The *Journal of Cotton Science 19:15–26 (2015)*
http://journal.cotton.org, © The Cotton Foundation 2015

BREEDING AND GENETICS

Construction of Genetic Linkage Map and QTL Analysis for Fiber Traits in Diploid Cotton (*Gossypium arboreum x Gossypium herbaceum*)

Ashok Badigannavar and Gerald O. Myers*

ABSTRACT

Diploid A genome cottons are grown for their natural fiber in drought prone areas of Asia. Although they possess inferior fiber qualities to cultivated Upland and Pima cotton, their value lies in their inherent resistance to pests and diseases. Molecular linkage maps provide essential tools for plant genetic research, facilitating quantitative trait locus (QTL) mapping, marker-assisted selection and map based cloning. In the present study, we used SSR, AFLP, and TRAP markers to construct a linkage map using 94 F_2 diploid individuals derived from a cross between *G. arboreum x G. herbaceum*. A total of 606 polymorphic markers were used to construct 37 linkage groups covering a total of 1109cM with an average distance of 7.92cM between two loci. Discriminant analysis identified three markers each associated for petal color and seed fuzziness, and four markers for petal spot. For quantitative traits, a total of 19 QTL were identified for five fiber traits using composite interval mapping. The identified QTL's such as qFL4-1, qFS4-2, qELO1-1 and qSI2-1 were found to be significantly linked to fiber length, strength, elongation and seed index, respectively. This work could lead to more such attempts in deploying markers onto the A genome, thus guiding QTL mapping in tetraploid cottons.

Cotton is a crop of global importance, mainly cultivated for natural fiber. The genus *Gossypium* consists of four cultivated species. Among the diploid species (2n=2X=26), *G. arboreum* and *G. herbaceum* are generally cultivated on marginal and drought prone regions of Asia. Phenotypically, they can be distinguished based on plant habit as well as leaf, bracteole and boll features (Fryxell, 1979). Long and narrow lobed leaves, bracteoles with fewer teeth, and round tapering bolls are the characteristics of *G. arboreum*, while constricted leaf lobes, wide bracteoles and round, and less pitted bolls are the common features of *G. herbaceum*. Within the A genome, *G. herbaceum* and *G. arboreum* diverged relatively recently. Cytologically these species can be distinguished by a reciprocal translocation (Gerstel, 1953), while the A_t (A subgenome in tetraploid) differs from the A, D and D_t (D subgenome in tetraploid) genomes by two reciprocal translocations. This suggests that *G. arboreum* arose as an incipient species with the origin through the fixation of the translocation (Endrizzi et al., 1985). Recently, draft sequences of the putative D genome parent *G. raimondii* (Wang et al., 2012) and A genome parent *G. arboreum* (Li et al., 2014) have been sequenced, providing new insights into the divergence among the polyploidy species.

Potentially valuable genetic variability has been observed for developmental traits, yield and fiber characters in *G. arboreum* (Singh and Singh, 1984) and *G. herbaceum* (Singh, 1983). Old world Asiatic diploid cottons were economically important during early global expansion of commercial cotton production. In the 1950's with the introduction of new world cotton, the area under diploid cotton cultivation was drastically reduced. Diploid cotton, however, is a model system for studying the genetics of fiber development compared to the more complicated system in tetraploid cottons. Therefore an understanding of the genetic inheritance and genomic regions controlling the fiber genes of diploid cotton species is critical (Li et al., 2014). In order to use the extant genetic diversity in the development of superior genotypes or transferring elite genes into cultivated tetraploids, molecular breeding techniques offer promising avenues compared to traditional breeding methods.

Molecular linkage maps provide essential tools for plant genetic research, facilitating quantitative trait locus (QTL) identification, marker-assisted selection, and map based cloning. To date, several genetic maps of cotton genomes have been constructed using diverse molecular markers and different mapping populations in tetraploid cottons (Reinisch et al., 1994;

A. Badigannavar, Nuclear Agriculture & Biotechnology Division, Bhabha Atomic Research Center, Mumbai-85, India and G.O. Myers*, School of Plant, Environmental, and Soil Science, LSU AgCenter, Baton Rouge, LA 70803
*Corresponding author: gmyers@agcenter.lsu.edu

Ulloa et al., 2002; Rong et al., 2004; Mei et al., 2004; Nguyen et al., 2004; Han et al., 2004 and Zhang et al., 2009). Comparatively few genetic maps have been developed in segregating populations involving diploid species. Interspecific linkage maps of diploid cottons have been constructed for the A genome (*G. herbaceum* × *G. arboreum*), the D genome (*G. trilobum* × *G. raimondii*) (Brubaker et al., 1999; Rong et al., 2004; Desai et al., 2006) and the G genome (*G. nelsonii* × *G. australe*) (Brubaker and Brown, 2003). An RFLP linkage map using an interspecific A genome diploid F_2 population mapped 275 loci (Desai et al.,2006). The 13 chromosomes of the A genome were represented by 12 large linkage groups reflecting an expected inter-chromosomal translocation between the parents. Though the diploid mapping parents are the closest living relatives of the allotetraploid A_t genome progenitor, two translocations and seven inversions were found between the A and A_t genomes (Desai et al., 2006).

Research on 'A' genome cotton has declined with the decrease in their cultivation during the last few decades. Although molecular tools have been available for a long time, only limited research has been carried out on the Asiatic cotton species (Brubaker et al., 1999). Understanding the molecular genetics of A genome cotton can be important for many reasons. They can foremost serve as a simple model system to study complex quantitative traits, yet only a limited number of genetic maps and QTL studies have been conducted. There is a significant opportunity for further mining the diploid genome with efficient marker systems to facilitate genetic mapping of fiber genes. In the present study, we used AFLP, SSR and TRAP markers to generate a framework genetic map of cultivated diploid cottons. The genetic map was used to identify QTL linked to fiber traits.

