

Molecular Mapping of QTLs for Fiber Quality Traits in *Gossypium hirsutum* Multi-parent Recombinant Inbred Lines

Sami Doğanlar (✉ samidoganlar@iyte.edu.tr)

Izmir Institute of Technology: Izmir Yuksek Teknoloji Enstitusu <https://orcid.org/0000-0002-5712-0826>

Asena Akköse Baytar

Izmir Institute of Technology: Izmir Yuksek Teknoloji Enstitusu

Ceng Peynircioğlu

Özaltın Agricultural Enterprises Industry and Commerce Inc.

Volkan Sezener

Nazilli Cotton Research Institute

Anne Fray

Izmir Institute of Technology: Izmir Yuksek Teknoloji Enstitusu

Research Article

Keywords: association mapping, fiber quality traits, quantitative trait loci (QTL), Upland cotton

Posted Date: March 31st, 2021

DOI: <https://doi.org/10.21203/rs.3.rs-309098/v1>

License:   This work is licensed under a Creative Commons Attribution 4.0 International License.

[Read Full License](#)

Abstract

Cotton is a valuable fiber crop which supplies raw material to more than 50 industries and is produced in more than 70 countries worldwide thus, it is worthy of its reputation as “white gold”. The superiority of cotton fiber over other crops is primarily dependent on its quality. However, further improvements in fiber length, strength and fineness are required for modern processing technology and for cotton to maintain its position in the global market. Association mapping enables identification of QTLs controlling fiber quality-related traits which can be useful in cotton breeding. In the present study, we performed genetic diversity, linkage disequilibrium and association mapping analyses in 157 *G. hirsutum* multi-parent recombinant inbred lines using a total of 102 SSR markers. The population had depressed genetic variability (14%), a result of inbreeding of modern cotton genotypes. Despite this, we identified 13 significant and stable marker-trait associations for seed cotton yield, lint percentage, fiber length, fiber strength and fiber fineness ($p < 0.005$). We also detected QTL co-localizations with positive and negative marker additive effects. Our results indicate that selection against negative alleles may be as important as selection for positive alleles. Analysis of the effects of allelic combinations at different QTLs revealed significant and stable marker clusters that can be selected for or against to provide maximum quality gains in cotton fiber quality.

Introduction

Gossypium is a complex and highly variable genus consisting of species with different ploidies. According to one estimate, the genus includes 50 species of which 45 are diploids ($n = x = 13$) and five are well-known allotetraploids ($n = 2x = 26$). *Gossypium* species have been used as a source of desired traits and cotton fiber for thousands of years (since 800-700 BC); however, only four species are domesticated and cultivated: the New World cottons, *G. hirsutum* L. and *G. barbadense* L. ($n = 2x = 26$), and the Old World cottons, *G. arboreum* L. and *G. herbaceum* L. ($n = x = 13$). *G. barbadense* has good fiber features such as long, stable, fine fibers while *G. hirsutum* does not. On the other hand, *G. barbadense* has narrow adaptability due to its requirement for a long growing season while *G. hirsutum* is adapted to a wider range of environments and produces higher yields (May and Lege 1999). Therefore, *G. hirsutum* is the main supplier for the textile industry accounting for 95% of global cotton production. *G. barbadense* has a small market share along with the other domesticated cottons in various parts of the world (Jenkins 2003; Lee 1984).

Upland cotton (*G. hirsutum*) is a valuable fiber crop which supplies raw material to more than 50 industries (textile, oil, paper, etc.) and is produced in more than 70 countries worldwide (USDA 2020) thus, it is worthy of its reputation as “white gold”. The market value and superiority of cotton is primarily dependent on fiber quality (Gordon and Hsieh 2007). The most important forces driving the improvement of cotton quality are technological developments in processing, increased quality of life, preference for environmentally friendly and organic products, population growth and increased competition with synthetic fibers. The natural structure of cotton fiber provides many advantages over synthetic fibers including air permeability, comfort and high absorbency (Baytar 2014; Kalabek and Babaarslan 2016);

however, increased fiber length, strength and fineness are needed for cotton fiber to adequately meet the requirements of modern processing technology. Therefore, significant improvements in fiber quality are essential for cotton to maintain its position in the global market.

Upland cotton has been subjected to selection and cultivation for several thousand years. As a result, domesticated cotton's genome and morphology have undergone dramatic changes compared to wild types (Seyoum et al. 2018; Grover et al. 2020). The use of closely related genotypes and inaccurate selection criteria have reduced the genetic diversity of modern cotton cultivars. This narrow genetic basis is the main obstacle to improvement of quality and productivity in cotton. Characterization of the genetic diversity of cotton germplasm is essential for its use in breeding because it allows selection of the most appropriate combinations of parental genotypes and effective use of genetic materials. Molecular analysis is an effective and efficient way to determine the genetic background of and relationships between individuals. The information gained from such analyses can also be used in combination with phenotypic characterization to map yield and quality parameters such as fiber traits.

Many association mapping studies have targeted fiber quality in Upland cotton using simple sequence repeat (SSR) markers. These include work that examined fiber traits (Shermotov et al. 2010; Baytar et al. 2018b; Zeng et al. 2009), yield traits (Li et al. 2017; Baytar et al. 2018a; Jia et al. 2014; Mei et al. 2013) yield and fiber traits (Abdurakhmonov et al. 2010) and fiber quality traits (Abdurakhmonov et al. 2008, 2009; Cai et al. 2014; Nie et al. 2016; Ademe et al. 2017; Iqbal and Rahman 2017; Dong et al. 2019). For example, Cai et al. (2014) identified 107 significant fiber trait-associations in 99 *G. hirsutum* accessions using 97 polymorphic SSR markers. They reported that 70 SSR loci of 107 associations were significant in more than one environment and, that 36 SSR loci had been identified in previous studies which indicated the stability of the loci. Huang et al. (2018) screened Upland cotton cross populations using 284 SSR markers. They revealed that 54 SSR loci were significantly associated with fiber quality and, that 14 of them matched previously reported quantitative trait loci (QTL). In another association study, 57 significant associations were determined for fiber quality in 305 Upland cotton accessions using 198 SSR markers. Of the significant loci, 34 were identified for more than one fiber trait (Ademe et al. 2017). These and similar association analysis studies will be helpful for fiber quality trait breeding of modern cotton cultivars.

In the present study, we aimed to (I) examine the genetic diversity and ancestral background of 157 *G. hirsutum* multi-parent recombinant inbred lines (RILs) using a total of 102 SSR markers to detect promising genotypes for breeding, (II) identify marker-trait associations for fiber quality traits in the RIL population, and (III) reveal significant and stable molecular marker clusters to select alleles that provide maximum quality gains in cotton breeding.

Materials And Methods

Plant material

A total of 157 advanced cotton lines (*G. hirsutum*) were provided by Özaltın Agricultural Enterprises Industry and Commerce Inc. (OAE), Aydın Turkey (Supplementary Information Table S1). The advanced cotton lines were developed by crossing 100 inbred lines with three donor parents, Candia, Carmen and N84-S which were selected for their fiber quality characteristics. The 300 F1 individuals were self-pollinated to produce F2 populations of approximately 100 plants which were grown in the field and observed for their agronomic features. Selected lines were self-pollinated with observation, selection and self-pollination of progeny continuing an additional six generations to produce the 157 F8 multi-parent recombinant inbred lines that were used in this study.

Field experiments and phenotypic data collection

Field experiments were performed over two growing seasons at (OAE), Aydın, Turkey. Plants were manually harvested in September. The experimental region had sandy, loam soil with 13-14% water capacity and 4-6% wilting point. Augmented design was used with 10 replicates for each of the 157 cotton lines. Each genotype was planted in a 12 m row with 0.7 m between rows and 0.2 m between individual plants.

Seed cotton yield was calculated as total weight of seed cotton (kg ha^{-1}) before ginning. A roller gin was used to remove seeds from fibers. Lint percentage was calculated based on the formula: $[\text{lint}(\text{g}) / (\text{lint}(\text{g}) + \text{seed}(\text{g})) \times 100\%]$. The fibers were incubated at 21 °C and 65% relative humidity for 48 h until they reached 7–8% moisture content before fiber traits were evaluated. Fiber length, fiber strength and fiber fineness were measured with a USTER-HVI machine based on HVI cotton standards.

Best linear unbiased predictions (BLUPs) were calculated to model the morphological data over two years based on the method described by Wen et al. (2014) using JMP software (JMP®, Version 1.4. SAS Institute Inc., Cary, NC, 1989-2019). BLUPs were used in heritability, bivariate pairwise correlation and association analyses. Broad sense heritability was estimated with the formula: $H^2 = V_g / V_p \times 100$, V_g (σ^2_g) = genotypic variance = $MS(G) - MS(E) / r$, V_p (σ^2_p) = phenotypic variance = $\sigma^2_g + \sigma^2_e$ where, V_e (σ^2_e) = error variance, i.e. $MS(E) / r$; and (r = number of replicates) (Bhagasara et al. 2017; Kruijer et al. 2014). $MS(G)$ and $MS(E)$ are the mean squares for genotype and residual error which were calculated with ANOVA analysis. Bivariate correlation coefficients were estimated with Pearson Correlation, two-tailed method. All statistical analyses were done with PAWS statistics software (SPSS Inc. Released 2009, PASW Statistics for Windows, Version 18.0, Chicago).

SSR genotyping

DNA was isolated from young, fresh leaves with the manual DNA isolation protocol introduced by Doyle and Doyle (1987). DNA was quantified using a Nanodrop ND-1000 spectrophotometer and adjusted to 50 ng/ μl for further analysis. A total of 102 SSR primer pairs (Supplementary Information Table S2) which were determined to be polymorphic in our previous study (Baytar et al. 2018b), were used to identify loci within the population. Primer information is available at Cotton Database Resources (www.cottongen.org).

Polymerase chain reactions (PCRs) were performed as described by Baytar et al. (2017). PCR amplicons were separated with a capillary electrophoresis Fragment Analyzer™ Automated CE System based on DNF-900-55-DNA-35–500 bp method. PROSize 2.0 analytical software was used to analyze the raw data. Allele sizes were determined by binning fragments into ± 2 base pair bins.

Genetic diversity and population structure

SSR data were scored dominantly: “1” for presence, “0” for absence and “9” for missing data. DARwin6 (Dissimilarity Analysis and Representation for Windows) (Perrier and Jacquemoud-Collet, 2006) was used to calculate pairwise distances between cultivars with the Dice coefficient and the unweighted neighbor-joining algorithm. Pairwise PhiPT values, analogous to F_{st} , were calculated among clusters in the population by molecular variance analysis with GenAlEx 6.503 (Peakall and Smouse, 2006, 2012). STRUCTURE software (Pritchard et al. 2000) was used to evaluate genetic structure in the population with a Bayesian iterative algorithm. Sub-sets of the entire population were estimated based on different allele frequencies in the data. Sub-set numbers (K) from 2 to 10 were tested with 20 iterations each. Admixture model was used with a burn-in of 50,000 iterations and 300,000 MCMC (Markov Chain Monte Carlo) replications for correct estimation. The best representative K was decided based on the highest ΔK acquired from the Structure Harvester web-based program (Earl and vonHoldt 2012) by implementing the Evanno method (Evanno et al. 2005). Individuals were assigned to the sub-sets based on the ancestral similarity pattern with a cut-off value of 60%.

Linkage disequilibrium and association mapping analysis

QTL analysis was performed with TASSEL software v2.1. SSR alleles with frequencies below 0.05 were removed before analysis because minor alleles can bias LD. General linear (GLM) and mixed linear (MLM) models were tested to determine the best fit for association analysis. At first, MLM was implemented with the kinship (K) matrix. Then, MLM (with K) and GLM were each corrected with Q matrix and principal components (PC), separately. They were also corrected with both Q and PC. Q matrix was acquired from population structure analysis. PCs and K were calculated by TASSEL software. The seven methods were evaluated based on the proportions of null p-values (p_0) estimated with QVALUE software (Storey 2002) (FDR = 0.05). The one with the highest proportion of significant p values ($p_1 = 1 - p_0$) was the best model for QTL detection. Significance level was set at $p \leq 0.005$ for QTL detection. False discovery rate (FDR) was calculated for p values with QVALUE ($p \leq 0.005$ and q-value < 0.1) (Weller et al. 1998; Benjamini and Yekutieli 2005).

