging from 0.69-0.93 (Zhang et al. 2005b). Although these studies provide information about the levels of genetic diversity in small subsets of Upland cotton accessions a comprehensive study is still needed to fully characterize the genetic structure and relationships in US cotton germplasm. A better understanding of genetic relationships and population structure would allow better utilization of germplasm resources and would help to incorporate un-used genetic variability that exists in Upland cotton. Objective of this study was to: (1) Estimate genetic diversity and population structure of Upland cotton germplasm and identify a core set of accessions which would represent the existing genetic diversity in a small subset of accessions (2) Identify the relationships of different yield components which affect hybrid yield from low yielding to high yielding environments.

## REFERENCES

- Abdalla AM, Reddy OUK, El-Zik KM, Pepper AE (2001) Genetic diversity and relationships of diploid and tetraploid cottons revealed using AFLP. *Theor Appl Genet* 102:222-229
- Abdurakhmonov IY, Kohel RJ, Yu JZ, Pepper AE, Abdullaev AA, Kushanov FN, Salakhutdinov LB, Buriev ZT, Saha S, Scheffler BE, Jenkins JN, Abdugarimov A (2008) Molecular diversity and association mapping of fiber quality traits in exotic *G. hirsutum* L. germplasm. *Genomics* 92:478-487
- Bassett DM, Heyer AA (1985) Acala cotton in California: 60 years of varietal improvement. In Proc Beltwide Cotton Prod Res Conf, New Orleans, LA. 6-11 Jan 1985 Natl Cotton Council Am, Memphis, TN pp 76-78
- Bayles MB, Verhalen LM, Johnson WM, Barnes BR (2005) Trends over time among cotton cultivars released by the Oklahoma Agricultural Experiment Station. *Crop Sci* 45:966-980
- Beasley JO (1940) The origin of American *Gossypium* species. *Amer Nat* 74:285-286
- Becelaere VG, Lubbers EL, Paterson AH, Chee PW (2005) Pedigree-vs. DNA marker-based genetic similarity estimates in cotton. *Crop Sci* 45:2281-2287
- Bertini CHCD, Schuster I, Sedyama T, Barros EG, Moreira MA (2006) Characterization and genetic diversity analysis of cotton cultivars using microsatellites. *Genet Mol Biol* 29:321-329
- Bird LS (1982) The MAR (Multi-Adversity Resistance) system. *Plant Disease* 66:172-176
- Bowman DT (1999) Public cotton breeder - Do we need them? *J Cotton Sci* 3:139-152
- Bowman DT, Gutierrez OA (2003) Sources of fiber strength in the US Upland cotton crop from 1980 to 2000. *J Cotton Sci* 7:164-169

Bowman DT, Gutierrez OA, Percy RG, Calhoun DS, May OL (2006) Pedigrees of upland and pima cotton cultivars released between 1970 and 2005. *Miss. Agric For Exp Stn Bull.* 1155

Bowman DT, May OL, Calhoun DS (1996) Genetic base of upland cotton cultivars released between 1970 and 1990. *Crop Sci* 36:577-581

Bridge RR, Meredith Jr WR (1983) Comparative performance of obsolete and current cultivars. *Crop Sci* 23:949-952

Bridge RR, Meredith Jr WR, Chism JF (1971) Comparative performance of obsolete varieties and current varieties of upland cotton. *Crop Sci* 11:29-32

Brubaker CL, Bourland FM, Wendel JF (1999) The origin and domestication of cotton. In CW Smith, JT Cothren, eds, *Cotton: Origin, History, Technology, and Production*. John Wiley & Sons, New York, pp 3–32

Campbell B, Williams V, Park W (2009) Using molecular markers and field performance data to characterize the Pee Dee cotton germplasm resources. *Euphytica* 169:285

Cole CB, Bowman DT, Bourland FM, Caldwell WD, Campbell BT, Fraser DE, Weaver DB (2009) Impact of heterozygosity and heterogeneity on cotton lint yield stability. *Crop Sci* 49:1577-1585

Culp T, Harrell D (1975) Influence of lint percentage, boll size, and seed size on lint yield of upland cotton with high fiber strength. *Crop Sci* 15:741-746

Culp TW, Green CC (1992) Performance of obsolete and current cultivars and Pee Dee germplasm lines of cotton. *Crop Sci* 32:35-41

Culp TW, Harrell DC (1973) Breeding methods for improving yield and fiber quality of Upland cotton (*Gossypium hirsutum* L.) *Crop Sci* 13:686-689

Dong J, Wu F, Jin Z, Huang Y (2006) Heterosis for yield and some physiological traits in hybrid cotton Chikangza 1. *Euphytica* 151:71-77

Fang DD, Hinze LL, Percy RG, Li P, Deng D, Thyssen G (2013) A microsatellite-based genome-wide analysis of genetic diversity and linkage disequilibrium in Upland cotton (*Gossypium hirsutum* L.) cultivars from major cotton-growing countries. *Euphytica* 1-11

Guinn G, Mauney JR (1984) Fruiting of cotton: II. Effects of plant moisture status and active boll load on boll retention. *Agron J* 76:94-98

Guo W, Cai P, Wang C, et al. (2007) A microsatellite-based, gene-rich linkage map reveals genome structure, function and evolution in *Gossypium*. *Genetics* 176:527–541

Hamblin MT, Warburton ML, Buckler ES (2007) Empirical comparison of simple sequence repeats and single nucleotide polymorphisms in assessment of maize diversity and relatedness. *PLoS One* 12:e1367

Iqbal MJ, Aziz N, Saeed NA, Zafar Y, Malik KA (1997) Genetic diversity evaluation of some elite cotton varieties by RAPD analysis. *Theor Appl Genet* 94:139-144

Jenkins JN, Gutierrez OA, McCarty JC, Wu J (2009) Genetic variance components and genetic effects among eleven diverse upland cotton lines and their F2 hybrids. *Euphytica* 167:397-408

Kerby TA, Hugie WV (2006) Yield and quality improvement of significant D&PL varieties during the last 25 years. In Proc Beltwide Cotton Conf, San Antonio, TX 3-6 Jan 2006 Natl Cotton Counc Am, Memphis, TN pp 858-865

Kerr T (1960) The trispecies hybrid ancestry of high strength cottons. *Proc Cotton Imp Conf* 21:82

Kerr T (1966) Yield components in cotton and their interrelations with fiber quality. In Cotton Improvement Conference, 11-12 Jan 1966, (18th) Natl Cotton Counc Am, Memphis, TN pp 276-287

Khan MA, Stewart JM, Zhang J, Myers GO, Cantrell RG (1999) Addition of new markers to the tri-specific cotton map. In Proc Beltwide Cotton Conf, Orlando, FL, 3-7 Jan 1999, Natl Cotton Counc Am, Memphis, TN pp 439

Kuraparthi V, Bowman DT (2013) Gains in breeding Upland cotton for fiber quality. J Cotton Sci (in press)

Liu D, Guo X, Lin Z, Nie Y, Zhang X (2006) Genetic diversity of asian cotton (*Gossypium arboreum* L.) in china evaluated by microsatellite analysis. Genet Res Crop Evol 53:1145-1152

Lu YQ, Wang B, Ren SM, He Y (1994) Development of genic male sterile line, Dong-A and its utilization in hybrid seed production. China cottons 21:8-9

Marani A (1967) Heterosis and combining ability in intraspecific and interspecific crosses of cotton. Crop Sci 7:519-522

May OL (2001) Registration of PD 94045 germplasm line of Upland cotton. Crop Sci 41:279-280

Meredith Jr WR (2002) Factors that contribute to lack of genetic progress. In Proc Beltwide Cotton Conf, Atlanta, GA 8-13 Jan 2002 Natl Cotton Counc Am, Memphis, TN pp 528-532

Meredith Jr WR (2006) Obsolete conventional vs. modern transgenic cultivars performance evaluations. In Proc Beltwide Cotton Conf, San Antonio, TX 3-6 Jan 2006 Natl Cotton Counc Am, Memphis, TN. pp 836-844

Meredith Jr WR, Heitholt JJ, Pettigrew WT, Rayburn ST (1997) Comparison of obsolete and modern cotton cultivars at two nitrogen levels. Crop Sci 37:1453-1457

Meredith WR (1992) RFLP association with varietal origin and heterosis. In Proc Beltwide Cotton Conf, Nashville, TN 6-10 Jan Natl Cotton Counc Am, Memphis, TN pp 607

Multani DS, Lyon BR (1995) Genetic fingerprinting of Australian cotton cultivars with RAPD markers. *Genome* 38:1005-1008.

Niles GA, Feaster CV (1984) Breeding. In R. J. Kohel and C. F. Lewis [eds.], *Cotton, Agronomy Monograph No. 24*, Crop Sci Society of Am, Madison, WI pp 201- 231

Paroda RS, Basu AK (1993) Cotton scenario with particular reference to hybrid cotton. In Proceedings of the FAO-ICAR regional expert consultation on hybrid cotton. Oct 22-25, 1990. Nagpur, India: CICR pp 21-46

Rafalski JA (1997) Randomly amplified polymorphic DNA (RAPD) analysis. In:Caetano-Anolles G, Gresshoff PM (eds.). *DNA markers: Protocols, Applications, and Overviews*. Wiley-Liss, Inc. USA

Reddy KR, Hodges HF, Reddy VR (1992) Temperature effects on cotton fruit retention. *Agron J* 84:26-30

Rong JK, Abbey C, Bowers JE et al. (2004) A 3347-locus genetic recombination map of sequence-tagged sites reveals types of genome organization, transmission and evolution of cotton *Gossypium*. *Genetics* 166:389–417

Saghai Maroof MA, Biyashev RM, Yang GP, Zhang Q, Allard RW (1994) Extraordinarily polymorphic microsatellite DNA in barley: species diversity, chromosomal locations, and population dynamics. *Proc Nat Acad Sci* 91:5466–5470

Smith CW, Cantrell RG, Moser HS, Oakley SR (1999) History of cultivar development in the United States. In: Smith CW, Cothren JT (eds) *Cotton: origin, history, technology, and production*. John Wiley & Sons Inc, New York, pp 99–171

- Song X, Wang K, Guo W, Zhang J, Zhang T (2005) A comparison of genetic maps constructed from haploid and BC1 mapping populations from the same crossing between *Gossypium hirsutum* L. and *Gossypium barbadense* L. *Genome* 48:378–390
- Staten G (1970) Breeding Acala 1517 cottons, Memoir Series No. 4, New Mexico State University, Las Cruces. 1926-1970
- Tatineni V, Cantrell RG, Davis DD (1996) Genetic diversity in elite cotton germplasm determined by morphological characteristics and RAPDs. *Crop Sci* 36:186-192
- Thaxton PM, El-Zik KM (2003) Registration of Eleven Multi-Adversity Resistant (MAR-7A) Germplasm Lines of Upland Cotton. *Crop Sci* 43:741-742
- Ulloa M, Meredith WR, Shappley ZW, Kahler AL (2002) RFLP genetic linkage maps from four F2.3 populations and a joinmap of *Gossypium hirsutum* L. *Theor Appl Genet* 104:200–208
- Van Esbroeck GA, Bowman DT (1998) Cotton improvement. Cotton germplasm diversity and its importance to cultivar development. *J of Cotton Sci* 2: 121-129
- Vos P, Hogers R, Bleeker M, Reijans M, Van T, Hornes M, Frijters A, Pot J, Peleman J, Kuiper M, Zabeau M (1995) AFLP: a new technique for DNA fingerprinting. *Nucleic Acids Res* 23:4407–4414
- 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
- Wang XD, Li YY (2002) Development of transgenic restorer of cytoplasmic male sterility in upland cotton. *Sci Agric Sin* 35:141-145
- Wendel J, Brubaker C, Percival A (1992) Genetic diversity in *Gossypium hirsutum* and the origin of upland cotton. *Am. J. Bot.* 79:1291-1310

Williams JGK, Kubelik AR, Livak KJ, Rafalski JA, Tingey SV (1990) DNA polymorphism amplified by arbitrary primers are useful as genetic markers. *Nucleic Acids Res* 18:6531-6535

Worley S, Ramey HH, Harrell DC, Culp TW (1976) Ontogenetic model of cotton yield. *Crop Sci* 16:30-34

Wu J, Gong Y, Cui M, Qi T, Guo L, Zhang J, Xing C (2011) Molecular characterization of cytoplasmic male sterility conditioned by *Gossypium harknessii* cytoplasm (CMS-D2) in upland cotton. *Euphytica* 181:17-29

Young E, Murray J (1966) Heterosis and inbreeding depression in diploid and tetraploid cottons. *Crop Sci* 6:436-438

Yu JZ, Fang DD, Kohel RJ, Ulloa M, Hinze LL, Percy RG, Zhang J, Chee P, Scheffler BE, Jones DC (2012) Development of a core set of SSR markers for the characterization of *Gossypium* germplasm. *Euphytica* 187:203-213

Zhang H, Li Y, Wang B, Chee PW (2008) Recent advances in cotton genomics. *Int J Plant Genomics* DOI:10.1155/2008/742304

Zhang JF, Lu Y, Adragna H, Hughs E (2005a) Genetic improvement of New Mexico Acala cotton germplasm and their genetic diversity. *Crop Sci* 45:2363-2373

Zhang JF, Lu Y, Cantrell RG, Hughs E (2005b) Molecular marker diversity and field performance in commercial cotton cultivars evaluated in the Southwest U.S. *Crop Sci* 45:1483-1490

## **CHAPTER I**

### **Genetic diversity and population structure in the US Upland cotton (*Gossypium hirsutum* L.)**

## ABSTRACT

Elite plant breeding programs could likely benefit from the unexploited standing genetic variation of obsolete cultivars without the yield drag typically associated with wild accessions. A set of 381 accessions comprising 378 Upland (*Gossypium hirsutum* L.) and 3 *Gossypium barbadense* L. accessions of the United States cotton belt were genotyped using 120 genome-wide SSR markers to establish the genetic diversity and population structure in tetraploid cotton. These accessions represent more than 100 years of Upland cotton breeding in the United States. Genetic diversity analysis identified a total of 546 alleles across 141 marker loci. Twenty two percent of the alleles in Upland accessions were unique specific to a single accession. Population structure analysis revealed extensive admixture and identified five subgroups corresponding to Southeastern, Midsouth, Southwest and Western zones of cotton growing areas in United States, with the three accessions of *G. barbadense* forming a separate cluster. Phylogenetic analysis supported the subgroups identified by STRUCTURE. Average genetic distance between *G. hirsutum* accessions was 0.195 indicating low levels of genetic diversity in Upland cotton germplasm pool. The results from both population structure and phylogenetic analysis were in agreement with pedigree information although there were a few exceptions. Further, core sets of different sizes representing different levels of allelic richness in Upland cotton were identified. Establishment of genetic diversity and population structure of the diversity panel and identification of core sets in this study could be useful for genetic and genomic analysis and systematic utilization of the standing genetic variation in Upland cotton.

## INTRODUCTION

Cotton is the leading natural fiber crop. Cotton belongs to the genus *Gossypium*, which has extensive phenotypic diversity among the approximately 50 species representing this genus (Campbell et al. 2010). Worldwide, four species are cultivated; two of these cultivated species are diploids ( $2n = 2x = 26$ ) and two are allotetraploids ( $2n = 4x = 52$ ). Most of the global cotton production comes from the two tetraploid species, *G.hirsutum* and *G. barbadense* (Wendel et al. 1992).

Allotetraploid cotton evolved approximately 1.5 million years ago from a hybridization event between Old world cotton *G. herbaceum* (A1 genome) and New world cotton *G.raimondii* (D5 genome) followed by subsequent diploidization, creating five tetraploid species (Wendel et al. 1992; Brubaker et al. 1999). *G. hirsutum* also called Upland cotton represents 95% of global cotton fiber production. *G. barbadense* (also known as Pima cotton) is valued for its higher fiber quality and contributes around 3% of global cotton production. The other three tetraploid species (*G. mustelinum*, *G. darwinii* and *G. tomentosum*) are wild and are not grown commercially (Wendel et al. 1994; Wendel and Percy 1990; DeJoode and Wendel 1992).

Domesticated in the Yucatan peninsula about 5,000 years ago, (Wendel et al. 1992) primitive *G. hirsutum* strains were photoperiod sensitive. During the course of domestication day-neutral stocks were selected which allowed cotton to be eventually grown in the U.S. Cotton cultivation in the U.S. dates back to the early seventeenth century (Smith et al. 1999). One of the most important events in US cotton breeding history was the introduction of

Mexican highland stocks in the early 1800s which contributed to the foundation of current Upland germplasm (Wendel et al. 1992). Several introductions of germplasm imported from outside the US were also incorporated into different breeding programs. Further development of cultivars was directed by the need for locally adapted cultivars and events like the Boll weevil (*Anthonomous grandis* Boh.) outbreak that necessitated the demand for early maturing varieties (Niles and Feaster, 1984). The cotton growing area in the US can be broadly divided into four regions: Western, Southwestern plains, Mid-south or the delta, and Southeast. Thus, cotton breeding programs developed cultivars that were specifically adapted to those regions.

Within species, *G. hirsutum* shows great phenotypic diversity (Wendel et al. 1992). Level of genetic diversity within *G. hirsutum* has been found to be higher than the other three cultivated cotton species (Wendel et al. 1992; Abdurakhmonov 2012). Yet studies have indicated that this diversity is not represented in the present cultivated germplasm of Upland cotton (Van Esbroeck et al. 1999). Apart from the initial bottleneck encountered during domestication process, cotton breeding has frequently involved crossing and re-selections within small sets of breeding materials which has led to the loss in genetic diversity (May et al. 1995; Bowman et al. 1996; Wendel et al. 1992; Brubaker et al. 1999). The narrow genetic base of Upland cotton has become a serious concern since limited genetic diversity translates to limited allelic availability for continued genetic gain (Brown, 1983). With a heightened risk of genetic vulnerability to disease epidemics and climate change, elite breeding programs could benefit from the unexploited standing genetic variation of obsolete cultivars

without the yield drag typically associated with wild accessions. It is also noted that even within the domesticated Upland cotton, unfavorable agronomic effects were observed when unadapted germplasm from a different area is used in a breeding program (Van Esbroeck and Bowman 1998). By characterizing genetic diversity between and within groups, breeding efforts can be greatly improved through better parental selection for generating segregating populations. Genetic diversity information is also helpful to identify heterotic groups, understand population structure and identify a core set of lines for genetic analysis studies. Thus, assessment of genetic diversity and population structure is important in the US Upland cotton.

Genetic diversity estimates have been made using pedigree and morphological data (May et al. 1995; Van Esbroeck et al. 1999), biochemical markers (Wendel et al. 1992) and DNA based molecular markers (Yu et al. 2012). In the pedigree-based studies, relationship between two accessions strongly depends on the availability of breeding records and validity of certain assumptions. In absence of such information, pedigree based methods cannot be used to accurately estimate genetic diversity. This is especially true of ancestral lines or introductions, for which detailed breeding records are not available. They are usually assumed to be equally unrelated even if otherwise, this often leads to overestimation of diversity (Bowman et al. 1996; Van Esbroeck et al. 1999). Molecular markers, on the other hand, are more reliable and informative since they can directly measure allelic diversity and give robust estimates of genetic distances. A multitude of DNA based markers systems including Restriction Fragment length polymorphism (RFLP) markers (Becelaere et al.

2005), Random Amplified Polymorphic DNA (RAPD) markers (Multani and Lyon, 1995; Iqbal et al. 1997), Amplified Fragment Length Polymorphism (AFLP) markers (Abdalla et al. 2001), Simple Sequence repeat (SSR) markers (Zhang et al. 2011), ISSR (inter-simple sequence repeat) markers (Liu and Wendel, 2001) and EST-SSRs (Mehboob et al. 2008) were used for measuring genetic diversity in cotton. SSRs have proven to be a very good marker system due to their codominant nature, reproducibility and convenience of use (Powell et al. 1996).

Although several studies employed marker-based estimation of genetic diversity in cotton, most of these studies are limited in number of accessions included or the number of markers used to characterize genetic diversity. Some of these studies have been conducted using germplasm specific for a breeding program, for example, the Pee Dee program (Campbell et al. 2009), New Mexico Acala program (Zhang et al. 2005) or cultivars from a specific geographical area such as cultivars grown in Greece (Kalivas et al. 2011), Brazil (Bertini et al. 2006), or China (Liu et al. 2006), whereas others are more focused towards exotic material (Abdurakhmonov et al. 2008) and interspecific relationships (Abdalla et al. 2001). Besides, all these studies used gel-based platforms for resolving marker allele diversity. However, capillary-based resolution of amplified product is more effective in separating different alleles than agarose gel-based systems, thus increasing the efficiency and utility of genetic diversity and population structure studies. Therefore, a comprehensive study involving a broad collection of germplasm and more efficient genotyping platforms is still needed to quantify overall genetic diversity in Upland cotton for its effective utilization

in breeding and genetical genomics in cotton. The objectives of this study were to (1) estimate US upland cotton genetic diversity; (2) analyze population structure; and (3) identify core sets of lines that maximize allelic diversity in the Upland cotton germplasm.

## **MATERIALS AND METHODS**

### **Plant material**

We sampled 378 *G. hirsutum* accessions, covering genotypes from all 14 states that constitute the US cotton belt. The selected accessions represent cotton cultivars from the early 1900s to 2005. The selected accessions include most of the important lines that have been used as parents in different breeding programs. Also included were 11 parents used to generate a Random mated population in Upland cotton (Jenkins et al. 2008). In addition to released public breeding lines and cultivars, the set also includes obsolete accessions from Delta and Pine Land Company, Stoneville, and Coker's Pedigreed Seed Company. Most of the seed material was obtained from the US National Cotton Germplasm Collection, USDA-ARS, College Station, TX. Three accessions of *G. barbadense* were also included as out-group in the complete panel. Detailed information about the 381 lines used in this study is provided in Supplemental Table S1. To reduce residual heterozygosity within the accessions, all entries were selfed for three generations with single plant selections at the Central Crops Research Station, Clayton, NC during the summers of 2010 and 2011.

## **SSR genotyping**

Leaf tissue was collected from a single plant of each accession and DNA was extracted using the procedure as described by Li et al. (2001). As a preliminary study we used a panel of 12 genotypes to identify SSR markers that gave reproducible amplification and could be confidently scored (data not shown). Out of 160 SSR markers initially tested, 135 were selected to genotype the whole panel. These selected markers are uniformly distributed across the genome, with a minimum of four markers per chromosome. Primer sequences for all SSR markers are publically available and were obtained from Cotton Marker Database now housed in CottonGen (<http://www.cottongen.org>). Supplemental Table S2 includes the list of 135 SSR primers with their repeat motif and chromosomal locations as reported in literature. All forward primers were modified by adding a M13 sequence of 19 bases to their 5'end. A fluorescent 6-FAM or HEX labeled M13 primer was separately added to the PCR mix to generate fluorescent-labeled amplified product. PCR reaction volume was 6  $\mu$ l and reaction mix included 20 ng DNA, 0.2 mM dNTP mix; 0.08  $\mu$ M modified forward primer, 0.6  $\mu$ M reverse primer and fluorescent-labeled M13 primer each and 0.6 unit of Taq polymerase with 1x reaction buffer containing 15mM MgCl<sub>2</sub>. All primers were amplified using a Touchdown protocol with amplification conditions as follows: 95° C for 5 min, 15 cycles of 94° C for 45 sec, 65° C for 45 sec with a reduction of 1° C per cycle, 72° C for 1 min, followed by 25 cycles of 95° C for 5 min, 50° C for 45 sec, 72° C for 1 min, with a final step of 72° C for 10 min. Amplified products were separated on an ABI 3730 capillary electrophoresis system (Applied Biosystems Inc., Foster City, CA, USA) with GeneScan™

500 LIZ<sup>®</sup> used as internal size standard. GeneMarker software version 1.91 (Softgenetics, LLC, State College, PA, USA) was used to analyze ABI output. Amplicons with different fluorescent labels were multiplexed during ABI runs to increase throughput.

### **Preliminary analysis of genotypic data and genetic diversity**

Since upland cotton is an allotetraploid crop with two different genomes it is possible that some markers could produce amplicons from both genomes giving rise to multilocus data. We employed two criteria for separating such multilocus data into different loci. First, if one of the amplicons was monomorphic across all entries it was considered to be an individual locus. Second, because all entries were selfed for three generations we did not expect high residual heterozygosity, so amplicons were separated into two different loci to reduce overall heterozygosity for the marker. Basic summary statistics for biallelic data were calculated using POWERMARKER software package version 3.25 (Liu and Muse, 2005). The polymorphism information content (PIC) of SSR marker was determined according to the method described by Botstein et al. (1980). A PIC value of 1 indicates that the marker can differentiate each line, and 0 indicates a monomorphic marker. Informative potential of a marker is high if its PIC value is more than 0.5, moderate if PIC is between 0.5 and 0.25 and only slightly informative if PIC value is below 0.25. Other statistics calculated were number of alleles, allele diversity and heterozygosity for each marker.

POWERMARKER software package version 3.25 was used to calculate pairwise genetic distance between the accessions using Nei et al.'s (1983)  $D_A$  distance. The distance matrix was used to construct a dendrogram using neighbor joining method in POWERMARKER

software. Dendroscope version 3.2.2 was used to visualize and edit the dendrogram. Further analysis of genetic structure was done by means of Principal co-ordinate analysis (PCA). Dominant data (0, 1 binary data) was used for PCA analysis in NTSYS-pc software version 2.2 using DCENTER and EIGEN functions (Rohlf 2000). Partitioning of genetic variance among and within groups was performed using Arlequin ver 3.5 software (Excoffier and Lischer 2010).

### **Analysis of genetic structure**

STRUCTURE software version 2.3.4 (Pritchard et al. 2000) which is a model-based Bayesian method was used to delineate 381 accessions into clusters of individuals based on co-dominant genotypic data. Admixture model was used with the option of correlated allele frequencies between populations for Structure analysis as recommended by Falush et al. (2003). An admixture model assumes that individuals may also have inherited a fraction of their genome from its ancestors in a different subpopulation thus having a mixed ancestry. The degree of admixture ( $\alpha$ ) was estimated from data. A value of  $\alpha$  near 0 indicates no admixture is present whereas  $\alpha$  more than 1 suggests that most of the individuals are admixed (Falush et al. 2003). Ten runs were conducted for each value of number of populations (K), with K ranging from 2 to 12. The length burn-in and number of replications were 10,000 each.

The number of sub-populations was estimated using the method proposed by Evanno et al. (2005) by plotting a distribution of  $\Delta K$ , an ad hoc quantity based on second order rate of change of the likelihood function with respect to K. The value of  $\Delta K$  was calculated as the

mean of absolute values of difference between successive likelihood values of  $K$  divided by the standard deviation of  $L(K)$ . The modal value of this distribution of  $\Delta K$  best represents the underlying value of  $K$ , which is the uppermost hierarchical level of population structure. This method is also successful in identifying the true value of population number when there is little genetic differentiation between populations (Evanno et al. 2005). Structure harvester software was used to estimation  $\Delta K$ . Accessions were assigned to a subgroup if the probability of membership was greater than 70%. If membership was less than 70%, then the accessions were assigned to the mixed subgroup.

### **Construction of core sets**

The genotypic data were used to identify core sets of accessions that maximize genetic diversity in a limited number of accessions. Phenotypic data were collected for days to flowering, node of first fruiting branch and fiber quality at three locations (data not presented). In order to identify a core set, any accessions with unfavorable traits like mutant phenotype, colored lint or late maturity were removed from analysis. Also, accessions with more than 1% introgression from *G. barbadense* based on membership probability estimates from population structure analysis were eliminated from analysis. This reduced the number of accessions in final set to 324. Core sets of lines were assembled by maximizing allelic richness using a simulated annealing algorithm in POWERMARKER software package using the following parameters,  $R=2500$ ,  $p=0.95$  and  $T_0 = 1$ . Core sets of different sizes, ranging from  $k=8$  to  $k=53$  were assembled, in increments of 5 accessions.