MATERIALS AND METHODS

Plant Material and Phenotypic Analysis. An interspecific F_2 population was developed from a cross between *G. arboreum* (acc. SMA-4, PI529740) *x G. herbaceum* (acc. A-97, PI529670), (provided by Dr. A.H. Paterson, University of Georgia, Athens, GA). The parents and 94 F_2 segregating plants were grown in the green house at the LSU Agricultural Center, Baton Rouge, LA. The phenotypic data on qualitative traits such as petal color (yellow or white), petal spot (absent or present) and seed hair (fuzzy or naked) was recorded for all the 94 F_2 individuals and parents. The parent SMA-4 possesses yellow flowers with petal spot and naked seeds while A-97 has white flowers without petal spot and fuzzy seeds. The quantitative traits namely, fiber length (inches), fiber strength (g/tex), short fiber index (SFI), fiber elongation (%), seed index (g), micronaire and uniformity ratio were measured on an individual plant basis. The fiber analysis was done using Uster 900 High Volume Instrumentation (HVI) system at the LSU Agricultural Center Cotton Laboratory. HVI measurements were repeated two times and mean values were used for analysis. Means of the phenotypic data from segregating individuals were used to test for normal distribution using PROC UNIVARIATE (SAS, 9.1.3, Cary, NC). Correlation analysis between pairs of traits was performed using PROC CORR in SAS. The correlation coefficients and a matrix plot were generated showing interrelationships among fiber traits.

Genotypic Analysis. The total genomic DNA from young leaves of the parents and F_2 plants was isolated using the cetyltrimethyl-ammonium bromide (CTAB) method as described previously (Zhang and Stewart, 2000). DNA concentration was measured using a NanoDrop-1000 spectrophotometer (NanoDrop, Wilmington, DE). Sixty-four primer combinations were used to generate AFLP data (Table 1) following the procedure given by Vos et al., (1995) with minor modifications. Sample DNA was digested with *EcoR*I (infrequent cutter with GAATTC recognition sequence) and *Mse*I (frequent cutter with TTAA recognition sequence) restriction enzymes and oligonucleotide adapters specific to restriction sites were ligated to the resulting fragments through incubation (180 min, 37°°C). Pre-amplifications were done using *EcoR* I+A and *Mse* I+C oligo primers and selective amplification was done using IR dye labeled EcoRI+ANN oligo primers (MWG Biotech, Germany). Touchdown PCR was used for selective amplification using the following profile: initial denaturing step at 94ºC for 2 min followed by initial 12 cycles at 94ºC for 30s, 65ºC for 30s (with 0.7ºC decrement every cycle) and 72ºC for 1 min, then followed by 23 cycles at 94ºC for 30s, 56ºC for 30s, and 72ºC for 1 min with a final extension step at 72ºC for 2 min. The PCR amplified products were run on a LI-COR 4300 sequencer (LI-COR Inc., Lincoln, NE). Gel images were saved and presence of a band was recorded as '1' and absence as '0', as per the typical dominant marker system. Ambiguous data that could not be resolved was discarded. The nomenclature of AFLP loci was followed according to Lacape et al.,(2003) and Myers et al., (2009), indicating the enzyme primer combinations with band size.

Table 1. Adapters and primers of AFLP marker system used for pre and selective amplification in diploid F_2 population.

Primer/adapter	Nomenclature[Z]	Sequence (5'-3')
ECORI primers:		
EcoRI linker 1	E-I	CTC GTA GAC TGC GTA CC
EcoRI linker 2	E-II	AAT TGG TAC GCA GTC TAC
EcoRI + A	E+A	GAC TGC GTA CCA ATT CA
E- AAC	E1	GACTGCGTACCAATTCAAC
E- AAG	E2	GACTGCGTACCAATTCAAG
E-ACA	E3	GACTGCGTACCAATTCACA
E-ACT	E4	GACTGCGTACCAATTCACT
E-ACC	E5	GACTGCGTACCAATTCACC
E-ACG	E6	GACTGCGTACCAATTCACG
E-AGG	E8	GACTGCGTACCAATTCAGG
E-AGA	E9	GACTGCGTACCAATTCAGA
MseI primers:		
MseI linker 1	M-I	GAC GAT GAG TCC TGA G
MseI linker 2	M-II	TAC TCA GGA CTC AT
MseI + C	M+C	GAT GAG TCC TGA GTA AC
M-CAA	M1	GATGAGTCCTGAGTAACAA
M-CAC	M2	GATGAGTCCTGAGTAACAC
M-CAG	M3	GATGAGTCCTGAGTAACAG
M-CAT	M4	GATGAGTCCTGAGTAACAT
M-CTA	M5	GATGAGTCCTGAGTAACTA
M-CTC	M6	GATGAGTCCTGAGTAACTC
M-CTG	M7	GATGAGTCCTGAGTAACTG
M-CTT	M8	GATGAGTCCTGAGTAACTT

[Z] Nomenclature is in accordance with the Lacape et al., 2003; Myers et al., 2009

In addition, 44 SSR/EST-SSR markers were used, which were potentially associated with fiber genes. The forward primer of these microsatellite markers was IR dye labeled (700 and 800) (MWG-Biotech, Germany). The PCR was carried out using an iCycler and the PCR products were separated using a LICOR 4300 sequencer. The gels were saved on a computer and scored manually as per co-dominant marker system. Four combinations of TRAP markers, a two primer PCR technique (Hu and Vick, 2003) were also tried utilizing sequence information on sucrose synthase (SuSy) and sucrose phosphate synthase (SuPS) genes (Genbank No. AF263384 and AB001338, respectively). The forward IR dye labeled TRAP primers were combined with arbitrary reverse primers. The PCR products were separated using a LICOR 4300 and the bands were scored similar to a dominant marker system. Allelic diversity at a given locus can be determined by Polymorphic Information Content (PIC) and it was calculated as 'PIC=1-$\sum f_i^{2}$' where, f_i is the frequency of the i^{th} allele (Weir, 1996). PROC ALLELE was used to calculate the PIC values and frequency estimate was done using PROC FREQ (SAS, 9.1.3, Cary, NC).

Linkage Map Construction. The segregation ratio for each marker was tested against the expected Mendelian ratio using Chi-square goodness of fit test. Only markers which are not significantly ($P \leq 0.05$) different from the expected 3:1 for dominant markers such as AFLP and TRAP and 1:2:1 for co-dominant markers such as SSR were utilized for map construction. Linkage map construction was performed using JOINMAP 3.0 (Stam and Oojien, 1995). The Kosambi map function was used (Kosambi, 1994) to convert recombination frequency to genetic map distance (cM). All the linkage groups were determined at logarithm of odds (LOD) scores ≥3.0 and a recombination frequency of 0.4 to provide evidence of linkage (Wu et al., 1992). The graphical representation of the LG was obtained using JOINMAP. Markers showing evidence of segregation distortion were marked specifically and used for mapping separately.

