Mol Breeding (2018) 38: 74
https://doi.org/10.1007/s11032-018-0831-0

# Genome-wide association mapping of yield components and drought tolerance-related traits in cotton

**Asena Akkose Baytar · Ceng Peynircioğlu · Volkan Sezener · Hüseyin Basal · Anne Frary · Amy Frary · Sami Doğanlar**

Received: 11 April 2017 / Accepted: 1 May 2018 / Published online: 24 May 2018


**Abstract** Drought causes serious yield losses in cotton production throughout the world. Association mapping allows identification and localization of the genes controlling drought-related traits which will be helpful in cotton breeding. In the present study, genetic diversity analysis and association mapping of yield and drought traits were performed on a panel of 99 upland cotton genotypes using 177 SSR (simple sequence repeat) markers. Yield parameters and drought tolerance-related traits were evaluated for two seasons under two watering regimes: water-stressed and well-watered. The traits included seed cotton yield (SCY), lint yield (LY), lint percentage (LP), water-use efficiency (WUE), yield potential (YP), yield reduction (YR), yield index (YI), drought sensitivity index (DSI), stress tolerance index (STI), harmonic mean (HM), and geometric mean productivity (GMP). The genotypes with the least change in seed cotton yield under drought stress were Zeta 2, Delcerro, Nazilli 87, and DAK 66/3 which were also the most water-use efficient cultivars. The average genetic diversity of the panel was 0.38. The linkage disequilibrium decayed relatively rapidly at 20–30 cM ($r^2 \geq 0.5$). We identified 30 different SSR markers associated with the traits. Fifteen and 23 SSR markers were linked to the traits under well-watered and water-stress conditions, respectively. To our knowledge, most of these quantitative yield and drought tolerance-associated loci were newly identified. The genetic diversity and association mapping results should facilitate the development of drought-tolerant cotton lines with high yield in molecular breeding programs.

**Keywords** Association mapping · *Gossypium hirsutum* L. · Simple sequence repeat · Drought tolerance

**Electronic supplementary material** The online version of this article (https://doi.org/10.1007/s11032-018-0831-0) contains supplementary material, which is available to authorized users.

A. A. Baytar · A. Frary · S. Doğanlar (✉)
Department of Molecular Biology and Genetics, Izmir Institute of Technology, Izmir, Turkey
e-mail: samidoganlar@iyte.edu.tr

C. Peynircioğlu
Özaltın Agricultural Enterprises Industry and Commerce Inc., Aydın, Turkey

V. Sezener
Nazilli Cotton Research Institute, Nazilli, Aydın, Turkey

H. Basal
Department of Field Crops, Faculty of Agriculture, Adnan Menderes University, Koçarlı, Aydın, Turkey

A. Frary
Department of Biological Sciences, Mount Holyoke College, South Hadley, MA 01075, USA

## Introduction

Humans have cultivated cotton (*Gossypium spp.*) for thousands of years. Every part of the cotton plant is useful. The seeds yield a high-quality oil, and the meal and hull byproducts of oil extraction serve as fodder and fertilizer. Paper can be made from the fibrous stems and the root bark is an ingredient in herbal medicine.

However, the plant's most valuable commodity is its seed fibers which are the main raw material of the world's textile industry. Demand for cotton is on the rise as industrial development increases standards of living worldwide. The top five cotton-producing countries are India, China, the USA, Pakistan, and Brazil; together they produce 76% of the world's cotton. While Turkey is eighth in cotton production, contributing only 3%, it ranks in the top three in terms of yield (approximately 1700 kg $ha^{-1}$ in 2016) (USDA-FAS 2016).