## RESULTS

### SSR marker analysis

Out of 135 SSR markers used for genotyping, 12 were found to be monomorphic, whereas 3 markers could not be scored with confidence. These 15 markers were dropped from analysis, leaving SSR data for 120 markers for 381 lines with less than 5% missing data. Due to a polyploid genome, SSR primer pairs generating multilocus markers are frequently observed in cotton (Fang et al. 2013). In the panel of accessions used in present study, 17.5% of the polymorphic primer pairs generated multilocus markers, thus identifying 141 polymorphic and 21 monomorphic loci across the panel. Data for monomorphic loci were excluded from data analysis.

Preliminary results show that these 381 lines (including *G. barabdense* accessions?) are highly homozygous (average  $H = 2.3\%$ ). Among 381 accessions a total of 546 SSR alleles were detected across 141 loci with an average of 3.9 alleles per SSR locus. Only 120 of the 141 SSR loci were found to be polymorphic exclusively among *G. hirsutum* accessions. For *G. hirsutum* accessions, observed heterozygosity was 2.6%. Total 367 alleles were generated at 120 loci with an average of 3.1 alleles per SSR locus. Average PIC value for SSRs was 0.17 for Upland cotton accessions and 0.16 for the complete panel. A summary of marker statistics for *G. hirsutum* accessions is presented in table 1.1.

## Unique alleles

Out of 546 alleles detected in the complete panel, 134 were unique, representing alleles found in only one accession (Table 1.2). Most (78%) of these unique alleles were present in *G. barbadense* accessions (Table 1.2). Most of the alleles had very low allele frequency (Fig. 1.1a and 1.1b). Among *G. hirsutum* accessions, 80 unique alleles were observed in 54 accessions (Table 1.3). The maximum number of unique alleles present in a single cultivar was seven, found in Rugose Indore. Eight of the 54 accessions that have unique alleles were introductions. Of the other 46 accessions, 14 were from mid-south or delta region, 14 from Southeast, 15 from Southwest or Texas plains and only three were from Western region representing Acala breeding program. Release dates of these accessions with unique alleles ranged from early 1900s to 1998. Interestingly, among accessions from delta and eastern region more accessions were released prior to 1980. However, out of 15 accessions from plains region that were carrying unique alleles, 10 were released after 1980. These results are indicative of an increase in genetic diversity in the accessions from plains over the years.

## Population Structure

Analysis of population structure was performed in the complete set of 381 accessions using software STRUCTURE. The number of subpopulations could not be identified from the plot of Ln (probability of data) for K (Fig. 1.2a). However, the number of clusters was estimated to be five (Fig. 1.2b) based on  $\Delta K$  value. Substantial admixture was found to occur between clusters (Fig. 1.3). Accessions with  $\geq 0.70$  membership probabilities were included in a cluster as members (Liu et al., 2003). Accessions without 70% membership to a cluster were

considered as mixed. Out of 381 accessions, 184 could be assigned to subgroups based on 70% membership threshold, meaning that more than half of cultivars are considered to have admixed parentage. Detailed description of membership probabilities of individual accessions is presented in Supplemental table S3.

Accessions from *G. hirsutum* were separated into four groups based on STRUCTURE results. (Fig. 1.3). Identified groups roughly correspond to the four zones of cotton breeding belt in United States. Group 1 (indicated red in figure 1.3) includes 36 accessions. These accessions have Acala germplasm in their pedigrees and most belong to the western zone of the Cotton Belt. Tashkent 1 which is an introduction from Uzbekistan was also included as a member of this cluster. Inclusion of Lone Star in group 1 was surprising and could not be explained. Lone Star which was developed from Jackson Round Boll, is a historically important cultivar being the founder line for Stoneville cultivars. Group 2 (indicated green in figure 1.3) includes 38 accessions mostly from eastern cotton belt. Most of the accessions included in group 2 were adapted to southeast zone of cotton growing are in United States and included Sealand accessions and lines from Coker breeding program. Group 2 had highest number of unique alleles detected within a group (Table 1. 4). Group 3 (indicated blue in figure 1.3) mostly includes accessions from the southwest region of cotton belt representing cotton breeding program for plains cotton. The accessions from Multi-Adversity Resistance program from Texas are included in group 3. Group 4 (indicated yellow in figure 1.3) primarily includes the accessions from midsouth or delta region. These include accessions from Delta and Pine Land Company and Stoneville Pedigreed Seed Company

which are the major breeding programs for this area. Group 5 (shown pink in figure 1.3) has the three accessions from *G. barbadense*, all grouped together in a distinct group with more than 99% membership probabilities.

### **Genetic diversity and phylogenetic analysis**

Neighbor joining analysis of genotypic data for *G. hirsutum* accessions indicated that the average genetic distance was 0.195 and ranged from 0.00 to 0.37. The highest genetic distance of 0.37 was between M4 and Tashkent 1. Among the four groups of accessions identified on based of membership probabilities, group 2, which included accessions from southeastern United States had the highest average genetic distance of 0.186 (Table 1.5). Using the distance matrix, a phylogenetic tree was constructed. Six major clusters were identified in the NJ tree. In order to see how STRUCTURE results correspond to the phylogenetic analysis, the dendrogram was colored to show STRUCTURE grouping (Fig. 1.4). Four clusters of *G. hirsutum* accessions identified in STRUCTURE formed distinct clusters in the tree. Overall there is good agreement between the two estimates. Clustering pattern was found to be in general agreement with relationships based on pedigree studies (Bowman et al. 2006). Although, in some cases lines that were selections or direct descendants from another lines were not grouped close to their parents, which is contradictory to the pedigree information. Still, in most cases they were still in the same major cluster. For example, Lankart 611 grouped with Stoneville 5A instead of grouping with Lankart 57 of which it is a direct descendent (Bowman et al. 2006). Deltapine 14 and TM1 (a selection from Deltapine 14) were also in the same major cluster but in different

subgroups. Phylogenetic tree with accession identifiers is provided online as Supplement Fig. S1.

Genetic relationships between *G. hirsutum* accessions were further studied by using Principal coordinate analysis (Fig. 1.5). The first two axes of PCO accounted for 59.2% of the variation. This indicates low level of genetic diversity in *G. hirsutum* germplasm with continuous variation between the subgroups. Analysis of molecular variance (AMOVA) revealed highly significant variation between the four groups identified by Structure analysis with 31.4% of the total variation contributed to between group differences (Table 1.6). However, a larger amount of variation (65.84%) was due to diversity within the groups (Table 1.6). Overall population differentiation estimate ( $F_{ST}$ ) among the groups was 0.342 ( $p < 0.0001$ ). Pairwise  $F_{ST}$  revealed that accessions from group 2 (eastern region) are closer to accessions from midsouth (group 4) and southwest region (group 3) as compared to western Upland accessions represented mostly by Acala germplasm (Table 1.7). Highest genetic differentiation was observed between accessions from western (group 1) and midsouth (group 4) regions of Upland cotton growing area with a pairwise  $F_{ST}$  of 0.4196 ( $p < 0.0001$ ) (Table 1.7).

### **Core sets of Upland cotton diversity panel**

Genotypic data from 120 SSR loci was analyzed to identify core sets of accessions based on allele number. Core group selection was constrained to positively include Acala Maxxa (as group I representative), Wannamaker Cleveland (as group II representative), Dixie King (as group III representative), Deltapine 14 (as group IV representative), DES56 (selected as

female parent for population development) and Paymaster HS200. A plot of percentage of alleles in 324 *G. hirsutum* accessions captured in different core set sizes is shown in Fig.1. 6. The smallest set with 8 accessions captures 64%, whereas the largest set of 53 accessions captures 96% of all alleles detected using 120 SSR markers on 324 accessions (Fig. 1.6). Complete list of accessions included in different sets is presented in table 1. 8. A core set of 23 accessions represents 84% of the alleles in 324 accessions and 74% of alleles identified in complete panel of 381 accessions.

## DISCUSSION

In the current study, 120 SSRs produced 141 polymorphic loci in the complete panel of 381 accessions and 120 polymorphic loci in a panel of 378 Upland cotton accessions. An average of 3.1 alleles was amplified per locus in *G. hirsutum* accessions and 3.9 alleles per loci for the complete panel. Similar results were observed by Hinze et al. (2012) on allele number using a smaller set of SSR markers on improved US cotton germplasm. However, other studies showed variable allele number per locus. For example, Bertini et al. (2006) used 31 SSR primers to characterize 53 cultivars and reported 2.13 alleles per SSR locus. While 80 SSR primer pairs amplified 4.2 alleles per SSR across 72 Pee Dee lines (Campbell et al. (2009). Few studies have reported more alleles amplified per marker, for example, 5.8 alleles per primer were detected in a panel of 59 cotton cultivars of China (Zhang et al. 2011), and 5.6 alleles per marker were detected in a study using landraces and wild accessions of *G. hirsutum* (Lacape et al. 2007). The higher allele number observed in these studies is more likely a result of diverse germplasm used. Overall, the number of alleles observed per marker

depends upon the selection of markers, collection of germplasm to be genotyped as well as the platform used for resolution of amplified products (Lacape et al. 2007). Fewer alleles per locus in Upland cotton is similar to the trend observed in other self-pollinated crops like rice (*Oryza sativa*) with 3.14 to 5.1 alleles per locus (Garris et al. 2005) and wheat (*Triticum aestivum* L.) with 4 alleles per locus (Oliveira et al. 2012). While the average alleles per locus for cross-pollinated crops are higher, for example, 14.1 alleles per locus reported in alfalfa (*Medicago sativa*) (Flajoulot et al. 2005) and 21.7 alleles per locus in maize (*Zea mays* L.) (Liu et al. 2003).

In the current study average PIC was 0.17, whereas in literature average PIC value for cotton SSRs can range from 0.122 (Abdurakhmonov et al. 2008) to 0.80 (Zhang et al. 2011). Lower number of alleles per locus and low PIC values in Upland cotton as shown in the current study further substantiates previous reports on narrow genetic base in cotton (Zhang et al. 2005; Campbell et al. 2009). We identified 22% unique alleles in the elite germplasm (Table 1.2) which is much higher than 3% reported in an earlier study (Abdurakhmonov et al. 2008). The higher amount of unique alleles observed in our study could be due to better resolution of marker alleles achieved on a capillary-based platform and use of a broader set of representative accessions. Among the cultivars released after 1998 none had any unique alleles (Table 1.3). Fang et al. (2013) also reported that the average number of unique alleles in US cultivars has declined from 0.53 in 1899-1950 to 0.24 in 1981-present. These findings indicate the trend in continued decline in genetic diversity over time in Upland cotton. It is interesting to note that the least number of unique alleles were found in Acala cultivars

(Table 1. 4) while previous studies have reported Acala cottons to be more diverse than delta or southeastern cultivars based on coefficient of parentage analysis (May et al. 1995).

Model based population structure analysis identified four differentiated sub-populations in Upland cotton accessions congruent with major cotton growing regions: Western, Southwestern, Midsouth and Eastern (Fig. 1.3, Supplementary table S3). One hundred and ninety one accessions were assigned to mixed group indicating significant admixture (Supplementary table S3). This admixture is possibly a result of germplasm sharing among different breeding programs. Another reason could be higher frequency of appearance of a few lines with favorable agronomic traits in multiple breeding programs (Van Esbroeck et al. 1998). For example, between 1970 and 1990, Stoneville, Coker and New Mexico lines were found in pedigrees of lines of other breeding programs more frequently than other lines (Bowman et al. 1996; Kuraparthi and Bowman 2013).

The Western group of accessions included lines that had Acala germplasm in their pedigree (Supplementary table S3). Original Acala accession was introduced in 1907 and breeding for better fiber quality traits helped shape the family of Acala cottons (Staten 1970). One of the Paymaster lines, PaymasterHS26 was also included in this group, its pedigree Acala SJ-4/5B9-184 shows that it has Acala germplasm and thus may explain this grouping (Bowman et al. 2006). The Southeastern group of accessions had representation of two different breeding programs, germplasm from Coker Seed Company and Pee Dee program. A large number of Pee Dee lines were included in this study. But most of the Pee Dee lines developed post boll weevil era could not be assigned to a group and were considered mixed.

This finding seems acceptable given the ancestry of these lines, since, Pee Dee germplasm was derived from a complex series of crosses using Triple hybrid, Sealand, Earlistaple, C 6-5, AHA, Dixie King, Auburn 56 and Coker 421 (Bowman et al. 2006). Early Sealand cultivars that were long staple Upland cotton cultivars adapted to southeastern US were also included in group 2 based on membership probability (Supplementary table S3). The third group included lines from region west to the Mississippi delta and Texas plains. Only two Paymaster cultivars which were bred for Texas high plains were included in this group. Paymaster HS26 clustered with group 1 whereas Paymaster 54 clustered with group 4. Other six cultivars from Paymaster Seed Company were assigned to the mixed group. These results were not surprising given that most of the Paymaster cultivars were mixed in ancestry (Smith et al. 1999). Historically, these cultivars were developed from a series of complex crosses between accessions from delta region (Stoneville), Acala cottons, Macha as well as Kekchi which was introduced from Guatemala (Smith et al. 1999).

A phylogenetic tree made using genotypic data broadly corroborated grouping of accessions detected by STRUCTURE (Fig. 1.4). The estimates of genetic distance ( $GD = 0.195$ ) reveal overall level of genetic diversity to be low, which is a finding similar to earlier reports (Zhang et al. 2005; Campbell et al. 2009; Fang et al. 2013). Most of the accessions in mixed group were located between major clusters in the Neighbor-joining tree (Fig. 1.4). There was good agreement between this study and pedigree information. However, for some accessions there were discrepancies between pedigree information and marker based relationships. Similar observations have been made in previous studies where discrepancies

were observed between pedigree information and genetic relationships based on SSR markers (Zhang et al. 2005; Fang et al. 2013). Genetic diversity within the group was lowest for Midsouth group and highest for eastern group (Table 1.5). High genetic diversity in the Eastern germplasm is probably due to several introductions as well as use of hybridization rather than only reselection for germplasm development. Eastern accessions in group 2 were closer to accessions from Midsouth or Southwest than Western accessions.

The differentiation between groups was further validated by high  $F_{ST}$  value, with 34% of the marker variation being explained by population structure of Upland cotton germplasm (Table 1.6).  $F_{ST}$  values for cotton observed in this study (0.29 – 0.42) are closer to another self-pollinating crop like rice (0.20-0.46) than to an out-crossing crop like corn (0.06 - 0.31) (Courtois et al. 2012; Garris et al. 2005; Liu et al. 2003). Presence of profound population differentiation could pose a challenge to successful Genome Wide Association Studies (GWAS) in Upland cotton germplasm for traits that are associated with population structure. The power of structure-based association studies to detect the effects of single genes would be reduced if a large fraction of variation were explained by population structure (Flint-Garcia et al. 2005). In such cases, alternative association mapping populations would be more useful (Flint-Garcia et al. 2005). Joint linkage-association mapping, especially the Nested Association Mapping populations developed from core sets of allelic richness could be used to detect and map agronomically desirable variation in crop plants (Wu and Zeng 2001; Meuwissen et al. 2002; Wu et al. 2002; Blott et al. 2003, Yu et al. 2008).

Core sets are a small subset of accessions that retain most of the genetic diversity present in an original collection of germplasm (Frankel 1984). They facilitate efficient utilization of overall genetic diversity while dealing with fewer accessions. Core sets are also excellent germplasm sets for developing association mapping populations. Molecular data based core sets have been identified in other crops, including maize (Liu et al. 2003), rice (Courtois et al. 2012), soybean (Kuroda et al. 2009) and Chinese wheat (Hao et al. 2008). In maize, the core sets identified from a panel of 260 lines lead to the development of largest genetic mapping populations, which have been used extensively in dissecting the genetic architecture of quantitative traits in corn (Liu et al. 2003; McMullen et al. 2009; Buckler et al. 2009). Genotypic values for agronomic traits have been used to identify core sets of *G. barbadense* accessions in China (Xu et al. 2006). However, no systematic efforts, utilizing molecular marker based genotyping methods, were made to identify core sets for the US Upland cotton. In the current study by using 324 accessions that represent 322 alleles in the US upland cotton, core sets were assembled from the cotton diversity panel with sizes ranging from 8 to 53 lines in increments of 5 lines by maximizing allelic richness. A core set of 23 accessions that capture 74% of the 322 alleles, was selected for developing NAM population in upland cotton for establishing the genetic architecture of quantitative traits in cotton. This population will be invaluable as a community resource for genome-wide association studies. Future work should involve measuring Linkage Disequilibrium (LD) in Upland cotton. In presence of extensive LD, the germplasm panel used in current study can be used for low resolution association analysis.

## REFERENCES

- Abdalla AM, Reddy OUK, El-Zik KM, Pepper AE (2001) Genetic diversity and relationships of diploid and tetraploid cottons revealed using AFLP. *Theor Appl Genet* 102:222-229
- Abdurakhmonov IY, Kohel RJ, Yu JZ, Pepper AE, Abdullaev AA, Kushanov FN, Salakhutdinov LB, Buriev ZT, Saha S, Scheffler BE, Jenkins JN, Abdugarimov A (2008) Molecular diversity and association mapping of fiber quality traits in exotic *G. hirsutum* L. germplasm. *Genomics* 92:478-487
- Abdurakhmonov IY, Buriev ZT, Shermatov SE, Abdullaev AA, Urmonov K, Kushanov F, Egamberdiev SS, Shapulatov U, Abdugarimov A, Saha S, Jenkins JN, Kohel RJ, Yu JZ, Pepper AE, Kumpatla SP, Ulloa M (2012) Genetic Diversity in *Gossypium* genus. In: Caliskan M (ed) *Genetic Diversity in Plants*, ISBN: 978-953-51-0185-7, InTech, pp 313-338. DOI: 10.5772/2640
- Bertini CHCD, Schuster I, Sedyama T, Barros EG, Moreira MA (2006) Characterization and genetic diversity analysis of cotton cultivars using microsatellites. *Genet Mol Biol* 29:321-329
- Blott S et al. (2003) Molecular dissection of a quantitative trait locus: a phenylalanine-to-tyrosine substitution in the transmembrane domain of the bovine growth hormone receptor is associated with a major effect on milk yield and composition. *Genetics* 16:253-266
- Botstein D, White RL, Skolnick M, and Davis RW (1980) Construction of a genetic linkage map in man using restriction fragment length polymorphisms. *Am J Hum Genet* 32:314-331
- Bowman DT, May OL, Calhoun DS (1996) Genetic base of upland cotton cultivars released between 1970 and 1990. *Crop Sci* 36:577-581

Bowman DT, Gutierrez OA, Percy RG, Calhoun DS, May OL (2006) Pedigrees of upland and pima cotton cultivars released between 1970 and 2005. *Miss Agric For Exp Stn Bull* 1155

Brown WL (1983) Genetic diversity and genetic vulnerability - an appraisal. *Econ Bot* 37:4-12

Brubaker CL, Bourland FM, Wendel JF (1999) The origin and domestication of cotton. In: Smith CW, Cothren JT (ed) *Cotton: Origin, History, Technology, and Production*. Wiley, New York, pp 3-32

Buckler ES et al. (2009) The genetic architecture of maize flowering time. *Science* 325:714-718

Campbell BT, Williams VE, Park W (2009) Using molecular markers and field performance data to characterize the Pee Dee cotton germplasm resources. *Euphytica* 169:285

Courtois B, Frouin J, Greco R, Bruschi G et al. (2012) Genetic Diversity and Population Structure in a European Collection of Rice. *Crop Sci* 52:1663-1675

Dejode D, Wendel J (1992) Genetic diversity and origin of the hawaiian-islands cotton, *Gossypium tomentosum*. *Am J Bot* 79:1311-1319

Dent AE and Bridgett MV (2012) STRUCTURE HARVESTER: a website and program for visualizing STRUCTURE output and implementing the Evanno method. *Conservation Genetics Resources* 4:359-361

Evanno G, Regnaut S, Goudet J (2005) Detecting the number of clusters of individuals using the software STRUCTURE: A simulation study. *Mol Ecol* 14:2611-2620

Excoffier L, Lischer HEL (2010) Arlequin suite ver 3.5: A new series of programs to perform population genetics analyses under linux and windows. *Mol Eco Res* 10:564-567

Falush D, Stephens M, Pritchard JK (2003) Inference of population structure using multilocus genotype data: Linked loci and correlated allele frequencies. *Genetics* 164:1567-1587

Fang DD, Hinze LL, Percy RG, Li P, Deng D, Thyssen G (2013) A microsatellite-based genome-wide analysis of genetic diversity and linkage disequilibrium in Upland cotton (*Gossypium hirsutum* L.) cultivars from major cotton-growing countries. *Euphytica* 1-11

Flajoulot S, Ronfort J, Baudouin P, Barre P, Huguet T, Huyghe C, Julier B (2005) Genetic diversity among alfalfa (*Medicago sativa*) cultivars coming from a breeding program, using SSR markers. *Theor Appl Genet* 111:1420-1429

Flint-Garcia SA et al. (2005) Maize association population: a high-resolution platform for quantitative trait locus dissection. *Plant J.* 44:1054–1064

Frankel OH (1984) Genetic perspectives of germplasm conservation. In: Arber W, Llimensee K, Peacock WJ, Starlinger P (ed) *Genetic manipulation: impact on man and society*. Cambridge University Press, Cambridge, pp 161–170

Garris AJ, Tai TH, Coburn J, Kresovich S, McCouch S (2005) Genetic Structure and Diversity in *Oryza sativa* L. *Genetics* 169:1631–1638

Hao C, Dong Y, Wang L, You G, Zhang H, Ge H, Jia J, Zhang X (2008) Genetic diversity and construction of core collection in Chinese wheat genetic resources. *Chinese Science Bulletin* 53:1518-1526

Hinze LL, Dever JK, Percy RG (2012) Molecular variation among and within improved cultivars in the U.S. cotton germplasm collection. *Crop Sci* 52:222-230

Iqbal MJ, Aziz N, Saeed NA, Zafar Y, Malik KA (1997) Genetic diversity evaluation of some elite cotton varieties by RAPD analysis. *Theor Appl Genet* 94:139-144

Jenkins JN, McCarty Jr. JC, Gutierrez OA, Hayes RW, Bowman DT, Watson CE, Jones DC (2008) Registration of RMUP-C5, a random mated population of upland cotton germplasm. *J Plant Registrations* 2:239-242

Kalivas A, Xanthopoulos F, Kehagia O, Tsiftaris AS (2011) Agronomic characterization, genetic diversity and association analysis of cotton cultivars using simple sequence repeat molecular markers. *Genet Mol Res* 10:208-217

Kuraparthy V, Bowman DT (2013) Gains in breeding Upland cotton for fiber quality. *J Cotton Sci* (in press)

Kuroda Y, Tomooka N, Kaga A, Wanigadeva SMSW, Vaughan DA (2009) Genetic diversity of wild soybean (*Glycine soja* Sieb. et Zucc.) and Japanese cultivated soybeans [*G. max* (L.) Merr.] based on microsatellite (SSR) analysis and the selection of a core collection. *Genet Res Crop Evol* 56:1045-1055

Lacape JM, Dessauw D, Rajab M, Noyer JL, Hau B (2007) Microsatellite diversity in tetraploid *Gossypium* germplasm: assembling a highly informative genotyping set of cotton SSRs. *Mol Breeding* 19:45-58

Li H, Luo J, Hemphill JK, Wang JT (2001) A rapid and high yielding DNA miniprep for cotton (*Gossypium* spp). *Plant Mol Biol Rep* 19:183a-183e

Liu B, Wendel JF (2001) Intersimple sequence repeat (ISSR) polymorphisms as a genetic marker system in cotton. *Mol Ecol Notes* 1:205-208

Liu D, Guo X, Lin Z, Nie Y, Zhang X (2006) Genetic diversity of asian cotton (*Gossypium arboreum* L.) in china evaluated by microsatellite analysis. *Genet Res Crop Evol* 53:1145-1152

Liu KJ, Goodman M, Muse S, Smith JS, Buckler E, Doebley J (2003) Genetic structure and diversity among maize inbred lines as inferred from DNA microsatellites.