Data Analysis And QTL Mapping

Discriminant Analysis for Qualitative Traits. Qualitative traits such as petal color, petal spot and seed fuzziness were analyzed using Discriminant Analysis (DA). Discriminant Analysis is used to classify cases into the values of a categorical dependent variable, usually dichotomous (Fisher, 1936). Discriminant analysis has two steps: (1) an F test (Wilk's lambda) is used to test if the discriminant model as a whole is significant, and (2) if the F test shows significance, then the individual independent variables are assessed to see which differ significantly in mean (by group) and these are used to classify the dependent variable. The smaller the Wilk's lambda value for an independent variable, the more that variable contributes to the discriminant function (Mcharo et al., 2004).

The qualitative traits were divided into two groups based on yellow or white petal color, presence or absence of petal spot, or seed fuzziness (present or absent). To identify the marker data that best differentiates training samples within each subpopulation, the parametric discriminant analysis (PROC STEPDISC of SAS 9.1.3) forward method was used in the first step. The nonparametric method within the PROC DISCRIM procedure was then performed considering only the selected markers to construct and validate the class prediction function and to predict group membership. An error rate defined by 'percent correct classification' was calculated to measure the ability of the markers to correctly assign individual lines to 5, 10 and 15% of the training samples. The CROSSVALIDATION option provides a better assessment of classification accuracy. This classification is also done for each observation; however, the discriminant function used in each case is constructed by taking that observation out of the data set. With high value of percent correct classification, an association between marker (s) and phenotype is inferred.

QTL Analysis for Fiber Traits. A diploid mapping population consisting of 94 F_2 individuals was evaluated for fiber length, strength, seed index, uniformity ratio, elongation percent and short fiber index. Using the linkage map and phenotypic information, QTL analysis was performed through interval and composite interval mapping in QTL Cartographer 2.5 (Basten et al., 2007). Composite interval mapping (CIM) was carried out using the Zmapqtl component of Cartographer (Zeng and Weir, 1996). The analysis was performed with a maximum of five background markers based on the forward-backward regression method of selection. Zmapqtl provides estimates for the square of the partial correlation coefficient (R^2) as well as the additive and the dominant effect. A LOD threshold of ≥ 2.5 (1000 permutations) was used to declare significant QTLs in the present investigation. A Chi-square test was performed to determine whether the allele frequency at each individual locus had normal segregation.

RESULTS

Phenotypic Trait Analysis. Phenotypic data for fiber traits of the parents and the F_2 individuals are summarized in Table 2. The two parents differed significantly for most of the fiber traits except fiber strength. Based on the wide range values and high variance estimates, it is evident that moderate to high phenotypic diversity was present in the population. Wide range values were observed for fiber traits, such as FL (1.78-2.59 cm; 0.7-1.02 inch), UNI (71.7-81.7%), SFI (9.9-29.8%), FS (155.66-302.51 kN m kg^{-1}; 15.9-30.9 g/tex) and SI (5-11.9 g) in the F_2 generation. Based on the degree of diversity present, it was concluded that the F_2 population possessed sufficient variation for QTL analysis. The correlation coefficients among the quantitative traits revealed that there was a significant positive relationship among the fiber traits (Table 3). The traits such as FL, UNI, FS and SFI were positively correlated with values ranging from 0.55-0.95. Significant negative correlation existed between ELO with FL, UNI and FS. Seed index (SI) was positively correlated with FL, UNI and FS but negatively associated with SFI and ELO. The binaries of phenotypic diversity for flower color, boll size and shape and seed fuzziness are illustrated in Fig 1. These qualitative traits fit the expected single gene inheritance. Petal color showed yellow and white types with or without petal spot. Boll shape also varied from round, pitted to oblong with pointed beak. Seed fuzziness showed extreme phenotypes, fuzzy and naked seeded.

Table 2. Univariate analyses of fiber traits and seed index in parental and diploid F_2 population.

Parameters	FL (cm; inch)[Z]	UNI	SFI	FS (kN m kg^{-1}; g/tex)	ELO	SI (g)
Range	1.78-2.59; 0.70-1.02	71.70-81.70	9.90-29.80	155.66-302.51; 15.90-30.90	4.20-5.70	5.01-11.90
Mean	2.18; 0.86	77.42	16.40	222.33; 22.71	4.89	7.79
Std. error	0.02	0.53	1.12	0.87	0.07	0.35
Variance	0.01	6.45	29.07	17.58	0.13	2.82
Std. deviation	0.08	2.54	5.39	4.19	0.35	1.68
Skewness	-0.36	-0.74	1.09	0.35	0.58	0.50
Kurtosis	-0.48	-0.44	0.08	-0.99	-0.13	-0.27
Parents						
PI529740	2.54; 1.00	80.6	12.1	259.44; 26.50	5.90	5.90
PI529670	2.18; 0.86	77.1	15.5	261.39; 26.70	8.45	8.45

[Z] FL= Fiber length, UNI= Uniformity index, SFI= Short fiber index, FS= Fiber strength, ELO= Elongation percentage, SI= Seed index, SD=Standard deviation, Var= Variance and SE= Standard error

Table 3. Pearson correlation coefficients among fiber traits of diploid F_2 population.

Trait	FL[Z]	UNI	SFI	FS	ELO	SI
FL	1					
UNI	0.94**	1				
SFI	0.86**	0.95**	1			
FS	0.55*	0.69**	-0.69**	1		
ELO	-0.52*	-0.52*	0.48*	-0.29	1	
SI	0.262	0.30	-0.51*	0.51*	-0.45*	1

*, ** Significant at $P \leq 0.05, 0.01$ respectively.

[Z] FL= Fiber length, UNI= Uniformity index, SFI= Short fiber index, FS= Fiber strength, ELO= Elongation percentage, SI= Seed index

Fig. 1. The phenotypic diversity present in the segregating F_2 population of A genome cottons, with respect to flower color and petal spot (center left: SMA-4 and center right: A-97 and F_2 are around), boll size and shape and seed fuzziness(extreme left: A-97 and extreme right: SMA-4 and center: F_2 segregants (clockwise from upper left).

Molecular Analysis. Sixty-four AFLP primer combinations were screened by selective amplification across the diploid mapping population. A total of 539 polymorphic bands were generated. In addition, SSR and TRAP markers generated 50 and 17 polymorphic loci, respectively. Among the different *Eco*RI primers tried, E2 (AAG) and E6 (ACG) generated the highest number of polymorphic bands across all the *Mse*I primers. The frequency of shared alleles among the F_2 population is presented in Fig 2. As expected, the F_2 segregants showed normal distribution with most of the individuals similar by 60-80%. Although the parents differed with respect to several distinguishable characters, the amount of genetic variability was moderate among the segregants. The polymorphic information content (PIC) is commonly used in genetics as a measure of polymorphism and to estimate the informativeness of a marker locus used in linkage analysis. In the present study, PIC values varied from 0.087 to 0.37 with an average of 0.253. The AFLP, TRAP and SSR markers produced moderate variability, mirroring the narrow genetic base of the characters in the selected parents.