A wide range of biotic and abiotic stresses causes serious yield losses in cotton including diseases such as *Verticillium* and *Fusarium* wilts; insect pests such as aphids, armyworms, and cutworms; and adverse soil conditions such as drought, salinity, and mineral toxicity (Saeed et al. 2011). Drought limits the movement of water from the soil into the root, thereby resulting in decreased osmotic potential of the plant. Plants have various adaptations (drought tolerance mechanisms) to overcome this abiotic stress. The three main mechanisms are stomatal closure to reduce transpiration and thereby sustain internal water potential (drought avoidance), early blooming and early maturity to shorten the life cycle (drought escape), and coping with water stress without altering physiological or developmental features (drought tolerance) (Iqbal et al. 2013).

Although cotton is known to have relatively good drought tolerance, its response to water stress depends on the developmental stage of the plant and the degree and timing of dry periods. Water stress can cause flower bud (square) shedding, reduced fiber elongation, altered fiber wall thickness, and reduced boll size, all of which result in poor fiber quality and decreased total cotton yield. Breeding for drought tolerance presents challenges. Conventional breeding of drought-tolerant cotton cultivars has been hampered by the complex genetics of drought tolerance mechanisms, a lack of adequate genetic variability for the trait in the crop and the susceptibility of drought-related traits to environmental conditions (Levi et al. 2009). Nevertheless, many breeders have applied classical breeding approaches to achieve water-stress adaptation in cotton (Levi et al. 2009). One strategy has been to generate cultivars with high yield under optimum irrigated conditions with the expectation that the cultivar will produce a relatively reasonable yield under limited irrigation (Quisenberry et al. 1980). A second method has entailed direct selection for high yield under water-stress conditions (Rosielle and Hamblin 1981). A third approach has been to start with a drought-tolerant line and to improve its yield (Saranga et al. 2004).

The development of molecular markers has provided new avenues for improvement of quantitative traits using a combination of molecular and traditional breeding methods. The genetic factors or quantitative trait loci (QTL) underlying traits of interest can be identified with DNA markers and an appropriate plant population. Markers linked with drought tolerance or yield can then serve as tools for rapid and efficient marker-assisted selection (MAS) in cotton (Shen et al. 2006).

Many QTL studies have examined morphological characters, fiber quality, and productivity traits in cotton (for example, Liang et al. 2014; Sun et al. 2012; Wu et al. 2009; Mei et al. 2004). However, fewer QTL analyses have looked at yield and physiological parameters under both water-limited and irrigated conditions. The most notable studies were performed using an interspecific population generated from a cross between inbred lines *G. hirsutum* cv. Siv'on and *G. barbadense* cv. F-177 (Saranga et al. 2001; Levi et al. 2009). Saranga et al. (2001) examined $F_2$ individuals from the population and found that distinct subsets of the 161 identified QTLs were specific to the degree of water availability. Thus 33 (20%) of the QTLs were detected only under water-limited conditions whereas 13 (8%) QTLs influenced the traits only under well-watered conditions. Levi et al. (2009) used a marker-assisted backcross strategy to generate near-isogenic lines (NILs) in which target QTLs for yield and physiological traits were introgressed from *G. hirsutum* cv. Siv'on into *G. barbadense* cv. F-177. The NILs and parents were tested under well-watered and water-limited treatments to assess the efficiency of marker-assisted selection (MAS) in improving cotton drought tolerance. The NILs displayed the expected phenotypes in many instances, illustrating the success of the marker-based QTL selection strategy. In another work, Saeed et al. (2011) mapped physiological and morphological traits in an $F_2$ intraspecific population derived from *G. hirsutum* cv. FH-901 (drought sensitive) and *G. hirsutum* cv. RH-510 (drought tolerant) under both well-irrigated and water-limited conditions. A total of seven QTLs were detected: three under the water-stress regime only and two under the well-watered regime only. The results of these QTL studies suggest that distinct sets of genetic loci control cotton productivity and physiological quality under different conditions of water availability. Combining alleles from these

independent loci into a single genotype could possibly produce a line adapted to both conditions. However given the number of QTLs influencing key cotton traits under water-limited conditions, breeding for drought tolerance remains a daunting task.