Pairwise linkage disequilibrium (LD) was estimated based on correlation coefficient (r^2) (Kruglyak 1999; Ardlie et al. 2002) between all locus pairs of SSRs with a rapid permutation test with 10,000 shuffles ($p \leq 0.01$) using TASSEL 2.1 (Bradbury et al. 2007). Significant LD ($p \leq 0.01$ and $r^2 \geq 0.03$) was used to generate the LD decay pattern. The chromosomal positions of the molecular markers were based on Blenda et al. (2012) and Fang and Yu (2012) (Supplementary Information Table S2).

Result And Discussion

Field experiments and phenotypic data collection

Advanced cotton lines were generated from F1 hybrids of 100 different inbred lines and three donor parents: Candia, Carmen and N84-S (Supplementary Information Table S1). The donors were selected for their good performance when grown in the Aegean region of Turkey. Candia was developed in Australia by crossbreeding and selection. It has good fiber quality with an average of 29.8 g/teks strength, 27.9 mm length, 4.5 micronaire fineness and 43.9% lint (Harem, 2014). Carmen, another Australian variety, was reported to have better quality than Candia with 30.9 – 36.6 g/teks strength, 30.3 – 32.0 mm length, 4.4 – 5.1 micronaire and 41.8% lint. N84-S (Nazilli 84-S) was selected in Turkey. Its fiber is of good quality with 78 – 84 (1000 Lb/inch²) strength, 28.5 – 29.5 mm length, 4.3 – 4.8 micronaire and a high ginning efficiency of 44 – 45% lint (Harem, 2014).

Seed cotton yield

Seed cotton yield of the 157 RILs ranged from 59 – 922 kg ha⁻¹ with an average of 482 kg ha⁻¹ for all genotypes (Table 1). Yield was more variable in the RILs than fiber characteristics (CV 29%). The majority of the RILs (59%) yielded from 1 to 92% more seed cotton than the population average. The four highest yielding genotypes (42 – 92% more than the average yield) were 13 OZ 034, 13 OZ 077, 13 OZ 104 and 13 OZ 090 whose donor parents were Carmen or Candia (Supplementary Information Table S1). These RILs may be used as high-yielding material in breeding programs. Seed cotton yield was significantly correlated only with fiber strength ($r = 0.23$, $p < 0.01$) indicating that strong-fiber varieties yielded more than weaker ones (Table 2). Thus, fiber strength is an important trait affecting yield and can be used as a selection criterion along with other parameters. Significant positive correlations were also reported between seed cotton yield and fiber strength in other work ($r = 0.63$; Yaqoob et al. 2016), ($r = 0.28$; Azhar et al. 2004). However, this is not consistent across the literature as negative correlations were detected for these traits in other work (Karademir et al. 2010; Mendez-Natera et al. 2012; Tang et al. 1996). Seed cotton yield is greatly affected by genotype, temperature and growing conditions. For example, Baytar et al. (2018a) detected a significant decrease in seed cotton yield (27%, $p < 0.001$) with insufficient soil moisture. These factors may account for the contradictions among fiber quality studies.

Lint percentage

Ginning turn-out is the percent of lint obtained from seed cotton and this is expected to be at the highest possible level for elite cultivars. Lint percentage significantly varied between genotypes ($p < 0.05$) (Supplementary Information Table S3) ranging from 37 to 47% with an average of 41% over two years (Table 1). In our study, almost half of the genotypes (45%) had more ginning yield with 1 to 10% more lint compared to the population average. There were statistically significant differences for lint percentage between the cross combinations using Carmen versus Candia ($p < 0.0001$) as well as Carmen versus N84-S ($p = 0.03$) (Supplementary Information Table S4) such that the Candia and N84-S combinations had slightly but significantly higher lint yield (Supplementary Information Table S5 and Table S6). Some

advanced lines stood out with the highest ginning efficiency: 13 OZ 111, 13 OZ 118, 13 OZ 110 and 13 OZ 109. Of these lines, three were from same cross (Candia x 06OZ70-4) which indicates the promise of this combination for improvement of lint yield. However, the fiber characteristics of these lines were poorer than average. In accordance with this result, we detected negative correlations between lint percentage and fiber length ($r = -0.33$, $p < 0.01$) and fiber strength ($r = -0.21$, $p < 0.01$) indicating that long fibers were associated with reduced ginning yield. Research has consistently indicated negative correlations between ginning efficiency and fiber parameters (fiber length and strength) (Karademir et al. 2010; Ulloa and Meredith, 2000; Rakshit et al. 2010; Stroman 1949). One exception is the work of Khalid et al. (2018) who saw a significant positive correlation between lint percentage and fiber length. Given the preponderance of evidence, it seems that selection for higher ginning efficiency is likely to be accompanied by negative linkage drag of undesired fiber quality. Moreover, we detected a moderate broad sense heritability (62%) for lint percentage with a significant year effect ($p = 8.4E-10$) (Supplementary Information Table S3). Low heritability was also determined in other studies (30%, Jarwar et al. 2018; 32%, Dhivya et al. 2014). Taking all of these results into consideration, phenotypic selection based on lint percentage alone may not be effective for improvement of productivity. This is consistent with the statement by Kearney (1912) that ginning efficiency should not be used directly to evaluate productivity as it can often lead to unintended consequences.

Fiber length

All genotypes had long or medium-long fibers ranging from 26.1 to 33.8 mm with an average of approximately 30.0 mm over both years. The majority of the RILs (87%) were classified as long-fiber types (≥ 29 mm) (Supplementary Information Table S1). There was a statistically significant difference between the averages for the three cross combinations ($p = 0.001$) such that the progeny of Carmen tended to have longer fibers than those of Candia. This was expected given that Carmen has significantly longer fibers than Candia. The genotypes with the longest fibers were 13 OZ 043, 13 OZ 132, 13 OZ 145 with average fiber length of 32 mm, similar to Carmen. Indeed, two of these RILs were from crosses with Carmen. In addition to its long fibers, RIL 13 OZ 043 also had strong fibers while 13 OZ 132 also had good fineness (4.1 micronaire) with fibers that were 11% thinner than the population average. Consistent with these findings, fiber length correlated positively with fiber strength ($r = 0.28$, $p < 0.01$) and negatively with fiber fineness ($r = -0.19$, $p < 0.01$) such that long fibers tended to have thinner diameter but were stronger than shorter ones. Similarly, other studies found significant correlations such that fiber length correlated positively with fiber strength ($r = 0.22-0.77$, $p < 0.01$) (Baytar et al. 2018b; Khalid et al. 2018; Karademir et al. 2010; Wan et al. 2007; Zhang et al. 2005; Asif et al. 2008) and negatively with fineness (-0.26 - -0.85 , $p < 0.01$) (Baytar et al. 2018b; Karademir et al. 2010; Khalid et al. 2018; Azhar et al. 2004; Asif et al. 2008).

Simultaneous improvement of fiber features is a highly desirable goal for cotton breeders and the correlations between these three features show that they can be improved together without unwanted linkage drag. Stronger, longer fibers positively affect spinning consistency as they can be twisted around each other with higher efficiency in processing (Chee et al. 2005; Moore, 1996). The early stages of fiber

growth are highly sensitive to differences in temperature, growth area and especially watering regime (Reynolds and Killough, 1933; Gibson and Joham 1969; Hanson et al. 1956). For example, we previously reported that reducing irrigation by 50% caused shorter fiber (5%, $p < 0001$) development (Baytar et al. 2018b). This sensitivity was reflected in our current study by a significant year effect ($p = 0.02$) and also a moderate heritability (43%) for fiber length which was similar in previous work (42%, Balci et al. 2020; 51%, Azhar et al. 2008; 64%, Khokhar et al. 2017). These results indicate that direct phenotypic selection may not be effective for improvement of fiber length.

Fiber strength

Fiber strength varied from 25.0 to 47.6 g teks⁻¹ with an average of approximately 33 g teks⁻¹ in the population over both years (Table 1). Significant differences between the genotypes were obtained for fiber strength ($p < 0.05$). However, there was no significant difference between cross combinations (Supplementary Information Table S4). Year effect was highest for fiber strength ($p < 0.001$) compared to the other traits with an average 15% increase in strength in the second year (Supplementary Information Table S1). In the first year, 4.5% of the population had intermediate and 95.5% had strong fibers (25.0 - 28.0 g teks⁻¹). However, in the second year all genotypes had either strong or very strong fibers (29.4 - 47.6 g teks⁻¹). The RILs 13 OZ 142, 13 OZ 065, 13 OZ 043 and 13 OZ 122 had the strongest fibers which were approximately 14% stronger than the population mean. Two of these lines were Carmen crosses; however, as already indicated, there were no statistically significant differences between the means for the three cross combinations.

Fiber strength is a complex feature and is reported to be influenced by environmental factors more than fiber fineness and length (Pope and Ware, 1945; Hanson et al. 1956) in agreement with our finding that this trait had the moderate heritability estimate (42%) and the highest year effect (Supplementary Information Table S3). Temperature differences, soil moisture level and especially sunlight intensity have great effects on fiber strength (Hanson et al. 1956). Accordingly, we previously found that insufficient soil moisture significantly reduced fiber strength (Baytar et al. 2018b). In this study, there was a 13.5% reduction in average precipitation during the second season's growing period (April-September) for Aydin (Supplementary Information Table S7). Moreover, the average number of sunny days in the spring-summer of this year was 17% higher than in the previous year (WWO, 2013), despite the fact that the average temperature remained similar for both years. Hanson et al. (1956) reported that sunlight may make the greatest contribution to fiber strength compared to other climatic factors. Thus, the improved fiber strength in the second year may be related to the greater number of sunny days during that year and unexpected rainfall during the previous growing season.

Fiber fineness

The thickness of cotton fibers was significantly different ($p < 0.05$) between genotypes ranging from 3.05 to 5.63 micronaire with an average of 4.6 micronaire over the two years. A total of 11.5% of the RILs had premium fibers (3.7 - 4.2 micronaire, USDA), an important breeding target which merits a higher market

price and greater spinning efficiency. The majority of the population (78.3%) had basic fibers (4.3 - 4.9) which are also acceptable for processing. However, 10.2% of the RILs had coarse fibers which are out of the acceptable range- (>5.0 micronaire) for fiber quality. The premium-fibered RILs were 13 OZ 086, 13 OZ 087, 13 OZ 146 and 13 OZ 028 with an average thinness that was 14.8% better than the mean of the population. Two of these lines were Candia progeny, however, the differences between cross combinations were not significant for fineness ($p>0.05$) (Supplementary Information Table S4).

In the spinning industry, premium fibers are desirable because finer fibers produce stronger yarn because they have more fibers per section. Moreover, thinner fibers are also soft and shiny and produce end-products that are more valuable and elegant. However, finer fibers must be processed more slowly to reduce fiber loss. Thicker fibers slightly increase ginning efficiency as fiber damage is decreased (Dever, 1986). This is consistent with the positive correlation between fiber fineness and lint percentage in both our (0.26, $p < 0.01$) and other studies (0.53, Yaqoob et al. 2016; 0.32%, Karademir et al. 2010, $p < 0.01$). In contrast, Khalid et al. (2018) found a weakly negative correlation between fineness and ginning efficiency (-0.19, $p < 0.05$). Fineness is considered to be more variety-dependent than other fiber characters which are greatly influenced by environmental conditions (Dever 1986). In agreement with this, we detected no year effect on fiber fineness ($p = 0.12$) (Supplementary Information Table S3). The trait had moderate heritability (50%) consistent with previous studies (41% Azhar et al. 2004; 42%, Balci et al. 2020; 69% Ali et al. 2008). In contrast, some studies found that fiber fineness was a highly heritable trait (90%, Percy 2006; 96% Farooq et al. 2014). Although fineness varies with cultivar and is reported to be controlled by dominant gene action, it is also influenced by environmental effects, such as moisture level, temperature, sunshine and fertilization management (Shakeel et al. 2012; Hanson et al. 1956). Interestingly, water stress was associated with increased fiber diameter in our previous study (Baytar et al. 2018b). Therefore, variable genetic materials and climatic conditions may be the underlying reasons for significant differences at the phenotypic level in different studies.