Genetics 165:2117–2128

Liu KJ, Muse SV (2005) PowerMarker: An integrated analysis environment for genetic marker analysis. Bioinformatics 21:2128-2129

May OL, Bowman DT, Calhoun DS (1995) Genetic diversity of U.S. upland cotton cultivars released between 1980 and 1990. Crop Sci 35:1570-1574

McMullen MD, Kresovich S, Villeda HS, Bradbury P, Li H, Sun Q, Flint-Garcia S, Thornsberry J, Acharya C, Bottoms C et al. (2009) Genetic properties of the maize nested association mapping population. Science 325: 737–740

Mehboob uR, Yasmin T, Tabbasam N, Ullah I, Asif M, Zafar Y (2008) Studying the extent of genetic diversity among *Gossypium arboreum* L. genotypes/cultivars using DNA fingerprinting. Genet Resour Crop Evol 55:331-339

Meuwissen THE, Karlsten A, Lien S, Olsaker I, Goddard ME (2002) Fine mapping of a quantitative trait locus for twinning rate using combined linkage and linkage disequilibrium mapping. Genetics 16:373–379

Multani DS, Lyon BR (1995) Genetic fingerprinting of Australian cotton cultivars with RAPD markers. Genome 38:1005-1008

Nei M, Tajima F, Tateno Y (1983) Accuracy of estimated phylogenetic trees from molecular data. J Mol Evol 19:153-170

Niles GA, Feaster CV (1984) Breeding. In Kohel RJ, Lewis CF (ed) Cotton, Agronomy Monograph No. 24, CSSA, Madison, WI. pp201- 231

Oliveira HR, Campana MG, Jones H, Hunt HV, Leigh F, Redhouse DI, Lister DL, Jones MK (2012) Tetraploid Wheat Landraces in the Mediterranean Basin: Taxonomy, Evolution and Genetic Diversity. PLoS One 7: e37063. doi:10.1371/journal.pone.0037063

Powell W, Machray GC, Provan J (1996) Polymorphism revealed by simple sequence repeats. Trends in Plant Sci 1:215-222

Pritchard JK, Stephens M, Donnelly P (2000) Inference of population structure using multilocus genotype data. Genetics 155:945-959

Rohlf FJ (2000) Numerical taxonomy and multivariate analysis system, ver. 2.11. Applied Biostatistics, New York

Smith CW, Cantrell RG, Moser HS, Oakley SR (1999) History of cultivar development in the United States. In: Smith CW, Cothren JT (ed) Cotton: origin, history, technology, and production. Wiley, New York, pp 99–171

Staten G (1970) Breeding Acala 1517 cottons, Memoir Series No. 4, New Mexico State University, Las Cruces. 1926-1970

Van Becelaere G, Lubbers EL, Paterson AH, Chee PW (2005) Pedigree- vs. DNA marker-based genetic similarity estimates in cotton. Crop Sci 45:2281-2287

Van Esbroeck GA, Bowman DT (1998) Cotton improvement. Cotton germplasm diversity and its importance to cultivar development. J Cotton Sci 2:121-129

Van Esbroeck GA, Bowman DT, May OL, Calhoun DS (1999) Genetic similarity indices for ancestral cotton cultivars and their impact on genetic diversity estimates of modern cultivars. Crop Sci 39:323–328

Wendel J, Percy R (1990) Allozyme diversity and introgression in the Galapagos-islands endemic *Gossypium darwinii* and its relationship to continental *Gossypium barbadense*. Biochem Syst Ecol 18:517-528

- Wendel J, Brubaker C, Percival A (1992) Genetic diversity in *Gossypium hirsutum* and the origin of Upland cotton. *Am J Bot* 79:1291-1310
- Wendel J, Rowley R, Stewart J (1994) Genetic diversity in and phylogenetic-relationships of the brazilian endemic cotton, *Gossypium mustelinum* (Malvaceae). *Plant Syst Evol* 192:49-59
- Wu R, Zeng ZB (2001) Joint linkage and linkage disequilibrium mapping in natural populations. *Genetics* 157:899–909
- Wu R, Ma CX, Casella G (2002) Joint linkage and linkage disequilibrium mapping of quantitative trait loci in natural populations. *Genetics* 160:779–792
- Xu H, Mei Y, Hu J, Zhu J, Gong P (2006) Sampling a core collection of Island cotton (*Gossypium barbadense* L.) based on the genotypic values of fiber traits. *Genet Res Crop Evo* 53:515–521
- Yu J, Holland JB, McMullen MD, Buckler ES (2008) Genetic design and statistical power of nested association mapping in maize. *Genetics* 178:539-551
- Yu JZ, Fang DD, Kohel RJ, Ulloa M, Hinze LL, Percy RG, Zhang J, Chee P, Scheffler BE, Jones DC (2012) Development of a core set of SSR markers for the characterization of *Gossypium* germplasm. *Euphytica* 187:203-213
- Zhang J, Lu Y, Cantrell R, Hughs E (2005) Molecular marker diversity and field performance in commercial cotton cultivars evaluated in the southwestern USA. *Crop Sci* 45:1483-1490
- Zhang Y, Wang XF, Li ZK, Zhang GY, Ma ZY (2011) Assessing genetic diversity of cotton cultivars using genomic and newly developed expressed sequence tag-derived microsatellite markers. *Genet Mol Res* 10:1462-1470

**Table 1.1** Summary statistics for SSR markers used to genotype *G. hirsutum* accessions.

<b>Marker</b>	<b>MAF</b>	<b>AlleleNo</b>	<b>GeneDiversity</b>	<b>Heterozygosity</b>	<b>PIC</b>
BNL3029_Fb	0.9836	2	0.03	0.00	0.03
BNL1434_F	0.5653	6	0.50	0.02	0.39
BNL1667_Fa	0.8733	2	0.22	0.00	0.20
BNL1667_Fb	0.7534	5	0.40	0.00	0.36
BNL226_Fa	0.9822	3	0.04	0.00	0.03
BNL226_Fb	0.9973	2	0.01	0.00	0.01
BNL1059_F	0.9973	2	0.01	0.00	0.01
BNL2572_F	0.6671	11	0.52	0.06	0.48
BNL3800_F	0.5804	6	0.57	0.04	0.51
BNL852_F	0.9973	2	0.01	0.00	0.01
BNL1440_F	0.8720	4	0.23	0.03	0.22
BNL1693_F	0.9046	3	0.17	0.01	0.16
BNL2440_Fb	0.9196	5	0.15	0.00	0.14
BNL2564_F	0.9973	2	0.01	0.00	0.01
BNL1145_F	0.9907	2	0.02	0.00	0.02
BNL1064_H	0.9866	3	0.03	0.00	0.03
BNL1665_F	0.9947	3	0.01	0.00	0.01
BNL830_H	0.5576	2	0.49	0.03	0.37
BNL3888_F	0.9960	3	0.01	0.00	0.01
BNL2882_H	0.9786	2	0.04	0.00	0.04
BNL3034_H	0.7609	2	0.36	0.02	0.30
BNL1597_F	0.9639	4	0.07	0.00	0.07
BNL1034_H	0.9880	3	0.02	0.00	0.02
BNL3008_F	0.9973	2	0.01	0.00	0.01
BNL2646_H	0.8445	3	0.26	0.01	0.23
BNL1721_F	0.7560	4	0.39	0.02	0.35
BNL3955_H	0.9920	2	0.02	0.00	0.02
BNL1423_F	0.8108	3	0.31	0.02	0.27
BNL119_H	0.6653	5	0.46	0.02	0.37
BNL1551_F	0.4198	6	0.71	0.03	0.66
BNL2553_H	0.9973	2	0.01	0.00	0.01
BNL2471_F	0.9760	3	0.05	0.01	0.05
BNL1673_F	0.9666	4	0.06	0.01	0.06
BNL3441_F	0.9906	2	0.02	0.00	0.02
BNL1231_H	0.5094	2	0.50	0.02	0.37
BNL1694_Fa	0.9960	2	0.01	0.00	0.01
BNL1694_Fb	0.5413	3	0.51	0.02	0.39
BNL1513_F	0.4000	6	0.66	0.05	0.60

Table 1.1 continued.

<b>Marker</b>	<b>MAF</b>	<b>AlleleNo</b>	<b>GeneDiversity</b>	<b>Heterozygosity</b>	<b>PIC</b>
BNL1061_Fa	0.9973	2	0.01	0.00	0.01
BNL1066_H	0.9024	2	0.18	0.01	0.16
BNL3627_H	0.9906	4	0.02	0.00	0.02
BNL1417_H	0.9590	2	0.08	0.00	0.08
BNL3511_F	0.9933	4	0.01	0.00	0.01
CIR376_Ha	0.8164	2	0.30	0.01	0.25
CIR376_Hb	0.9919	2	0.02	0.00	0.02
CIR171_Fa	0.9973	2	0.01	0.00	0.01
CIR171_Fb	0.9370	2	0.12	0.01	0.11
BNL4007_H	0.9946	3	0.01	0.00	0.01
BNL1045_Fb	0.9973	2	0.01	0.00	0.01
BNL2960_H	0.6721	3	0.45	0.02	0.35
BNL3479_F	0.9866	2	0.03	0.00	0.03
BNL3976_Hb	0.9051	3	0.17	0.00	0.16
CIR289_F	0.9947	3	0.01	0.00	0.01
BNL3510_F	0.9946	2	0.01	0.00	0.01
BNL2499_F	0.7099	4	0.45	0.02	0.41
BNL2812_Fa	0.9892	2	0.02	0.00	0.02
BNL2812_Fb	0.9716	4	0.06	0.00	0.05
BNL3778_F	0.9827	2	0.03	0.00	0.03
BNL3649_H	0.5389	2	0.50	0.02	0.37
BNL3474_Fa	0.7487	2	0.38	0.02	0.31
BNL3449_Hb	0.9946	2	0.01	0.00	0.01
BNL3418_H	0.9365	2	0.12	0.00	0.11
BNL3590_Fa	0.7682	2	0.36	0.01	0.29
BNL3590_Fb	0.5377	2	0.50	0.01	0.37
BNL3650_F	0.7895	4	0.34	0.01	0.30
BNL3379_F	0.9719	2	0.05	0.00	0.05
CIR143_F	0.9987	2	0.00	0.00	0.00
CIR272_Hb	0.9893	2	0.02	0.00	0.02
CIR372_Fa	0.5799	2	0.49	0.03	0.37
TMB1750_Fa	0.9973	2	0.01	0.00	0.01
TMB1750_Fb	0.9973	2	0.01	0.00	0.01
BNL569_H	0.6599	3	0.51	0.28	0.45
BNL2667_F	0.6247	3	0.47	0.00	0.37
TMB1489_H	0.9987	2	0.00	0.00	0.00
BNL2655_H	0.7857	4	0.34	0.01	0.29
BNL252_F	0.9824	3	0.03	0.00	0.03

Table 1.1 continued.

<b>Marker</b>	<b>MAF</b>	<b>AlleleNo</b>	<b>GeneDiversity</b>	<b>Heterozygosity</b>	<b>PIC</b>
BNL358_F	0.9893	3	0.02	0.00	0.02
BNL2495_H	0.5587	2	0.49	0.02	0.37
BNL3031_F	0.5147	3	0.51	0.03	0.39
BNL3537_F	0.8514	3	0.25	0.29	0.22
CIR030_F	0.6895	2	0.43	0.01	0.34
BNL409_F	0.9945	3	0.01	0.00	0.01
BNL580_Fa	0.9973	2	0.01	0.00	0.01
BNL580_Fb	0.6250	5	0.50	0.04	0.41
BNL256_F	0.9705	2	0.06	0.00	0.06
BNL786_F	0.6381	3	0.47	0.02	0.37
CIR228_H	0.5709	7	0.58	0.01	0.51
BNL1666_Fb	0.8235	2	0.29	0.01	0.25
BNL1395_H	0.8577	4	0.25	0.01	0.22
BNL827_F	0.5860	2	0.49	0.02	0.37
BNL3482_H	0.6800	3	0.44	0.01	0.35
BNL3835_F	0.8054	3	0.32	0.01	0.28
BNL285_F	0.9773	3	0.04	0.00	0.04
BNL834_F	0.9707	3	0.06	0.00	0.06
BNL686_Ha	0.8492	4	0.26	0.02	0.23
TMB1645_H	0.2938	12	0.80	0.03	0.78
BNL1605_Ha	0.6332	4	0.52	0.03	0.46
BNL1605_Hb	0.9912	4	0.02	0.00	0.02
BNL1404_H	0.9973	2	0.01	0.00	0.01
CIR218_Hb	0.5000	3	0.50	1.00	0.38
BNL946_H	0.7907	2	0.33	0.01	0.28
GH459_H	0.9946	3	0.01	0.00	0.01
BNL1153_Fb	0.9973	2	0.01	0.00	0.01
BNL1161_Ha	0.9973	2	0.01	0.00	0.01
BNL1161_Hb	0.9611	3	0.08	0.00	0.07
BNL3594_Fb	0.5000	8	0.65	0.04	0.59
BNL3987_F	0.9880	2	0.02	0.00	0.02
BNL836_F	0.9920	3	0.02	0.00	0.02
BNL1162_H	0.9947	3	0.01	0.00	0.01
BNL1017_F	0.9947	2	0.01	0.00	0.01
BNL4041_F	0.9532	2	0.09	0.00	0.09
BNL1604_Fa	0.8476	3	0.26	0.00	0.23
BNL1604_Fb	0.7973	4	0.33	0.00	0.28
BNL1047_F	0.9973	2	0.01	0.00	0.01

<b>Marker</b>	<b>MAF</b>	<b>AlleleNo</b>	<b>GeneDiversity</b>	<b>Heterozygosity</b>	<b>PIC</b>
CIR97_Hb	0.9505	3	0.09	0.01	0.09
CIR110_F	0.9973	2	0.01	0.00	0.01
BNL1672_F	0.4652	5	0.65	0.54	0.58
BNL1122_Ha	0.8579	3	0.25	0.00	0.22
BNL1122_Hb	0.6475	4	0.46	0.02	0.36
CIR187_Fa	0.9973	2	0.01	0.00	0.01
<b>Mean</b>	<b>0.8495</b>	<b>3</b>	<b>0.20</b>	<b>0.03</b>	<b>0.17</b>

**Table 1.2** Summary of unique (present in only one accession) and rare alleles observed in a combined panel of *G. hirsutum* and *G. barbadense* accessions vs. only the Upland cotton accessions.

<b>Panel</b>	<b>Total alleles</b>	<b>Unique alleles</b>	<b>Rare alleles (freq &lt;5%)</b>
Combined panel	546	134 (24 %)	199 (36%)
<i>G. hirsutum</i> panel	367	80 (22 %)	94 (25%)

**Table 1.3** List of *G. hirsutum* accessions with unique (present in only one accession) alleles

<b>Sr. No.</b>	<b>Accession name</b>	<b>Number of unique alleles</b>
1	RUGOSE_INDORE	7
2	BULGARIA_432	4
3	LANKART_311	3
4	MU8B_UA7_44	3
5	M_4	3
6	AK_DJURA_VIR	2
7	BLIGHTMASTER	2
8	CLEVELAND_54	2
9	LBBCDBOAKH190	2
10	PD93034	2
11	OR27	2
12	AUBURN 634RNR	2
13	SEALAND7_WHITEFL	2
14	SEALAND7_YELWFL	2
15	SOUTHLAND_M1	2
16	TOOLE	2
17	ACALA_1517D	1
18	AK_DJURAFREGO	1
19	ARIZONA_RED	1
20	ARKANSAS_10	1
21	ARKOT_8102	1
22	BJAGLNECT	1
23	CA_17	1
24	CABD3CABCH189	1
25	CS_8610	1
26	DELTAPINE_SR383	1
27	EARLISTAPLE_7	1
28	EMPIRE_WR	1
29	EXPRESS_121	1
30	FJA	1
31	GSA_74	1
32	H1220	1
33	HOPI_MOENCOPI	1
34	LA_850082FN	1
35	LOCKETT_88	1
36	MULTIPLE_MARKER	1
37	NEW_BOYKIN	1

Table 1.3 continued

<b>Sr. No.</b>	<b>Accession name</b>	<b>Number of unique alleles</b>
38	PAYMASTER_101	1
39	PAYMASTER_266	1
40	PAYMASTER_HS26	1
41	PD113	1
42	PD2164	1
43	PD4461Q	1
44	PD781	1
45	PD93009	1
46	ROGERS_LG10	1
47	ROWDEN_41BTPSA	1
48	SEALAND1	1
49	SEALAND_472	1
50	STATIONMILLER	1
51	Tashkent1	1
52	TIDEWATER_29	1
53	TIDEWATER_E372_4	1
54	WILDS_34_4T85_2	1

**Table 1.4** Summary of rare alleles found in Upland cotton accessions grouped in clusters based on STRUCTURE analysis. The color name in bracket shows the color assigned to the group in STRUCTURE Q-plot.

<b>Group</b>	<b>Number of lines</b>	<b>Total alleles</b>	<b>Rare alleles (present in less than 5% lines)</b>
West (red)	36	233	29 (12%)
Southeast (green)	38	288	60 (21%)
Southwest (blue)	57	237	32 (14%)
Midsouth (yellow)	50	228	24 (11%)

**Table 1.5** Genetic distance estimates calculated using Nei's 1983 distance within and between *G. hirsutum* groups identified by model based population structure analysis

	West	Southeast	Southwest	Midsouth
West	0.124			
Southeast	0.245	0.186		
Southwest	0.207	0.232	0.131	
Midsouth	0.209	0.212	0.193	0.109

**Table 1.6** Analysis of molecular variance for upland cotton accessions between and within four groups corresponding to four major regions of cotton production in United States as identified by STRUCTURE.

Source of variation	<i>df</i>	Sum of squares	Variance components	Percentage of variation
Among groups	3	1221.62	4.45879 ***Va	34.16
Within groups	358	3076.46	8.59347 ***Vb	65.84
Total	361	4298.09	13.05226	

\*\*\*significant at  $p < 0.00001$

**Table 1.7** Pairwise  $F_{ST}$  estimates for the four groups corresponding to four major regions of cotton production in United States as identified by STRUCTURE

	West	Southeast	Southwest
Southeast	0.3456		
Southwest	0.3581	0.2995	
Midsouth	0.4196	0.2936	0.3494

**Table 1.8** Core sets of Upland cotton accessions identified by simulated annealing algorithm using Powermarker software.

Complete panel has 375 accessions whereas reduced panel contains 324 accessions after excluding agronomically unfavorable accessions. Acala Maxxa, Wannamaker Cleveland, Dixie King, Deltapine 14, DES 56 and Paymaster HS200 are used as positive constraints in core sets of all sizes.

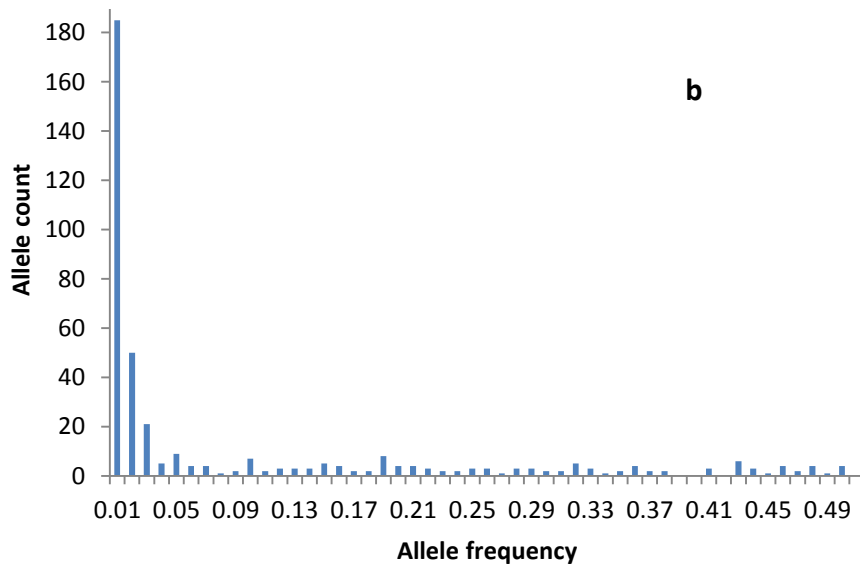
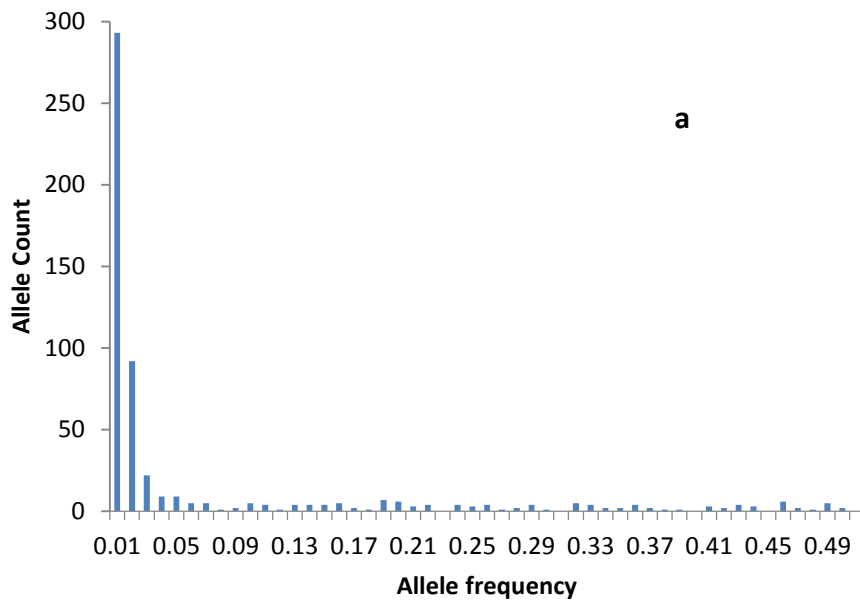
<b>Set size</b>	<b>Additional accessions in core set</b>	<b>Allele number</b>	<b>% of alleles in reduced panel</b>	<b>% of alleles in complete <i>G. hirsutum</i> panel</b>
8	Acala 5 , M.U.8B UA 7-44	207	64	56
13	Acala 5 , M.U.8B UA 7-44, NC 88-95, PD 0113, PD 785, Auburn 634RNR, Toole	237	74	65
18	Acala 5 , M.U.8B UA 7-44, NC 88-95, PD 0113, PD 785, Auburn 634RNR, Toole, Allen 33, Arkansas 10, PD 2164, Southland M1, Tashkent 1	257	80	70
23	Acala 5 , Allen 33, CD3HCABCUH-1-89, DES 24, FJA, M.U.8B UA 7-44, NC 88-95, Paymaster HS26, PD 2164, PD 785, Auburn 634RNR, Sealand #2, Southland M1, Station miller, Tashkent 1, Tidewater 29 , Toole	270	84	74
28	Acala 5 , Allen 33, BJAGL NECT, CD3HCABCUH-1-89, CS-8610, DES 24, Earlistaple 7, FJA, LBBCDBOAKH-1-90, M.U.8B UA 7-44, NC 88-95, Paymaster 101, Paymaster 266, PD 2164, PD 785, Auburn 634RNR, Sealand #2, Sealand #7 Yellow Flower , Southland M1, Station miller, Tashkent 1, Toole	280	87	76

Table 1.8 continued.

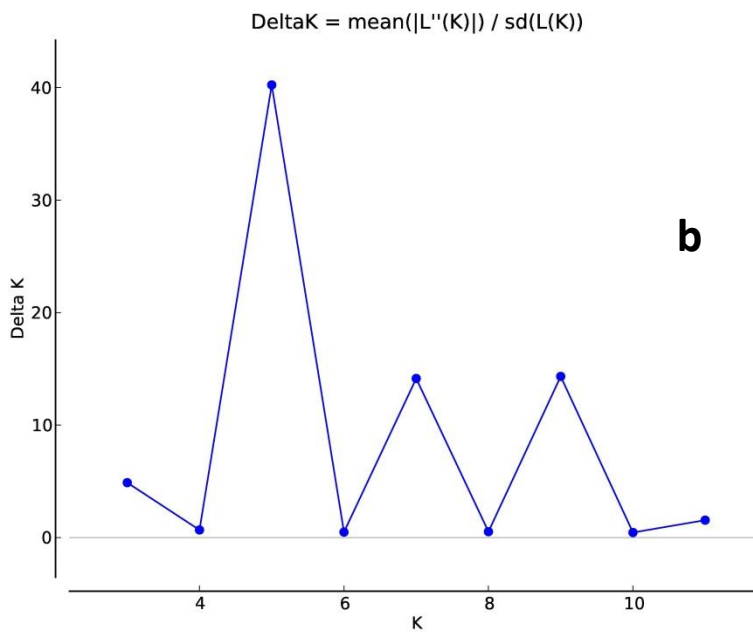
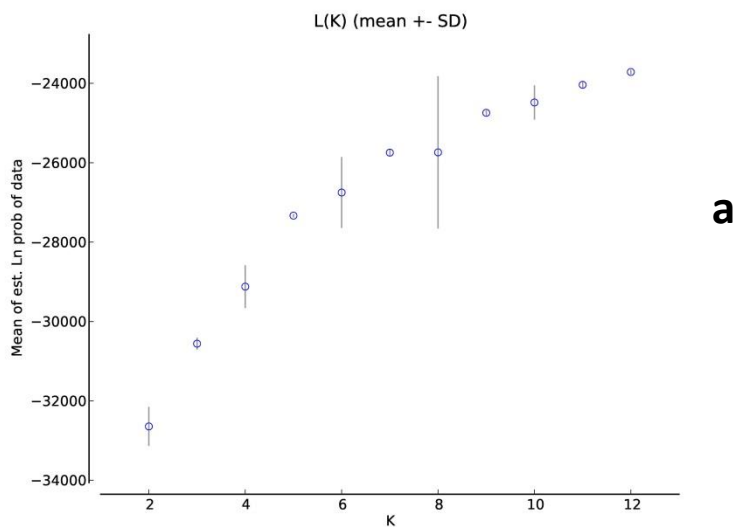
33	Acala 5 , Allen 33, BJAGL NECT, CA17, CABD3CABCH-1-89, CS-8610, Earlistaple 7, FJA, GSA 74, LBBCDBOAKH-1-90, M.U.8B UA 7-44, NC 88-95, Paymaster 101, Paymaster HS26, PD 0113, PD 2164, PD 785, Auburn 634RNR, Sealand #2, Sealand #7 Yellow Flower , Southland M1, SPNXCHGLBH-1-94, Station miller, Tashkent 1, Tidewater 29 , Toole, Wilds 18	289	90	79
38	Acala 5 , Allen 33, BJAGL NECT, CA23, CABD3CABCH-1-89, Coker’s Wilds #2, CS-8610, Earlistaple 7, FJA, GSA 74, La.850082FN, LBBCDBOAKH-1-90, M.U.8B UA 7-44, NC 88-95, Paymaster 266, PD 0113, PD 2164, PD 785, PD 93009, PD 93021, Auburn 634RNR, Sealand #2 , Sealand #7 Yellow Flower , Southland M1, SPNXCHGLBH-1-94, Station miller, Tashkent 1, Tidewater 29, Toole, Western Stormproof, Wilds 18, Wilds 34-4(411) T85-2	294	91	80
43	Acala 5 , Allen 33, BJAGL NECT, CA23, CD3HCABCUH-1-89, Coker 201, CS-8610, Earlistaple 7, Empire WR, FJA, GSA 74, La.850082FN, LBBCDBOAKH-1-90, Lightning Express, Lockett 88, M.U.8B UA 7-44, NC 88-95, New Boykin, Paymaster 266, Paymaster HS26 , PD 0113, PD 2, PD 2164, PD 785, Piedmont Cleveland, Rowden 41B TPSA, Auburn 634RNR, Sealand #2 , Sealand #7 Yellow Flower , Southland M1, SPNXCHGLBH-1-94, SPNXHQBPIS-1-94, Station miller, Stoneville 20, Tidewater 29 , Toole, Wilds 18	299	93	81

Table 1.8 continued.