In the present study, we performed association mapping to identify QTLs controlling yield and drought tolerance traits for two seasons under both water-limited and irrigated conditions. A panel of 99 upland cotton genotypes (mostly cultivars used in commercial production) was screened with 177 simple sequence repeat (SSR) markers. Our findings should be useful for developing drought-resistant cotton cultivars by marker-assisted selection.

## Materials and methods

### Plant material

A germplasm panel composed of 99 upland cotton genotypes (*G. hirsutum* L.) (Table S1) was provided by Nazilli Cotton Research Station (Aydın, Turkey). The panel consisted of cultivars bred and registered in Turkey and those developed elsewhere. The genotypes of the panel were selected based on their high geometric mean productivity and low drought sensitivity index as assessed by a previous agro-morphological analysis under drought stress (Sezener et al. 2015).

### Phenotypic evaluation

Field experiments were performed during the 2011 and 2012 growing seasons at the Agricultural Research Station of Adnan Menderes University (ADU) and at Özaltın Agricultural Enterprises Industry and Commerce Inc. (OAE), both of which are in Kocarli, Aydin, Turkey. The region has sandy, loamy soil with average annual precipitation of 657 mm. The field capacity (water content) and wilting point of the experimental fields ranged from 20.3 to 27.6% and from 7.2 to 9.7%, respectively, at ADU; and from 12.7 to 14.1% and from 4.1 to 5.8%, respectively, at OAE. The 99 cotton genotypes (94 cultivars and five controls: BA 119, Carmen, Claudia, GSN 12, and Sahin 2000) were planted at 0.70-m row width and 0.20-m spacing between individuals on 19 May 2011 and 3 May 2012. Each genotype occupied a single 12-m row with four replications in an augmented experimental design. Two watering regimes were applied using drip irrigation: well-watered (100%, full irrigation) and water-limited (50%, deficit irrigation). Both treatments were irrigated when 50% of available soil moisture was consumed in the 1.20-m root zone in the well-watered treatment. The full irrigation treatment received about 626 mm while the water-limited treatment received 313 mm water during the growing period. Soil water content of the plots was measured using the gravimetric method. Before planting, a compound fertilizer (NPK 15-15-15) was applied at a rate of 60 kg $ha^{-1}$. Additional nitrogen was applied before the first irrigation in the form of 33% ammonium nitrate. Hand harvesting was conducted on 29 September 2011 and 14 September 2012.

Yield traits and drought-related parameters were measured under both well-watered (control) and water-limited (water stress) field conditions. The traits included seed cotton yield (SCY) (kg $ha^{-1}$), lint yield (LY) (kg $ha^{-1}$), lint percentage (LP) (%), and water-use efficiency (WUE). In addition, seven parameters were calculated for each genotype by using combined data from both locations and both well-watered and water-limited trials to measure the effect of drought: yield potential (YP), yield reduction (YR) (%), yield index (YI), drought sensitivity index (DSI), stress tolerance index (STI), harmonic mean (HM), and geometric mean productivity (GMP).

Seed cotton yield was calculated as total weight of seed cotton (kg) $ha^{-1}$. Lint yield was measured as total weight of lint (kg) $ha^{-1}$. Lint percentage [g lint / (g lint + g seed) × 100%] was determined after ginning the cotton lint using a roller gin. Water-use efficiency was calculated using the formula: yield ($Y$) in kg $ha^{-1}$ / total applied water (mm) (Howell and Hiler 1975). Yield potential was calculated as ($\hat{Y}s + \hat{Y}p$) / 2, where $\hat{Y}s$ and $\hat{Y}p$ are the means of all genotypes under well-watered and water-limited conditions, respectively (Rosielle and Hamblin 1981). Percentage yield reduction was calculated as $100 - (Ys/Yp \times 100)$. Yield index was calculated as $Ys/\hat{Y}s$ (Gavuzzi et al. 1997). Drought sensitivity index was calculated as $(1 - Ys/Yp)/D$, where $D$ is 1 − (mean yield of all cotton cultivars under water-limited condition/mean yield of all cotton cultivars under well-watered condition) (Fischer and Maurer 1978). Stress tolerance index was calculated as $(Yp \times Ys)/(\hat{Y}p)^2$ (Fernandez 1992; Kristin et al. 1997). Harmonic mean was calculated as 2 ($Yp \times Ys$) / ($Yp + Ys$) (Kristin et al. 1997) where $Yp$ and $Ys$ are mean yields of a given