It is important to highlight individuals with multiple favorable traits. Such individuals are good candidates for simultaneous improvement of fiber quality parameters. The genotypes 13 OZ 120 (Carmen X 06OZ569) and 13 OZ 104 (Candia X 06OZ653) had good seed yield, fiber length and fiber strength; while 13 OZ 146 (Carmen X 06OZ653) and 13 OZ 087 (N84-S X 06OZ720) had longer, stronger and finer fibers than the population averages.

Genetic diversity and population structure

A total of 102 SSR primer pairs produced 379 loci with an average of 3.7 alleles per marker across 157 multi-parent Upland cotton RILs. Eight individuals were discarded from genetic analysis due to a high proportion of missing data in pairwise comparisons (>50%): 13 OZ 035, 13 OZ 042, 13 OZ 044, 13 OZ 076, 13 OZ 101, 13 OZ 123, 13 OZ 130 and 13 OZ 139. Unweighted neighbor-joining (NJ) dendrogram analysis revealed three main sub-clusters (Supplementary Information Fig. S1): Cluster1 with 78 (52.3%), Cluster2 with 62 (41.6%) and Cluster3 with 9 (6%) individuals. There was a strong correlation between the Dice coefficient dissimilarity matrix and the relationships demonstrated by the dendrogram ($r= 0.89$)

based on a Mantel test. The progeny of cross combinations were distributed over the three main clusters without any tendency to group based on cultivated parent (Supplementary Information Table S8) indicating that recipient parents contributed greatly to the genetic differentiation of the RILs.

We used intraspecific multi-parent RILs in our study because the results of interspecific population QTL analysis are often not effective for breeding of Upland cotton (Fang et al. 2014). For example, favorable alleles often come from the other species' genome and must be introgressed into *G. hirsutum*. This process adds several generations to the breeding process. Moreover, polymorphic markers which were detected in interspecific populations may not be polymorphic in *G. hirsutum* resulting in markers which are unusable in selection for Upland cotton (Fang et al. 2014). On the other hand, intraspecific populations often have low variation within the population. In this study, we used a three parent-based RIL population in an attempt to obtain higher genetic variability in the population, however, genetic analysis revealed low diversity with a mean dissimilarity of only 14%. This indicates a genetic bottleneck within current cotton genotypes due to thousands of years of selection and over-cultivation of limited genetic material (Noumkina et al. 2019; Chaundry et al. 2010; May OI 1999). In contrast, when germplasm instead of a bi-parental population was used, we observed more genetic diversity (38%) (Baytar et al. 2017). Low and moderate levels of genetic diversity were also reported for Upland cotton in many other studies regardless of population model (Tyagi et al. 2014; Abdurahkmanaov et al. 2008; Lapece et al. 2007; Fang et al. 2013; Fang et al. 2014; Hinze et al. 2012).

Molecular genetic analysis reveals existing diversity among individuals as well as the ancestral background of the population. Such knowledge is particularly useful in guiding selection of desirable genotypes especially when pedigree information is lacking. We identified the highest dissimilarities between lines from Candia x 06 OZ 770-4 and Candia x 06 OZ 602 (27-29%) (Supplementary Information Table S9). Progeny from Candia x 06 OZ 770-4 were also found to have the highest ginning efficiency which indicates that these lines have breeding potential. The least diversity was between 13 OZ 013 and 13 OZ 010 with 95% similarity. Within the cross combination groups, genetic difference was low with a mean of 14% dissimilarity for Candia combinations and 13% for both Carmen and N84-S combinations. AMOVA analysis revealed that nearly all of the variation in the RILs (98%) resided within the cross groups rather than among groups (2%) ($\Phi_{IPT} = 0.019$, $p = 0.001$) (Supplementary Information Table S10).

Based on the evaluation of population structure, the best representative number of sub-groups (K) was inferred to be three for the population because the highest ΔK peaked at $K = 3$ (Fig. 2 and Supplementary Information Fig. S2a-b). There was also a meaningful signal at $K = 6$ which was also distinguishable in the NJ tree supporting that three main clusters of the population could be further separated into six smaller groups. Similar subtle sub-structures were also observed in other clustering analyses (Zhao et al. 2014; Ersoz et al. 2007; Kuzay et al. 2020). However, the Q matrix for K3 was used for association analysis as it is recommended that the minimum K is used in order to include as many loci as possible. The individuals were assigned to the three sub-groups based on a membership probability cutoff value of 60% (Supplementary Information Table S11). Sub-groups 1, 2 and 3 contained 23 (14.6%), 31 (19.8%), and 39 (24.8%) individuals, respectively. Sixty-four (40.8%) individuals remained admixed with probability

values less than 60%. Admixed individuals were not assigned clearly to a group as they possess several ancestral sources indicating that they are more genetically variable than the other lines. The best example of this was 13 OZ 103 (Candia x 06 OZ 602) which was admixed and the most divergent line based on diversity analysis (Supplementary Information Table S9). The results of both structure and diversity analysis were compared to judge their consistency (Fig. 1). This comparison showed that 31.8% of the population was similarly grouped in both clustering methods (Supplementary Information Table S11). While this value is low, it is not unexpected given the fact that the two methods use different approaches to determine genetic relationships: the dendrogram was drawn using a distance-based method while population structure was determined with a Bayesian approach.

Linkage disequilibrium and association mapping analysis

Linkage disequilibrium

Linkage disequilibrium (LD) refers to statistically non-random associations between pairs of loci in the genome. Genome-wide LD must be calculated because it determines the quality of association mapping and reveals the effects of genetic components (recombination, genetic migration, natural selection, etc.) on the population through evolutionary history (Shen et al. 2019; Silva-Junior and Grattapaglia 2015). In this study, 8.4% and 3.1% of SSR loci were in significant LD at $p \leq 0.05$ and $p \leq 0.01$ with average r^2 of 0.13 and 0.18, respectively. Approximately 6.4% of SSR loci remained in significant linkage at $r^2 \geq 0.05$ with an average r^2 of 0.16. In addition, 2.5% of loci were in LD with $r^2 \geq 0.1$ (average $r^2 = 0.30$). The extent of LD was low and similar to results of previous studies that used different panels of Upland cotton germplasm: 1.83% ($p < 0.001$) of loci in 158 elite germplasm lines (Zhao et al. 2014), 2.93% ($p < 0.01$) of loci in 81 accessions (Zhang et al. 2013), 9.4% ($p \leq 0.05$) of loci in 241 cotton lines (Qin et al. 2015) and 4% ($r^2 \geq 0.05$) of loci in 335 varieties (Abdurakmanov et al. 2009). Although LD values vary considerably with the different thresholds and germplasm used, an overall low level of LD was observed and is hypothesized to result from the high recombination rate in allopolyploid cotton (Brubaker et al. 1999). Because of this, it is important to include population structure (Q matrix) and kinship (K) in the association models for this crop (Qin et al. 2015). As a result, we corrected GLM with both Q and PC to remove spurious linkages in association mapping.

A reduced rate of recombination is expected between loci in significant LD due to their physical adjacency on the same chromosome. However, 6.6% of unlinked loci were in significant LD ($p \leq 0.05$) with an average r^2 of 0.35 which was higher than the average LD of linked loci (0.30). Similar findings were previously reported in cotton (Qin et al. 2015; Mei et al. 2013; Abdurakhamanov et al. 2008; Fang et al. 2013) indicating that not only physical linkage but also other factors produce significant LD in this species (Abdurakmanov et al. 2009; Nordborg et al. 2002). These factors include selection for certain traits and the concomitant indirect selection of other loci through generations of breeding, use of common parents in the population and genetic bottlenecks (Abdurakmanov et al. 2009; Strich et al. 2005; Huttley et al. 1999). Molecular explanations for this phenomenon include the co-segregation of protein complex subunit genes which must be inherited together through generations in order to be functional in

the cell (Miglani 2002). In addition, multi-stage pathways in which different genomic loci contribute to the same mechanism generate LD between unlinked loci in the genome.

LD decay is another useful parameter for association analysis. The distance for which the LD clearly decayed gives information about how the linkage between SSR loci was conserved over generations. LD decay is also used to deduce theoretically marker density and the fitness of the QTL detection method. In this study, LD decay was plotted based on genetic distance at a significance level of $p \leq 0.05$ and $r^2 > 0.03$. It decayed around 60 cM at $r^2 = 0.1$ (Supplementary Information Fig. S3). The tetraploid cotton genome is 5,200 cM (Paterson and Smith, 1999) and, based on the extent of LD decay, 90 - 100 markers are sufficient for association mapping in this population. Therefore, the 102 markers used in this study to detect QTLs should be sufficient. The LD decay observed in our population was similar to that seen by Jia et al. (2014) with 55-60 cM at $r^2 = 0.1$.

Association mapping

Both Q and PC-corrected GLM produced the highest proportion of significant p values (π_1) among the seven models tested (Supplementary Information Table S12). Therefore, the GLM (Q + PC) method was applied to detect QTLs at a significance level of $p \leq 0.005$ and $q < 0.1$. As a result, association analysis produced 13 SSR loci significantly associated with five traits: seed cotton yield, lint percentage, fiber length, fiber strength and fiber fineness (Table 3). PVE values of individual SSR loci were relatively low ranging from 0.07 to 0.12. This was expected due to the complex nature of quantitative traits which are under the control of multiple genes that have small effects on the trait phenotype and are highly sensitive to environmental factors (Mackay, 2009).

BNL2495₂₁₅ was the only SSR locus which was significantly associated with seed cotton yield ($p = 1.8 \times 10^{-4}$). It had a positive effect, increasing seed yield approximately 51 kg ha⁻¹ and accounting for 9% of the phenotypic variation in this trait.

Five marker alleles were associated with lint percentage with PVE values ranging from 8 to 11%. BNL1495₂₃₆ was the most statistically significant locus ($p = 6.3 \times 10^{-5}$) and had a positive allelic effect with the highest phenotypic variance. Two alleles of DPL322 marker, DPL322₂₂₇ and DPL322₂₁₈, had positive and negative effects on lint percentage, respectively. This indicates that if this marker is used for selection, the allele with negative effects should also be considered (i.e. selected against) for improvement of lint percentage.

Three marker loci were linked to fiber length. BNL3545₁₈₁ was associated with this trait and had higher significance ($p = 3.3 \times 10^{-5}$) and PVE value (12%) than the other loci. The allelic loci of BNL3545 had opposite additive effects on fiber length: BNL3545₁₈₁ was associated with increased length while BNL3545₁₇₀ was associated with decreased length.

Two SSR loci were significantly associated with fiber strength with positive effects on the trait. BNL1495₂₄₂ and BNL3545₁₂₀ had 8% and 7% PVE.

Fiber fineness was associated with two allelic-marker loci: JESPR274₁₄₄ and JESPR274₁₂₀. Both alleles had relatively high significance and JESPR274₁₄₄ was the most significant marker locus in this study (3.15×10^{-5}). Both alleles accounted for average amounts of phenotypic variation (9%). JESPR274₁₄₄ was associated with an increased value for fineness whereas JESPR274₁₂₀ had the opposite effect. For this trait, the negative effect-markers should be considered in improvement of fiber fineness because lower micronaire values are desired in the cotton industry.

Marker co-localization

Several QTLs were significantly associated with multiple traits in this study. This may indicate that a single gene/locus has a role in controlling more than one trait (pleiotropy) or that linked genes/loci are responsible for biological pathways inherited together in cotton. Such QTL hotspots harbor potential for deeper investigation to reveal complicated QTL networks and the mechanisms underlying traits (Wang et al. 2019).