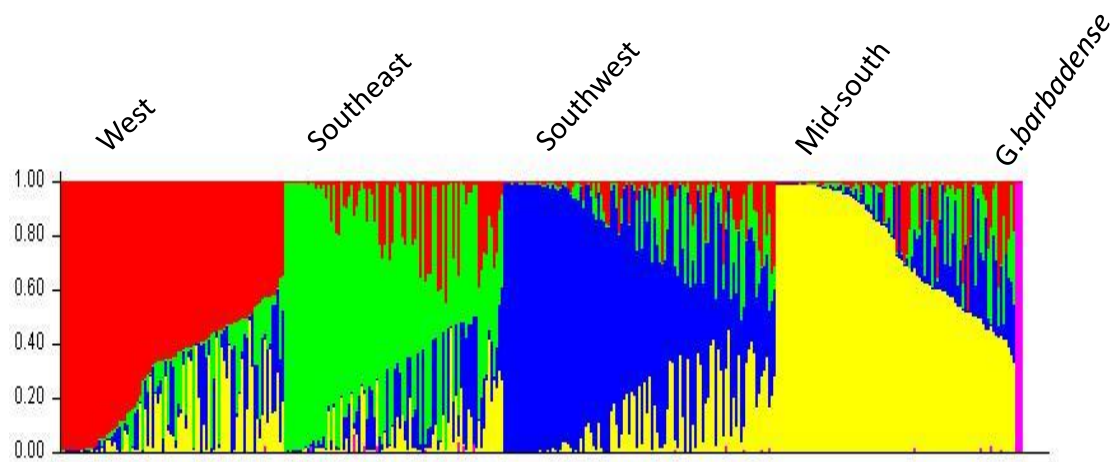
48	<p>Acala 5 , Allen 33, Arkansas 10, BJAGL NECT, CA17, CABD3CABCH-1-89, Coker 201, Coker’s Wilds #2, CS-8610, Earlistaple 7, Express 121, FJA, FOX 4, GSA 74, H1220, La.850082FN, LBBCDBOAKH-1-90, M.U.8B UA 7-44, NC 88-95, NC-4-M(3), Paymaster 266, PD 0113, PD 2, PD 2164, PD 785, PD 93009, PD 93021, PSC 355, Rogers LG-10, Auburn 634RNR, Sealand #2 , Sealand #7 Yellow Flower , Sealand 883 , Southland M1, SPNXCHGLBH-1-94, Station miller, tamcot SP-23, Tashkent 1, Tidewater 29 , Toole, Wilds 18, Wilds 34-4(411) T85-2</p>	304	94	83
53	<p>Acala 111 Rogers, Acala 5 , Allen 33, Arkansas 10, Arkot 8102, BJAGL NECT, CA23, CABD3CABCH-1-89, CAHUGLBBCS-1-88, Coker 201, CS-8610, Earlistaple 7, Empire, Express 121, FJA, Gregg 35, GSA 74, H1220, Half and half, Hopi moencopi, La.850082FN, LBBCDBOAKH-1-90, LOCKETT 88, M.U.8B UA 7-44, NC 88-95, New boykin, Paymaster 101, Paymaster HS26 , PD 0113, PD 2164, PD 2165, PD 781, PD 785, PD 93009, PD 93030, Auburn 634RNR, Sealand #2 , Sealand #7 Yellow Flower , Southland M1, SPNXCHGLBH-1-94, Station miller, Tamcot luxor, Tamcot SP-23, Tashkent 1, Tidewater 29 , Toole, Wilds 18</p>	309	96	84



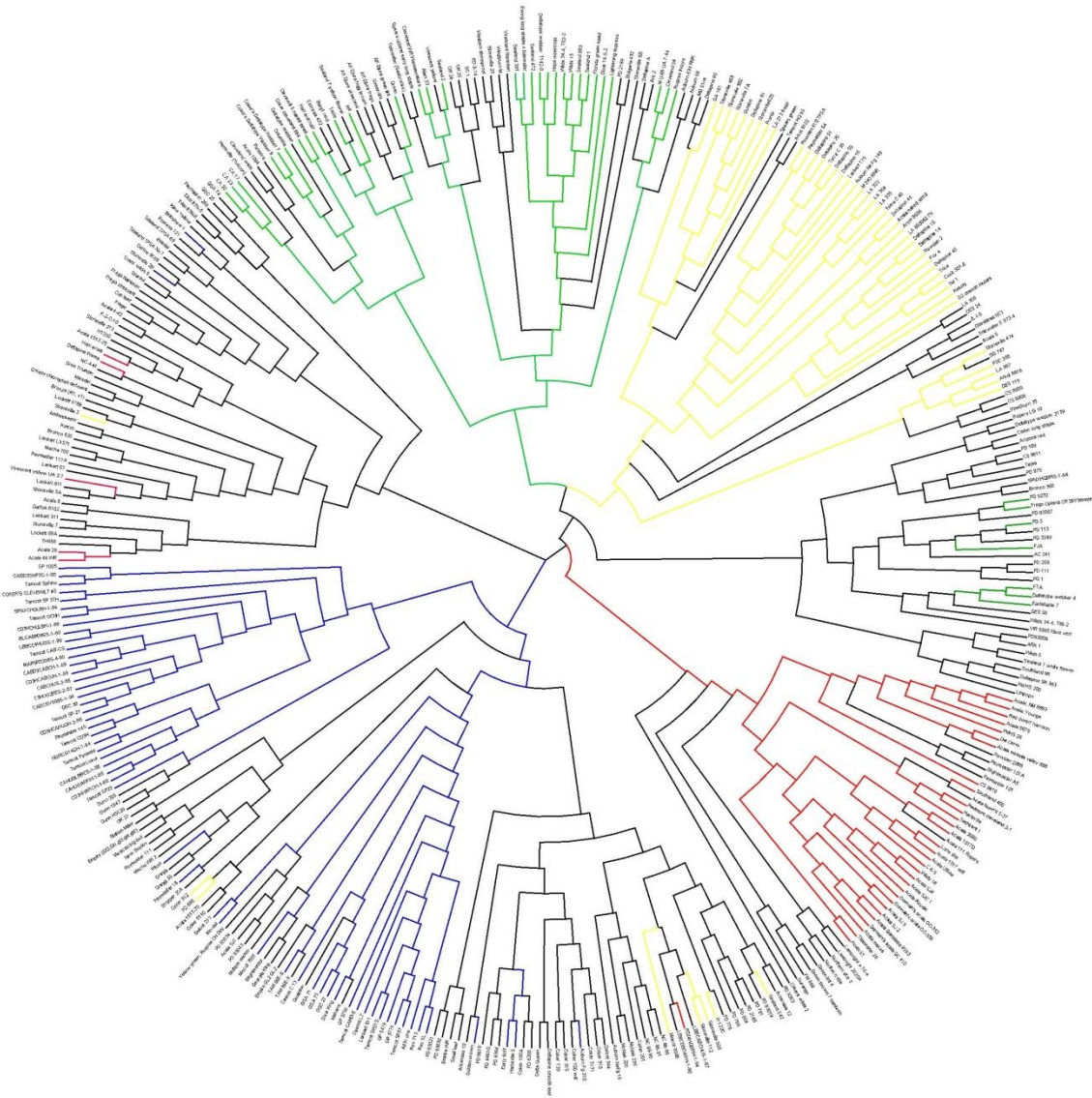
**Figure 1.1** (a) Histogram of allele frequencies for complete panel of *G. hirsutum* and *G. barbadense* accessions. (b) Histogram of allele frequencies for alleles amplified in *G. hirsutum* accessions



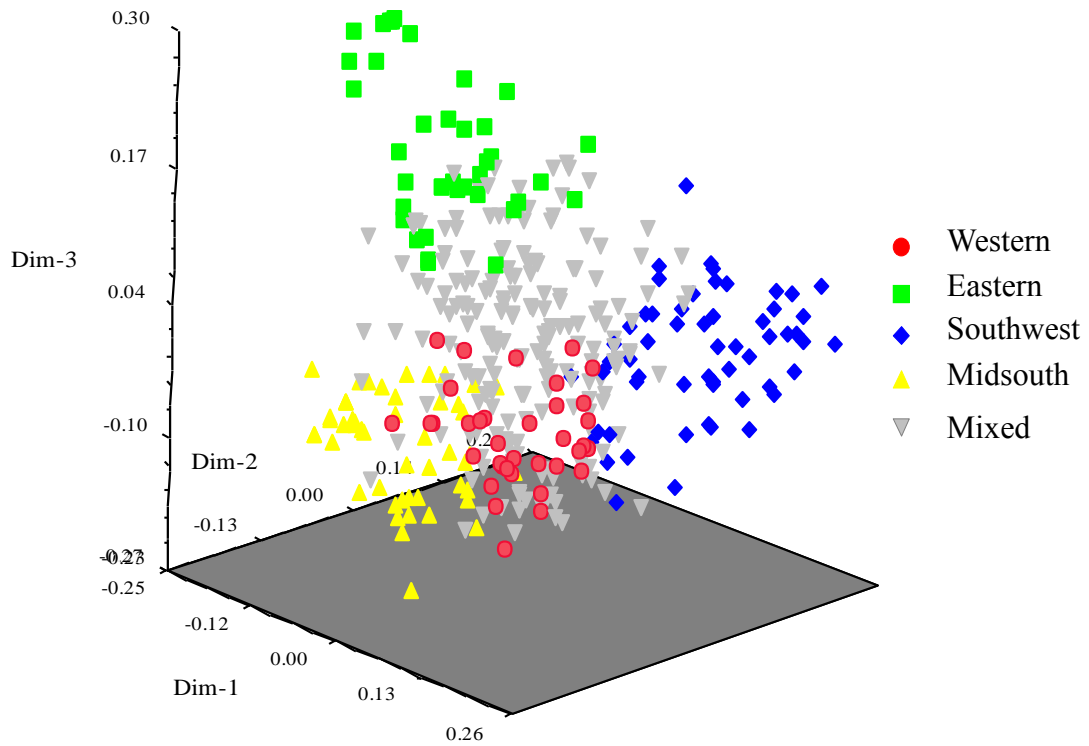
**Figure 1.2** (a) Ln (probability of data) for K ranging from 2 to 12 (b) Estimating number of subpopulations using delta K values for K ranging from 2 to 12 using method proposed by Evanno et al. (2005)



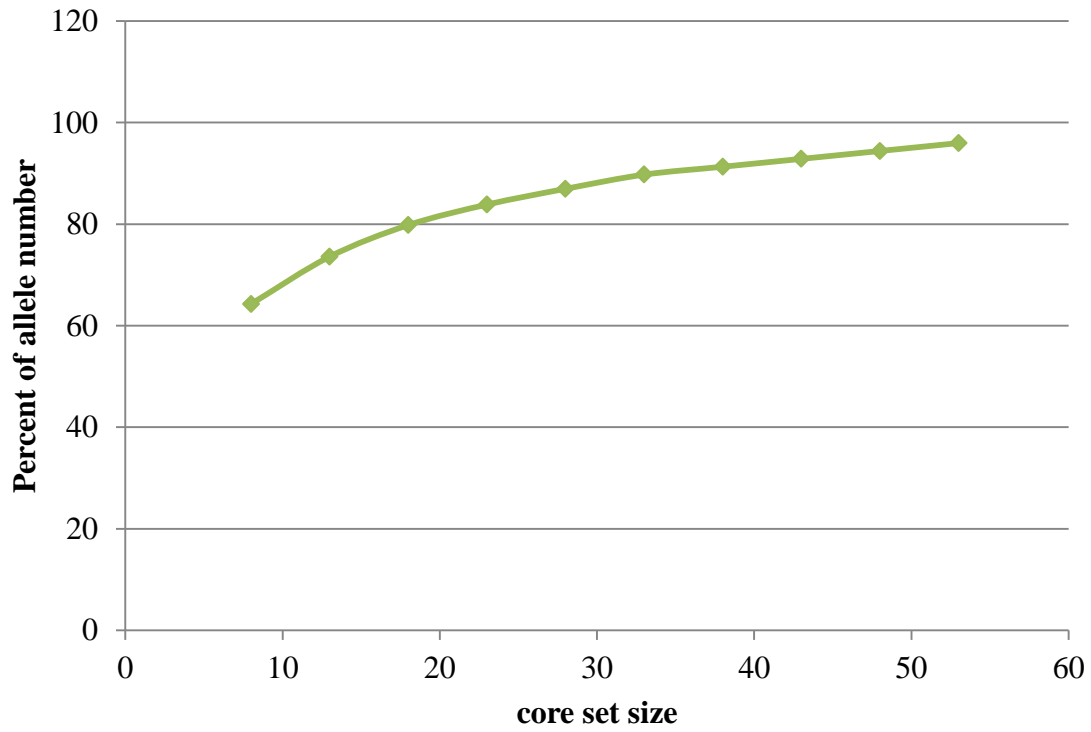
**Figure 1.3** Q-plot showing clustering of 381 *Gossypium* accessions based on analysis of genotypic data using STRUCTURE. Each accession is represented by a vertical bar. The colored subsections within each vertical bar indicate membership coefficient (Q) of the accession to different clusters.



**Figure 1.4** Dendrogram of 381 *G. hirsutum* accessions by NJ analysis. Colors in the dendrogram correspond to population structure as identified in Structure analysis. Membership threshold of 70% was used to assign accessions to different cluster. Accessions with less than 70% membership to any cluster were considered as mixed and are indicated in black in this dendrogram



**Figure 1.5** Three dimensional Principal Coordinate analysis (PCA) of Upland cotton diversity panel genotyped with SSR markers



**Figure 1.6** Plot showing percentage of alleles included in core sets of different sizes ranging from 8 to 53 accessions of *G. hirsutum*

## **CHAPTER II**

### **Components of hybrid vigor in upland cotton (*Gossypium hirsutum* L.) and their relationship with environment**

## ABSTRACT

Upland cotton (*Gossypium hirsutum* L.) hybrids display commercially useful levels of heterosis for lint yield. Cotton lint yield is primarily a multiplicative product of boll number and lint per boll, both of which can be further dissected into sub-components. Relationships among the yield components are complex where they further interact with the environment. To identify different yield components of hybrid cotton lint yield and their relationship with environment, three cotton varieties, DP51, STV474 and LA887 and, their respective hybrids were evaluated for lint yield components across five environments. Heterosis was observed for lint yield as well as yield components. The relationship between heterosis and mean environmental yield was negative for two hybrid combinations, whereas one hybrid showed increased heterosis from low to high yielding environments. Boll number was the major yield component contributing to lint yield. However, yield components contributing to the change in heterosis from low to high yielding environments were different for the three hybrids. The change in lint yield heterosis across environments was correlated with lint per boll and lint per seed for DP51 x STV474 and LA887 x DP51. Whereas, bolls number and boll retention were the main yield components contributing to higher lint yield heterosis of STV474 x LA887 in low yielding environments. Parental entries had a significant effect on the relationship between heterosis and environment.

## INTRODUCTION

Upland cotton (*Gossypium hirsutum* L.) hybrids express a commercially useful level of heterosis for yield at both mid-parent and high parent levels (Campbell et al. 2008; Meredith and Bridge 1972). Studies have indicated significant heterosis for lint yield and yield components in inter- and intra-specific crosses of cotton (Jones and Loden 1951; Marani 1967; Cole et al. 2009). Unlike maize (*Zea mays* L.), where hybrid vigor has been successfully exploited to achieve yield increases (Duvick 1992), practical limitations of producing hybrid seed have limited the progress towards establishing cotton as a hybrid crop.

Cotton lint yield components form a complex network of relationships. Kerr (1966) proposed a multiplicative model to explain the contribution of yield components towards lint yield as:  $\text{Yield} = \text{Boll number} \times \text{seeds boll}^{-1} \times \text{lint seed}^{-1}$ . Further, boll number can be expressed as a multiplicative product of the number of fruiting sites and fruit retention. Through the growing season, a cotton plant produces far more fruiting sites than it can sustain to fully utilize the environmental potential and overcome any periods of stress. However not all initiated fruiting sites result in mature harvestable bolls. There is a competition for resources between bolls at later development stages and recently initiated squares and young bolls, causing the plant to abort immature bolls (Ehlig and Lemert 1973; Guinn and Mauney 1984). Thus, an increase in boll number leads to a decrease in boll retention (Ehlig and Lemert 1973).

Of all yield components, boll number has been found to be the major yield component affecting yield for inbreds as well as hybrids (Turner 1953; Worley et al. 1974; Zeng and Wu 2012). However, boll weight, which is the cumulative product of lint per boll

and seed per boll, has been found to contribute towards higher lint yield of hybrids derived from obsolete cultivars (Campbell et al. 2008). Other studies indicated that that higher yield of hybrids is associated with boll weight rather than boll number (Wells and Meredith, 1986). In other cases, yield heterosis has also been found to be associated with both boll number and boll size (Marani 1968). Although both boll number and boll size contribute to lint yield, they are negatively correlated (Turner 1953).

Heterotic effects for cotton yield were often variable (Young and Murray 1966) because yield is a quantitative trait, expression of which is greatly affected by environment. Jenkins et al. (2009) studied F<sub>2</sub> hybrids generated from eleven diverse upland cotton lines and found that boll number, seed cotton yield and lint yield were highly dependent on specific environments. They reported that only a small fraction of phenotypic variance for lint yield (38%) and boll number (30%) could be explained by genotypic and genotype x environment variance. However, 83% of the variance for boll weight could be attributed to genotype and genotype x environment variance. These results show the effect of environment on yield components and in turn yield. Further, Cole et al. (2009) showed that there is a significant negative correlation between percent heterosis for lint yield and mean environment yield. In both studies, boll number was significantly correlated with seed cotton yield and lint yield whereas boll size was not correlated with lint yield (Jenkins 2009; Cole et al. 2009). Yield heterosis in cotton was more pronounced in low yielding environments (Bauer and Green 1996; Marani 1968). Similar observations have been made in maize, where higher yield heterosis was observed in low yielding or stressed environments, with kernel

number being the primary yield component contributing to yield heterosis (Liu and Tollenaar 2009).

A multitude of environmental factors like water availability and temperature have a profound effect on cotton yield and yield components (Guinn and Mauney 1984; Reddy et al. 1992). It has been reported that factors promoting photosynthesis improve boll retention and accelerate flowering (Guinn 1974). Suboptimal water availability and high temperature both cause a decrease in the number of harvestable bolls. A study to test the effect of three levels of irrigation on boll number revealed that with abundant irrigation, cotton plants increase boll number by producing more fruiting sites, while in a moderate irrigation regime plants produce fewer fruiting sites but retain more fruits (Stockton et al. 1961). Hybrids can also produce higher yields by setting bolls in a more favorable period, usually early in the season so that bolls have a better likelihood of reaching maturity. Thus, even with similar duration of flowering season, hybrids in comparison to parents, can achieve a larger number of harvestable bolls (Galal 1996; Wells and Meredith 1986). Early fruit development stage is significantly associated with heterosis in the growth cycle of a cotton plant (Wells and Meredith 1986). Wells (1988) showed that hybrids have a faster rate of development than inbred lines during seedling stage leading to greater vegetative growth which gives an advantage to hybrids by being able to accumulate more photosynthates. This may enable hybrids to be better prepared to confront stress and perform better than their parents in stressed environments by having more bolls per unit area.

Few attempts were made to understand the components of lint yield heterosis in cotton. Galal et al. (1966) reported that both boll retention and increased number of fruiting

sites was responsible for increased boll number in cotton hybrids. While in another study, heterosis for boll number was found to be associated with boll retention in *G. hirsutum* and increased number of fruiting sites in *G. barbadense* (Marani 1968). Previous studies also indicated that boll number was closely associated with change in percent heterosis from low to high yielding environments (Cole et al. 2009). However, it is unknown if the higher boll number of hybrids in low yielding environments is a result of the production of more fruiting sites or increased boll retention by hybrids. The objective of the current study was to determine which yield components are associated with increased hybrid lint yield in poor yielding environments versus high yielding environments.

#### **MATERIAL AND METHODS**

Three commercial cultivars ‘Deltapine 51’ (DP 51), ‘Stoneville 474’ (STV 474) and ‘LA 887’ (LA887) were selected based on a previous study on hybrid cotton yield by Cole et al. (2009). The coefficients of parentage for these cultivars were 0.15 for DP51 and STV474, 0.17 for DP51 and LA887, and 0.37 for STV474 and LA887 (Cole et al. 2009). DP51 and STV474 are early maturing cultivars whereas LA887 is a full season cultivar. These three cultivars were crossed in a half-diallel without reciprocals to give three hybrid combinations: DP 51 x STV 474, STV 474 x LA 887, and LA 887 x STV 474. Hybrids were generated by hand pollinations during summers of 2008 and 2009 at Central Crops Research Station, Clayton, NC. Hybrids and the three parents were tested at five environments during summer of 2010. The environments were selected to represent different growing conditions and a range of soil types. Plots at two of the locations were irrigated and the other three locations were rain fed to increase environment variability and sample high yielding as well as low

yielding environments. All entries were grown as two-row plots with three replicates in a randomized complete block design. Plot lengths ranged from 12.19m to 13.4m and row spacing ranged from 91.4 cm to 96.5 cm. Standard cultural practices were implemented for each location.

In each plot, an area of 1 m<sup>2</sup> with uniform plant stand was selected for plant mapping data collection. Data for plant height, node of first fruiting (sympodial) branch, number of fruiting sites and boll retention was collected for every plant in the sampled area to evaluate plant structure and distribution and retention of bolls on the plants prior to harvesting (Bourland and Watson 1990). The number of fruiting sites was calculated as combined number of open bolls, closed bolls and scars for each entry at each location. Percent fruit retention was calculated as a ratio of fruits retained to total number of fruiting sites (Bourland and Watson 1990).

Harvesting was performed using a mechanical cotton picker at each location and total plot yields for seed cotton were measured. From each plot a random sample of 25 bolls was collected manually before mechanical harvesting. The 25 boll samples were ginned on a 10 saw laboratory gin. Measurements of seed weight and lint weight from the samples were used to calculate lint percent, lint yield, boll size and lint per boll (Worley et al. 1976). Seed index was calculated as weight of 100 fuzzy seed, and lint index was calculated as weight of lint obtained from 100 seeds.

Mid-parent heterosis was calculated for each hybrid using the formula,  $[(F_1 - MP)/MP] * 100$ . Data analysis for number of fruiting sites, boll retention and percent heterosis was performed using General linear model procedure of SAS version 9.3.1 (SAS Institute

Inc. Cary, NC). The model used for data analysis incorporated location, reps (location), entry and entry x location effects. Genotypes were considered fixed effects, whereas environments and replications were considered random effects. The least significant difference (LSD) at 0.05 level of probability was used to separate the means. Pearson correlations between yield components were calculated using PROC CORR in SAS.

## **RESULTS**

### **Environment effects on entry means**

Analysis of variance revealed significant interactions between entries and environments for lint yield (Table 2.1). Irrigation regime affected lint yield, and both irrigated environments produced more lint than the non-irrigated environments (Table 2.2). In a comparison of dry vs. irrigated environments, significant differences were observed for all yield components except lint percent (data not shown). Across locations, irrigated test site at Arkansas had the highest overall mean lint yield. Based on least significant difference of means for lint yield, environments were arranged in a gradient from low yielding to high yielding environments. Both test sites at Arkansas were in the high yielding group with group mean as 1407 Kg/ha. Moderate yielding group had a mean of 1148 Kg/ha and included the tests at Florence, SC and Rocky Mount, NC. The correlation between average environment lint yield and boll per m<sup>2</sup> was 0.81 ( $p < 0.005$ ) whereas the correlation between mean environment yield and boll size was -0.55 ( $p < 0.005$ ). A negative trend of decreasing means was observed for lint index, seed index and boll size. High yielding environments had higher bolls number and fruiting sites but lower boll size, seed and lint index (Table 2.2).

### **Entry Means across locations**

Of the three parental entries, LA887 was the lowest yielding and STV474 was the highest yielding entry across locations (Table 2.3). Even though LA887 had the largest bolls with highest lint per boll and more fruiting sites per meter square than the other two inbreds, poor boll retention caused a decrease in the number of harvestable bolls leading to lower lint yield. STV474 had the highest lint percent with smaller bolls and less seeds per boll than DP51 and LA887. DP51 had the shortest plants and started producing fruiting branches earlier than other two parents (Table 2.3). Despite having fewer fruiting sites, DP51 was able to produce more harvestable bolls than other two parents due to higher boll retention (Table 2.3).

Hybrids had significantly higher group means for lint yield and lint per boll than parents (Table 2.3). There were no significant differences between hybrids and parent group means for any other traits including boll number. Both mid-parent and high-parent heterosis was observed for lint yield in all three hybrids. DP51 x STV474 had the smallest bolls and least seeds per boll among the three hybrid entries. The highest lint per boll was observed for STV474 x LA887 with more lint per seed than other two hybrid entries (Table 2.3).

### **Correlations between means of yield components**

Boll number was the main yield component positively correlated with lint yield ( $r^2 = 0.76$ ,  $p < 0.001$ ) (Table 2.4). Boll retention and number of fruiting sites were positively correlated to boll number. The correlation between boll number and number of fruiting sites was 0.63 while the correlation between boll number and boll retention was 0.51. Boll number was negatively correlated with all within boll yield components (Table 2.4). Within boll components were negatively correlated with lint yield. As lint yield increased, boll size

decreased with a decrease in lint per boll and lint per seed. Lint per boll was positively correlated with lint per seed and seed per boll (Table 2.4). An increase of within-boll yield components, seed per boll and lint per seed was correlated with a decrease in boll retention and bolls number (Table 2.4).

### **Correlation between mid-parent heterosis for lint yield, yield components and environmental mean lint yield**

When mid-parent heterosis for lint yield was plotted against an environment gradient from low to high yielding environments, two out of the three hybrids showed decreasing level of lint yield heterosis with increasing environmental mean (Fig. 2.1). Hybrids DP51 x STV474 and STV474 x LA887 had a negative relationship between lint yield heterosis and lint yield environment mean with a correlation of -0.94 and -0.97 respectively ( $p < 0.05$ ). In contrast, the  $F_1$  hybrid LA887 x DP51 had a correlation of 0.92 ( $p$ -value = 0.03) between lint yield heterosis and lint yield environment mean suggesting that lint yield heterosis was lower in poor environments than in higher yielding environments for this hybrid. Because different hybrid combinations showed different correlations between lint yield heterosis and environmental mean, individual hybrid combinations were studied for the components of hybrid cotton lint yield and their relationship with environment. All three hybrids were studied separately to identify the yield components associated with this relationship. In addition, the role of parental combinations in the relationship between lint yield heterosis and environmental yield were also studied for individual hybrids.

### **DP51 x STV474**

For hybrid DP51 x STV474, lint per boll was the primary yield component associated with a change in lint yield heterosis across increasing mean environment yield (Table 2.5). Boll number was not correlated to lint yield heterosis. As environment yield increased, lint per boll heterosis decreased with a correlation of -0.90 ( $p < 0.05$ ). Lint per boll was also the only yield component to be positively correlated with lint yield ( $r^2 = 0.89$ ,  $p < 0.05$ ). Lint per boll is a multiplicative product of lint per seed and seeds per boll. Of these two components lint per seed heterosis was correlated with lint per boll heterosis. STV474 had a strong influence on the relationship of lint yield heterosis with the environment gradient for lint yield. In high yielding environments, STV474 produced lint yield comparable to the hybrid (Fig. 2.2A). This resulted in more mid-parent lint yield and less percent heterosis for lint yield. STV474 had less lint per boll than the hybrid in low yielding environment but produced more lint per boll in high yielding environments leading to increased lint yield (Fig. 2.2C).

### **STV474 x LA887**

Of all the yield components studied, Boll number heterosis showed a negative correlation with environment gradient for lint yield for hybrid STV x LA887 (Table 2.5). Boll number heterosis increased as lint yield heterosis increased. This indicated that hybrids have more bolls per hectare as compared to parents in poor yielding environments resulting in higher hybrid lint yield. More bolls on a plant are a result of number of fruiting sites produced and total bolls retained prior to harvest. There was a negative relationship between boll retention and environment yield gradient ( $r^2 = -0.98$ ,  $p < 0.05$ ). As boll number heterosis decreased with increase in mean lint yield of environment, boll retention heterosis also decreased (Table

2.5). Boll retention heterosis was positively correlated to lint yield heterosis ( $r^2 = 0.94$ ,  $p < 0.05$ ) and Boll number heterosis ( $r^2 = 0.95$ ,  $p < 0.05$ ). The number of fruiting sites was not significantly correlated to boll number heterosis. None of the within boll yield components showed significant correlations to environment gradient for lint yield. However, lint per boll heterosis was negatively correlated to lint yield heterosis (Table 2.5). Similar to the first hybrid, the relationship of lint yield heterosis and increasing mean environment lint yield for STV474 x LA887 was mostly affected by STV474 (Fig. 2.3A). Although in this case, STV474 produced lint yield equivalent to hybrid by having more bolls in high yielding environments to increase yield (Fig. 2.3B).

#### **LA887 x DP51**

Unlike the first two hybrids lint yield heterosis for LA887 x DP51 increased as environmental mean lint yield increased with a correlation of 0.92 (p-value 0.03) between lint yield heterosis and environment gradient for lint yield (Table 2.5). The correlation coefficient between boll number heterosis and environment mean lint yield was positive but not significant. But boll number heterosis had a strong correlation with boll retention heterosis (Table 2.5). Lint per boll heterosis and lint per seed decreased with increase in mean environment yield although the significance level for correlation between lint per boll and environment was very low. Lint yield heterosis was negatively correlated with lint per seed heterosis (Table 2.5). None of the other yield components had significant correlations with lint yield heterosis. None of the parental lines appeared to be driving the increase in lint yield heterosis with an increase in mean lint yield of the environment (Fig. 2.4).

## DISCUSSION

In this study, three cultivars and three hybrids were tested across multiple locations to find the differences in contribution of yield components towards lint yield of inbreds and hybrids and to understand how they respond to different environments. In order to sample a greater variation of growing conditions for the current study trials at two out of five locations were irrigated.