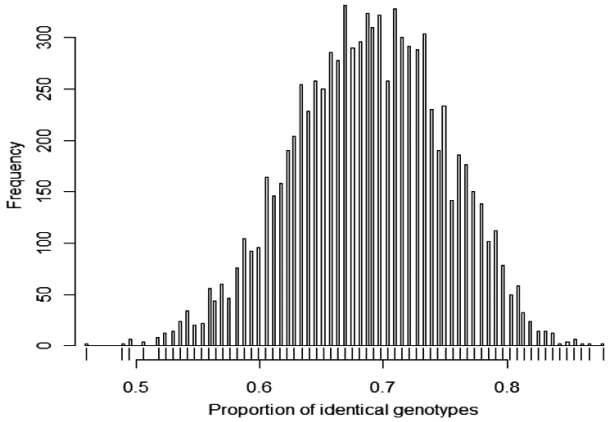

Fig: 2. Frequency of shared alleles among the diploid segregating population. X axis: proportion of shared alleles; Y axis: frequency values

Construction of Genetic Map. Using the three marker systems, total of 606 polymorphic loci were amplified from 94 F_2 individuals. Significant departures from the expected 3:1 (for AFLP and TRAP) and 1:2:1 (for SSR) segregation ratios were detected for 146 loci at $P \leq 0.05$, accounting for 24% of the polymorphic loci detected. A genetic linkage map was constructed with 37 linkage groups of AFLP and SSR markers ranging from to 11 to 98 cM in length per linkage group, LG (Fig. 3). The linkage groups were numbered from LG1 to LG37 in descending order of length. The map covered a total of 1109 cM with an average distance of 7.92 cM between loci. Nine linkage groups were considered as major ones (hosting more than 4 markers/LG) and the remaining were minor groups. The number of markers ranged from 2 to 21 per linkage group.

QTL Analysis

Qualitative traits. In the present study, a survey of floral and seed morphology was done in the segregating F_2 population of a cross between the A_1 (SMA-4) and A_2 (A-97) genomes. A total of 606 markers including AFLP, SSR and TRAP were used to discriminate population for two floral characters and seed fuzziness. The number of markers selected by the STEPDISC procedure applied after DA and the percent correct classification of F_2 individuals based on the selected markers is presented in Table 4. DA identified three markers each for petal color and seed fuzziness and four markers for petal spot. The percent correct classification (obtained by cross-validation) was 100% with no error rate estimate. For petal color, DA identified AFLP markers E5M2_60, E5M6_205 and E9M1_560, which were able to discriminate the F_2 individuals with 100% correct classification in each training sample (5, 10 and 15%). Similar markers showed significant correlations with qualitative traits across different training samples selected for the study.

The 'Wilk's lambda' P values were significant for the model as a whole and E9M1_560 was found to contribute more variation to the discriminant function. Similarly, DA identified E2M3_342, E8M8_510 and E9M3_440 as suitable marker for discriminating the population for seed fuzziness. The marker E2M3_342 is located on LG 27 and the marker E9M3_440 on LG 21 where it is also associated with SFI. For the petal spot, DA identified SUSTRAP1_175 (LG 18), E6M1_410, E9M2_520 and E2M5_520 as significant markers.

Quantitative Traits. The locations, LG, LOD scores and additive and dominant effect of major QTL for all fiber traits of interest, and seed index are given in Table 5. A total of 19 QTL were identified and linked with five fiber traits or seed index by composite interval mapping. Ninteen QTL were identified, of these LG4 and LG1 hosted markers linked with more than three traits.

A total of four QTL were detected for fiber length, which were located on linkage groups, 4, 17, 22 and 24 (Table 5). The qFL4-1 and qFL17-2 had high LOD values and explained 11.58 and 7.55% of the phenotypic variation, respectively. Three QTL were detected for uniformity ratio in this population located on LG 1, 22 and 25 with R^2 values ranging from 5.6-9.6%. The major QTL (LOD=9.4) was found in the interval of E1M8_90 - CIR-199. For SFI, five QTL on LG 1, 2, 4, 18 and 21 were identified which had R^2 values ranging from 1.1-2.4. The major QTL (qSFI 2-2) was in the interval E9M1_505-E1M3_400, with an LOD value of 8.8 and dominant effect (17.2). Two QTL affecting fiber strength were identified, one, qFS4-2 on LG 4 and the other one, qFS1-1 on LG 1, had LOD value of 2.9 and 2.5 respectively. A major QTL for uniformity index also mapped in the same region as one for fiber strength. A QTL located on E4M2_145 (qELO1-1) recorded a 2.9 LOD and explained 9.6% of phenotypic variation for elongation. Although there were two QTL for seed index, only qSI2-1 located on LG2, possessed 4.1 LOD score, explaining 10.09% of the phenotypic variation. Significant QTLs identified in this study showed additive and dominant effect across various fiber traits.