Different alleles of three markers were associated with more than one trait: BNL1495, BNL3545 and JESPR274, each with three alleles (Table 3). BNL1495 was linked to fiber length with a negative effect and, to both lint percentage and fiber strength with positive effects. This is compatible with the finding of a negative correlation between lint percentage and fiber length in our study. This may indicate that there is unwanted linkage drag between lint percentage and fiber length which makes it challenging to increase fiber quality without sacrificing ginning yield.

JESPR274 was linked to both lint percentage and fiber fineness with adverse marker effects. This was consistent with the finding of a positive correlation between lint percentage and fiber fineness in this study. This may be an example of pleiotropy as finer fibers may be more difficult to gin. However, there another allele of JESPR274 was associated with increased fiber fineness indicating even more complex interactions between these traits.

Two alleles of BNL3545 were associated with fiber length and strength with positive marker effects indicating a pleiotropic interaction that was consistent with the positive correlation identified between them in this study. On the other hand, a third allele of BNL3545 had a negative allelic effect on fiber length. On account of this, improvement of fiber features could be done through simultaneous elimination of negative alleles and selection of positive alleles in molecular breeding.

QTL comparisons with previous reports

It is difficult to compare QTL number and effect across different studies due to factors such as population model, experimental environments, cutoff value for significance levels, p value-correction methods and lack of common markers (Fang et al. 2013), all of which can affect the outcomes of such

work. For example, we obtained 100 significant QTLs for all traits at $p \leq 0.05$; however, after controlling for FDR, this dropped to 13 QTLs at $p \leq 0.005$. High numbers of marker-trait associations are neither effective nor practical for use in molecular breeding. A manageable number of QTLs is as important as reliable and strong QTLs for each trait.

Because few common markers have been used in different QTL studies, it is difficult to compare results. We compared our results with a reference genetic map (Fang and Yu, 2012) to find significant QTLs in close proximity with our markers (less than approximately 10 cM) for the same trait or other fiber features. This is useful as it compiles the QTL results for multiple cotton traits which must be simultaneously improved by breeders (Supplementary Information Table S13).

BNL2495 marker loci (73.26 cM on chromosome 26) linked to seed cotton yield in our study were also detected for seed cotton yield by Yu et al. (2013) indicating this is a reliable QTL for fiber yield. Moreover, this region was also associated with fiber strength (Wang et al. 2015) and micronaire (Zhang et al. 2012). Liang et al. (2013) also identified a QTL for fiber uniformity at both BNL2495 and the nearby marker DPL070 (0.57 cM away).

BNL1495 which was linked to lint percentage, fiber length and strength in this study was similarly reported as linked to lint yield and percentage (Wu et al. 2009; Ma et al. 2019), fiber length and strength (Frelichowski et al. 2006; Liang et al. 2013; Abdullaev et al. 2017), as well as, fiber elongation (Park et al. 2005; Frelichowski et al. 2006). Thus, this genomic region houses a common, stable QTL cluster for fiber quality. BNL1495 was mapped on chromosome 13 (53.7 cM) and within 5.8 cM there is another QTL cluster for fiber strength and length harboring DPL0687 (55.6 cM) (Liang et al. 2013) and GH678 (59.2 cM, fiber strength only) (Yu et al. 2013).

JESPR274 on chromosome 9 was associated with lint percentage and fiber fineness in this study. We previously identified JESPR274 as associated with plant structural features (Baytar et al. 2018b) as well as lint and seed cotton yield (Baytar et al. 2018a). This marker was also linked to lint index (Wang et al. 2007) and fiber maturity (Qin et al. 2015). Together with two adjacent markers, DPL783 (Zhang et al. 2016) and MUSB958 (Yu et al. 2014), the region of JESPR274 is an informative QTL for lint percentage within an interval of 5.7 cM.

DPL322, associated with lint percentage in this study, was linked to lint percentage and lint yield in previous work (Baytar et al. 2018a) and also proposed as a useful marker for improvement of fiber traits (Saeed and Elçi 2017). DPL322 is on chromosome 15 (51.4 cM) where marker-fiber trait associations clustered within 3.1 cM: DPL003 was linked to seed cotton weight (Zhang et al. 2016), MUSB1267 and TMB2931 linked to fiber fineness and short fiber content (Yu et al. 2014) and BNL2496 associated with seed cotton weight, seed index (Zhang et al. 2016) and fiber uniformity (Wang et al. 2017).

BNL3545 (119 cM on chromosome 2), identified for fiber length and strength in our study, was also linked to seed cotton weight (Liu et al. 2018). Moreover, it mapped in relatively close proximity with BNL3972 (at 106.5 cM) and BNL1434 (at 96.4 cM) which were associated with fiber elongation (Wang et al. 2016)

indicating a QTL for fiber quality in this region. The intervals between BNL3545 and its neighboring markers related to fiber traits are more than 10 cM due to the insufficient saturation of this region of chromosome 2. However, this distance is within the range of significant LD level determined by this study. BNL3545 was also mapped on chromosome 14 (7.6 cM) which is homeologous to chromosome 02. This marker mapped close to C2-00118 (16.4 cM) which was linked to fiber length and micronaire (Huang et al. 2018) and TMB1931 (11.49 cM) linked to 2.5 span length and micronaire (Shang et al. 2016). BNL3545 is also close to TMB0071 (6.3 cM) which is linked to micronaire and lint index (Shang et al. 2016). All of these loci contribute to a 9 cM interval for fiber quality traits.

Identification of common and stable QTLs is important for marker-mediated selection in breeding. In this respect, our results showed that the study was successful in validating common markers.

QTL confirmation

The population was divided into subsets containing different combinations of favorable SSR marker alleles for each trait. The trait means of these subsets were compared to gain insight into how to use the alleles effectively in both positive and negative molecular selection for breeding of improved fiber quality.

For seed cotton yield, the individuals were divided into two groups based on presence or absence of SSR locus BNL2495₂₁₅. ANOVA analysis indicated significantly higher yield in the group which had the allele linked to improved seed cotton yield ($p = 0.021$) (Fig. 3) (Supplementary Information Table S14A).

For lint percentage, the individuals were first divided into three groups based on the presence of positive alleles (BNL1495₂₃₆ and DPL322₂₂₇): groups 1 and 2 with one and two positive alleles, respectively, and the third group (0) with no positive alleles. The statistical analysis showed that there was a significant increase in lint percentage as the number of positive alleles increased ($p = 0.001$) (Fig. 4) (Supplementary Information Table S14B). The individuals were then divided into four groups based on the absence of negative alleles (DOW050₂₀₅, JESPR274₁₆₁ and DPL322₂₁₈): groups 1 and 2 lacked one and two adverse alleles (from any of the linked loci), respectively; while group 3 lacked all three adverse alleles. There was a statistically significant decrease in lint percentage as the number of adverse alleles increased ($p = 0.0001$) (Fig. 4) (Supplementary Information Table S14C) with no significant differences between alleles. In other words, each allele was equally effective at reducing lint yield. These results indicate that negative selection over each adverse allele had a positive impact on lint percentage, and that, eliminating all adverse alleles resulted in higher yields.

For fiber length, the population was separated into two groups. Individuals having the favorable allele (BNL3545₁₈₁) and lacking two adverse alleles (BNL3545₁₇₀ and BNL1495₂₃₁) were clustered in one group while all other allelic combinations remained in the second group. As a result, the first group had significantly longer fibers than the other group ($p = 0.029$) (Fig. 5) (Supplementary Information Table S14D). We also tested for significant differences when individuals with either presence of the positive allele or absence of any negative allele were clustered in the desired group with the remaining individuals making up the second group. In this case, the first group had significantly longer fiber than the other

group ($p < 0.0001$) indicating that positive and negative selection were equally effective (Fig. 5) (Supplementary Information Table S14E). In addition, negative selection of BNL3545₁₇₀ and BNL1495₂₃₁ gave similar results ($p > 0.05$) indicating that selection over either of the two adverse allele resulted in significantly longer fiber length. The desired group also showed statistically significant finer fibers than the other group ($p = 0.01$) (Supplementary Information Table S14F) indicating that these marker loci could be effective for co-selection of fiber length and fineness.

There were two favorable alleles for fiber strength (BNL1495₂₄₂ and BNL3545₁₂₀). Therefore, individuals that had both alleles were combined and the remaining individuals formed the second group. Expectedly, the first group had statistically significant stronger fibers than the second one ($p = 0.004$) (Fig. 6) (Supplementary Information Table S14G). In addition, we grouped the individuals that had any two favorable alleles (BNL1495₂₄₂ or BNL3545₁₂₀) while the rest of the individuals were assigned to the opposing group. Similarly, there was significantly greater fiber strength in the first group ($p = 0.003$) (Fig. 6) (Supplementary Information Table S14H) indicating that the presence of at least one positive allele contributed to stronger fibers.

Conclusion

Conventional breeding has encountered many challenges in the development of elite varieties such as inbreeding depression, unpredictable interaction of genotype versus environment and the complex nature of quantitative traits that do not follow the laws of Mendelian genetics. Breeding is also time-consuming and laborious work. Molecular genetic advances promise an additional tool that can be leverage along with traditional breeding methods. Molecular marker-based strategies save time and workload and help achieve success in breeding.

In this study, we clearly detected the depression of genetic variability that has resulted from inbreeding of modern cotton genotypes indicating that genetic diversity must be urgently broadened. Our three-parent RIL population contained individuals with promise for breeding of fiber quality traits which are highly complex features that are affected by indirect interactions between genetic components, pleiotropy and the environment. It is not feasible, nor necessarily desirable, to reveal all minor effect-genes for fiber traits. Thus, we identified a manageable number of reliable and stable QTL for fiber quality traits and confirmed the significance of positive and negative selection based on these QTLs. We also revealed important QTL hotspots for fiber related traits which can be useful in breeding. Based on this work, we recommend that: I. fiber quality traits are improved simultaneously rather than individually due to the effects of negative linkage drag; II. molecular markers are integrated into classical breeding programs to direct improvement; and III. negative selection can be as important as positive selection in marker-assisted breeding.

Abbreviations

BLUPs, Best linear unbiased predictions; GLM, General linear model; LD, Linkage disequilibrium; MLM, Mixed linear model; QTL, Quantitative trait loci; RIL, Recombinant inbred line; SSR, Simple sequence

repeat.

Declarations

*Ethics approval and consent to participate

The authors declare that they have followed all necessary ethical guidelines. Human and animal subjects were not used in this work

*Consent for publication

All authors and associated institutes have consented to publication of this work.

*Availability of data and materials

Data are available upon request.

*Conflict of interest/competing interests

The authors declare that they have no conflict of interest or competing interests.

*Funding

This research was funded by grants from The Scientific and Technological Research Council of Turkey to Sami Doğanlar (TÜBİTAK 1190677) and Ceng Peynircioğlu (TÜBİTAK TEYDEP 3060450).

*Authors' contributions

AAB: molecular characterization, data analysis, interpretation of data, manuscript drafting and revision; CP, VS: field experiments; AF: conception and design, interpretation of data, manuscript revision; SD: conception and design, manuscript revision; All: final approval of the version to be published.

*Acknowledgements

Not applicable.

*Authors' information (optional)

Not applicable.

References

Abdullaev AA, Salakhutdinov IB, Egamberdiev SS., Khurshut EE, Rizaeva SM, Ulloa M, Abdurakhmonov IY (2017) Genetic diversity, linkage disequilibrium, and association mapping analyses of *Gossypium barbadense* L. germplasm. PLoS ONE 12(11): e0188125. <https://doi.org/10.1371/journal.pone.0188125>

Abdurakhmonov IY, Buriev ZT, Shermatov SE, Kushanov FN, Makamov A, Shopulatov U, Turaev O, Norov T, Akhmedov C, Mirzaakhmedov M, Abdukarimov A (2010) Utilization of natural diversity in Upland cotton (*G. hirsutum*) germplasm collection for pyramiding genes via marker-assisted selection program. International Cotton Advisory Committee.