We found that boll number was the primary yield component contributing to lint yield. Previous studies have also reported the importance of boll number towards cotton yield (Turner 1953; Worley et al. 1974; Zeng and Wu 2012). Number of fruiting sites and boll retention, which are the yield components contributing to boll number were negatively correlated to each other. Since fruit initiation and retention is dependent upon availability of resources and accumulation of photosynthates by the plant, a higher boll load restrains the availability of photosynthates to developing bolls, leading to their abscission (Guinn 1985). Lint yield as well as boll number was negatively correlated with lint per boll and boll size. Similar relationships among these yield components have been reported in previous studies (Zeng and Wu 2012; Culp and Harrell 1975). Culp and Harrell (1975) recommended that yield improvements could be made by selecting for medium to small bolls with more seeds per boll to increase lint per boll.

Physiological growth and lint yield of cotton is significantly affected by water stress (Pettigrew 2004). In the present study, a marked increase in vegetative growth was observed with an increase in water availability as indicated by increase in plant height and higher

location of node of first fruiting branch (sympodial branch) on the plant axis. Likewise lint yield was higher for all entries in irrigated locations as compared to dry locations (Table 2.2). Increased lint yield in irrigated environments was most closely related to an increase in boll number. This finding is in agreement with previous studies stating that yield reduction in dry environments is generally due to a reduction in the total number of harvestable bolls (Stockton et al. 1961). Pettigrew (2004) reported a 25% lint yield decrease in water stressed environments associated with a reduction in the number of bolls. The phase of growth in which water stress occurs also affects yield. Yield losses are higher if water stress occurs in the peak flowering period as compared to other stages in the flowering period (Grimes et al. 1970). Thus, yield reduction across non-irrigated environments can differ due to timing and the amount of rainfall during the growing period along with other factors like heat stress. Variability observed among the non-irrigated locations in the current study is possibly a result of similar environment variation. Of the three non-irrigated locations, Lewiston produced the lowest lint yield along with fewer bolls per meter square (Table 2.2).

In absence of water stress, cotton plants develop a larger canopy leading to an increase in photosynthetic area and greater photosynthetic accumulation. Higher photosynthate production allows the plant to produce more bolls (Guinn 1985). Plants in irrigated environments are also able to sustain flower production for a longer time as compared to plants in dry environments as well as support fruit production (McMichael and Hesketh 1982; Pettigrew 2004). However, an increase in irrigation above a certain threshold results in the production of fewer bolls due poor boll retention (Grimes et al. 1969). This demonstrates the impact of water availability on yield. There was a significant interaction

between entries and environment (Table 2.1). Therefore all hybrid and parent combinations were studied separately.

Since LA887 is a full season cultivar, a higher number of fruiting sites is likely the result of floral initiation that continues for a time longer than the early maturing cultivars DP51 and STV474. Early maturity cultivars had higher boll retention, which lead to higher yield. This indicated more effective utilization of plant reserves by early maturity cultivars in stressed environments.

Hybrids produced higher lint yield than parental entries and all hybrids out-yielded their respective high yielding parent (Table 2.3). When expressed as a fraction of the mid-parent, two of the three hybrids showed a negative relationship between lint yield heterosis and environmental mean lint yield. Better hybrid performance over parents in low yielding environments may be a result of hybrids having vegetative advantage over parents during seedling growth stage as shown in earlier studies (Wells et al. 1988); however this component was not measured in the current study. Even though both hybrids produced higher lint yields than the mid-parent values in low yielding environments, the yield components associated with yield advantage of hybrids over the mid-parent were different in both cases. In the case of DP51 x STV474, lint per boll was the primary yield component to affect lint yield heterosis. Of the two sub-components, lint per seed was the main component contributing to higher lint per boll in low yielding environments. For STV474 x LA887, Bolls  $m^{-2}$  was the main yield component to affect lint yield with retention being more important that number of fruiting sites towards increased hybrid yield in low yielding environments. None of the yield components could explain the change in level of heterosis

for lint yield of hybrid LA887 x DP51 using data from the current study, even though lint per boll and lint per seed were correlated to the change in mean environment yield. The relationship between mid-parent heterosis and environment is affected by parental combinations.

Generally, high yielding parents produce higher yielding hybrids (Khorgade et al. 2000). However, the level of heterosis is much higher when two low yielding inbreds are crossed to produce a hybrid. It has been reported that parents with lighter bolls contributed positive additive effects to boll number and parents with heavier bolls contributed negative additive effects to boll number (Jenkins et al. 2009). An increase in mid-parent yield compared to hybrid yield decreases the mid-parent heterosis value. Miller and Lee (1964), tested hybrids across different environments and reported that the change in yield heterosis across environments was not accompanied by a change in magnitude of heterosis on a pound per acre basis. Thus, with the same magnitude of heterosis, the estimation of heterosis would be higher if the mid-parent value is lower in a low yielding environment and vice versa. A combined study using data from cotton as well as maize concluded that the decrease in percent heterosis over time was due to faster improvement in parental inbred yields than hybrid yield (Troyer and Wellin 2009). In a related study of how high yielding parental lines affect hybrid yield, Miller and Lee (1964) examined 22 entries and their hybrids across multiple locations. Higher mid-parent values could be driven by either one or both the parents. In the current study, STV474 was the highest yielding parental entry and both the hybrids with STV474 as a parent produced more lint than LA887 x DP51.

LA887 x DP51 showed a positive relationship between lint yield heterosis and environment mean contrary to the other two hybrids. In a previous study, the bulk of single and reciprocal crosses made using DP51 and LA887 showed a decrease in lint yield heterosis with an increase in environmental mean (Cole et al. 2009). In this case, the hybrid and its parents were tested in 18 environments, much larger than the number of environments included in current study. It is therefore possible that the difference in the relationship between heterosis and environment gradient in the two studies resulted from differences in the environments sampled.

The results of this study show that expression of heterosis for lint yield in cotton is affected by environment. The relationship of lint yield heterosis and environment is also influenced by parental combination. Boll number, boll retention and lint per boll were found to be the major yield components affecting relationship between midparent heterosis for lint yield and environment. However, no single yield component was associated with changes in heterosis across different environments for all three hybrid combinations. Further studies are required to understand the effect of parental combinations on relationship of mid-parent heterosis with environment.

## REFERENCES

- Bauer P, Green C (1996) Evaluation of F-2 genotypes of cotton for conservation tillage. *Crop Sci* 36:655-658
- Bourland FM, Watson Jr. CE (1990) COTMAP, a technique for evaluating structure and yield of cotton plants. *Crop Sci* 30:224-226
- Campbell BT, Bowman DT, Weaver DB (2008) Heterotic effects in topcrosses of modern and obsolete cotton cultivars. *Crop Sci* 48:593-600
- Cole CB, Bowman DT, Bourland FM, Caldwell WD, Campbell BT, Fraser DE, Weaver DB (2009) Impact of heterozygosity and heterogeneity on cotton lint yield stability. *Crop Sci* 49:1577-1585
- Culp T, Harrell D (1975) Influence of lint percentage, boll size, and seed size on lint yield of upland cotton with high fiber strength. *Crop Sci* 15:741-746
- Duvick DN (1992) Genetic contributions to advances in yield of US maize. *Maydica* 37:69-79
- Ehlig CF, Lemert RD (1973) Effects of fruit load, temperature, and relative humidity on boll retention of cotton. *Crop Sci* 13:168-171
- Galal HES, Miller PA, Lee JA (1966) Heterosis in relation to development in upland cotton *Gossypium hirsutum* L. *Crop Sci* 6:555-559
- Grimes DW, Dickens WL, Anderson WD (1969) Functions for Cotton (*Gossypium hirsutum* L.) production from irrigation and nitrogen fertilization Variables: II. Yield components and quality characteristics. *Agron J* 61:773-776
- Grimes DW, Miller R, Dickens L (1970) Water stress during flowering of cotton. *Calif Agric* 24:4-6

Guinn G (1974) Abscission of cotton floral buds and bolls as influenced by factors affecting photosynthesis and respiration. *Crop Sci* 14:291-293

Guinn G, Mauney JR (1984) Fruiting of cotton: II. Effects of plant moisture status and active boll load on boll retention. *Agron J* 76:94-98

Guinn G (1985) Fruiting of cotton: III. Nutritional stress and cutout. *Crop Sci* 25:981-985

Jenkins JN, Gutierrez OA, McCarty JC, Wu J (2009) Genetic variance components and genetic effects among eleven diverse upland cotton lines and their F2 hybrids. *Euphytica* 167:397-408

Jones JE, Loden HD (1951) Heterosis and combining ability in upland cotton. *Agron J* 43:514-516

Kerr T (1966) Yield components in cotton and their interrelations with fiber quality. In *Cotton Improvement Conference (18th) Memphis, Tennessee (National Cotton Council)*. pp 276-287

Khorgade PW, Stange IV, Mesham LD (2000) Diallel analysis in American cotton (*Gossypium hirsutum* L.). *Indian J Agric Res* 34:172-5

Liu W, Tollenaar M (2009) Physiological mechanisms underlying heterosis for shade tolerance in maize. *Crop Sci* 49:1817-1826

Marani A (1967) Heterosis and combining ability in intraspecific and interspecific crosses of cotton. *Crop Sci* 7:519-522

Marani A (1968) Heterosis and inheritance of quantitative characters in interspecific crosses of cotton. *Crop Sci* 8:299-303

McMichael BL, Hesketh JD (1982) Field investigations of the response of cotton to water deficits. *Fields Crops Res* 5:319-333

Meredith WR, Bridge RR (1972) Heterosis and gene action in cotton, *Gossypium hirsutum* L. Crop Sci 12:304-310

Miller PA, Lee JA (1964) Heterosis and combining ability in varietal top crosses of upland cotton, *Gossypium hirsutum* L. Crop Sci 5:646-649

Pettigrew WT (2004) Moisture deficit effects on cotton lint yield, yield components, and boll distribution. Agron J 96:377-383

Reddy KR, Hodges HF, Reddy VR (1992) Temperature effects on cotton fruit retention. Agron J 84:26-30

Stockton JR, Doneen LD, Walhood VT (1961) Boll shedding and growth of the cotton plant in relation to irrigation frequency. Agron J 53:272-275

Troyer AF, Wellinb EJ (2009) Heterosis decreasing in hybrids: Yield test inbreds. Crop Sci 49:1969-1976

Turner JH (1953) A study of heterosis in upland cotton: 1. Yield of hybrids compared with varieties. Agron J 45:484-486

Wells R, Meredith WR (1986) Heterosis in upland cotton: 1. Growth and leaf-area partitioning. Crop Sci 26:1119-1123

Wells R, Meredith WR, Williford JR (1988) Heterosis in upland cotton: 2. Relationship of leaf-area to plant photosynthesis. Crop Sci 28:522-525

Worley S, Culp TW, Harrell DC (1974) Relative contributions of yield components to lint yield of upland cotton, *Gossypium hirsutum* L. Euphytica 23:399-403

Worley S, Ramey HH, Harrell DC, Culp TW (1976) Ontogenetic model of cotton yield. Crop Sci 16:30-34

Young E, Murray J (1966) Heterosis and inbreeding depression in diploid and tetraploid cottons. *Crop Sci* 6:436-438

Zeng L, Wu J (2012) Germplasm for genetic improvement of lint yield in Upland cotton: genetic analysis of lint yield with yield components. *Euphytica* 187:247-261

Table 2.1 ANOVA table showing source of variation, degrees of freedom, mean squares and significance for main effects and interaction for lint yield tested in five locations

<b>Source</b>	<b>DF</b>	<b>Mean squares</b>	<b>F-value</b>	<b>P-value</b>
Locations	4	135895	40.75	<.0001
Reps(locations)	10	23336	2.07	0.044
Entry	5	185580	8.73	0.0002
Entry X Location	20	21250	1.89	0.035
Error	50	11248		

Table 2.2 Mean values for yield components for the five locations averaged across six entries

<b>Location</b>	<b>HT</b> <b>(cm)</b>	<b>FN</b>	<b>LY</b> <b>(kg/ha)</b>	<b>BPM</b>	<b>RE</b> <b>T</b>	<b>NFSM</b>	<b>LPB</b> <b>(g)</b>	<b>SPB</b>	<b>LPS</b> <b>(g)</b>	<b>LP</b>	<b>BS</b> <b>(g)</b>	<b>SI</b>	<b>LI</b>
<i>Irrigated</i>													
Keiser, AR, Site I	104.6	7.9	1433	86	30.1	291	1.81	28.7	0.063	40.0	4.5	9.5	6.3
Keiser, AR, Site II	109.5	8.5	1381	92	28.8	323	1.73	26.6	0.065	40.2	4.2	9.3	6.4
<i>Non-irrigated</i>													
Florence, SC Rocky Mt, NC	100.5	6.0	1178	54	18.2	308	2.39	30.0	0.080	42.4	5.5	10.4	7.8
Lewiston, NC	78.5	5.7	739	45	21.2	213	2.16	29.0	0.074	40.0	5.2	10.6	7.2
LSD $\alpha=0.05$	6.7	1.0	71	7	2.4	26	0.08	0.7	0.002	1.3	0.2	0.3	0.3

Abbreviations: HT – Plant height, FN – First fruiting node, LY – Lint yield, BPM – Bolls per meter square, RET – Percent of bolls retained, NFSM – Number of fruiting sites per meter square, LPB – Lint per boll, SPB – Seeds per boll, LPS – Lint per seed, LP – lint percent, BS – Boll size and SI – Seed Index

Table 2.3 Mean values for entries for lint yield and its components across locations

<b>Entry</b>	<b>HT (cm)</b>	<b>FN</b>	<b>LY (kg/ha)</b>	<b>BPM</b>	<b>RET</b>	<b>NFSM</b>	<b>LPB (g)</b>	<b>SPB</b>	<b>LPS (g)</b>	<b>LP</b>	<b>BS (g)</b>	<b>SI</b>	<b>LI</b>
DP51	82.0	6.1	1087	71	29.9	244	1.90	29.2	0.064	39.0	4.7	9.6	6.3
STV474	89.4	7.6	1182	67	27.5	248	1.93	25.4	0.076	42.0	4.4	9.7	7.3
LA887	92.5	7.3	995	58	21.9	272	2.24	30.4	0.073	39.6	5.5	10.7	7.2
DP51 x STV474	92.7	6.3	1274	75	28.6	274	2.00	27.6	0.072	40.7	4.7	10.0	7.0
STV474 x LA887	88.9	6.5	1282	65	26.4	254	2.21	28.6	0.077	41.8	5.1	10.2	7.5
LA887 x DP51	91.7	6.5	1198	60	24.0	258	2.10	30.4	0.069	40.1	5.2	10.1	6.8
LSD $\alpha$ =0.05	5.6	0.6	111	11	3.5	31	0.12	0.9	0.003	1.6	0.3	0.3	0.4
Parents	88.0	7.0	1088	65	26.4	255	2.02	28.3	0.071	40.2	4.9	10.0	6.9
Hybrids	91.1	6.4	1251	67	26.3	262	2.10	28.9	0.073	40.9	5.0	10.1	7.1
LSD $\alpha$ =0.05	3.5	0.6	44	7	2.7	17	0.07	0.7	0.002	1.4	0.1	0.1	0.4

Abbreviations: HT – Plant height, FN – First fruiting node, LY – Lint yield, BPM – Bolls per meter square, RET – Percent of bolls retained, NFSM – Number of fruiting sites per meter square, LPB – Lint per boll, SPB – Seeds per boll, LPS – Lint per seed, LP – lint percent, BS – Boll size and SI – Seed Index

Table 2.4 Correlation of lint yield with yield components over entries

<b>Trait</b>	<b>LY</b>	<b>BPM</b>	<b>Ret</b>	<b>NFSM</b>	<b>LPB</b>	<b>SPB</b>	<b>LPS</b>	<b>LP</b>	<b>BS</b>	<b>SI</b>
BPM	0.76*									
Ret	0.45*	0.51*								
NFSM	0.50*	0.63*	-0.31*							
LPB	-0.47*	-0.79*	-0.56*	-0.28						
SPB	-0.32	-0.46*	-0.45*	-0.05	0.61*					
LPS	-0.39*	-0.70*	-0.39*	-0.35	0.84*	0.09				
LP	0.26	-0.17	-0.28	0.13	0.37*	-0.20	0.60*			
BS	-0.55*	-0.78*	-0.57*	-0.29	0.95*	0.79*	0.66*	0.13		
SI	-0.66*	-0.82*	-0.48*	-0.46*	0.89*	0.53*	0.76*	0.11	0.91*	
LI	-0.31	-0.68*	-0.46*	-0.27	0.85*	0.14	0.97*	0.72*	0.67*	0.75*

\* Significantly different at 0.05 level of probability

Abbreviations: LY – Lint yield, BPM – Bolls per meter square, Ret – Percent of bolls retained, NFSM – Number of fruiting sites,

LPB – Lint per boll, SPB – Seeds per boll, LPS – Lint per seed, LP – lint percent, BS – boll size and SI – Seed Index

Table 2.5 Correlation between environment mean lint yield and heterosis values for lint yield and yield components for three hybrids.

	ENV	LY	BPM	RET	NFSM	LPB	SPB	LPS	LP	BS	SI
<b>DP51 x STV474</b>											
LY	-0.94										
BPM	0.52	-0.37									
RET	0.31	-0.19	0.62								
NFSM	0.43	-0.38	0.66	-0.17							
LPB	-0.90*	0.89*	-0.35	0.12	-0.62						
SPB	0.05	0.24	0.72	0.45	0.38	0.10					
LPS	-0.82†	0.67	-0.76	-0.17	-0.79	0.83†	-0.48				
LP	-0.71	0.74	-0.36	0.33	-0.82†	0.93*	0.06	0.79			
BS	-0.71	0.76	-0.05	0.46	-0.55	0.94*	0.31	0.64	0.92*		
SI	0.44	-0.63	-0.33	0.11	-0.40	-0.34	-0.78	0.13	-0.11	-0.35	
LI	-0.68	0.64	-0.51	0.25	-0.91*	0.88†	-0.17	0.88†	0.97*	0.83†	0.10
<b>STV474 x LA887</b>											
LY	-0.97*										
BPM	-0.91*	0.85†									
RET	-0.98*	0.94*	0.95*								
NFSM	-0.16	0.14	0.52	0.30							
LPB	0.79	-0.88†	-0.50	-0.69	0.24						
SPB	0.72	-0.65	-0.73	-0.83†	-0.31	0.38					
LPS	0.27	-0.42	0.04	-0.08	0.52	0.75	-0.33				
LP	-0.08	0.22	-0.03	0.15	0.15	-0.35	-0.37	-0.10			
BS	0.61	-0.68	-0.28	-0.57	0.42	0.85†	0.56	0.49	-0.56		
SI	-0.32	0.21	0.67	0.41	0.82†	0.24	-0.28	0.50	-0.39	0.52	
LI	-0.23	0.19	0.50	0.42	0.84†	0.16	-0.67	0.67	0.45	0.06	0.58
<b>LA887 x DP51</b>											
LY	0.92*										
BPM	0.53	0.43									
RET	0.45	0.45	0.97*								
NFSM	0.55	0.24	0.81†	0.64							
LPB	-0.86†	-0.71	-0.42	-0.28	-0.56						
SPB	0.01	0.12	-0.18	-0.06	-0.26	0.48					
LPS	-0.97*	-0.84†	-0.40	-0.29	-0.54	0.94*	0.16				
LP	-0.35	-0.01	-0.10	0.09	-0.59	0.25	-0.27	0.38			
BS	-0.72	-0.56	-0.56	-0.41	-0.67	0.95*	0.67	0.81†	0.15		
SI	-0.72	-0.47	-0.86†	-0.71	-0.97*	0.71	0.28	0.69	0.47	0.78	
LI	-0.83†	-0.55	-0.50	-0.31	-0.81	0.77	0.04	0.86†	0.77	0.69	0.84†

\* Significant at 0.05 and † significant at 0.1 level of probability

Abbreviations: ENV- Environment mean lint yield, LY – Lint yield, BPM – Bolls per meter square, RET – Percent of bolls retained, NFSM – Number of fruiting sites, LPB – Lint per boll, SPB – Seeds per boll, LPS – Lint per seed, LP – lint percent, BS – boll size and SI – Seed Index

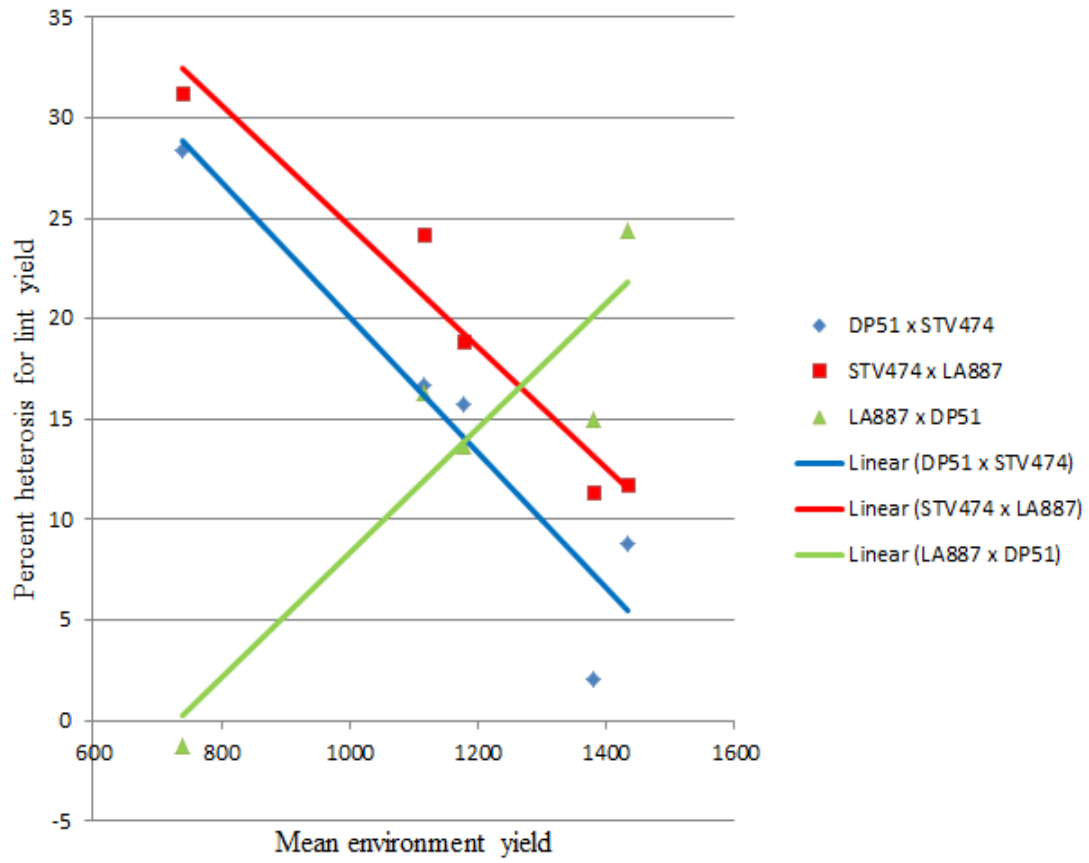


Figure 2.1 Relationship of percent heterosis for lint yield with mean environmental yield (Kg/ha) from low yielding to high yielding environments

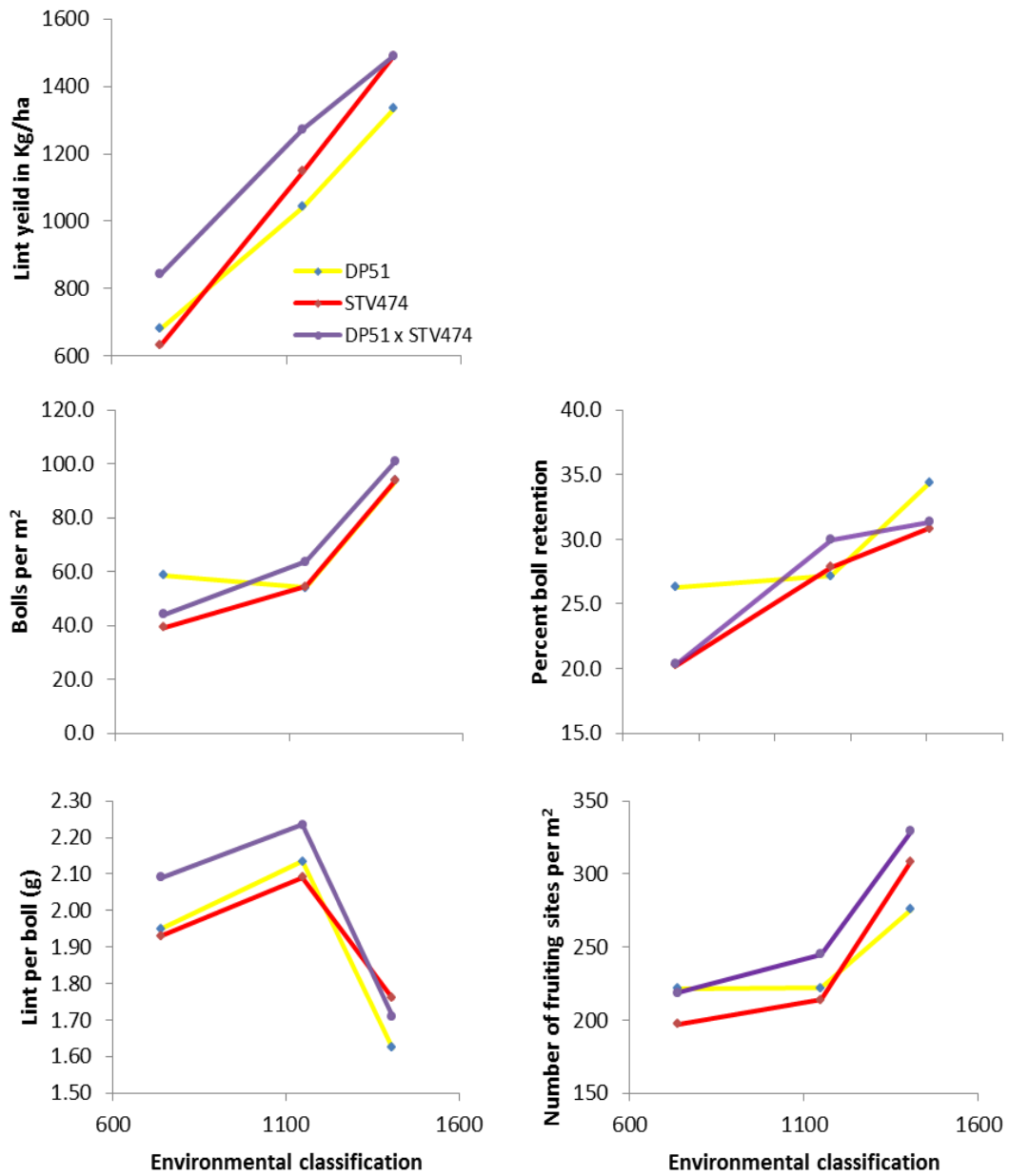


Figure 2.2 Lint yield and major yield components over low to high yielding environments for DP51 x STV474 and its parents