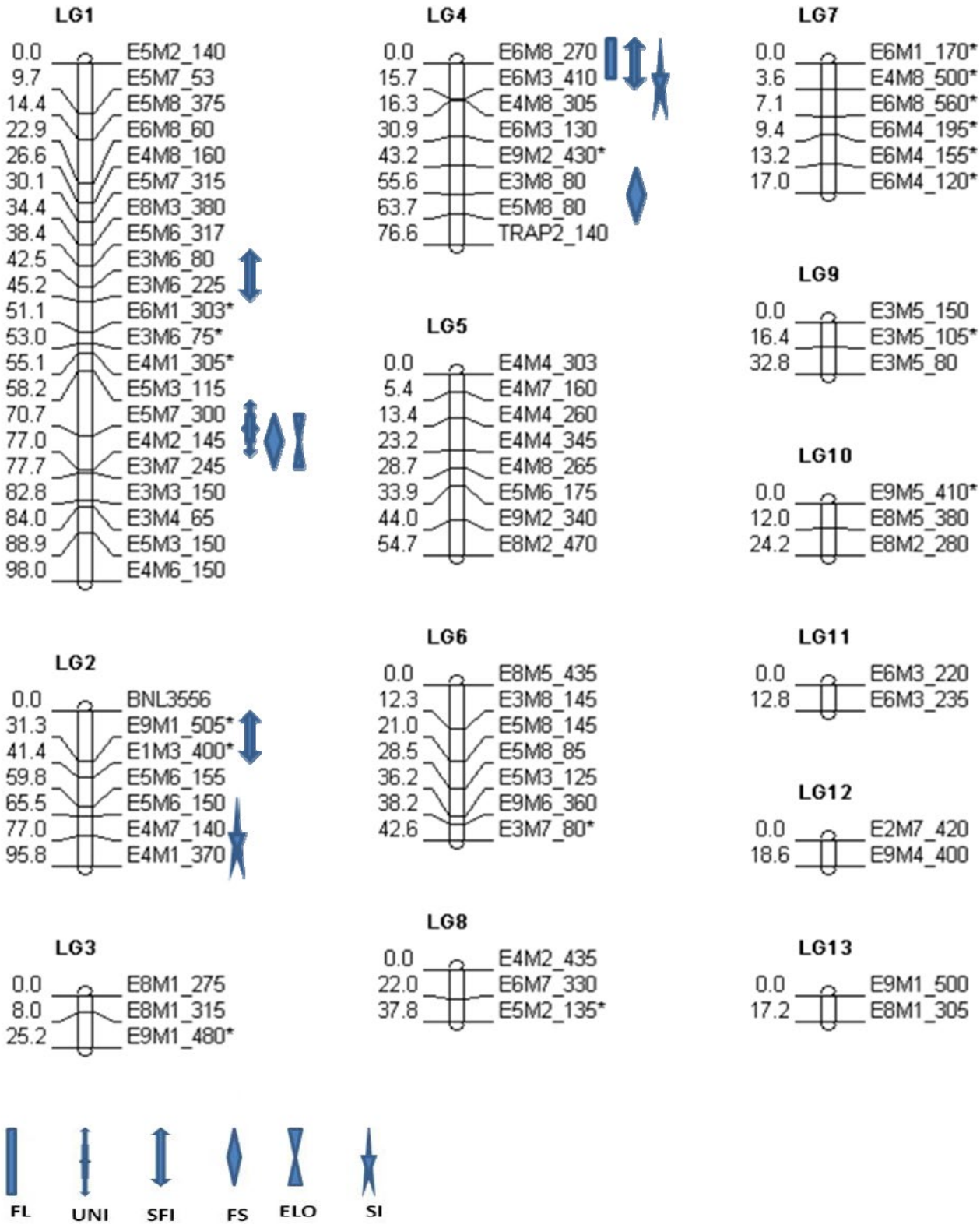

Fig. 3. A genetic linkage map of the A genome diploid cotton based on the AFLP, SSR and TRAP markers. The map contains 37 linkage groups covering 1109cM with an average of 7.92 cM between loci. A total of 146 markers were identified as distorted ones, departing from the Mendelian segregation. They were represented with asterisk (*). QTL for fiber traits and seed index are represented as boxes to the right side of each LG.

Fig. 3. Continued

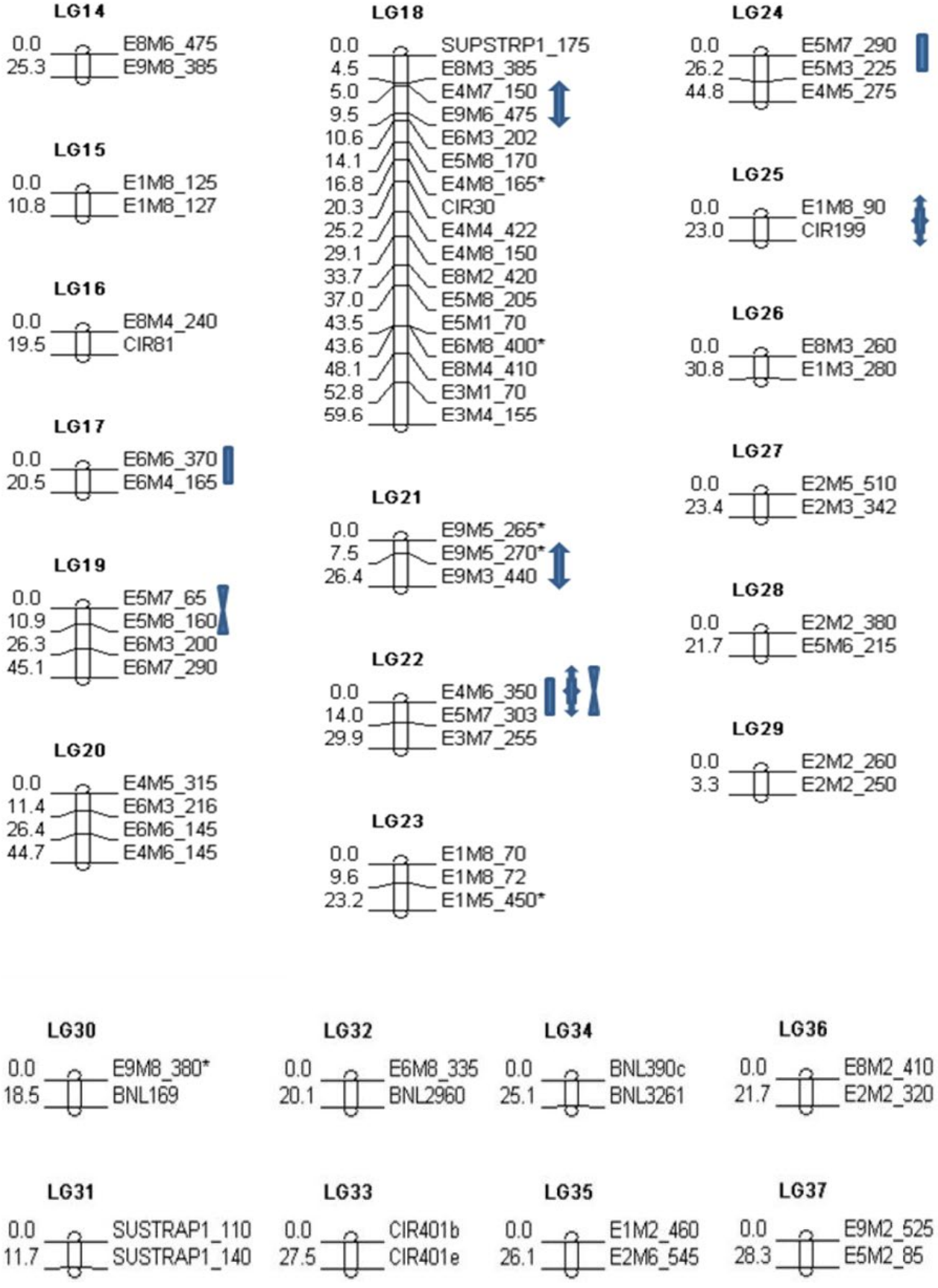

Table 4. Discriminant analyses selected markers for petal color, seed fuzziness and petal spot in a diploid F_2 population of cotton