cultivar under well-watered and water-limited conditions, respectively. Geometric mean productivity was calculated as $(Yp \times Ys)^{1/2}$ (Fernandez 1992; Kristin et al. 1997). PAWS statistics software (SPSS Inc. Released 2009, PASW Statistics for Windows, Version 18.0, Chicago: SPSS Inc) with Pearson correlation, two-tailed method, was employed to evaluate bivariate correlation coefficients between traits.

## DNA isolation

Young leaves were harvested from plants at the 4–5 leaf stage and genomic DNA extraction was performed as described by Doyle and Doyle (1987). DNA concentrations were quantified using a Nanodrop ND-1000 spectrophotometer and adjusted to 50 ng/μl for further analysis.

## SSR analysis

A total of 177 pairs of SSR primers (DPL, BNL, DOW, JESPR, TMB, CIR, MUSS, GH, MGHES, NAU, STV) (Table S2) were used to detect polymorphic loci within the population. The primer collection was selected to span the entire genome with at least three markers per chromosome. A core SSR primer set developed by Yu et al. (2012) was also included. Primer information was obtained from the Cotton Database Resources (www.cottongen.org).

Polymerase chain reaction (PCR) was conducted in a total volume of 25 μl, containing 2.5 μl 10× PCR buffer (50 mM KCl, 10 mM Tris-HCl, 1.5 mM $MgCl_2$, pH 8.3), 1.5 μl $MgCl_2$ (25 mM), 0.5 μl dNTP (0.2 mM), 0.5 μl forward and 0.5 μl reverse primers (10 pmol), 0.3 μl Taq polymerase (0.25 U), 18.2 μl sterile ultra-distilled water, and 1 μl DNA (~ 50 ng/μl). PCR conditions were optimized as follows: 1 cycle of 3 min at 94 °C for denaturation, 35 cycles with 1 min at 94 °C, 45 s at 55–60 °C annealing temperature (depending on primer pair), 1 min at 72 °C for extension, and a final extension step of 10 min at 72 °C in BIO-RAD Thermal Cycler™. A Fragment Analyzer™ Automated CE System was used to separate DNA fragments at high resolution with the DNF-900-55-DNA-35-500 bp separation method. The data were analyzed using PROSize 2.0 analytical software. Allele sizes were determined by binning fragments into ±2 base pair bins.

## Diversity and population sub-structure analysis

Allelic data were scored dominantly with "1" for presence, "0" for absence, and "9" for missing data. Gene diversity values for the markers were calculated using Gene Diversity software (GDdom) (Abuzayed et al. 2016). To identify genomic distances between cultivars, DARwin5 (Dissimilarity Analysis and Representation for Windows) (Perrier and Jacquemoud-Collet 2006) was used with the Dice coefficient and the unweighted neighbor-joining algorithm. To detect the sub-structure of the population, STRUCTURE 2.3.4 software was used (Pritchard et al. 2000). This program uses a model-based clustering method with an admixture model to determine ancestry. For clustering, the length of the burn-in period was 50,000 and MCMC (Markov Chain Monte Carlo) replication after burn-in was 300,000 to achieve accurate parameter estimation. For sub-structure determination, cluster numbers ($K$) from 1 to 10 were tested with 20 iterations each. The Q matrix showing the proportion of assignment to the most correct cluster for each individual was obtained from the analysis and processed with the STRUCTURE HARVESTER program (Earl and vonHoldt 2012) to visualize STRUCTURE results and for implementation of the Evanno method (Evanno et al. 2005) to decide the best $K$. The cut-off value for assignment to subpopulations was determined as 60%. Individuals with an assignment probability lower than 60% were described as "admixed."