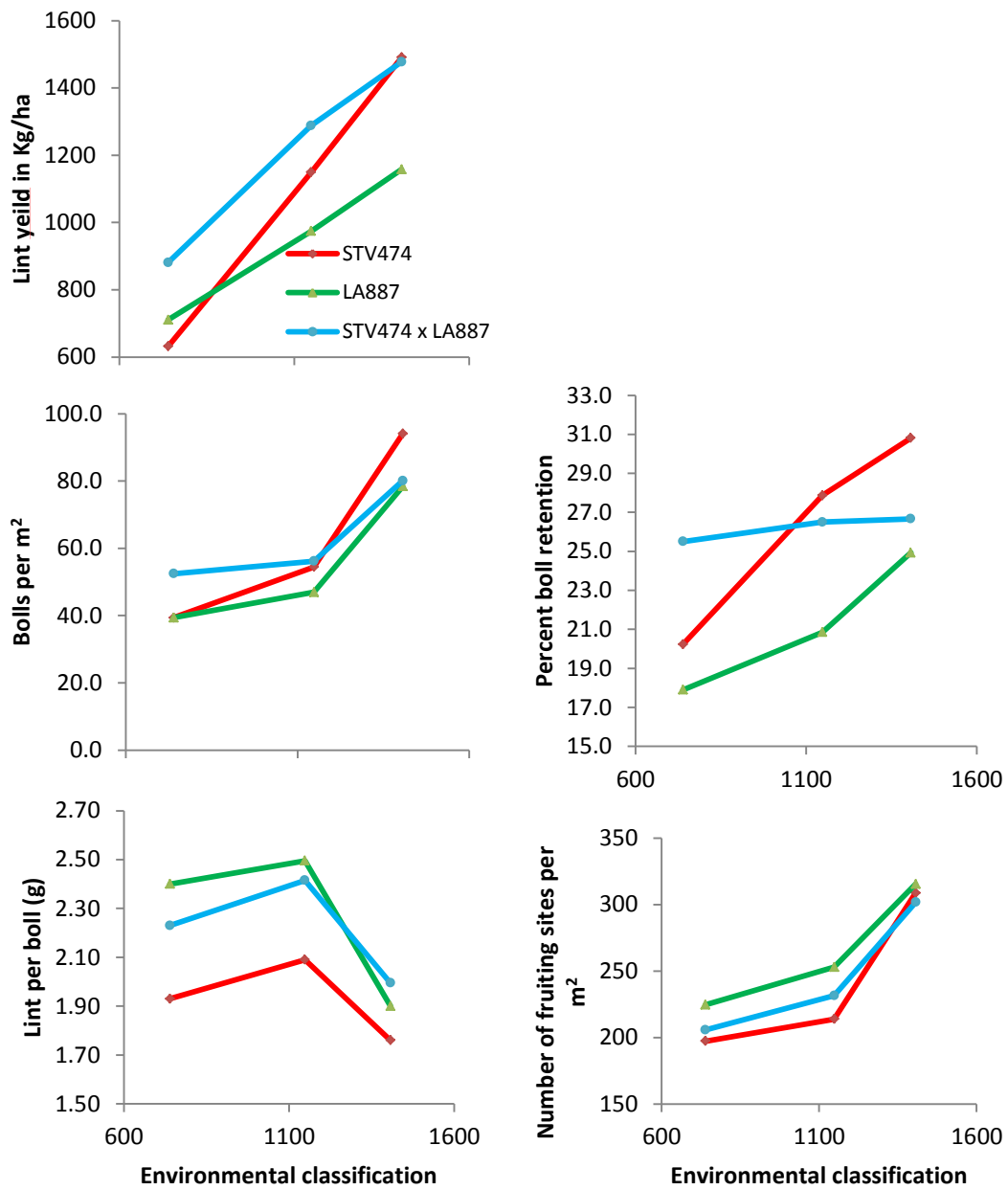


Figure 2.3 Lint yield and major yield components over low to high yielding environments for STV474 x LA887 and its parents

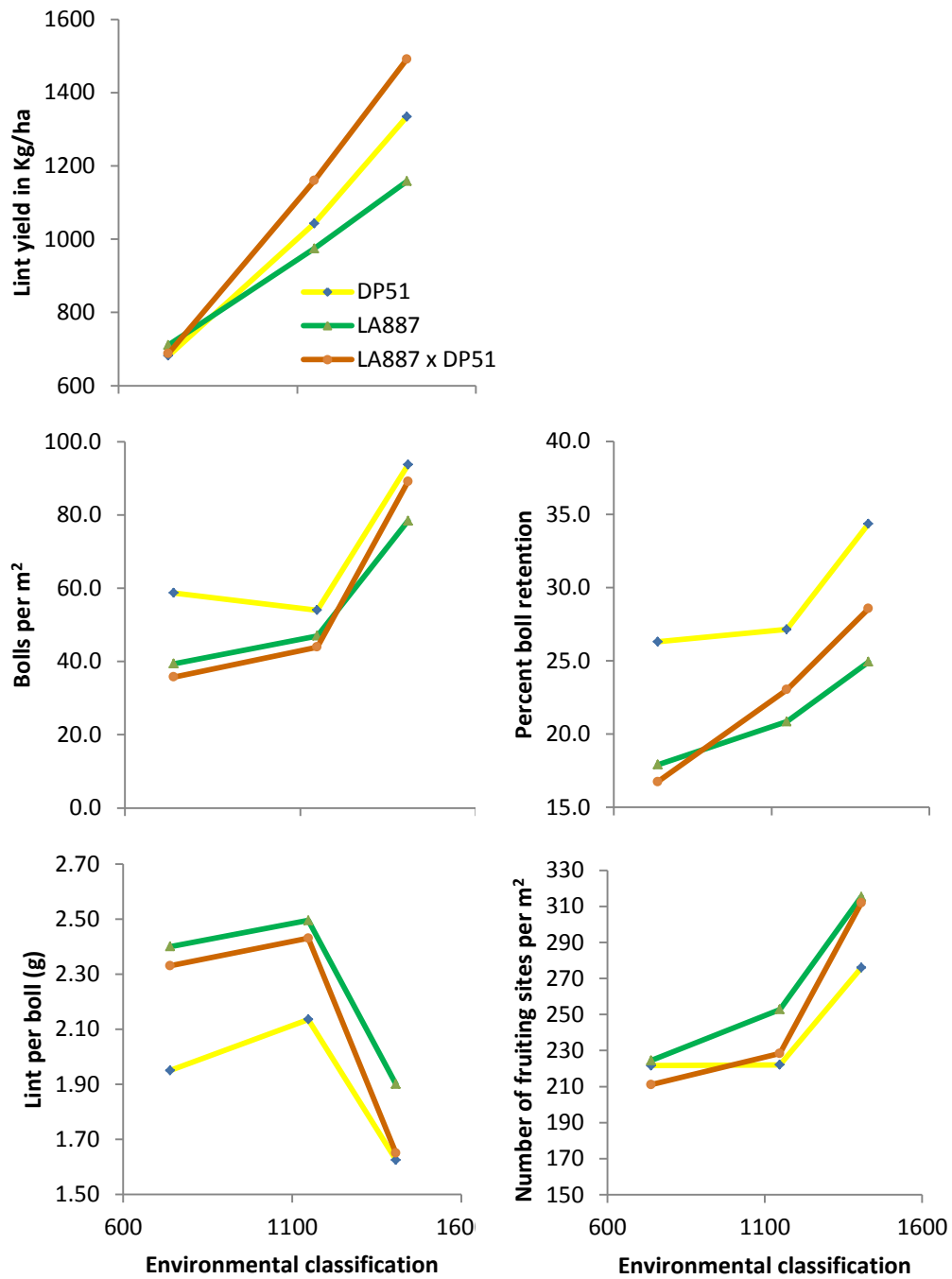


Figure 2.4 Lint yield and major yield components over low to high yielding environments for LA887 x DP51 and its parents

## APPENDICES

Appendix A. List of 381 cotton accessions with identification number (PI number)

<b>No.</b>	<b>Accession name</b>	<b>PI #</b>	<b>Species</b>
1	AC 241	PI 529534	<i>G. hirsutum</i>
2	ACALA #111, ROGERS	PI 528816	<i>G. hirsutum</i>
3	ACALA 1064(New Mexico)	PI 528606	<i>G. hirsutum</i>
4	ACALA 1517 WILT	PI 528758	<i>G. hirsutum</i>
5	ACALA 1517-70	PI 529290	<i>G. hirsutum</i>
6	ACALA 1517-75	PI 529542	<i>G. hirsutum</i>
7	ACALA 1517D	PI 529243	<i>G. hirsutum</i>
8	ACALA 29	PI 529427	<i>G. hirsutum</i>
9	ACALA 3080	PI 529543	<i>G. hirsutum</i>
10	ACALA 4-42	PI 529116	<i>G. hirsutum</i>
11	ACALA 44WR	PI 529118	<i>G. hirsutum</i>
12	ACALA 5	PI 529169	<i>G. hirsutum</i>
13	ACALA 51	PI 529428	<i>G. hirsutum</i>
14	ACALA 5675	PI 528674	<i>G. hirsutum</i>
15	ACALA 8	PI 528967	<i>G. hirsutum</i>
16	ACALA GLANDLESS 8160	PI 529536	<i>G. hirsutum</i>
17	ACALA MAXXA	PI540885	<i>G. hirsutum</i>
18	ACALA NAKED SEED	PI 528609	<i>G. hirsutum</i>
19	ACALA NUNN'S #5-37	PI 528603	<i>G. hirsutum</i>
20	ACALA ROYALE	PI542049	<i>G. hirsutum</i>
21	ACALA SJ-1	PI 529246	<i>G. hirsutum</i>
22	ACALA SJ-2	PI606810	<i>G. hirsutum</i>
23	ACALA SJ-3	PI 529537	<i>G. hirsutum</i>
24	ACALA SJ-4	PI 529538	<i>G. hirsutum</i>
25	ACALA SJC-1	PI 529539	<i>G. hirsutum</i>
26	Acala Ultima	PI 603079	<i>G. hirsutum</i>
27	ACALA YOUNG'S	PI 528755	<i>G. hirsutum</i>
28	ACALA, MESSILLA VALLEY 898	PI 529112	<i>G. hirsutum</i>
29	ACALA, N.M.8893	PI 529147	<i>G. hirsutum</i>
30	AK-DJURA FREGO	PI 528931	<i>G. hirsutum</i>
31	AK-DJURA GREEN LINT	PI 528929	<i>G. hirsutum</i>
32	AK-DJURA HIGG BROWN	PI 529165	<i>G. hirsutum</i>
33	AK-DJURA VIRESCENT	PI 528927	<i>G. hirsutum</i>

Appendix A continued.

<b>No.</b>	<b>Accession name</b>	<b>PI #</b>	<b>Species</b>
34	ALLEN 33	PI 529318	<i>G. hirsutum</i>
35	ALL-IN-ONE	PI 529018	<i>G. hirsutum</i>
36	AMBASSADOR	PI 528973	<i>G. hirsutum</i>
37	ARIZONA RED	PI 528427	<i>G. hirsutum</i>
38	ARK-1	PI 529547	<i>G. hirsutum</i>
39	ARK-2	PI 529548	<i>G. hirsutum</i>
40	ARKANSAS 10	PI 528912	<i>G. hirsutum</i>
41	ARKANSAS 12	PI 528914	<i>G. hirsutum</i>
42	ARKOT 8102	PI595852	<i>G. hirsutum</i>
43	ARKOT 8606	PI628634	<i>G. hirsutum</i>
44	ARKOT 8918	PI628638	<i>G. hirsutum</i>
45	AUBURN 56	PI 529215	<i>G. hirsutum</i>
46	Auburn 634 RNR	PI607237	<i>G. hirsutum</i>
47	Auburn Fg-310	PI 607242	<i>G. hirsutum</i>
48	Auburn Ne Fg-149	PI 607248	<i>G. hirsutum</i>
49	Auburn Ne Fg-16	PI 607247	<i>G. hirsutum</i>
50	BJAGL NECT	PI 529382	<i>G. hirsutum</i>
51	BLCABPD86S-1-90	PI603008	<i>G. hirsutum</i>
52	BLIGHT MASTER	PI 529202	<i>G. hirsutum</i>
53	BLIGHT MASTER A5	PI 529549	<i>G. hirsutum</i>
54	BOB SHAW 1	PI 528662	<i>G. hirsutum</i>
55	BOBDEL	PI 528669	<i>G. hirsutum</i>
56	BRONCO 360	PI601663	<i>G. hirsutum</i>
57	BRONCO 625	PI600977	<i>G. hirsutum</i>
58	Bronze	PI 528448	<i>G. hirsutum</i>
59	BULGARIA 432	PI 529463	<i>G. hirsutum</i>
60	C5HUG2BES-2-87	PI595762	<i>G. hirsutum</i>
61	C6-5	PI 529425	<i>G. hirsutum</i>
62	CA17	PI 529322	<i>G. hirsutum</i>
63	CA23	PI 529323	<i>G. hirsutum</i>
64	CA30	PI 529324	<i>G. hirsutum</i>
65	CABCHUS-2-86	PI595758	<i>G. hirsutum</i>
66	CABCSV506S-1-94	PI634320	<i>G. hirsutum</i>

## Appendix A continued.

<b>No.</b>	<b>Accession name</b>	<b>PI #</b>	<b>Species</b>
67	CABD3CABCH-1-89	PI603002	<i>G. hirsutum</i>
68	CABD3SHP3S-1-90	PI603007	<i>G. hirsutum</i>
69	CAHUGARPIH-1-88	PI602998	<i>G. hirsutum</i>
70	CAHUGLBBCS-1-88	PI603005	<i>G. hirsutum</i>
71	CARTER'S LONG STAPLE	PI 528722	<i>G. hirsutum</i>
72	CASCOT C-13	PI600895	<i>G. hirsutum</i>
73	CASCOT L-7	PI607181	<i>G. hirsutum</i>
74	CD3HCABCUH-1-89	PI603003	<i>G. hirsutum</i>
75	CD3HCAHUGH-2-88	PI603000	<i>G. hirsutum</i>
76	CD3HCHULBH-1-88	PI603001	<i>G. hirsutum</i>
77	CD3HHARCIH-1-88	PI602999	<i>G. hirsutum</i>
78	CLEVELAND 54	PI 528612	<i>G. hirsutum</i>
79	CLEVEWILT 6 NAKED SEED	PI 528611	<i>G. hirsutum</i>
80	COKER 100 WILT	PI 528761	<i>G. hirsutum</i>
81	COKER 100A (WR)	PI 529216	<i>G. hirsutum</i>
82	COKER 139	PI601389	<i>G. hirsutum</i>
83	COKER 201	PI 529247	<i>G. hirsutum</i>
84	COKER 310	PI 529249	<i>G. hirsutum</i>
85	COKER 312	PI 529278	<i>G. hirsutum</i>
86	COKER 3131	PI 529531	<i>G. hirsutum</i>
87	COKER 315	PI 529530	<i>G. hirsutum</i>
88	COKER 5110	PI 529279	<i>G. hirsutum</i>
89	COKER'S CLEVELAND 884	PI 528614	<i>G. hirsutum</i>
90	COKER'S CLEVEWILT 3	PI 528617	<i>G. hirsutum</i>
91	COKER'S DELTATYPE WEBBER #7	PI 528620	<i>G. hirsutum</i>
92	COKER'S DELTATYPE WEBBER #9	PI 528619	<i>G. hirsutum</i>
93	COKER'S WILDS #2	PI 528626	<i>G. hirsutum</i>
94	COKER'S WILDS #4	PI 528625	<i>G. hirsutum</i>
95	COKER'S WILDS #9	PI 528624	<i>G. hirsutum</i>
96	COLUMBIA	PI 528743	<i>G. hirsutum</i>
97	COOK-307-6	PI 528997	<i>G. hirsutum</i>
98	CS-8608	PI 513390	<i>G. hirsutum</i>
99	CS-8609	PI 513391	<i>G. hirsutum</i>

Appendix A continued.

<b>No.</b>	<b>Accession name</b>	<b>PI #</b>	<b>Species</b>
100	CS-8610	PI 513392	<i>G. hirsutum</i>
101	CS-8611	PI 513393	<i>G. hirsutum</i>
102	CUP LEAF	PI 529014	<i>G. hirsutum</i>
103	D2 SMOOTH MUTANT	PI 529170	<i>G. hirsutum</i>
104	DEL CERRO	PI 529358	<i>G. hirsutum</i>
105	DELCOT 277	PI 529258	<i>G. hirsutum</i>
106	DELCOT 344	PI601344	<i>G. hirsutum</i>
107	DELFOS 6102	PI 528958	<i>G. hirsutum</i>
108	DELFOS 9169 (ORIGINAL)	PI 528655	<i>G. hirsutum</i>
109	DELTA QUEEN	PI 529220	<i>G. hirsutum</i>
110	DELTAPINE 14	PI 528970	<i>G. hirsutum</i>
111	DELTAPINE 15	PI 528770	<i>G. hirsutum</i>
112	DELTAPINE 16	PI 529251	<i>G. hirsutum</i>
113	DELTAPINE 20	PI529567	<i>G. hirsutum</i>
114	DELTAPINE 41	PI 529527	<i>G. hirsutum</i>
115	DELTAPINE 45	SA-3607.01	<i>G. hirsutum</i>
116	DELTAPINE 50	PI529566	<i>G. hirsutum</i>
117	DELTAPINE 51	SA-3138	<i>G. hirsutum</i>
118	DELTAPINE 61	PI607174	<i>G. hirsutum</i>
119	DELTAPINE 90	PI 529529	<i>G. hirsutum</i>
120	DELTAPINE A	PI 528767	<i>G. hirsutum</i>
121	DELTAPINE PREMA	SA 1669	<i>G. hirsutum</i>
122	DELTAPINE SMOOTH LEAF	PI 529219	<i>G. hirsutum</i>
123	DELTAPINE SR-383	PI600905	<i>G. hirsutum</i>
124	DELTATYPE WEBBER	PI 528717	<i>G. hirsutum</i>
125	DELTATYPE WEBBER #4	PI 528628	<i>G. hirsutum</i>
126	DELTATYPE WEBBER (253-1) T142-8	PI 528844	<i>G. hirsutum</i>
127	DELTATYPE WEBBER 2139	PI 528598	<i>G. hirsutum</i>
128	DES 119	PI606809	<i>G. hirsutum</i>
129	DES 24	PI 529522	<i>G. hirsutum</i>
130	DES 56	PI 529520	<i>G. hirsutum</i>
131	DIXIE 14-5-2	PI 528629	<i>G. hirsutum</i>
132	DIXIE KING	PI 529021	<i>G. hirsutum</i>

## Appendix A continued.

<b>No.</b>	<b>Accession name</b>	<b>PI #</b>	<b>Species</b>
133	DIXIE TRIUMPH	PI 528956	<i>G. hirsutum</i>
134	DUNN 1047	PI601196	<i>G. hirsutum</i>
135	DUNN 325	PI601199	<i>G. hirsutum</i>
136	DUNN HS 120	PI601460	<i>G. hirsutum</i>
137	DURANGO	PI 529057	<i>G. hirsutum</i>
138	EARLISTAPLE 7	PI529570	<i>G. hirsutum</i>
139	EARLY FLOFF	PI 529047	<i>G. hirsutum</i>
140	EMPIRE	PI 529179	<i>G. hirsutum</i>
141	EMPIRE GL2 GL2	SA-1113	<i>G. hirsutum</i>
142	EMPIRE WR	SA-1158	<i>G. hirsutum</i>
143	EWINGS LONG STAPLE	PI 528630	<i>G. hirsutum</i>
144	EWINGS LONG STAPLE x TIDEWATER	PI 528726	<i>G. hirsutum</i>
145	EXPRESS 121	PI 528977	<i>G. hirsutum</i>
146	EXPRESS 432	PI 528702	<i>G. hirsutum</i>
147	F-2-C-10	PI 529181	<i>G. hirsutum</i>
148	Fibermax 966	N/A	<i>G. hirsutum</i>
149	FJA	PI529572	<i>G. hirsutum</i>
150	FLORIDA GREEN SEED	PI 528694	<i>G. hirsutum</i>
151	FOX 4	PI 529225	<i>G. hirsutum</i>
152	FREGO	PI 528584	<i>G. hirsutum</i>
153	FREGO NANKEEN	PI 528937	<i>G. hirsutum</i>
154	FREGO UPLAND CR.DW.MEA	PI 528934	<i>G. hirsutum</i>
155	FREGO VIRESCENT	PI 528932	<i>G. hirsutum</i>
156	FTA	PI529573	<i>G. hirsutum</i>
157	GA 161	PI612959	<i>G. hirsutum</i>
158	GEORGIA KING	PI552537	<i>G. hirsutum</i>
159	GERMAIN'S ACALA GC-352	PI601180	<i>G. hirsutum</i>
160	GERMAINS ACALA GC-356	PI601474	<i>G. hirsutum</i>
161	Germain's Acala GC-410	PI601390	<i>G. hirsutum</i>
162	GLANDLESS NC-1	PI 529183	<i>G. hirsutum</i>
163	GOLDEN CROWN	PI 529015	<i>G. hirsutum</i>
164	GP 1005	PI600961	<i>G. hirsutum</i>
165	GP 3755	PI607178	<i>G. hirsutum</i>

Appendix A continued.

<b>No.</b>	<b>Accession name</b>	<b>PI #</b>	<b>Species</b>
166	GP 3774	PI607177	<i>G. hirsutum</i>
167	GP 5479	PI600894	<i>G. hirsutum</i>
168	GREEN	PI601708	<i>G. hirsutum</i>
169	GREEN BROWN 7(NANKEEN)	PI 528784	<i>G. hirsutum</i>
170	GREEN LINT	SA-3142	<i>G. hirsutum</i>
171	GREEN LINT 4	PI 528787	<i>G. hirsutum</i>
172	GREGG	PI 529094	<i>G. hirsutum</i>
173	GREGG 35	PI 529189	<i>G. hirsutum</i>
174	GSA 74	PI600741	<i>G. hirsutum</i>
175	GSA 75	SA-3625	<i>G. hirsutum</i>
176	GSA71	PI529577	<i>G. hirsutum</i>
177	GSC 25	PI601109	<i>G. hirsutum</i>
178	GSC 27	PI601351	<i>G. hirsutum</i>
179	GSC 30	PI601484	<i>G. hirsutum</i>
180	GUMBO	PI529578	<i>G. hirsutum</i>
181	H1220	PI578226	<i>G. hirsutum</i>
182	H1330	PI583875	<i>G. hirsutum</i>
183	HALF AND HALF	PI 528964	<i>G. hirsutum</i>
184	HARTSVILLE	PI 528741	<i>G. hirsutum</i>
185	HARTSVILLE #5	PI 528632	<i>G. hirsutum</i>
186	HARTSVILLE (TUSCON)	PI 528633	<i>G. hirsutum</i>
187	HGPICG14QH-1-94	PI634321	<i>G. hirsutum</i>
188	HOPI ACALA	PI 529244	<i>G. hirsutum</i>
189	HOPI MOENCOPI	PI 528635	<i>G. hirsutum</i>
190	JL-1-S(MS)	PI 529180	<i>G. hirsutum</i>
191	KASCH	PI 528983	<i>G. hirsutum</i>
192	KEKCHI	PI 528634	<i>G. hirsutum</i>
193	KIME YELLOW	PI 528990	<i>G. hirsutum</i>
194	LA 213-FREGO	PI 529411	<i>G. hirsutum</i>
195	LA 304 (LA RN 910)	PI630663	<i>G. hirsutum</i>
196	LA 306 (LA RN 4-4)	PI630665	<i>G. hirsutum</i>
197	LA 322 (LA RN 910)	PI630666	<i>G. hirsutum</i>
198	LA 333 (LA RN 1032)	PI630667	<i>G. hirsutum</i>

Appendix A continued.

<b>No.</b>	<b>Accession name</b>	<b>PI #</b>	<b>Species</b>
199	LA 887	PI547084	<i>G. hirsutum</i>
200	La.850082FN	PI 572268	<i>G. hirsutum</i>
201	LAMBRIGHT 2020A	PI592517	<i>G. hirsutum</i>
202	LAMBRIGHT X-15-4	PI529584	<i>G. hirsutum</i>
203	LANKART	PI601147	<i>G. hirsutum</i>
204	LANKART 311	PI601392	<i>G. hirsutum</i>
205	LANKART 511	PI601302	<i>G. hirsutum</i>
206	LANKART 57	PI 528822	<i>G. hirsutum</i>
207	LANKART 611	SA-1006	<i>G. hirsutum</i>
208	LANKART LX571	PI606808	<i>G. hirsutum</i>
209	LBBCABCHUS-1-87	PI595760	<i>G. hirsutum</i>
210	LBCC4HUGS-1-89	PI603006	<i>G. hirsutum</i>
211	LBBCDBOAKH-1-90	PI603004	<i>G. hirsutum</i>
212	LIGHTNING EXPRESS	PI 528978	<i>G. hirsutum</i>
213	Limpopo	PI607199	<i>G. hirsutum</i>
214	LOCKET 4789	PI 529188	<i>G. hirsutum</i>
215	LOCKETT 88	SA-1017	<i>G. hirsutum</i>
216	LONE STAR	PI 528636	<i>G. hirsutum</i>
217	M.U.8B UA 7-44	PI 528560	<i>G. hirsutum</i>
218	M240 RNR	PI 592511	<i>G. hirsutum</i>
219	M4	PI 529123	<i>G. hirsutum</i>
220	MACHA 700 (J. GANNAWAY)	PI607220	<i>G. hirsutum</i>
221	MACHA WR 2 (J. GANNAWAY)	PI607219	<i>G. hirsutum</i>
222	MAR5PD208S-4-90	PI603009	<i>G. hirsutum</i>
223	MCNAIR 220	PI 529525	<i>G. hirsutum</i>
224	MCNAIR 235	PI 529526	<i>G. hirsutum</i>
225	MD51NE	PI 566941	<i>G. hirsutum</i>
226	MEBANE	PI 528985	<i>G. hirsutum</i>
227	MEXICAN BIG BOLL	PI 528692	<i>G. hirsutum</i>
228	MISCOT 7801	PI556978	<i>G. hirsutum</i>
229	MISCOT 8006	PI564681	<i>G. hirsutum</i>
230	MISSDEL	PI 528976	<i>G. hirsutum</i>
231	MO-DEL	PI 529257	<i>G. hirsutum</i>

Appendix A continued.