Abdurakhmonov IY, Kohel RJ, Yu JZ, Pepper AE, Abdullaev AA, Kushanov FN, Salakhutdinov IB, 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, Saha S, Jenkins JN, Buriev ZT, Shermatov SE, Scheffler BE, Pepper AE, Yu JZ, Kohel RJ, Abdugarimov A (2009) Linkage disequilibrium based association mapping of fiber quality traits in *G. hirsutum* L. variety germplasm. *Genetica* 136:401–417.29.

Ademe MS, He S, Pan Z, Sun J, Wang Q, Qin H, ... Du X (2017) Association mapping analysis of fiber yield and quality traits in Upland cotton (*Gossypium hirsutum* L.). *Mol Genet Genomics* 292(6):1267-1280.
<https://doi.org/10.1007/s00438-017-1346-9>

Ali MA, Khan IA, Awan SI, Ali S, Niaz S (2008) Genetics of fibre quality traits in cotton (*Gossypium hirsutum* L.). *Aust J Crop Sci* 2:10-17

Asif M, Mirza JI, Zafar Y (2008) Genetic analysis for fiber quality traits of some cotton genotypes. *Pak J Bot* 40: 1209-1215.

Ardlie KG, Kruglyak L, Seielstad M (2002) Patterns of linkage disequilibrium in the human genome. *Nat Rev Genet* 3, 299–309.

Azhar FM, Naveed M, Ali A (2004) Correlation analysis of seed cotton yield with fiber characteristics in *Gossypium hirsutum* L. *Int J Agri Biol* Vol 6, No. 4.

Azhar FM, Naeem M (2008) Assessment of cotton (*Gossypium hirsutum*) germplasm for combining abilities in fiber traits. *J Agric Soc Sci* 4:129-131

Balcı Ş, Çınar V, Ünay A (2020) A Study on genetic advance and heritability for quantitative traits in cotton (*Gossypium hirsutum* L.). *Adnan Menderes Üniversitesi Ziraat Fakültesi Dergisi* 17 (1), 81-84. DOI: 10.25308/aduziraat.709514

Benjamini Y, Yekutieli D (2005) Quantitative trait loci analysis using the false discovery rate. *Genetics* 171 (2), 783–790. <https://doi.org/10.1534/genetics.104.036699>.

Bradbury PJ, Zhang Z, Kroon DE, Casstevens TM, Ramdoss Y, Buckler ES (2007) TASSEL: software for association mapping of complex traits in diverse samples. *Bioinformatics* 23, 2633–2635.

Baytar AA (2014) Molecular mapping of quantitative trait loci conferring resistance to *Verticillium* wilt in cotton. Master Thesis. Izmir Institute of Technology, Izmir, July 2014.

Baytar AA, Peynircioğlu C, Sezener V, Basal H, Frary A, Frary A, Doğanlar S (2018b) Identification of stable QTLs for fiber quality and plant structure in Upland cotton (*G. hirsutum* L.) under drought stress. *Ind Crop Prod* 124, 776–786. doi: 10.1016/j.indcrop.2018.08.054

- Baytar AA, Peynircioğlu C, Sezener V, Basal H, Frary A, Frary A, Doğanlar S (2018a) Genome-wide association mapping of yield components and drought tolerance-related traits in cotton. *Mol Breed* 38(6). doi:10.1007/s11032-018-0831-0
- Baytar AA, Erdogan O, Frary A, Frary A, Doğanlar S (2017) Molecular diversity and identification of alleles for *Verticillium* wilt resistance in elite cotton (*Gossypium hirsutum* L.) germplasm. *Euphytica* 213, 31 <https://doi.org/10.1007/s10681-016-1787-y>
- Bhagasara VK, Ranwah BR, Meena BL, Khan R (2017) Estimation of GCV, PCV, heritability and genetic gain for yield and its related components in sorghum [*Sorghum bicolor* (L.) Moench]. *Int J Curr Microbiol App Sci* 6(5): 1015-1024. doi: doi.org/10.20546/ijcmas.2017.605.110
- Blenda A, Fang DD, Rami JF, Garsmeur O, Luo F, Lacape JM (2012) A high density consensus genetic map of tetraploid cotton that integrates multiple component maps through molecular marker redundancy check. *PLoS One* 7 (9), e45739.
- Brubaker CL, Bourland FM, Wendel JE (1999) The origin and domestication of cotton. In CW Smith, JT Cothren, eds *Cotton: Origin, History, Technology, and Production*. John Wiley and Sons, Inc., New York, 3-31.
- Cai C, Ye W, Zhang T, Guo W (2014) Association analysis of fiber quality traits and exploration of elite alleles in Upland cotton cultivars/accessions (*Gossypium hirsutum* L.). *J. Integr Plant Biol* 56(1), 51–62. doi:10.1111/jipb.12124
- Chaudhary L, Sindhu A, Kumar M, Kumar R, Saini M (2010) Estimation of genetic divergence among some cotton varieties by RAPD analysis. *J. Plant Breed. Crop Sci* 2(3):039–43.
- Chee P, Draye X, Jiang CX, Decanini L, Delmonte T, Bredhauer R, Smith CW, Paterson AH (2005) Molecular dissection of interspecific variation between *Gossypium hirsutum* and *Gossypium barbadense* (cotton) by a backcross-self approach: III. Fiber length. *Theor Appl Genet* 111: 772-781.
- Dever JGK (1986) The influence of cotton fiber properties on the effectiveness of lint cleaning in ginning (Doctoral thesis). Graduate Faculty of Texas Tech University, USA, 1986
- Dong C, Wang J, Yu Y, Ju L, Zhou X, Ma X, ... Zhang T (2019) Identifying functional genes influencing *Gossypium hirsutum* fiber quality. *Front Plant Sci* 9,1968. doi: 10.3389/fpls.2018.01968
- Doyle JJ, Doyle JL (1987) A rapid DNA isolation procedure for small quantities of fresh leaf tissue. *Phytochem Bull* 19, 11–15.
- Dhivya R, Amalabalu P, Pushpa R, Kavithamani D (2014) Variability, heritability and genetic advance in upland cotton (*Gossypium hirsutum* L.). *Afr J Plant Sci* 8(1), 1–5. doi:10.5897/ajps2013.1099

- Earl DA, vonHoldt BM (2012) STRUCTURE HARVESTER: a website and program for visualizing STRUCTURE output and implementing the Evanno method. *Conserv Genet Resour* 4, 359–361.
- Ersoz ES, Yu J, Buckler ES (2007) Applications of linkage disequilibrium and association mapping in crop plants. *Genomics-assisted crop Improvement-Springer Dordrecht* 97–120.
- 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
- 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* 191(3), 391–401. doi:10.1007/s10681-013-0886-2
- Fang DD, Jenkins JN, Deng DD, McCarty JC, Li P, Wu J (2014) Quantitative trait loci analysis of fiber quality traits using a random-mated recombinant inbred population in Upland cotton (*Gossypium hirsutum* L.). *BMC Genomics* 15, 397. <https://doi.org/10.1186/1471-2164-15-397>
- Fang DD, Yu JZ (2012) Addition of 455 Microsatellite marker loci to the high-density *Gossypium hirsutum* tm-1 x g. barbadense 3-79 genetic map. *J Cotton Sci* 16:229-248
- Farooq J, Anwar M, Riaz M, Farooq A, Mahmood A, Shahid MTH, Rafiq MS, Ilahi F (2014) Correlation and path coefficient analysis of earliness, fiber quality and yield contributing traits in cotton (*Gossypium hirsutum* L.). *JAPS* 24(3), 781–790.
- Frelichowski JE, Palmer MB, Main D, Tomkins JP, Cantrell RG, Stelly DM, Yu J, Kohel RJ, Ulloa M (2006) Cotton genome mapping with new microsatellites from Acala “Maxxa” BAC-ends. *Mol. Genet. Genomics* 275:479–491. doi: [10.1007/s00438-006-0106-z](https://doi.org/10.1007/s00438-006-0106-z)
- Gipson JR, Joham HE (1969) Influence of night temperature on growth and development of cotton (*Gossypium hirsutum* L.) III. fiber elongation. *Crop Sci* 9: 127–129
- Gordon S, Hsieh YL (2007) (Eds.), *Cotton Science and Technology*. CRC Press
- Grover CE, Yoo MJ, Lin M, Murphy MD, Harker DB, Byers RL, Lipka AE, Hu G, Yuan D, Conover JL, Udall JA, Paterson AH, Gore MA, Wendel JF (2020) Genetic analysis of the transition from wild to domesticated cotton (*G. hirsutum* L.). *G3- Genes Genomes Genet* g3.400909.2019. doi:10.1534/g3.119.400909
- Hanson RG, Ewing EC, Ewing Jr EC (1956) Effect of environmental factors on fiber properties and yield of deltapine cottons. *Agron J* 48: 546–581, 1956.
- Harem E (2014) Türkiye pamuk çeşit kataloğu. Tarımsal Araştırmalar ve Politikalar Genel Müdürlüğü Pamuk Araştırma İstasyonu Müdürlüğü, Yayın No:74 Nazilli

- Hinze L, Dever J, Percy R (2012) Molecular variation among and within improved cultivars in the U.S. Cotton Germplasm Collection. *Crop Sci* 52:222–230.
- Huang C, Shen C, Wen T, Gao B, Zhu D, Li X, ... Lin Z (2018) SSR-based association mapping of fiber quality in upland cotton using an eight-way MAGIC population. *Mol Genet Genomics* 293, 793. doi: 10.1007/s00438-018-1419-4
- Huttley GA, Smith MW, Carrington M, O'Brien SJ (1999) A scan for linkage disequilibrium across the human genome. *Genetics* 152:1711–1722
- Iqbal MA, Rahman M (2017) Identification of marker-trait associations for lint traits in cotton. *Front Plant Sci* 8:86. doi: 10.3389/fpls.2017.00086
- Jarwar AH, Wang X, Wang L, Ma Q, Fan S (2018) Genetic advancement, variability and heritability in upland cotton (*Gossypium hirsutum* L.). *J Environ Agric Sci* 16: 24-31.
- Jenkins JN (2003) Cotton. Traditional Crop Breeding Practices: An Historical Review to Serve As a Baseline for Assessing the Role of Modern Biotechnology. OECD pp.61–70.
- Jia YH, Sun JL, Wang XW, Zhou ZL, Pan ZE, He SP, Pang BY, Wang LR, Du XM (2014) Molecular diversity and association analysis of drought and salt tolerance in *Gossypium hirsutum* L. germplasm. *J Integr Agric* 13:1845–1853
- Jia Y, Sun X, Pan Z, Wang X, He S, Xiao S, Shi W, Zhou Z, Pang B, Wang L, Liu J, Ma J, Du X, Zhu J (2014) Association mapping for epistasis and environment interaction of yield traits in 323 cotton cultivars under 9 different environments. *PLoS One* 9:1-8, e95882. DOI: 10.1371/journal.pone.0095882
- Kalabek NA, Babaarslan O (2016) Fiber Selection for the Production of Nonwovens. *Non-Woven Fabrics*. doi:10.5772/61977
- Karademir E, Karademir C, Ekininci R, Gencer O (2010) Relationship between yield, fiber length and other fiber-related traits in advanced cotton strains. *Not Bot Horti Agrobo* 38(3), 111-116. <https://doi.org/10.15835/nbha3834889>
- Kearney TH (1912). Lint Index and lint percentage in cotton breeding. *J Hered* 7(1) pp 25–29, <https://doi.org/10.1093/jhered/os-7.1.25>
- Khalid MA, Malik TA, Fatima N, Shakeel A, Karim I, Arfan M, Merrium S, Khanum P (2018) Correlation for economic traits in Upland cotton. *ASAG* 2.10 59-62.
- Khokhar ES, Shakeel A, Maqbool MA, Anwar MW, Tanveer Z, Irfan MF (2017) Genetic study of cotton (*Gossypium hirsutum* L.) genotypes for different agronomic, yield and quality traits. *Pak J Agric Res* 30(4): 363-372.