## Linkage disequilibrium analysis and LD decay

Pairwise linkage disequilibrium (LD) was estimated as the correlation coefficient ($r^2$) between all pairs of SSRs using TASSEL 2.1 (Bradbury et al. 2007). Before conducting LD analysis, it was important to filter genotype alignment data to remove minor alleles which could bias LD estimations. Thus, SSR alleles with frequencies below 0.05 were removed using the site filtration function. LD analysis was then performed on the filtered dataset using the squared allele-frequency correlations between marker pairs using the rapid permutation test with 10,000 shuffles ($p \leq 0.01$). The LD decay pattern was generated for significant data ($p \leq 0.01$ and $r^2 \geq 0.01$). LD analysis was performed with $r^2$, as it is considered a better LD parameter than $D'$ (Kruglyak 1999; Ardlie et al. 2002; Terwilliger et al. 2002). Chromosomal positions of molecular markers were based on

Blenda et al. (2012) and Yu et al. (2012). A and D sub-genome chromosome assignments were based on Wang et al. (2006).

### QTL analysis

Linkage analysis was performed with TASSEL 2.1 software using the general linear model (GLM) ($Q$) and mixed linear model (MLM) ($Q$ and $K$) methods to identify QTLs for the yield and drought-related traits (Bradbury et al. 2007). Significance levels were determined at $p \leq 0.01$. Association analysis of the phenotypes (morphological data) and genotypes (SSR allelic data) was performed using the $Q$ matrix calculated by STRUCTURE 2.3.4 and the relative kinship among individuals ($K$ matrix) determined by TASSEL 2.1.

## Results

### Phenotypic evaluation

Yield traits and water-use parameters were evaluated under both well-watered (Table S3) and water-limited (Table S4) conditions. In addition, drought-related traits were calculated using the well-watered and water-limited data from both locations combined to determine the response of cotton genotypes to water stress (Table S5). Phenotypic distributions showed that all traits segregated in a quantitative fashion and therefore were suitable for QTL analysis (Figs. 1 and 2).

Seed cotton yield (SCY) ranged from 2440 to 6520 kg ha$^{-1}$ with a mean of 4080 kg ha$^{-1}$ under well-watered conditions. In contrast, under water-limited conditions, SCY ranged from 1790 to 3990 kg ha$^{-1}$ with a mean of 2980 kg ha$^{-1}$. Lint yield ranged from 920 to 2370 kg ha$^{-1}$ under the well-watered regime and from 650 to 1530 kg ha$^{-1}$ under water-limited regime with mean values of 1490 and 1100 kg ha$^{-1}$, respectively. Thus the mean values of both traits were reduced by ~ 25% under drought conditions (SCY, 27%; LY, 26%). Lint percentage showed no significant variation between the watering regimes, ranging from 31 to 40% with a mean of 36% under well-watered conditions; and from 32 to 42% with a mean of 37% under water-limited conditions. Water-use efficiency (WUE) varied between 4 and 10.7 kg ha$^{-1}$ mm$^{-1}$ with a mean of 6.6 kg ha$^{-1}$ mm$^{-1}$ under well-watered regime. In contrast, WUE ranged from 6 to 13.3 kg ha$^{-1}$ with a 53% increase in the mean under water stress.

Drought parameters were calculated under water stress (deficit irrigation 50%) conditions. Yield potential of the different genotypes ranged from 212 to 526 kg ha$^{-1}$ with a mean of 353 kg ha$^{-1}$. Yield reduction ranged from 0.5 to 51% with a mean of 26%. Yield index ranged from 0.6 to 1.3 with a mean of 1. Drought sensitivity index ranged from 0.1 to 1.9 with a mean of 0.95. Stress tolerance index varied between 0.3 and 1.6 with a mean of 0.7. Harmonic mean ranged from 207 to 495 kg ha$^{-1}$ with a mean of 343 kg ha$^{-1}$. Geometric mean productivity varied between 209 and 510 kg ha$^{-1}$ with a mean of 348 kg ha$^{-1}$.