Markers entered	Model R^2	Pr > F	Wilk's λ	Pr < λ^Z
Petal Color				
E5M2_60	0.39	0.0001	0.60	0.0001
E5M6_205	0.75	<0.0001	0.24	<0.0001
E9M1_560	1.00	<0.0001	0.00	<0.0001
Seed fuzziness				
E2M3_342	0.58	<0.0001	0.41	<0.0001
E8M8_510	0.84	<0.0001	0.15	<0.0001
E9M3_440	1.00	<0.0001	0.00	<0.0001
Petal Spot				
SUPSTRP1_175	0.32	0.0008	0.67	0.0008
E6M1_410	0.77	<0.0001	0.22	<0.0001
E9M2_520	0.87	<0.0001	0.12	<0.0001
E2M5_520	1.00	<0.0001	0.00	<0.0001

[Z] λ = Wilk's lambda used to test the significance of the discriminant model; Partial R^2 and Model R^2 were calculated from multiple regression (PROC REG, SAS Institute, ver. 9.1.2); % correct classification were calculated by leave one out validation within the training samples.

Table 5. Composite Interval Mapping for fiber traits using F_2 diploid cotton population from a cross between *G. arboreum* and *G. herbaceum*. QTLs are listed traitwise along with their position on LG, LOD, additive and dominant effects.

QTL[Z]	LG	Position	LOD	Marker Interval	Additive[Y]	Dominant	R^2
FL							
qFL4-1	4	4.0	5.1	E6M8_270-E6M3_410	-0.05	0.42	11.58
qFL17-2	17	14.0	5.1	E6M6_370-E6M4_165	-0.02	0.47	7.55
qFL22-3	22	2.0	2.7	E4M6_350-E5M7_303	0.14	-0.14	9.65
qFL24-4	24	20.0	4.3	E5M7_290-E5M3_225	0.03	0.78	4.80
UNI							
qUNI1-1	1	76.8	2.9	E5M7_300-E4M2_145	11.98	-14.4	9.56
qUNI22-2	22	2.0	2.5	E4M6_350-E5M7_303	12.40	-14.3	8.44
qUNI25-3	25	14.0	9.4	E1M8_90-CIR199	-1.31	25.6	5.60
SFI							
qSFI1-1	1	44.5	3.4	E3M6_80-E3M6_225	-0.36	18.44	1.41
qSFI2-2	2	33.1	8.8	E9M1_505-E1M3_400	0.37	17.17	1.90
qSFI4-3	4	12.0	7.5	E6M8_270-E6M3_410	-0.11	16.97	1.10
qSFI18-4	18	9.1	3.8	E8M3_385-E9M6_475	-1.50	19.74	2.36
qSFI21-5	21	23.5	3.4	E9M5_270-E9M3_440	0.30	-0.51	1.70
FS							
qFS1-1	1	77.0	2.5	E4M2_145	3.56	-5.31	9.50
qFS4-2	4	60.0	2.9	E5M3_115-E5M7_300	-1.35	-8.51	1.07
ELO							
qELO1-1	1	77.0	2.9	E4M2_145	0.76	-0.90	9.60
qELO19-2	19	10.0	2.5	E5M7_65-E5M8_160	0.86	-1.37	9.10
qELO22-3	22	2.0	2.5	E4M6_350-E5M7-303	0.78	-0.89	8.51
SI							
qSI2-1	2	91.0	4.1	E4M7_140-E4M1_370	1.61	0.45	10.09
qSI4-2	4	15.7	2.9	E6M8_270-E6M3_410	-1.32	7.60	6.80

[Z] Nomenclature for the QTL was followed as per Shen et al., 2005.

[Y] Positive or negative additive effect leads to increased/decreased trait value with reference to SMA-4

DISCUSSION

Qualitative traits such as petal color, petal spot and seed fuzziness showed single gene mendelian inheritance. The *G. arboreum* petals were white without petal spot, while *G. herbaceum* had yellow flower with petal spot. These traits fit a monogenic inheritance model with the presence of petal spot as dominant over its absence and yellow petal color as dominant over white (Desai et al., 2006). The A genome cotton also exhibited correlation between petal size and petal color with white color petals mostly associated with small flowers (Hutchinson, 1931).

Phenotypic diversity for fiber traits showed a wide range of values except for the ELO and UNI. The traits FL (1.78-2.59 cm; 0.7-1.02 inch), SFI (9.9-29.8), FS (155.66-302.51 kN m kg^{-1}; 15.90-30.90 g/tex) and SI (4.90-11.90g) showed a wide range of values in the segregants. Most of the traits such as FL, UNI, SFI, FS and ELO were significantly correlated. Our results showed that approximately 75% of the F_2 individual plants were identical by 60-70% at the molecular level for loci measured by markers. According to Kebede et al., (2007), microsatellite analysis revealed a low to moderate interspecific and intraspecific genetic diversity in *G. herbaceum* and *G. arboreum* accessions. The three marker systems used in this study showed marker polymorphism up to 0.37 with an average of 0.25. The narrow genetic base of the parents as reflected in the segregants could be the reason for low levels of polymorphism.

Among the 606 AFLP, SSR and TRAP markers used in this study, 460 markers were used to construct a diploid genetic map. Excluding 24% of the distorted markers, the map was assembled into 37 linkage groups. The map covers 1109 cM with each loci at an average of 7.92 cM. A map from a similar cross, based on 274 RFLP loci covered a map length of 1147 cM with an average distance of 4.2 cM between adjacent markers (Desai et al., 2006). Obviously more markers would help in saturating this map. Segregation distortion, the deviation of segregation ratios from expected Mendelian ratios has been reported in a wide range of plant species (Jenczewski et al., 1997). As many as 140 markers exhibited segregation distortion and accounted for 24% of the polymorphic markers scored. Segregation distortion may be due to the presence of lethal genes and/or overlapping fragments consisting of identically sized fragments (Hansen et al., 1999). It could also be related to different size of the parent genomes or to distorting factors, such as self-incompatible alleles (Bert et al., 1999). Population size also influences the segregation distortion when the two markers are separated by more than 10 cM (Hackett and Broadfoot, 2003).