- Kruglyak L (1999) Prospects for whole-genome linkage disequilibrium mapping of common disease genes. *Nat Genet* 22, 139–144.
- Kruijer W, Boer MP, Malosetti M, Flood PJ, Engel B, Kooke R, ... Eeuwijk FA van (2014) Marker-based estimation of heritability in immortal populations. *Genetics* 199(2), 379–398.
doi:10.1534/genetics.114.167916
- Kuzay S**, Conaty PH, Palkovic A, Gepts P (2020) Is the USDA core collection of common bean representative of genetic diversity of the species, as assessed by SNP diversity? *Crop Sci* Doi: 10.1002/csc2.20032
- 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 Breed* 19: 45–58.
- Lee JA (1984) Cotton as a world crop. In: Kohel RJ, Lewis CL (eds) *Cotton agronomy monograph*, vol 24. Crop Science Society of America, Madison, pp 1–25
- Li X, Wu M, Liu G, Pei W, Zhai H, Yu J, Zhang H, Yu S (2017) Identification of candidate genes for fiber length quantitative trait loci through RNASeq and linkage and physical mapping in cotton. *BMC Genomics* 18(427):3-12. DOI: 10.1186/s12864-017-3812-5
- Liang Q, Hu C, Hua H, Li Z, Hua J (2013) Construction of a linkage map and QTL mapping for fiber quality traits in upland cotton (*Gossypium hirsutum* L.). *Chinese science bulletin Kexue tongbao*. 58(26):3233-3243.
- Liu R, Gong J, Xiao X, Zhang Z, Li J, Liu A, Lu Q, Shang H, Shi Y, Ge Q, Iqbal MS, Deng X, Li S, Pan J, Duan L, Zhang Q, Jiang X, Zou X, Hafeez A, Chen Q, Geng H, Gong W, Yuan Y (2018) GWAS Analysis and QTL identification of fiber quality traits and yield components in Upland cotton using enriched high-density SNP markers. *Front Plant Sci* (2018) 9:1067.
- Ma L, Wang Y, Ijaz B, Hua J (2019) Cumulative and different genetic effects contributed to yield heterosis using maternal and paternal backcross populations in Upland cotton. *Sci Rep* 9(1). doi:10.1038/s41598-019-40611-9
- Mackay TF (2009) Q&A: Genetic analysis of quantitative traits. *J Biol* 8(3), 23. doi:10.1186/jbiol133
- May OL (1999) Genetic variation in fiber quality. In: Basra AS, editor. *Cotton fibers: developmental biology, quality improvement, and textile processing*. NewYork: Haworth Press 183–229
- May OL, Lege KE (1999) *Development of the World Cotton Industry*. C.W. Smith, J.T. Cothren (Eds.). John Wiley & Sons. pp: 70-79.
- Mei H, Zhu X, Zhang T (2013) Favorable QTL alleles for yield and its components identified by association mapping in Chinese Upland cotton cultivars. *PLoS One* 8(12): e82193. DOI:

Mendez-Natera JR, Rondon A, Hernandez J, Merazo-Pinto JF (2012) Genetic study in Upland cotton. Three genetic parameters, correlation and path analysis. *SABRAO J Breed Genet* 44:112-128.

Miglani GS (2002) *Advanced Genetics*. CRC Press.

Moore JF (1996) Cotton Classification and Quality. p. 51–57. In E.H. Glade Jr., L.A. Meyer, and H. Stults (ed.) *The cotton industry in the United States*. USDA-ERS Agric Econ Rep 739. U.S. Gov. Print. Office, Washington, DC.

Naoumkina M, Thyssen GN, Fang DD, Jenkins JN, McCarty JC, Florane CB (2019) Genetic and transcriptomic dissection of the fiber length trait from a cotton (*Gossypium hirsutum* L.) MAGIC population. *BMC Genomics* 20(1). doi:10.1186/s12864-019-5427-5

Nie X, Huang C, You C, Li W, Zhao W, Shen C, Zhang B, Wang H, Yan Z, Dai B, Wang M, Zhang Z, Lin Z (2016) Genome-wide SSR-based association mapping for fiber quality in nation-wide upland cotton inbred cultivars in China. *BMC Genomics* 17:352. DOI: 10.1186/s12864-016-2662-x

Nordborg M, Borevitz JO, Bergelson J, Berry CC, Chory J, Hanblad J, Kreitman M, Maloof JN, Noyes T, Oefner PJ, Stahl EA, Weigel D (2002) The extent of linkage disequilibrium in *Arabidopsis thaliana*. *Nat Genet* 30:90–193. doi:10.1038/ng813

Park YH, Alabady MS, Ulloa M, Sickler B, Wilkins TA, Yu J, Stelly DM, Kohel RJ, El-Shihy OM, Cantrell RG (2005) Genetic mapping of new cotton fiber loci using EST-derived microsatellites in an interspecific recombinant inbred line cotton population. *Mol Genet Genomics* 274:428–441. doi: 10.1007/s00438-005-0037-0

Peakall R, Smouse PE (2006) GENALEX 6: Genetic analysis in Excel. Population genetic software for teaching and research. *Mol Ecol Notes* 6, 288–295.

Peakall R, Smouse PE (2012) GenAEx 6.5: genetic analysis in Excel. Population genetic software for teaching and research – an update. *Bioinformatics* 28, 2537–2539.

Paterson AH, Smith RH (1999) Future horizons: biotechnology for cotton improvement. In: Smith, C.W., Cothren, J.T. (Eds.), *Cotton: Origin, History, Technology, and Production*. Wiley, Inc., New York, pp. 415–432.

Percy RG, Cantrell RG, Zhang J (2006) Genetic variation for agronomic and fiber properties in an introgressed recombinant inbred population of cotton. *Crop Sci* 46(3), 1311. doi:10.2135/cropsci2005.08-0284

Perrier X, Jacquemoud-Collet JP (2006) **DARwin** software <http://darwin.cirad.fr/>

- Pritchard JK, Stephens M, Donnelly P (2000) Inference of population structure using multilocus genotype data. *Genetics* 155, 945–959.
- Pope OA, Ware JO (1945) Effect of variety, location, and season on oil, protein, and fuzz of cottonseed and on fiber properties of lint. United States Department of Agriculture, Tech Bull No 903.
- Qin H, Chen M, Yi X, Bie S, Zhang C, Zhang Y, Lan J, Meng Y, Yuan Y, Jiao C (2015) Identification of associated SSR markers for yield component and fiber quality traits based on frame map and Upland cotton collections. *PLoS ONE* 10(1): e0118073. doi: 10.1371/journal.pone.0118073
- Rakshit A, Rakshit S, Singh J, Chopra SK, Balyan HS, Gupta PK, Bhat SR (2010) Association of AFLP and SSR markers with agronomic and fibre quality traits in *Gossypium hirsutum* L. *J Genet* 89, 155–162]
- Reynolds EB, Killough DT (1933) The effect of fertilizers and rainfall on the length of the cotton fiber. *J Am Soc Agron* 25:756-764.
- Saeed A, Elçi E (2017) Microsatellite-based characterization of cotton genotypes for *Verticillium* wilt and fiber quality traits. *Turk J Biochem* 43(3), 277–288. doi:10.1515/tjb-2017-0169
- Seyoum M, Du XM, He SP, Jia, YH, Pan Z, Sun JL (2018) Analysis of genetic diversity and population structure in Upland cotton (*Gossypium hirsutum* L.) germplasm using simple sequence repeats. *J Genet* 97(2), 513–522. doi:10.1007/s12041-018-0943-7
- Silva-Junior OB, Grattapaglia D (2015) Genome-wide patterns of recombination, linkage disequilibrium and nucleotide diversity from pooled resequencing and single nucleotide polymorphism genotyping unlock the evolutionary history of *Eucalyptus grandis*. *New Phytol* 208(3), 830–845. doi:10.1111/nph.13505
- Shakeel A, Ahmad S, Naeem M, Tahir MHN, Saleem MF, Freed S, Nazeer W (2012) Evaluation of *Gossypium hirsutum* L. genotypes for combining ability studies of yield and quality traits. *Iğdır Univ J Inst Sci Tech* 2(1): 67-74
- Shang L, Abduweli A, Wang Y, Hua J (2016) Genetic analysis and QTL mapping of oil content and seed index using two recombinant inbred lines and two backcross populations in upland cotton. *Plant Breed* 135:224–231.
- Shen C, Wang N, Huang C, Wang M, Zhang X, Lin Z (2019) Population genomics reveals a fine-scale recombination landscape for genetic improvement of cotton. *Plant J* 99 (3) 494-505
- Shermotov SE, Buriev ZT, Makamov A, Shopultov U, Kushanov FN, Mavlonov GT, Abdurakhmonov (2010) Proceedings of International Cotton advisory committee 69th Plenary Meeting; 20-25th September 2010; Lubbock, Texas USA

- Stich B, Melchinger AE, Frisch M, Maurer HP, Heckenberger M, Reif JC (2005) Linkage disequilibrium in European elite maize germplasm investigated with SSRs. *Theor Appl Genet* 111:723–730. doi:10.1007/s00122-005-2057-x
- Storey JD (2002) A direct approach to false discovery rates. *J R Stat Soc Ser B Stat. Methodol.* 64, 479–498.
- Stroman GN (1949) Variability and correlation in a cotton breeding program. *J Agric Res* 78:353–364.
- Tang B, Jenkins JN, Watson CE, McCarty JC, Creech RG (1996) Evaluation of genetic variances and correlation for yield and fiber traits among cotton F2 hybrid populations. *Euphytica* 91:315-322.
- Tyagi P, Gore MA, Bowman DT, Campbell BT, Udall JA, Kuraparthi V (2014) Genetic diversity and population structure in the US Upland cotton (*Gossypium hirsutum* L.). *Theor Appl Genet* 127 283-295. 10.1007/s00122-013-2217-3.
- Ulloa M, Meredith JWR (2000) Genetic linkage map and QTL analysis of agronomic and fiber quality traits in an intraspecific population. *J Cotton Sci* (2000) 4:161-170.
- USDA (2020) <https://apps.fas.usda.gov/psdonline/app/index.html#/app/advQuery> (accessed 16 June 2020)
- USDA (2018) (The United States Department of Agriculture), Agricultural Marketing Service. The Classification of Cotton Agricultural Handbook, 566.
- Wang B, Liu L, Zhang D, Zhuang Z, Guo H, Qiao X, Wei L, Rong J, May OL, Paterson AH, Chee PW (2016) A genetic map between *Gossypium hirsutum* and the Brazilian endemic *G. mustelinum* and its application to QTL mapping. *G3- Genes Genomes Genet* 6 (6) 1673-1685, <https://doi.org/10.1534/g3.116.029116>
- Wang B, Guo W, Zhu X, Wu Y, Huang N, Zhang T (2007) QTL mapping of yield and yield components for elite hybrid derived-RILs in upland cotton. *J Genet Genomics* 34 (1), 35–45. Doi: 10.1016/S1673-8527(07)60005-8.
- Wang B, Draye X, Zhuang Z, Zhang Z, Liu M, Lubbers EL, Jones D, May OL, Paterson AH, Chee PW (2017) QTL analysis of cotton fiber length in advanced backcross populations derived from a cross between *Gossypium hirsutum* and *G. mustelinum*. *Theor Appl Genet* 130:1297–1308. Doi. [10.1007/s00122-017-2889-1](https://doi.org/10.1007/s00122-017-2889-1)
- Wang H, Jin X, Zhang B, Shen C, Lin Z (2015) Enrichment of an intraspecific genetic map of upland cotton by developing markers using parental RAD sequencing. *DNA research: an international journal for rapid publication of reports on genes and genomes.* 2015 Feb 5.
- Wang M, Tu L, Yuan D, Zhu D, Shen C, Li J (2019) et al. Reference genome sequences of two cultivated allotetraploid cottons, *Gossypium hirsutum* and *Gossypium barbadense*. *Nat Genet* 51, 224–229. doi:

10.1038/s41588-018-0282-x

Wan Q, Zhang Z, Hu M, Chen L, Liu D, Chen X, ... Zheng J (2007) T1 locus in cotton is the candidate gene affecting lint percentage, fiber quality and spiny bollworm (*Earias* spp.) resistance. *Euphytica* 158(1-2), 241–247. doi:10.1007/s10681-007-9446-y

Weller JI, Song JZ, Heyen DW, Lewin HA, Ron M (1998) A new approach to the problem of multiple comparisons in the genetic dissection of complex traits. *Genetics* 150: 1699–1706.