<b>No.</b>	<b>Accession name</b>	<b>PI #</b>	<b>Species</b>
232	MULTIPLE MARKER	PI 528950	<i>G. hirsutum</i>
233	NC 88-90	PI 583374	<i>G. hirsutum</i>
234	NC 88-91	PI 583375	<i>G. hirsutum</i>
235	NC 88-95	PI 583376	<i>G. hirsutum</i>
236	NC-4-M(3)	PI 529182	<i>G. hirsutum</i>
237	NEW BOYKIN	PI 528984	<i>G. hirsutum</i>
238	NORTHERN STAR	PI 528814	<i>G. hirsutum</i>
239	NORTHERN STAR 5	PI 529082	<i>G. hirsutum</i>
240	O'kelly chlorophyll defi	PI 528449	<i>G. hirsutum</i>
241	OR 25	PI607194	<i>G. hirsutum</i>
242	OR 26	PI607195	<i>G. hirsutum</i>
243	OR 27	PI607196	<i>G. hirsutum</i>
244	PANDORA	PI 528685	<i>G. hirsutum</i>
245	PAYMASTER 101	SA-1021	<i>G. hirsutum</i>
246	PAYMASTER 101A	PI 529206	<i>G. hirsutum</i>
247	PAYMASTER 111	PI 529259	<i>G. hirsutum</i>
248	PAYMASTER 111-A	SA-3741.01	<i>G. hirsutum</i>
249	PAYMASTER 145	PI529602	<i>G. hirsutum</i>
250	PAYMASTER 18	PI529600	<i>G. hirsutum</i>
251	PAYMASTER 266	PI529604	<i>G. hirsutum</i>
252	PAYMASTER 54	PI 528820	<i>G. hirsutum</i>
253	PAYMASTER HS200	PI542974	<i>G. hirsutum</i>
254	PAYMASTER HS26	PI606814	<i>G. hirsutum</i>
255	PD 0109	PI529611	<i>G. hirsutum</i>
256	PD 0111	PI529612	<i>G. hirsutum</i>
257	PD 0113	PI529613	<i>G. hirsutum</i>
258	PD 0259	PI529614	<i>G. hirsutum</i>
259	PD 0695	PI529615	<i>G. hirsutum</i>
260	PD 0875	PI529616	<i>G. hirsutum</i>
261	PD 1	PI606805	<i>G. hirsutum</i>
262	PD 2	PI606806	<i>G. hirsutum</i>
263	PD 2164	PI529617	<i>G. hirsutum</i>
264	PD 2165	PI529618	<i>G. hirsutum</i>

## Appendix A continued.

No.	Accession name	PI #	Species
265	PD 3246	PI529619	<i>G. hirsutum</i>
266	PD 3249	PI529620	<i>G. hirsutum</i>
267	PD 4381	PI 529621	<i>G. hirsutum</i>
268	PD 4461Q	PI529630	<i>G. hirsutum</i>
269	PD 6208 (FORMALLY PD-3)	PI511353	<i>G. hirsutum</i>
270	PD 778	PI 533642	<i>G. hirsutum</i>
271	PD 781	PI 533643	<i>G. hirsutum</i>
272	PD 785	PI 533644	<i>G. hirsutum</i>
273	PD 804	PI 533645	<i>G. hirsutum</i>
274	PD 8619	PI529625	<i>G. hirsutum</i>
275	PD 9232	PI529627	<i>G. hirsutum</i>
276	PD 93007	PI591418	<i>G. hirsutum</i>
277	PD 93009	PI591419	<i>G. hirsutum</i>
278	PD 93019	PI591420	<i>G. hirsutum</i>
279	PD 93021	PI591421	<i>G. hirsutum</i>
280	PD 93030	PI591422	<i>G. hirsutum</i>
281	PD 93034	PI591423	<i>G. hirsutum</i>
282	PD 93043	PI591424	<i>G. hirsutum</i>
283	PD 9363	PI529628	<i>G. hirsutum</i>
284	PD 9364	PI529629	<i>G. hirsutum</i>
285	PD24HQBPIH-1-94	PI634324	<i>G. hirsutum</i>
286	PD-3-14	PI591417	<i>G. hirsutum</i>
287	PIEDMONT CLEVELAND	SA-0366	<i>G. hirsutum</i>
288	PRONTO	PI529594	<i>G. hirsutum</i>
289	PSC 355	PI 612974	<i>G. hirsutum</i>
290	QUAPAW	PI607169	<i>G. hirsutum</i>
291	Red dwarf harrison	PI 528518	<i>G. hirsutum</i>
292	REX 713	PI529583	<i>G. hirsutum</i>
293	REX SL	PI 529226	<i>G. hirsutum</i>
294	RILCOT	PI 529096	<i>G. hirsutum</i>
295	ROGERS LG-10	PI600731	<i>G. hirsutum</i>
296	ROWDEN	SA-0300	<i>G. hirsutum</i>
297	ROWDEN 2088	PI 528689	<i>G. hirsutum</i>

Appendix A continued.

<b>No.</b>	<b>Accession name</b>	<b>PI #</b>	<b>Species</b>
298	ROWDEN 41B, TPSA	PI 528818	<i>G. hirsutum</i>
299	Rugose indore	PI 528510	<i>G. hirsutum</i>
300	SC-1	PI529598	<i>G. hirsutum</i>
301	SEALAND #1	PI 528871	<i>G. hirsutum</i>
302	SEALAND #2	PI 528872	<i>G. hirsutum</i>
303	SEALAND #7 WHITE FLOWER	PI 528874	<i>G. hirsutum</i>
304	SEALAND #7 YELLOW FLOWER	PI 528873	<i>G. hirsutum</i>
305	SEALAND 391	PI 528727	<i>G. hirsutum</i>
306	SEALAND 472	PI 528729	<i>G. hirsutum</i>
307	SEALAND 542	PI 528730	<i>G. hirsutum</i>
308	SEALAND 883	PI 528875	<i>G. hirsutum</i>
309	SG 747 (Sure-Grow 747)	PI 656375	<i>G. hirsutum</i>
310	SMALL LEAF	PI 529396	<i>G. hirsutum</i>
311	SOUTHLAND 400	PI540872	<i>G. hirsutum</i>
312	SOUTHLAND M1	PI601652	<i>G. hirsutum</i>
313	SPEARS GREEN	PI 528987	<i>G. hirsutum</i>
314	SPEARS UPLAND EARLY LONG	PI 529043	<i>G. hirsutum</i>
315	SPNXCHGLBH-1-94	PI634326	<i>G. hirsutum</i>
316	SPNXHQBPIS-1-94	PI634327	<i>G. hirsutum</i>
317	STARDEL	PI 529040	<i>G. hirsutum</i>
318	STATION MILLER	PI 528965	<i>G. hirsutum</i>
319	STONEVILLE 112	PI529631	<i>G. hirsutum</i>
320	STONEVILLE 2	PI 528751	<i>G. hirsutum</i>
321	STONEVILLE 20	PI 528671	<i>G. hirsutum</i>
322	STONEVILLE 213	PI 529229	<i>G. hirsutum</i>
323	STONEVILLE 2B	SA-0308	<i>G. hirsutum</i>
324	STONEVILLE 302	PI600952	<i>G. hirsutum</i>
325	STONEVILLE 453	PI601544	<i>G. hirsutum</i>
326	STONEVILLE 506	PI 529523	<i>G. hirsutum</i>
327	STONEVILLE 5A	PI 528658	<i>G. hirsutum</i>
328	STONEVILLE 5B	PI 528753	<i>G. hirsutum</i>
329	STONEVILLE 7	PI 529041	<i>G. hirsutum</i>
330	STONEVILLE 7A	PI 529228	<i>G. hirsutum</i>

## Appendix A continued.

<b>No.</b>	<b>Accession name</b>	<b>PI #</b>	<b>Species</b>
331	STONEVILLE 825	PI 529524	<i>G. hirsutum</i>
332	STRIPPER 31A	PI529593	<i>G. hirsutum</i>
333	STV 474	PI 578877	<i>G. hirsutum</i>
334	TAM 86E-8	PI 540306	<i>G. hirsutum</i>
335	TAM 86E-9	PI 540307	<i>G. hirsutum</i>
336	TAM 87N-5	PI 540299	<i>G. hirsutum</i>
337	TAM 87N-6	PI 540300	<i>G. hirsutum</i>
338	TAMCOT CAB-CS	PI564768	<i>G. hirsutum</i>
339	TAMCOT CAMD-E	PI529633	<i>G. hirsutum</i>
340	TAMCOT CD3H	PI513381	<i>G. hirsutum</i>
341	TAMCOT GCNH	PI564769	<i>G. hirsutum</i>
342	TAMCOT HQ95	PI 538033	<i>G. hirsutum</i>
343	TAMCOT LUXOR	PI610194	<i>G. hirsutum</i>
344	TAMCOT PYRAMID	PI617042	<i>G. hirsutum</i>
345	TAMCOT SP-21	PI529634	<i>G. hirsutum</i>
346	TAMCOT SP-21S	PI529635	<i>G. hirsutum</i>
347	TAMCOT SP-23	PI529636	<i>G. hirsutum</i>
348	TAMCOT SP-37	PI529637	<i>G. hirsutum</i>
349	TAMCOT SP-37H	PI529638	<i>G. hirsutum</i>
350	TAMCOT SPHINX	PI592801	<i>G. hirsutum</i>
351	TASHKENT 1	PI 529447	<i>G. hirsutum</i>
352	TEJAS	PI591047	<i>G. hirsutum</i>
353	TERRA C-30	PI601224	<i>G. hirsutum</i>
354	TERRA C-40	PI601225	<i>G. hirsutum</i>
355	TH 458(TRIPLE HYBRID)	PI 529434	<i>G. hirsutum</i>
356	TIDELAND T.P.S.A.NO.1	PI 529087	<i>G. hirsutum</i>
357	TIDELAND, TPSA-69	PI 529194	<i>G. hirsutum</i>
358	TIDEWATER (SEABROOKS) (G.B. INTO G.H.)	PI 528642	<i>G. hirsutum</i>
359	TIDEWATER 29 (G.B. X G.H.)	PI 528839	<i>G. hirsutum</i>
360	TIDEWATER E372-4 (G.B. X G.H.)	PI 528728	<i>G. hirsutum</i>
361	TM1	PI 607172	<i>G. hirsutum</i>
362	TOOLE	PI 528963	<i>G. hirsutum</i>
363	TRICE	PI 528955	<i>G. hirsutum</i>

Appendix A continued.

<b>No.</b>	<b>Accession name</b>	<b>PI #</b>	<b>Species</b>
364	VIR-5895 FIBRE VERT (GREEN FIBER)	SA-3235	<i>G. hirsutum</i>
365	Virescen yellow	PI 528447	<i>G. hirsutum</i>
366	VIRESCENT NANKEEN	PI 528936	<i>G. hirsutum</i>
367	VIRESCENT YELLOW UA 2-7	PI 528567	<i>G. hirsutum</i>
368	WANNAMAHER CLEVELAND	SA-0296	<i>G. hirsutum</i>
369	WESTBURN 70	PI 529503	<i>G. hirsutum</i>
370	WESTBURN M	PI 529504	<i>G. hirsutum</i>
371	WESTERN STORMPROOF	PI 529088	<i>G. hirsutum</i>
372	WIELDS CLEVELAND	PI 528960	<i>G. hirsutum</i>
373	WILD'S 15	PI 528780	<i>G. hirsutum</i>
374	WILD'S 18	PI 528781	<i>G. hirsutum</i>
375	WILDS 34-4(411), T82-2	PI 528842	<i>G. hirsutum</i>
376	WILDS 34-4(411), T85-2	PI 528843	<i>G. hirsutum</i>
377	WILDS 5	PI 528979	<i>G. hirsutum</i>
378	YELLOW GREEN,RUGOSE CH	PI 529072	<i>G. hirsutum</i>
379	AZAS 830	PI607290	<i>G. barbadense</i>
380	Peru 83-07-10-9	PI607336	<i>G. barbadense</i>
381	PUERTO RICAN REGULAR	PI 528898	<i>G. barbadense</i>

Appendix B. List of SSR primers used to genotype a panel of 378 cotton accessions

S.N	Primer name	Chromosome	Repeat motif
1	BNL1017	AD_chr08, AD_chr16	(CA)14
2	BNL1030	AD_chr05, AD_chr09, AD_chr23	(GT)16, (CA)13
3	BNL1034	AD_chr11, AD_chr17, AD_chr21	(CT)16
4	BNL1045	AD_chr12, AD_chr22, AD_chr26	(AG)16, (CA)10
5	BNL1047	AD_chr25	(CA)12
6	BNL1059	AD_chr14	(CA)16, (CA)11
7	BNL1061	AD_chr22, AD_chr25	(CA)12, (GT)11
8	BNL1064	AD_chr06, AD_chr26	(CA)15, (GT)13
9	BNL1066	AD_chr11	(GT)10+(GA)9
10	BNL1079	AD_chr18	(CA)11, (GT)11
11	BNL1122	AD_chr16	(AG)16
12	BNL1145	AD_chr02, AD_chr20	(GA)12
13	BNL1153	AD_chr06, AD_chr25	(GA)11+(GT)7
14	BNL1160	AD_chr10	(AG)10+G+(GA)3
15	BNL1161	AD_chr09, AD_chr23, AD_chr10	(AG)24
16	BNL1162	AD_chr09	(GA)14
17	BNL119	AD_chr20	(AG)10
18	BNL1231	AD_chr11, AD_chr21, AD_chr25	(AG)15
19	BNL1395	AD_chr07, AD_chr16	(AT)11+(AG)10
20	BNL1404	AD_chr11, AD_chr25	(AG)2+(TG)+(AG)3+T+(GA)11
21	BNL1408	AD_chr05, AD_chr11	(AG)17
22	BNL1414	AD_chr09, AD_chr23	(AG)16
23	BNL1417	AD_chr25	(AG)15
24	BNL1421	AD_Ch13, AD_Ch18	(AG)29, (AG)14
25	BNL1423	AD_chr09	(AG)12
26	BNL1434	AD_chr02	(AG)13
27	BNL1438	AD_chr13	(AG)13
28	BNL1440	AD_chr05, AD_chr06, AD_chr25	(AG)15
29	BNL1495	AD_Ch13, AD_Ch18	(AG)14
30	BNL1513	AD_chr24	(GA)17
31	BNL1551	AD_chr05, AD_chr16, AD_chr21	(AG)22
32	BNL1597	AD_chr07, AD_chr16	(GA)13

Appendix B continued.

<b>S.N</b>	<b>Primer name</b>	<b>Chromosome</b>	<b>Repeat motif</b>
33	BNL1604	AD_chr07, AD_chr16	(AG)25
34	BNL1605	AD_chr12	(AG)25
35	BNL1646	AD_chr08, AD_chr24	(AG)20
36	BNL1665	AD_chr09, AD_chr10, AD_chr20	(AG)16
37	BNL1666	AD_chr07, AD_chr15	(AG)14
38	BNL1667	AD_chr02, AD_chr14, AD_chr15	(AG)19
39	BNL1672	AD_chr09, AD_chr23	(AG)14
40	BNL1673	AD_chr12, AD_chr22	(AG)24
41	BNL1693	AD_chr01, AD_chr15, AD_chr21	(CT)12+(CA)9
42	BNL1694	AD_chr16	(AG)19, (TC)19
43	BNL1721	AD_chr18	(AG)17
44	BNL226	AD_chr03, AD_chr14	(GA)16
45	BNL2440	AD_chr01, AD_chr15	(AT)11+(AG)18
46	BNL2471	AD_chr06, AD_chr17, AD_chr18	(AG)12
47	BNL2495	AD_chr26	(AG)14, (TC)14
48	BNL2499	AD_chr23, AD_chr24	(GA)14, (CT)15
49	BNL252	AD_chr24	(CT)21
50	BNL2553	AD_chr20	(GA)10
51	BNL256	AD_chr09, AD_chr10	(GA)17
52	BNL2564	AD_chr01, AD_chr15	(AG)16
53	BNL2572	AD_chr04, AD_chr10	(GA)23
54	BNL2646	AD_chr07, AD_chr09, AD_chr15	(GA)3+G+A2+(AG)4+(GA)4, (TC)4+(CT)17
55	BNL2655	AD_chr24	(CT)14
56	BNL2667	AD_Ch13, AD_Ch18	(GA)21
57	BNL2812	AD_Ch11, AD_Ch21	(AAT)8
58	BNL285	AD_chr19	(GA)12, (GA)3+GC+(GA)12+A+(AG)2
59	BNL2882	AD_Ch03, AD_Ch14	(GA)12
60	BNL2895	AD_Ch11, AD_Ch21	(GA)10
61	BNL2960	AD_chr10	(GA)10
62	BNL2967	AD_chr12	(GA)17, (TC)17

Appendix B continued.

<b>S.N</b>	<b>Primer name</b>	<b>Chromosome</b>	<b>Repeat motif</b>
63	BNL3008	AD_chr16	(GA)13
64	BNL3029	AD_chr05, AD_chr19	(AG)12
65	BNL3031	AD_chr09, AD_chr23	(AG)27
66	BNL3034	AD_chr01, AD_chr14	(AG)12
67	BNL3084	AD_chr08, AD_chr24	(GA)12
68	BNL3379	AD_chr12, AD_chr20	(GA)14, (CT)14
69	BNL3418	AD_chr11, AD_chr21	(AC)13
70	BNL3441	AD_chr03	(AC)18, (AT)2(AC)18(AT)4
71	BNL3449	AD_Ch11, AD_Ch11	(CA)12, (CT)6TA(CA)12
72	BNL3452	AD_chr05, AD_chr10, AD_chr19	(CA)13
73	BNL3474	AD_chr08, AD_chr24	(CA)16
74	BNL3479	AD_Ch13, AD_Ch13	(AC)15, (TC)6T(AC)15G(CA)2
75	BNL3482	AD_chr20, AD_chr26	(AC)12
76	BNL3502	AD_chr14	(AC)12+(AT)2
77	BNL3510	AD_chr12, AD_chr26	(AC)15 + (TC)20
78	BNL3511	AD_chr23	(AC)11
79	BNL3537	AD_chr12, AD_chr26	(AC)11
80	BNL358	AD_chr22	(AG)12
81	BNL3590	AD_chr02, AD_chr17	(CA)20
82	BNL3594	AD_chr06, AD_chr25	(TC)37
83	BNL3599	AD_chr12, AD_chr26	(TC)15
84	BNL3627	AD_chr03, AD_chr08, AD_chr24	(TC)17
85	BNL3649	AD_chr11, AD_chr21	(TC)20
86	BNL3650	AD_chr06	(TC)15+(TA)6
87	BNL3778	AD_chr01	(GT)11
88	BNL3800	AD_chr08, AD_chr24	(GT)3+A+T+(TG)3+(TA)2+(TG)21
89	BNL3835	AD_chr04, AD_chr12	(TG)18
90	BNL3888	AD_Ch14, AD_Ch14	(TG)15
91	BNL3955	AD_chr05, AD_chr17, AD_chr22	(CA)12, (GT)13
92	BNL3976	AD_chr05, AD_chr21	(TC)17
93	BNL3987	AD_chr06	(CT)14
94	BNL3993	AD_Ch10, AD_Ch10, AD_Ch10	(TC)16, (TC)15

Appendix B continued.

<b>S.N</b>	<b>Primer name</b>	<b>Chromosome</b>	<b>Repeat motif</b>
95	BNL4007	AD_chr13	(TG)11
96	BNL4029	AD_Ch13, AD_Ch18	(TG)12, (AC)10
97	BNL4041	AD_chr12	(AT)5+(GT)14
98	BNL4049	AD_Ch04, AD_Ch09, AD_Ch22	(AC)11
99	BNL409	CH13_08	(GT)12
100	BNL569	AD_Ch13, AD_Ch18	(AG)20
101	BNL580	AD_chr16	(CT)7+(GT)+(CT)14
102	BNL686	AD_chr23, AD_chr09	(GA)22
103	BNL786	AD_chr15	(AG)14
104	BNL827	AD_chr25	(CA)19
105	BNL830	AD_chr15	(AC)10
106	BNL834	AD_chr17	(CA)13
107	BNL836	AD_Ch07, AD_Ch11	(TG)11
108	BNL852	AD_Ch05, AD_Ch19	(CA)13
109	BNL946	AD_Ch10, AD_Ch20	(GA)14
110	CIR009	AD_chr01, AD_chr15	(TG)6(N)1(TATG)6
111	CIR018	AD_chr01, AD_chr10	(TG)13
112	CIR030	AD_chr03, AD_chr14	(C)8(TC)6(CA)8
113	CIR097	AD_chr14	(GT)7+(GA)7
114	CIR110	AD_chr15	(AC)7+(N)9+(A)10
115	CIR119	AD_chr08, AD_chr24	(GT)8
116	CIR143	AD_chr15, AD_chr26	(AC)7(TA)5
117	CIR171	AD_chr09, AD_chr10, AD_chr20	(TG)16+(AG)6
118	CIR181	AD_chr03, AD_chr14	(TG)7
119	CIR187	AD_Ch10, AD_Ch20	(CA)8
120	CIR202	AD_chr03, AD_chr12	(AC)9
121	CIR218	AD_Ch04, AD_Ch22	(GT)9
122	CIR228	AD_chr03, AD_chr14	(TG)12(N)3(TGTA)11
123	CIR253	AD_Ch05	(TC)15(N)8(AC)5(N)7(CA)8
124	CIR272	AD_chr12, AD_chr26	(CA)8
125	CIR289	AD_chr24	(TG)7
126	CIR372	AD_chr10	(GT)12

Appendix B continued.

<b>S.N</b>	<b>Primer name</b>	<b>Chromosome</b>	<b>Repeat motif</b>
127	CIR376	AD_Ch02, AD_Ch05, AD_Ch08	(CA)15
128	CIR381	AD_Ch02, AD_Ch04, AD_Ch14	(AC)7
129	CIR47	AD_chr14	(CA)7
130	GH354	AD_chr19	AGA(17)
131	GH459	AD_chr19	TCT(13)
132	TMB1295	AD_chr19	(GA)22
133	TMB1489	AD_chr19	(GA)13
134	TMB1645	AD_chr19	(GA)36+(GA)12
135	TMB1750	AD_chr05	(GA)12+(TAA)6

Appendix C. Proportional Membership of cotton accessions to clusters as determined by model based analysis using STRUCTURE. Lines were assigned to a group based on membership probability higher than 0.70. The identified clusters roughly correspond to following geographical areas of cotton belt: Cluster 1-Western, cluster 2-Eastern, cluster 3-southwest, cluster 4-midsouth and cluster 5-G. barbadense

<b>Line ID</b>	<b>Cluster 1</b>	<b>Cluster 2</b>	<b>Cluster 3</b>	<b>Cluster 4</b>	<b>Cluster 5</b>	<b>Group</b>
AC 241	0.04	0.49	0.31	0.16	0.00	mixed
ACALA #111, ROGERS	0.66	0.29	0.05	0.00	0.00	mixed
ACALA 1064 (NEW MEXICO)	0.34	0.57	0.08	0.01	0.00	mixed
ACALA 1517 WILT	0.99	0.00	0.00	0.01	0.00	GP1
ACALA 1517-70	0.21	0.14	0.26	0.39	0.00	mixed
ACALA 1517-75	0.67	0.10	0.01	0.22	0.00	mixed
ACALA 1517D	0.85	0.15	0.00	0.00	0.00	GP1
ACALA 29	0.92	0.02	0.01	0.05	0.00	GP1
ACALA 3080	0.88	0.11	0.00	0.01	0.00	GP1
ACALA 4-42	0.53	0.07	0.01	0.40	0.00	mixed
ACALA 44WR	0.79	0.02	0.01	0.18	0.00	GP1
ACALA 5	0.23	0.65	0.07	0.05	0.00	mixed
ACALA 51	0.91	0.02	0.04	0.04	0.00	GP1
ACALA 5675	0.95	0.01	0.04	0.00	0.00	GP1
ACALA 8	0.65	0.03	0.26	0.06	0.00	mixed
ACALA GLANDLESS 8160	0.98	0.00	0.01	0.00	0.00	GP1
ACALA MAXXA	0.98	0.00	0.01	0.01	0.00	GP1
ACALA NAKED SEED	0.04	0.01	0.03	0.92	0.00	GP4
ACALA NUNN'S #5-37	0.66	0.24	0.03	0.08	0.00	mixed
ACALA ROYALE	0.98	0.01	0.00	0.01	0.00	GP1
ACALA SJ-1	0.52	0.09	0.39	0.01	0.00	mixed
ACALA SJ-2	0.99	0.00	0.01	0.00	0.00	GP1
ACALA SJ-3	0.99	0.00	0.00	0.00	0.00	GP1
ACALA SJ-4	0.99	0.00	0.00	0.01	0.00	GP1
ACALA SJC-1	0.99	0.01	0.00	0.01	0.00	GP1
ACALA ULTIMA	0.98	0.00	0.01	0.01	0.00	GP1

Appendix C continued.

<b>Line ID</b>	<b>Cluster 1</b>	<b>Cluster 2</b>	<b>Cluster 3</b>	<b>Cluster 4</b>	<b>Cluster 5</b>	<b>Group</b>
ACALA YOUNG'S	0.99	0.00	0.00	0.00	0.00	GP1
ACALA, MESSILLA VALLEY 898	0.86	0.04	0.04	0.07	0.00	GP1
ACALA, N.M. 8893	0.99	0.00	0.00	0.01	0.00	GP1
AK-DJURA FREGO (R2p, fg)	0.12	0.36	0.36	0.15	0.02	mixed
AK-DJURA GREEN LINT (R2p)	0.23	0.64	0.12	0.01	0.01	mixed
AK DJURA HIGG BROWN (R2p)	0.00	0.64	0.34	0.01	0.00	mixed
AK-DJURA VIRESCENT (v1, R2p)	0.01	0.49	0.47	0.01	0.03	mixed
ALLEN 33	0.00	0.78	0.01	0.21	0.00	GP2
ALL-IN-ONE	0.01	0.01	0.95	0.04	0.00	GP3
AMBASSADOR	0.50	0.12	0.03	0.35	0.00	mixed
ARIZONA RED (less R1)	0.37	0.60	0.02	0.01	0.01	mixed
ARK-1	0.01	0.18	0.24	0.57	0.00	mixed
ARK-2	0.00	0.51	0.24	0.24	0.00	mixed
ARKANSAS 10	0.35	0.12	0.34	0.19	0.00	mixed
ARKANSAS 12	0.01	0.21	0.13	0.65	0.00	mixed
ARKOT 8102	0.20	0.04	0.16	0.60	0.00	mixed
ARKOT 8606	0.00	0.01	0.12	0.86	0.00	GP4
ARKOT 8918	0.01	0.01	0.02	0.97	0.00	GP4
AUBURN 56	0.14	0.37	0.45	0.04	0.00	mixed
BJAG1 NECT	0.02	0.56	0.32	0.11	0.00	mixed
BLCABPD86S-1-90	0.01	0.00	0.99	0.00	0.00	GP3
BLIGHT MASTER	0.11	0.06	0.52	0.28	0.03	mixed
BLIGHTMASTER A5	0.50	0.06	0.43	0.00	0.00	mixed
BOB SHAW 1	0.15	0.01	0.77	0.07	0.00	GP3
BOBDEL	0.07	0.06	0.41	0.45	0.00	mixed
BRONCO 360	0.12	0.26	0.60	0.02	0.00	mixed
BRONCO 625	0.36	0.01	0.03	0.60	0.00	mixed
BRONZE (R1, v1)	0.39	0.14	0.04	0.43	0.00	mixed
BULGARIA 432	0.35	0.52	0.08	0.02	0.04	mixed
C5HUG2BES-2-87	0.00	0.00	0.99	0.00	0.00	GP3
C6-5	0.97	0.01	0.01	0.01	0.00	GP1
CA 17	0.01	0.74	0.01	0.25	0.00	GP2

Appendix C continued.