Based on the discriminant analysis (DA), our study revealed that markers E5M2_60 and E5M6_205 were associated with petal color, while SUSTRAP1_175 and E6M1_410 were associated with petal spot. E2M3_342 was able to discriminate seed fuzziness and was found to be located on LG 27. Seed fuzziness or naked seed was categorically discriminated by this marker and its parallel association with SFI indicated its role in suppressing seed fiber growth. In another *G. arboreum x G. herbaceum* segregating population, Rong et al., 2005 mapped the naked seed phenotype, *sma-4(fz)*, near the terminus of LG A.

Based on composite interval mapping, 19 QTL were identified on 10 linkage groups for the five fiber quality, and one seed related trait. The phenomenon of QTL clustering has been reported earlier in cotton (Shappley et al., 1998; Ulloa and Meredith, 2000; Qin et al., 2008). A total of four intervals were found to be involved in the control of more than two traits and located on LG 1, 2, 4 and 22. Not only were these fiber traits highly correlated, but also influenced by tight linkage (Qin et al., 2008). For the significant QTL identified, alleles associated with an increase in the trait value came from both parents. QTL having high LOD values, affecting UNI and FS were located on LG1 and had significant additive effects, explaining 9.56% and 9.50% of the phenotypic variation, respectively. Similarly, Zhang et al., (2005) detected additive effects for UNI and FL on chromosome 5 of the upland cotton map, explaining 25% of trait variation. The marker E4M2_145 was found to be significantly associated with UNI, FS and ELO with high LOD values explaining up to 12% of the phenotypic variation. The significance of this marker was confirmed using composite interval mapping. Although the QTL detected in this study have moderate genetic effects and their number is limited, the findings will help in validation and comparison to the tetraploid map (Badigannavar, 2010).

The present study revealed a moderate level of genetic diversity for the F_2 segregants. The study explored diploid cotton genome as a model system to map floral, fiber quality and seed related traits. DA was effective in identifying potential markers, which can differentiate among the floral traits. Composite

interval mapping confirmed the association of QTL with fiber and seed traits. The construction of an A genome diploid map, combing AFLP, TRAP and SSR markers can serve as a model for the advancement of cotton genetics, including the understanding of inheritance of fiber genes. With the availability of A_2 genome sequences (Li et al., 2014), it would be possible to find chromosomal locations of the polymorphic markers. Adding additional markers to the existing map will assist in map based cloning efforts and gene discovery.

ACKNOWLEDGEMENTS

We thank Dr. A.H. Paterson, UGA for providing the diploid mapping population and Cotton Inc. for financial support. We also thank the Cotton Fiber Lab, LSU AgCenter, Baton Rouge for HVI fiber analysis.

REFERENCES

Badigannavar, A.M. 2010. Genetic linkage mapping and marker trait associations in diploid and Association genetics for quantitative traits in tetraploid cottons. PhD thesis submitted to Louisiana State University, LA, USA.

Basten, C.J., B.S. Weir, and Z.B. Zeng. 2007. QTL Cartographer 2.5: A Reference manual and Tutorial for QTL Mapping. Department of Statistics, North Carolina State University, Raleigh, NC.

Bert, P.F., G. Charmet, P. Sourdille, M.D. Hayward, and Balfourier. 1999. A high-density molecular map for ryegrass (*Lolium perenne*) using AFLPmarkers. Theor Appl Genet 99: 445-452.

Brubaker, C.L., and A.H. Brown, 2003. The use of multiple alien chromosome addition aneuploids facilitates genetic linkage mapping of the *Gossypium* genome. Genome 46, 774-791.

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

Desai, A., P.W. Chee, J. Rong, O.L. May, and A.H,. Paterson. 2006. Chromosome structural changes in diploid and tetraploid A genomes of *Gossypium*. Genome 49, 336-345.

Endrizzi, J.E., E.L. Turcotte, and R.J., Kohel. 1985. Genetics, cytology, and evolution of *Gossypium*. Adv Genet. 23: 271-375.

Fisher, R.A., 1936. The use of multiple measurements in taxonomic problems. Ann. Eugenics. 7:179–188.

Fryxell, P.A., 1979. The Natural History of the Cotton Tribe. Texas A&M University Press, Collage Station, Texas.

Gerstel, D.U., 1953. Chromosomal translocations in interspecific hybrids of the genus *Gossypium*. Evolution, **7,** 234–244.

Hackett, C.A., and L.B. Broadfoot. 2003. Effect of genotyping errors, missing values and segregation distortion in molecular marker data on the construction of linkage maps. Heredity, 90: 33-38.

Han, Z.G., W.Z. Guo, X.L. Song, and T.Z. Zhang. 2004. Genetic mapping of EST-derived microsatellites from the diploid *Gossypium arboreum* in allotetraploid cotton. Mol. Genet. Genomics 272, 308-327.

Hansen, M., T. Kraft, M. Christiansson, and N.O. Nilsson. 1999. Evaluation of AFLP in *Beta*. Theor. Appl. Genet. 98, 845–852.

Hu, J., and B.A. Vick. 2003. TRAP target region amplified polymorphism, a novel marker technique for plant genotyping. Plant Mol. Biol. Rep. 21: 289–294.

Hutchinson, J.B. 1931. Studies on the inheritance of corolla color and petal size in Asiatic cottons. J. Genet. 24: 325-353.

Jenczewski, E.M., I. Gherardi, J.M. Bonnin, I. Prosperi, and Oliveri. 1997. Insight on segregation distortions in two intraspecific crosses between annual species of *Medicago* (*Leguminosae*). Theor. Appl. Genet. 94: 682–691.

Kebede, H., G. Burow, R.G. Dani, and R.D. Allen. 2007. A genome cotton as a source of genetic variability for upland cotton (*Gossypium hirsutum*). Genet. Resources Crop Evol. 54: 885-895.

Kosambi, D. 1944. The estimation of map distance from recombination values. Annals of Eugenics, 12, 172–175.

Lacape, J.M., T.B. Nguyen, S. Thibivilliers, B. Bojinov, B. Courtois, R.G. Cantrell, B. Burr, and B. Hau. 2003. A combined RFLP–SSR–AFLP map of tetraploid cotton based on a *Gossypium hirsutum* ×*Gossypium barbadense* backcross population. Genome. 46: 612–626.