Wen Z, Tan R, Yuan J, Bales C, Du W, Zhang S, ... Wang D (2014) Genome-wide association mapping of quantitative resistance to sudden death syndrome in soybean. *BMC Genomics* 15, 809
<https://doi.org/10.1186/1471-2164-15-809>

Wu J, Gutierrez OA, Jenkins JN, McCarty JC, Zhu J (2009) Quantitative analysis and QTL mapping for agronomic and fiber traits in an RI population of upland cotton. *Euphytica* 165:231–245. doi:
[10.1007/s10681-008-9748-8](https://doi.org/10.1007/s10681-008-9748-8)

WWO (2013) <https://www.worldweatheronline.com/aydin-weather-history/aydin/tr.aspx> (accessed 28 July 2020)

Yaqoob M, Fiaz S, Ijaz B (2016) Correlation analysis for yield and fiber quality traits in upland cotton. *Commun Plant Sci* 6: 55-60.

Yu J, Zhang K, Li S, Yu Y, Zhai H, Wu M, Li X, Fan S, Song M, Yang D, Li Y, Zhang J (2013) Mapping quantitative trait loci for lint yield and fiber quality across environments in a *Gossypium hirsutum* × *Gossypium barbadense* backcross inbred line population. *Theor Appl Genet* 126:275–287.

Yu JZ, Ulloa M, Hoffman SM, Kohel RJ, Pepper AE, Fang DD, Percy RG, Burke JJ (2014) Mapping genomic loci for cotton plant architecture, yield components, and fiber properties in an interspecific (*Gossypium hirsutum* L. × *G. barbadense* L.) RIL population. *Mol Genet Genomics*:MGG.

Zeng LH, Meredith WR, Gutierrez OA, Boykin DL (2009) Identification of associations between SSR markers and fiber traits in an exotic germplasm derived from multiple crosses among *Gossypium* tetraploid species. *Theor Appl Genet* 119:93-103. <https://doi.org/10.1007/s00122-009-1020-7>

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

Zhang K, Zhang J, Ma J, Tang SY, Liu DJ, Teng ZH, Liu DX, Zhang ZS (2012) Genetic mapping and quantitative trait locus analysis of fiber quality traits using a three-parent composite population in upland cotton (*Gossypium hirsutum* L.). *Mol Breed* 29(2) 335-348

Zhang SW, Zhu XF, Feng LC, Gao X, Yang B, Zhang TZ, Zhou BL (2016) Mapping of fiber quality QTLs reveals useful variation and footprints of cotton domestication using introgression lines. *Sci Rep*

6:31954.

Zhang T, Qian N, Zhu X, Chen H, Wang S, Mei H, Zhang Y (2013) Variations and Transmission of QTL Alleles for Yield and Fiber Qualities in Upland Cotton Cultivars Developed in China. PLoS ONE 8(2): e57220. doi: 10.1371/journal.pone.0057220

Zhao Y, Wang H, Chen W, Li Y (2014) Genetic Structure, Linkage Disequilibrium and Association Mapping of *Verticillium* Wilt Resistance in Elite Cotton (*Gossypium hirsutum* L.) Germplasm Population. PLoS ONE 9(1): e86308. doi: 10.1371/journal.pone.0086308

Tables

Tables 1-3 are available in the Supplementary Files.

Figures

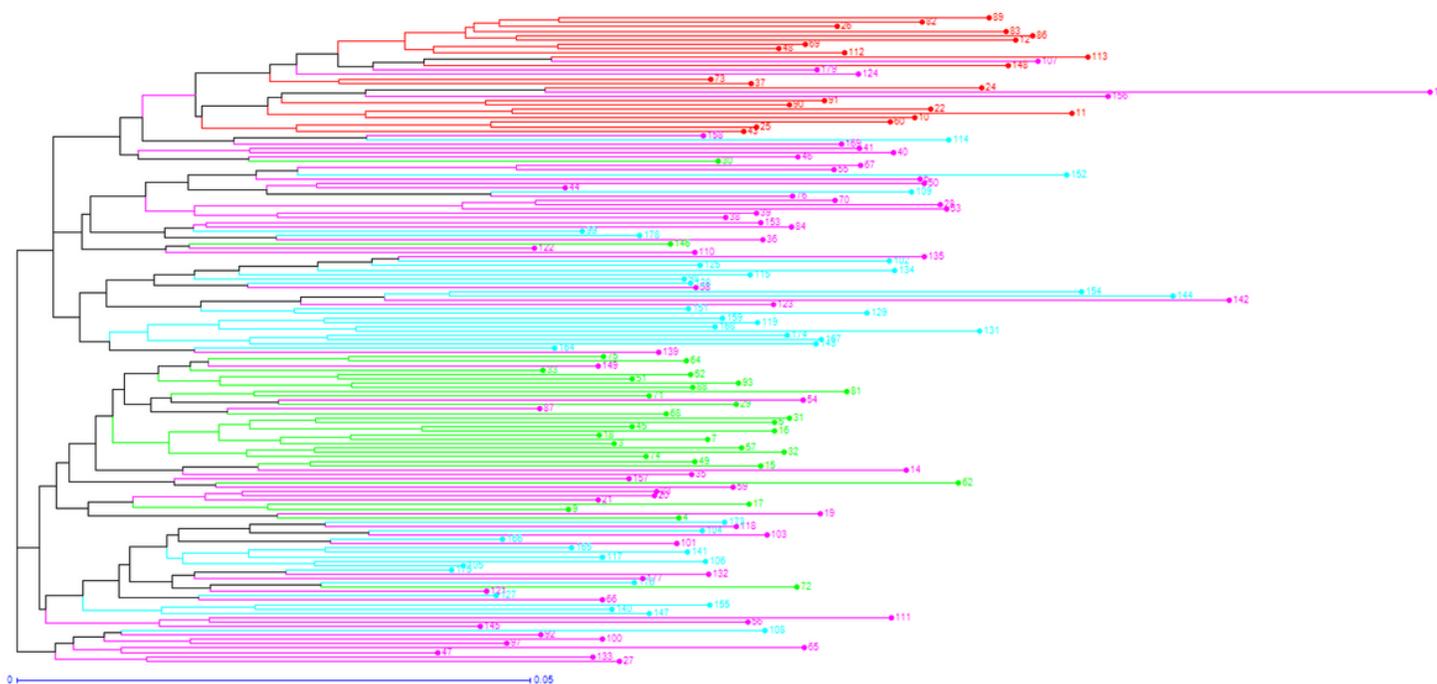


Figure 1

Genetic diversity of 157 *G. hirsutum* multi-parent recombinant inbred lines (RILs). Colors are based on the results of structure analysis. Red represents sub-group 1 (Q1), green represents sub-group 2 (Q2) and blue represents sub-group 3 (Q3). “Admixed” individuals are in pink

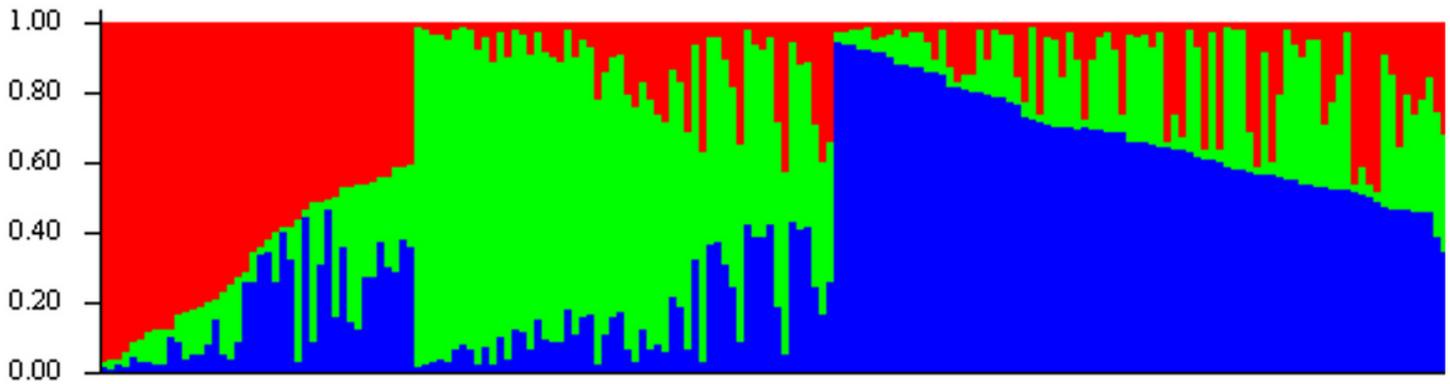


Figure 2

Bar plot demonstration of population structure for K = 3

Seed cotton yield (kg ha⁻¹)

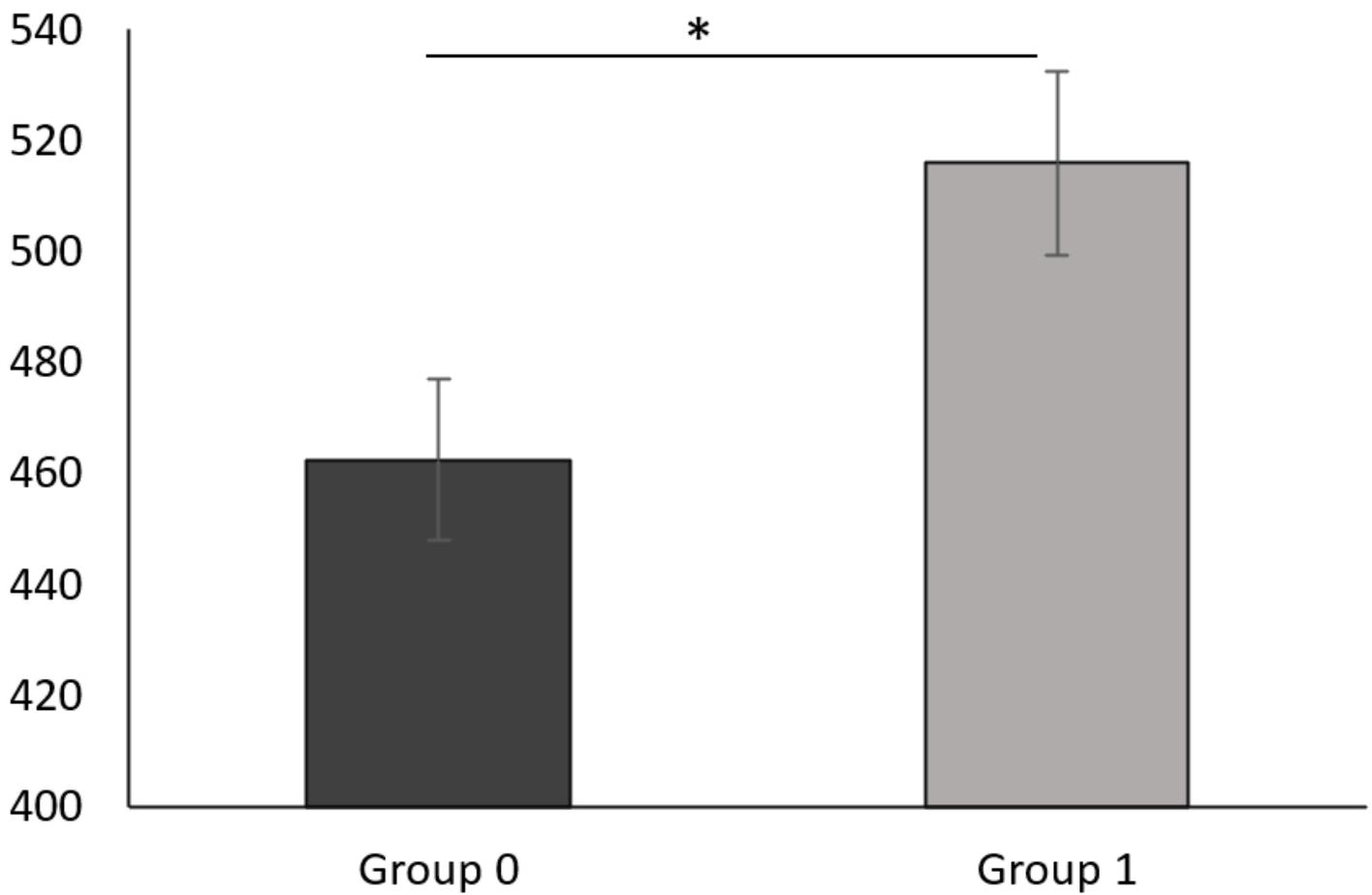


Figure 3

Significantly higher yield in Group 1 which had the allele (BNL2495215) linked to improved seed cotton yield as determined by AMOVA. Group 0 represents individuals without the favorable allele, BNL2495215