<b>Line ID</b>	<b>Cluster 1</b>	<b>Cluster 2</b>	<b>Cluster 3</b>	<b>Cluster 4</b>	<b>Cluster 5</b>	<b>Group</b>
CA 23	0.01	0.76	0.01	0.22	0.00	GP2
CA 30	0.01	0.80	0.01	0.17	0.00	GP2
CABCHUS-2-86	0.07	0.03	0.90	0.00	0.00	GP3
CABCSV506S-1-94	0.00	0.00	0.99	0.00	0.00	GP3
CABD3CABCH-1-89	0.01	0.00	0.98	0.00	0.00	GP3
CABD3SHP3S-1-90	0.00	0.01	0.81	0.18	0.00	GP3
CAHUGARPIH-1-88	0.02	0.01	0.94	0.03	0.00	GP3
CAHUGLBBCS-1-88	0.01	0.00	0.98	0.01	0.00	GP3
CARTER LONG STAPLE	0.02	0.36	0.62	0.01	0.00	mixed
CASCOT C-13	0.02	0.04	0.93	0.01	0.00	GP3
CASCOT L-7	0.01	0.01	0.97	0.02	0.00	GP3
CD3HCABCUH-1-89	0.01	0.00	0.99	0.00	0.00	GP3
CD3HCAHUGH-2-88	0.01	0.00	0.99	0.00	0.00	GP3
CD3HCHULBH-1-88	0.06	0.09	0.78	0.07	0.00	GP3
CD3HHARCIH-1-88	0.01	0.01	0.98	0.01	0.00	GP3
CLEVELAND 54	0.03	0.74	0.01	0.20	0.03	GP2
CLEVEWILT #6 NAKED SEED (1 YM)	0.02	0.94	0.02	0.02	0.00	GP2
COKER 100 WILT	0.01	0.04	0.61	0.35	0.00	mixed
COKER 100A (WR)	0.02	0.28	0.67	0.03	0.00	mixed
COKER 139	0.01	0.17	0.44	0.39	0.00	mixed
COKER 201	0.01	0.02	0.69	0.28	0.00	mixed
COKER 310	0.00	0.01	0.65	0.34	0.00	mixed
COKER 312	0.01	0.00	0.27	0.72	0.00	GP4
COKER 3131	0.01	0.05	0.44	0.50	0.00	mixed
COKER 315	0.01	0.24	0.62	0.14	0.00	mixed
COKER 5110	0.01	0.04	0.37	0.58	0.00	mixed
COKER'S CLEVELAND 884	0.00	0.99	0.01	0.00	0.00	GP2
COKER'S CLEVEWILT #3	0.11	0.04	0.84	0.01	0.00	GP3
COKER'S DELTATYPE WEBBER #7	0.01	0.95	0.03	0.02	0.00	GP2
COKER'S DELTATYPE WEBBER #9	0.01	0.93	0.04	0.02	0.00	GP2
COKER'S WILDS #2	0.36	0.30	0.30	0.04	0.00	mixed

## Appendix C continued.

<b>Line ID</b>	<b>Cluster 1</b>	<b>Cluster 2</b>	<b>Cluster 3</b>	<b>Cluster 4</b>	<b>Cluster 5</b>	<b>Group</b>
COKER'S WILDS #4	0.00	0.99	0.00	0.01	0.00	GP2
COKER'S WILDS #9	0.27	0.00	0.31	0.42	0.00	mixed
COLUMBIA	0.03	0.73	0.24	0.01	0.00	GP2
COOK 307-6	0.00	0.00	0.00	0.99	0.00	GP4
CS-8608	0.24	0.68	0.01	0.07	0.00	mixed
CS-8609	0.23	0.67	0.01	0.10	0.00	mixed
CS-8610	0.92	0.02	0.04	0.02	0.00	GP1
CS-8611	0.52	0.34	0.02	0.13	0.00	mixed
CUP LEAF	0.50	0.07	0.27	0.16	0.00	mixed
D2 SMOOTH MUTANT	0.00	0.01	0.07	0.92	0.00	GP4
DEL CERRO	0.94	0.01	0.01	0.04	0.00	GP1
DELCOT 277	0.01	0.09	0.80	0.10	0.00	GP3
DELCOT 344	0.01	0.02	0.60	0.37	0.00	mixed
DELFOSS 6102	0.17	0.24	0.49	0.09	0.00	mixed
DELFOSS 9169 (ORIGINAL)	0.30	0.10	0.54	0.06	0.00	mixed
DELTA QUEEN	0.01	0.07	0.54	0.39	0.00	mixed
DELTAPINE 14	0.00	0.00	0.00	0.99	0.00	GP4
DELTAPINE 15	0.00	0.00	0.00	0.99	0.00	GP4
DELTAPINE 16	0.00	0.01	0.01	0.97	0.00	GP4
DELTAPINE 20	0.00	0.00	0.00	0.99	0.00	GP4
DELTAPINE 41	0.00	0.00	0.01	0.99	0.00	GP4
DELTAPINE 45	0.02	0.03	0.01	0.94	0.00	GP4
DELTAPINE 50	0.00	0.00	0.00	0.99	0.00	GP4
DELTAPINE 51	0.00	0.02	0.00	0.97	0.00	GP4
DELTAPINE 61	0.01	0.00	0.00	0.99	0.00	GP4
DELTAPINE 90	0.14	0.25	0.01	0.61	0.00	mixed
DELTAPINE A	0.02	0.25	0.06	0.67	0.00	mixed
Deltapine Prema	0.71	0.01	0.01	0.28	0.00	GP1
DELTAPINE SMOOTH LEAF	0.01	0.03	0.57	0.40	0.00	mixed
DELTAPINE SR-383	0.55	0.01	0.02	0.43	0.00	mixed
DELTATYPE WEBBER	0.34	0.58	0.07	0.02	0.00	mixed
DELTATYPE WEBBER #4	0.01	0.97	0.01	0.02	0.00	GP2

Appendix C continued.

<b>Line ID</b>	<b>Cluster 1</b>	<b>Cluster 2</b>	<b>Cluster 3</b>	<b>Cluster 4</b>	<b>Cluster 5</b>	<b>Group</b>
DELTATYPE WEBBER(253-1)T142-8	0.00	0.99	0.00	0.00	0.00	GP2
DELTATYPE WEBBER 2139	0.02	0.57	0.35	0.07	0.00	mixed
DES 119	0.01	0.07	0.03	0.89	0.00	GP4
DES 24	0.14	0.01	0.23	0.63	0.00	mixed
DES 56	0.01	0.37	0.12	0.50	0.00	mixed
DIXIE 14-5-2	0.03	0.95	0.01	0.02	0.00	GP2
DIXIE KING	0.01	0.01	0.97	0.01	0.00	GP3
DIXIE TRIUMPH	0.27	0.44	0.02	0.27	0.00	mixed
DUNN 1047	0.02	0.27	0.69	0.01	0.00	mixed
DUNN 325	0.06	0.07	0.69	0.18	0.00	mixed
DUNN HS 120	0.03	0.13	0.54	0.31	0.00	mixed
DURANGO	0.65	0.31	0.03	0.02	0.00	mixed
EARLISTAPLE 7	0.14	0.81	0.04	0.01	0.00	GP2
EARLY FLOFF	0.05	0.09	0.85	0.01	0.00	GP3
EMPIRE (Gl2,Gl1,g 3,g 4,g 5)	0.39	0.01	0.31	0.29	0.00	mixed
EMPIRE GL2 GL2	0.30	0.00	0.68	0.02	0.00	mixed
EMPIRE WR	0.01	0.01	0.36	0.62	0.00	mixed
EWING LONG STAPLE X TIDEWATER	0.19	0.79	0.01	0.01	0.00	GP2
EXPRESS 121	0.20	0.07	0.38	0.35	0.00	mixed
EXPRESS 432	0.21	0.66	0.00	0.14	0.00	mixed
F-2-C-10 (Gl1,g 2,g 3,g 4,g 5)	0.66	0.07	0.05	0.22	0.00	mixed
FJA	0.15	0.72	0.12	0.01	0.00	GP2
FLORIDA GREEN SEED	0.00	0.99	0.00	0.00	0.00	GP2
FM966	0.43	0.15	0.28	0.14	0.00	mixed
FOX 4	0.00	0.08	0.02	0.90	0.00	GP4
FREGO	0.49	0.01	0.19	0.32	0.00	mixed
FREGO NANKEEN	0.31	0.09	0.32	0.29	0.00	mixed
FREGO UPLAND CR. DW.	0.06	0.92	0.01	0.01	0.00	GP2
MEADE (no fg)						
FREGO VIRESCENT	0.54	0.01	0.12	0.34	0.00	mixed

Appendix C continued.

<b>Line ID</b>	<b>Cluster 1</b>	<b>Cluster 2</b>	<b>Cluster 3</b>	<b>Cluster 4</b>	<b>Cluster 5</b>	<b>Group</b>
FTA	0.01	0.97	0.01	0.01	0.00	GP2
GA 161	0.26	0.01	0.02	0.71	0.00	GP4
GEORGIA KING	0.01	0.01	0.63	0.36	0.00	mixed
GERMAIN'S ACALA GC-352	0.84	0.14	0.01	0.02	0.00	GP1
GERMAINS ACALA GC-356	0.90	0.01	0.08	0.01	0.00	GP1
GLANDLESS NC- 1(gl1,gl2,gl3,gl4,gl5	0.66	0.09	0.06	0.18	0.00	mixed
GOLDEN CROWN ?	0.01	0.02	0.76	0.21	0.00	GP3
GP 1005	0.01	0.01	0.89	0.10	0.00	GP3
GP 3755	0.00	0.12	0.86	0.02	0.00	GP3
GP 3774	0.00	0.01	0.99	0.01	0.00	GP3
GP 5479	0.00	0.01	0.98	0.01	0.00	GP3
GREEN	0.04	0.72	0.22	0.01	0.00	GP2
GREEN BROWN #7 (NANKEEN)	0.46	0.22	0.32	0.01	0.00	mixed
Green Lint	0.36	0.61	0.02	0.01	0.00	mixed
GREEN LINT #4	0.56	0.42	0.02	0.00	0.00	mixed
GREGG	0.03	0.32	0.64	0.02	0.00	mixed
GREGG 35	0.01	0.29	0.69	0.02	0.00	mixed
GSA 74	0.08	0.30	0.61	0.01	0.00	mixed
GSA 75	0.02	0.00	0.78	0.20	0.00	GP3
GSA 71	0.03	0.12	0.71	0.14	0.00	GP3
GSC 25	0.25	0.25	0.05	0.45	0.00	mixed
GSC 27	0.02	0.01	0.73	0.25	0.00	GP3
GSC 30	0.00	0.00	0.99	0.00	0.00	GP3
GUMBO (Lo)(Okra)	0.00	0.00	0.00	0.99	0.00	GP4
H 1220	0.01	0.00	0.01	0.98	0.00	GP4
H 1330	0.59	0.01	0.01	0.39	0.00	mixed
HALF AND HALF	0.05	0.77	0.18	0.01	0.00	GP2
HARTSVILLE	0.83	0.02	0.13	0.01	0.00	GP1
HARTSVILLE #5	0.06	0.07	0.85	0.01	0.00	GP3
HARTSVILLE (TUCSON)	0.01	0.52	0.27	0.21	0.00	mixed
HGPICG14QH-1-94	0.01	0.00	0.98	0.00	0.00	GP3

## Appendix C continued

<b>Line ID</b>	<b>Cluster 1</b>	<b>Cluster 2</b>	<b>Cluster 3</b>	<b>Cluster 4</b>	<b>Cluster 5</b>	<b>Group</b>
HOPI ACALA	0.95	0.02	0.01	0.02	0.00	GP1
HOPI MOENCOPI	0.00	0.99	0.00	0.01	0.00	GP2
JL-I-S (MS)(gl1,gl2,gl3,gl4,gl5)	0.32	0.40	0.02	0.27	0.00	mixed
KASCH	0.68	0.01	0.30	0.02	0.00	mixed
KEKCHI	0.01	0.15	0.01	0.83	0.00	GP4
KIME YELLOW (KY)	0.22	0.01	0.59	0.19	0.00	mixed
LA 213-FREGO	0.01	0.12	0.01	0.85	0.00	GP4
LA 304 (LA RN 910)	0.00	0.02	0.00	0.97	0.00	GP4
LA 306 (LA RN 4-4)	0.01	0.01	0.34	0.63	0.00	mixed
LA 322 (LA RN 910)	0.00	0.02	0.00	0.98	0.00	GP4
LA 333 (LA RN 1032)	0.00	0.01	0.02	0.97	0.00	GP4
LA-887	0.02	0.08	0.02	0.88	0.00	GP4
LA. 850082FN	0.01	0.00	0.19	0.79	0.00	GP4
LAMBRIGHT 2020A	0.42	0.03	0.40	0.15	0.00	mixed
LAMBRIGHT X-15-4 (SI)	0.55	0.17	0.27	0.01	0.00	mixed
LANKART 175	0.00	0.00	0.00	0.99	0.00	GP4
LANKART 311	0.43	0.13	0.02	0.39	0.03	mixed
LANKART 511	0.16	0.01	0.83	0.01	0.00	GP3
LANKART 57	0.65	0.01	0.27	0.07	0.00	mixed
LANKART 611	0.58	0.01	0.05	0.36	0.00	mixed
LANKART LX571	0.42	0.18	0.02	0.38	0.00	mixed
LBBCABCHUS-1-87	0.07	0.01	0.60	0.32	0.00	mixed
LBCC4HUGS-1-89	0.00	0.00	0.99	0.00	0.00	GP3
LBBCDBOAKH-1-90	0.01	0.49	0.48	0.02	0.00	mixed
LIGHTNING EXPRESS	0.09	0.76	0.09	0.07	0.00	GP2
LOCKETT 4789	0.30	0.01	0.01	0.69	0.00	mixed
LOCKETT 88A	0.59	0.01	0.19	0.21	0.00	mixed
LONE STAR	0.99	0.00	0.00	0.01	0.00	GP1
M.U.8B UA 7-44	0.09	0.78	0.12	0.00	0.01	GP2
M-240 RNR	0.01	0.19	0.01	0.80	0.00	GP4
M4	0.01	0.72	0.01	0.24	0.03	GP2
MACHA 700 (J. GANNAWAY)	0.49	0.01	0.01	0.49	0.00	mixed

Appendix C continued.

<b>Line ID</b>	<b>Cluster 1</b>	<b>Cluster 2</b>	<b>Cluster 3</b>	<b>Cluster 4</b>	<b>Cluster 5</b>	<b>Group</b>
MACHA WR 2 (J. GANNAWAY)	0.01	0.30	0.39	0.30	0.00	mixed
MAR5PD208S-4-90	0.01	0.00	0.99	0.00	0.00	GP3
MCNAIR 220	0.04	0.49	0.34	0.13	0.00	mixed
MCNAIR 235	0.01	0.22	0.57	0.20	0.00	mixed
MD 51 ne	0.18	0.14	0.00	0.68	0.00	mixed
MEBANE	0.01	0.01	0.65	0.33	0.00	mixed
MEXICAN BIG BOLL	0.38	0.09	0.01	0.52	0.00	mixed
MISCOT 7801	0.01	0.01	0.77	0.22	0.00	GP3
MISCOT 8006	0.02	0.02	0.11	0.85	0.00	GP4
MISSDEL	0.14	0.16	0.50	0.21	0.00	mixed
MO-DEL	0.00	0.03	0.85	0.11	0.00	GP3
MULTIPLE MARKER	0.05	0.26	0.64	0.04	0.01	mixed
GERMAINS ACALA GC-410	0.99	0.00	0.01	0.00	0.00	GP1
NC 88-90	0.00	0.00	0.40	0.59	0.00	mixed
NC 88-91	0.01	0.16	0.43	0.41	0.00	mixed
NC 88-95	0.19	0.01	0.28	0.52	0.00	mixed
NC-4-M (3)(G11,g12,g13,G14,g15)	0.59	0.01	0.29	0.11	0.00	mixed
NEW BOYKIN	0.27	0.01	0.03	0.69	0.00	mixed
NORTHERN STAR	0.62	0.01	0.05	0.33	0.00	mixed
NORTHERN STAR 5	0.43	0.19	0.28	0.11	0.00	mixed
O'KELLY CHLOROPHYLL DEF.	0.11	0.11	0.41	0.37	0.00	mixed
PANDORA	0.01	0.54	0.31	0.14	0.00	mixed
PAYMASTER 101	0.59	0.02	0.39	0.00	0.00	mixed
PAYMASTER 101A	0.52	0.09	0.38	0.00	0.00	mixed
PAYMASTER 111	0.54	0.02	0.05	0.39	0.00	mixed
PAYMASTER 111-A	0.66	0.01	0.05	0.28	0.00	mixed
PAYMASTER 145	0.01	0.02	0.94	0.04	0.00	GP3
PAYMASTER 18	0.07	0.19	0.73	0.02	0.00	GP3
PAYMASTER 266	0.34	0.10	0.35	0.21	0.00	mixed
PAYMASTER 54	0.00	0.02	0.05	0.93	0.00	GP4
PAYMASTER HS200	0.53	0.01	0.19	0.28	0.00	mixed
PAYMASTER HS26	0.95	0.00	0.04	0.01	0.00	GP1

## Appendix C continued

<b>Line ID</b>	<b>Cluster 1</b>	<b>Cluster 2</b>	<b>Cluster 3</b>	<b>Cluster 4</b>	<b>Cluster 5</b>	<b>Group</b>
PD 0109	0.45	0.43	0.07	0.05	0.00	mixed
PD 0111	0.17	0.42	0.22	0.19	0.00	mixed
PD 0113	0.40	0.56	0.03	0.01	0.00	mixed
PD 0259	0.29	0.69	0.02	0.01	0.00	mixed
PD 0695	0.01	0.13	0.06	0.81	0.00	GP4
PD 0875	0.51	0.26	0.01	0.22	0.00	mixed
PD 1	0.01	0.68	0.29	0.02	0.00	mixed
PD 2	0.23	0.71	0.05	0.01	0.00	GP2
PD 2164	0.44	0.54	0.01	0.01	0.00	mixed
PD 2165	0.13	0.40	0.06	0.42	0.00	mixed
PD 3249	0.26	0.47	0.19	0.09	0.00	mixed
PD 4461Q	0.01	0.51	0.45	0.01	0.03	mixed
PD 6208 (FORMALY PD-3)	0.06	0.37	0.51	0.06	0.00	mixed
PD 778	0.03	0.36	0.08	0.54	0.00	mixed
PD 781	0.11	0.35	0.24	0.31	0.00	mixed
PD 785	0.04	0.42	0.45	0.08	0.01	mixed
PD 804	0.05	0.24	0.23	0.47	0.00	mixed
PD 8619	0.01	0.22	0.69	0.07	0.00	mixed
PD 9232	0.09	0.78	0.08	0.05	0.00	GP2
PD 93007	0.01	0.63	0.36	0.01	0.00	mixed
PD 93009	0.02	0.17	0.36	0.45	0.00	mixed
PD 93019	0.01	0.64	0.15	0.21	0.00	mixed
PD 93021	0.12	0.43	0.03	0.42	0.00	mixed
PD 93030	0.01	0.35	0.27	0.36	0.00	mixed
PD 93034	0.01	0.16	0.15	0.66	0.02	mixed
PD 93043	0.11	0.31	0.41	0.17	0.00	mixed
PD 9363	0.01	0.34	0.03	0.63	0.00	mixed
PD 9364	0.01	0.61	0.38	0.01	0.00	mixed
PD24HQBPIH-1-94	0.01	0.23	0.25	0.52	0.00	mixed
PD-3-14	0.04	0.66	0.21	0.09	0.00	mixed
PIEDMONT CLEVELAND 2-1	0.81	0.03	0.15	0.01	0.00	GP1
PRONTO (Ls)(Laciniate leaf)	0.01	0.00	0.01	0.99	0.00	GP4

## Appendix C continued

<b>Line ID</b>	<b>Cluster 1</b>	<b>Cluster 2</b>	<b>Cluster 3</b>	<b>Cluster 4</b>	<b>Cluster 5</b>	<b>Group</b>
PSC 355	0.14	0.01	0.01	0.84	0.00	GP4
QUAPAW	0.16	0.13	0.60	0.10	0.00	mixed
RED DWARF HARRISON (Rd)	0.88	0.02	0.09	0.01	0.00	GP1
REX 713	0.01	0.00	0.88	0.11	0.00	GP3
REX SL	0.00	0.00	0.99	0.01	0.00	GP3
RILCOT	0.01	0.24	0.72	0.03	0.00	GP3
ROGERS LG-10	0.25	0.13	0.40	0.22	0.00	mixed
ROWDEN #2	0.00	0.00	0.00	0.99	0.00	GP4
ROWDEN 2088	0.58	0.06	0.34	0.01	0.00	mixed
ROWDEN 41B, TPSA	0.01	0.03	0.00	0.95	0.01	GP4
RUGOSE INDORE (CB 1385)	0.13	0.76	0.02	0.02	0.07	GP2
OR 25	0.46	0.22	0.20	0.12	0.00	mixed
OR 26	0.62	0.22	0.02	0.15	0.00	mixed
OR 27	0.17	0.36	0.01	0.45	0.03	mixed
LIMPOPO	0.42	0.04	0.39	0.15	0.00	mixed
AUBURN 634 RNR	0.04	0.35	0.34	0.27	0.00	mixed
AUBURN Fg-310	0.00	0.01	0.76	0.23	0.00	GP3
AUBURN Ne Fg-16	0.02	0.27	0.17	0.54	0.00	mixed
AUBURN Ne Fg-149	0.01	0.06	0.01	0.92	0.00	GP4
SC-1	0.21	0.63	0.14	0.02	0.00	mixed
SEALAND #1	0.00	0.99	0.00	0.00	0.00	GP2
SEALAND #2	0.14	0.73	0.11	0.01	0.00	GP2
SEALAND #7 WHITE FLOWER	0.01	0.25	0.24	0.48	0.02	mixed
SEALAND #7 YELLOW FLOWER	0.29	0.68	0.01	0.02	0.00	mixed
SEALAND 391	0.19	0.79	0.01	0.01	0.00	GP2
SEALAND 472	0.01	0.96	0.00	0.00	0.02	GP2
SEALAND 542	0.02	0.01	0.13	0.84	0.00	GP4
SEALAND 883	0.00	0.99	0.00	0.01	0.00	GP2
SG 747	0.01	0.13	0.26	0.60	0.00	mixed
SMALL LEAF	0.00	0.29	0.28	0.43	0.00	mixed
SOUTHLAND 400	0.88	0.01	0.03	0.09	0.00	GP1
SOUTHLAND M1	0.61	0.02	0.02	0.35	0.01	mixed

Appendix C continued

<b>Line ID</b>	<b>Cluster 1</b>	<b>Cluster 2</b>	<b>Cluster 3</b>	<b>Cluster 4</b>	<b>Cluster 5</b>	<b>Group</b>
SPEARS GREEN	0.04	0.26	0.10	0.59	0.00	mixed
SPEARS UPLAND EARLY LONG	0.01	0.53	0.00	0.46	0.00	mixed
STAPLE						
SPNXCHGLBH-1-94	0.01	0.03	0.81	0.15	0.00	GP3
SPNXHQBPIS-1-94	0.01	0.36	0.59	0.04	0.00	mixed
STARDEL	0.21	0.01	0.45	0.33	0.00	mixed
STATION MILLER	0.26	0.35	0.15	0.24	0.00	mixed
STONEVILLE 112	0.01	0.01	0.03	0.95	0.00	GP4
STONEVILLE 2	0.26	0.00	0.01	0.73	0.00	GP4
STONEVILLE 20	0.32	0.48	0.18	0.02	0.00	mixed
STONEVILLE 213	0.47	0.00	0.00	0.52	0.00	mixed
STONEVILLE 2B (ORIGINAL)	0.19	0.01	0.78	0.02	0.00	GP3
STONEVILLE 302	0.01	0.01	0.01	0.97	0.00	GP4
STONEVILLE 453	0.01	0.00	0.01	0.98	0.00	GP4
STONEVILLE 506	0.01	0.01	0.03	0.95	0.00	GP4
STONEVILLE 5A	0.61	0.01	0.04	0.34	0.00	mixed
STONEVILLE 5B	0.60	0.38	0.01	0.01	0.00	mixed
STONEVILLE 7	0.18	0.02	0.15	0.65	0.00	mixed
STONEVILLE 7A	0.00	0.00	0.01	0.99	0.00	GP4
STONEVILLE 825	0.00	0.00	0.01	0.99	0.00	GP4
STRIPPER 31A	0.07	0.11	0.67	0.16	0.00	mixed
STONEVILLE 474	0.02	0.01	0.02	0.95	0.00	GP4
TAM 86E-8	0.02	0.00	0.72	0.26	0.00	GP3
TAM 86E-9	0.01	0.01	0.57	0.42	0.00	mixed
TAM 87N-5	0.31	0.01	0.26	0.42	0.01	mixed
TAM 87N-6	0.31	0.01	0.40	0.27	0.01	mixed
TAMCOT CAB-CS	0.02	0.01	0.97	0.00	0.00	GP3
TAMCOT CAMD-E	0.00	0.01	0.99	0.00	0.00	GP3
TAMCOT CD3H	0.00	0.00	0.99	0.01	0.00	GP3
TAMCOT GCNH	0.03	0.02	0.87	0.08	0.00	GP3
TAMCOT HQ95	0.01	0.01	0.48	0.51	0.00	mixed
TAMCOT LUXOR	0.01	0.05	0.93	0.01	0.00	GP3

Appendix C continued

<b>Line ID</b>	<b>Cluster 1</b>	<b>Cluster 2</b>	<b>Cluster 3</b>	<b>Cluster 4</b>	<b>Cluster 5</b>	<b>Group</b>
TAMCOT PYRAMID	0.01	0.01	0.93	0.06	0.00	GP3
TAMCOT SP-21	0.00	0.02	0.98	0.00	0.00	GP3
TAMCOT SP-21S	0.00	0.00	0.99	0.00	0.00	GP3
TAMCOT SP-23	0.04	0.01	0.94	0.01	0.00	GP3
TAMCOT SP-37	0.00	0.06	0.93	0.00	0.00	GP3
TAMCOT SP-37H	0.16	0.01	0.83	0.00	0.00	GP3
TAMCOT SPHINX	0.02	0.01	0.96	0.01	0.00	GP3
TASHKENT I	0.71	0.06	0.20	0.03	0.00	GP1
TEJAS	0.55	0.43	0.01	0.02	0.00	mixed
TERRA C-30	0.00	0.00	0.00	0.99	0.00	GP4
TERRA C-40	0.01	0.00	0.01	0.98	0.00	GP4
TH 458 (TRIPLE HYBRID)	0.61	0.25	0.04	0.09	0.00	mixed
TIDELAND T.P.S.A. NO. 1	0.05	0.02	0.37	0.56	0.00	mixed
TIDELAND, TPSA-69	0.01	0.02	0.51	0.45	0.00	mixed
TIDEWATER (SEABROOKS)	0.39	0.48	0.01	0.12	0.00	mixed
TIDEWATER 29	0.73	0.26	0.01	0.00	0.00	GP1
TIDEWATER E 372-4	0.63	0.20	0.04	0.13	0.01	mixed
TM 1	0.01	0.00	0.00	0.99	0.00	GP4
TOOLE	0.03	0.76	0.21	0.01	0.00	GP2
TRICE	0.00	0.00	0.00	0.99	0.00	GP4
VIR-5895 FIBRE VERT (GREEN FIBER)	0.04	0.39	0.55	0.02	0.00	mixed
VIRESCENT YELLOW (seg v1	0.40	0.55	0.01	0.04	0.00	mixed
VIRESCENT NANKEEN (v1)	0.18	0.27	0.21	0.33	0.00	mixed
VIRESCENT YELLOW UA 2-7 (v1)	0.73	0.01	0.11	0.16	0.00	GP1
CLEVELAND W.R. WANNAMAHER'S	0.01	0.81	0.02	0.16	0.00	GP2
WESTBURN 70	0.03	0.28	0.12	0.58	0.00	mixed
WESTBURN M	0.01	0.53	0.35	0.11	0.00	mixed
WESTERN STORMPROOF	0.28	0.53	0.15	0.04	0.00	mixed
CLEVELAND WILD'S	0.08	0.50	0.41	0.01	0.00	mixed
WILDS 15	0.00	0.95	0.04	0.01	0.00	GP2

Appendix C continued.

<b>Line ID</b>	<b>Cluster 1</b>	<b>Cluster 2</b>	<b>Cluster 3</b>	<b>Cluster 4</b>	<b>Cluster 5</b>	<b>Group</b>
WILDS 18	0.99	0.00	0.00	0.00	0.00	GP1
WILDS 34-4 (411), T82-2	0.00	0.99	0.00	0.00	0.00	GP2
WILDS 34-4 (411), T85-2	0.01	0.67	0.20	0.13	0.00	mixed
WILDS 5	0.02	0.67	0.30	0.02	0.00	mixed
YELLOW GREEN, RUGOSE CH DW	0.18	0.01	0.46	0.36	0.00	mixed
PUERTO RICAN REGULAR	0.00	0.00	0.00	0.00	1.00	GP5
GB 602	0.00	0.00	0.00	0.00	1.00	GP5
GB 758	0.00	0.00	0.00	0.00	1.00	GP5