Li F., G. Fan, K. Wang, F. Sun, Y. Yuan, G. Song, Q. Li, Z. Ma, C. Lu, C. Zou, W. Chen, X. Liang, H. Shang, W. Liu, C. Shi, G. Xiao, C. Gou, W. Ye, X. Xu, X. Zhang, H. Wei, Z. Li, G. Zhang, J. Wang, K. Liu, R.J. Kohel, R.G. Percy, J.Z. Yu, Y. Zhu, J. Wang, and S. Yu. 2014. Genome sequence of the cultivated cotton *Gossypium arboreum*. Nat. Genet. 46 (6): 567-574.

Mcharo, M., D.R. LaBonte, J.H. Oard, S.J. Kays, and W.L.McLaurin. (2004). Linking quantitative traits with AFLP markers in sweet potato using discriminant analysis. Acta Hort. 637:285–93.

Mei, M., N.H. Syed, W. Gao, P.M. Thaxton, C.W. Smith, D.M. Stelly, and Z.J. Chen. 2004. Genetic mapping and QTL analysis of fiber-related traits in cotton (*Gossypium*). Theor. Appl. Genet. 108: 280-291.

Myers, G.O., B. Jiang, M.W. Akash, A.M. Badigannavar, and S. Saha. 2009. Chromosomal assignment of AFLP markers in upland cotton (*Gossypium hirsutum* L.) Euphytica, 165(2): 391-399.

Nguyen, T.B., M. Giband, P. Brottier, A.M. Risterucci, and J. M. Lacape. 2004. Wide coverage of the tetraploid cotton genome using newly developed microsatellite markers. Theor. Appl. Genet. 109:167–175.

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

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

Rong, J., C. Abbey, J.E. Bowers, C.L. Brubaker, C. Chang, P.W. Chee, T.A. Delmonte, X. Ding, J.J. Garza, B.S. Marler, C. Park, G.J. Pierce, K.M. Rainey, V.K. Rastogi, S.R. Schulze, N.L. Trolinder, J.F. Wendel, T.A. Wilkins, D.W. Coplin T, R.A. Wing, R.J. Wright, X. Zhao, L. Zhu, and A. H., Paterson. 2004. A 3347-locus genetic recombination map of sequence tagged sites reveals features of genome organization, transmission and evolution of cotton. Genetics. 166: 389-417.

Rong, J., G.J. Pierce, V.N. Waghmare, C.J. Rogers, A. Desai, P.W. Chee, O.L. May, J.R. Gannaway, J.F. Wendel, T.A. Wilkins, and A.H. Paterson. 2005. Genetic mapping and comparative analysis of seven mutants related to seed fiber development in cotton. Theor. Appl. Genet. 111: 1137-1146.

SAS. 9.1.3. 2009. SAS. Statistical Analysis Software for Windows, 9.1.3 edition Cary, NC. USA.

Shappley, Z.W., J.N. Jenkins, J. Zhu, and J.C. McCarty. 1998. Quantitative trait loci associated with agronomic and fiber traits of Upland cotton. J. Cot. Sci. 4: 153-163.

Shen, X.L, W.Z. Guo, X.F. Zhu, Y.L. Yuan, J.Z. Yu, R.J. Kohel, and T.Z. Zhang. 2005. Molecular mapping of QTL for fiber qualities in three diverse lines in Upland cotton using SSR markers. Mol. Breed. 15:169–181.

Singh, P., and J. Singh. 1984. Variability for some economic characters in the genetic stocks of *Gossypium arboreum* and *G. barbadense* cottons. Cotton Development, 14: 15–16.

Singh, V.V., 1983. Range of variability in *Gossypium herbaceum* germplasm. Cot Dev, 14, 45

Stam, P., and J.W. Van Oojien. 1995. JoinMap version 2.0: software for the calculation of genetic linkage maps. CPRO-DLO, Wageningen, The Netherlands.

Ulloa, M., and W.R. Meredith. 2000. Genetic linkage map and QTL analysis of agronomic and fiber quality traits in an intraspecific population. J. Cot. Sci., 4: 161-170.

Ulloa, M., W.R. Meredith, Z.W. Shappley, and A.L. Kahler. 2002. RFLP genetic linkage maps from four $F_{2,3}$ populations and a joinmap of *Gossypium hirsutum* L. Theor. Appl. Genet., 104: 200–208.

Vos, P., R. Hogers, M. Bleeker, M. Reijans, T. Van de Lee, M. Hornes, A. Freijters, J. Pot, J. Peleman, M. Kuiper, and M. Zabeau. 1995. AFLP: a new technique for DNA fingerprinting. Nucleic Acids Res. 23: 4407–4414.

Wang W., Z. Wang, F. Li, W. Ye, J. Wang, G. Song, Z. Yue, L. Cong, H. Shang, Shilin Zhu, C. Zou, Q. Li, Y. Yuan, C. Lu, H. Wei, C. Gou, Zequn Zheng, Y. Yin, X. Zhang,K. Liu, B. Wang, Chi Song, Nan Shi, R. J. Kohel, R. G. Percy, J. Z. Yu, Y. Zhu, J. Wang and S. Yu. 2012. The draft genome of a diploid cotton *Gossypium raimondii*. Nat. Genet., 44, 1098–1103.

Weir, B.S., 1996. Genetic data analysis II, Sinauer, Sunderland, MA.

Wu, K.K., W. Burnquist, M.E. Sorrells, T.L. Tew, P. H. Moore, and S. Tanksley. 1992. The detection and estimation of linkage in polyploids using single-dose restriction fragments, Theor. Appl. Genet. 83: 294-300.

Zeng, Z.B., and B.S. Weir. 1996. Statistical methods for mapping Quantitative Trait Loci. Acta Agronomica Sinica. 22: 535-549.

Zhang, J., and J. McD Stewart. 2000. Economical and rapid method for extracting cotton genomic DNA. J. Cot. Sci. 4: 193-201.

Zhang, Z.S., M.C. Hu, J. Zhang, D.J. Liu, J. Zhang, K. Zhang, W. Wang, and Q. Wan. 2009. Construction of comphrensive PCR based marker linkage map and QTL mapping for fiber quality traits in upland cotton (*G. hirsutum*, L.). Euphytica. 24: 49-61.

Zhang, Z.S., Y.H. Xiao, M. Luo, X.B. Li, X.Y. Luo, L. Hou, D.M. Li, and Y. Pei. 2005. Construction of a genetic linkage map and QTL analysis of fiber-related traits in upland cotton (*Gossypium hirsutum* L.). Euphytica. 144: 91-99.