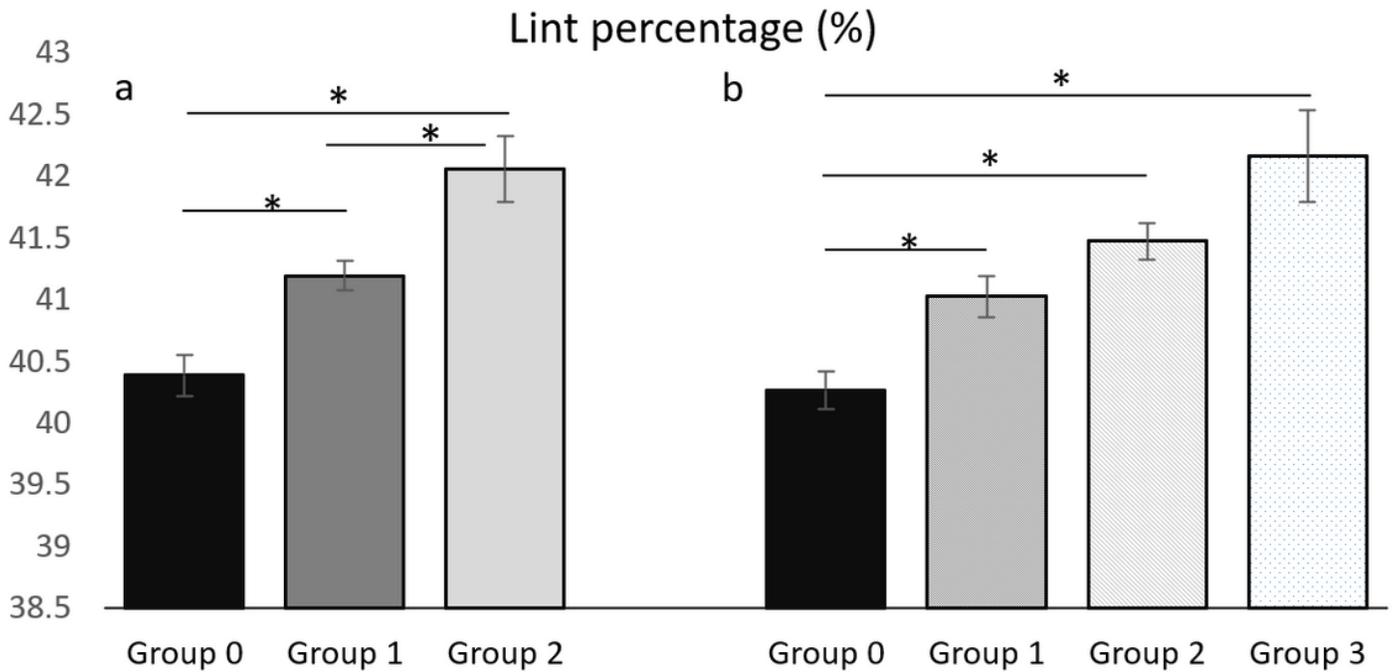


Figure 4

(a) Significant increase in lint percentage as the number of positive alleles increased. (b) Significant decrease in lint percentage as the number of adverse alleles increased with no significant differences between alleles

Fiber length (mm)

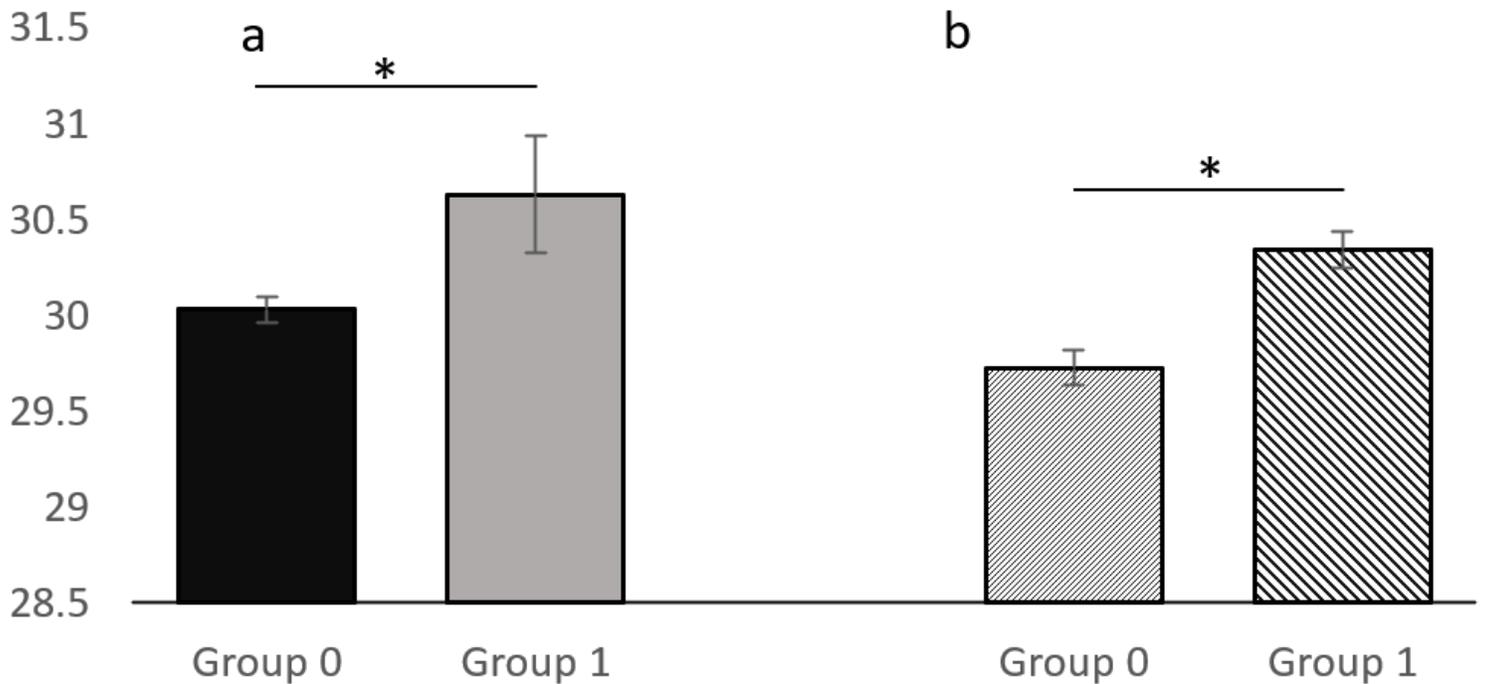


Figure 5

(a) Significantly longer fibers in Group 1 which contained individuals having the favorable allele (BNL3545181) and lacking two adverse alleles (BNL3545170 and BNL1495231) compared to the rest of the population (Group 0). (b) Significantly longer fibers in Group 1 which contained individuals with either presence of the positive allele or absence of any negative allele as compared to the result of the population

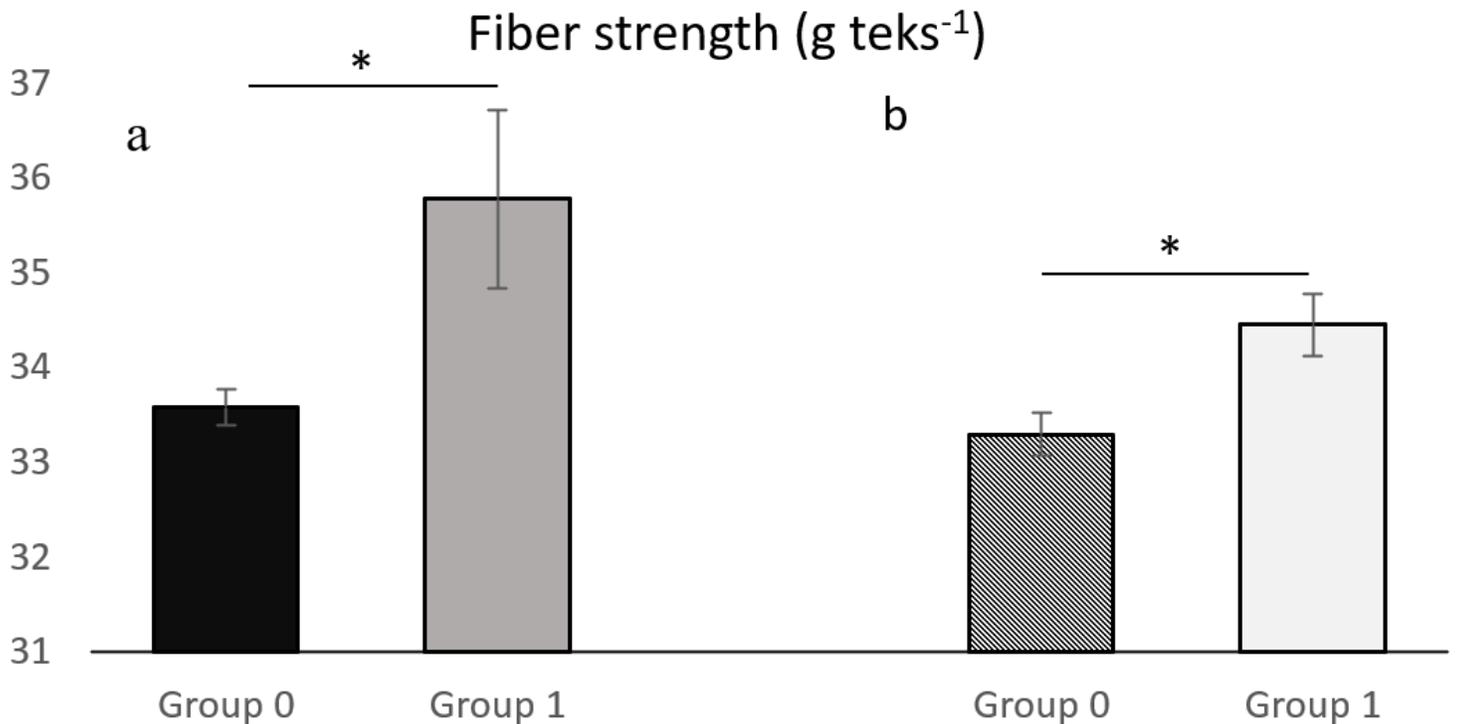


Figure 6

(a) Significantly stronger fibers in Group 1 which contained individuals with two favorable alleles for fiber strength (BNL1495242 and BNL3545120) compared to the rest of the population (Group 0). (b) Significantly stronger fibers in Group 1 which contained individuals that had any two favorable alleles (BNL1495242 or BNL3545120) as compared to the rest of the individuals (Group 0)

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- [Tables.docx](#)
- [cottonRILSupplementaryInformationFigures.pdf](#)
- [cottonRILSupplementaryInformationTables.xlsx](#)