Strong positive correlations were found between many of the traits as expected because some of them used the same measurements for calculation. Under the well-watered regime, seed cotton yield was correlated ($p < 0.01$) with water-use efficiency ($r = 0.99$) and lint yield ($r = 0.94$) (Table S6). Similarly, lint yield was correlated with water-use efficiency ($r = 0.94$). Under the water-limited regime, drought parameters (geometric mean productivity, harmonic mean, stress tolerance index, water-use efficiency) showed highly significant positive correlations ($r > 0.80$, at $p < 0.01$) with each other. Drought sensitivity index was positively correlated with yield reduction ($p < 0.01$). The stress tolerance and yield indices were also correlated ($r = 0.86$) ($p < 0.01$). Lint percentage did not show significant correlation with any traits except yield potential ($r = 0.52$, at $p < 0.01$). Negative correlations between traits tended to be much weaker ($r \leq 0.34$) (Table S7).

### Diversity and population sub-structure analysis

The 177 SSR markers revealed a total of 967 fragments among the 99 cultivars with an average of 5.5 alleles per marker. Fragment lengths ranged from 76 to 434 bp. The overall average genetic diversity of the SSR markers ranged from 0.01 to 0.50 with a mean of 0.28. Diversity analysis of the population was performed with DARwin5 software. Because 50% of valid data was required for each unit pair, the cultivars Delta Diamond, Gloria, Nazilli 143, and Niab 111 were discarded from the diversity analysis. The unweighted neighbor-joining tree yielded four different sub-groups for the population (Fig. 3 and Table S1). This distribution was confirmed with principal coordinate analysis (PCoA) (data not shown). These four sub-groups, described as Group 1,

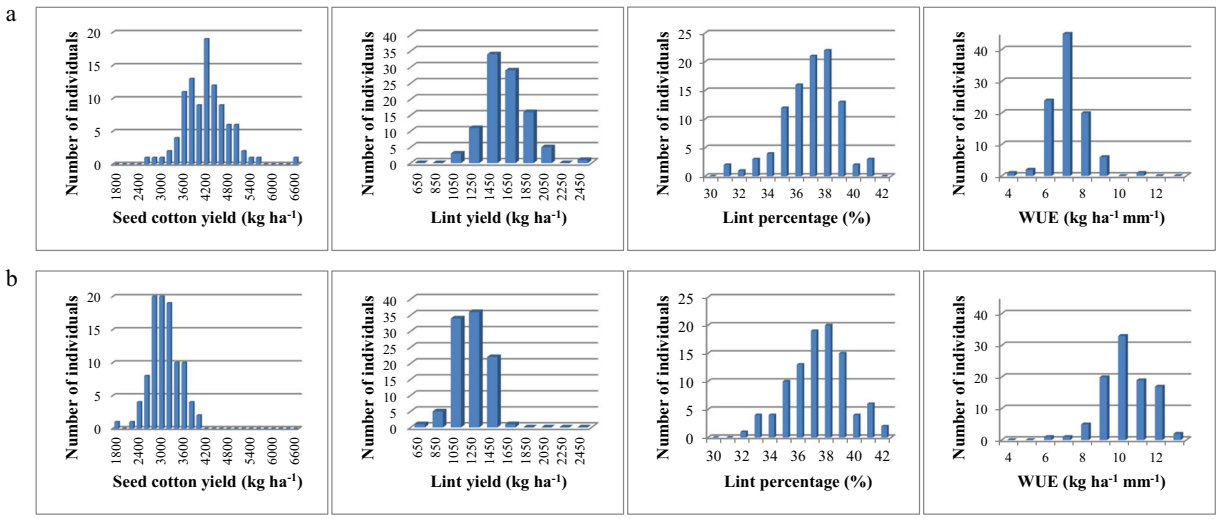


**Fig. 1** Distributions of yield and drought traits within the germplasm panel under well-watered (**a**) and water-limited (**b**) regimes

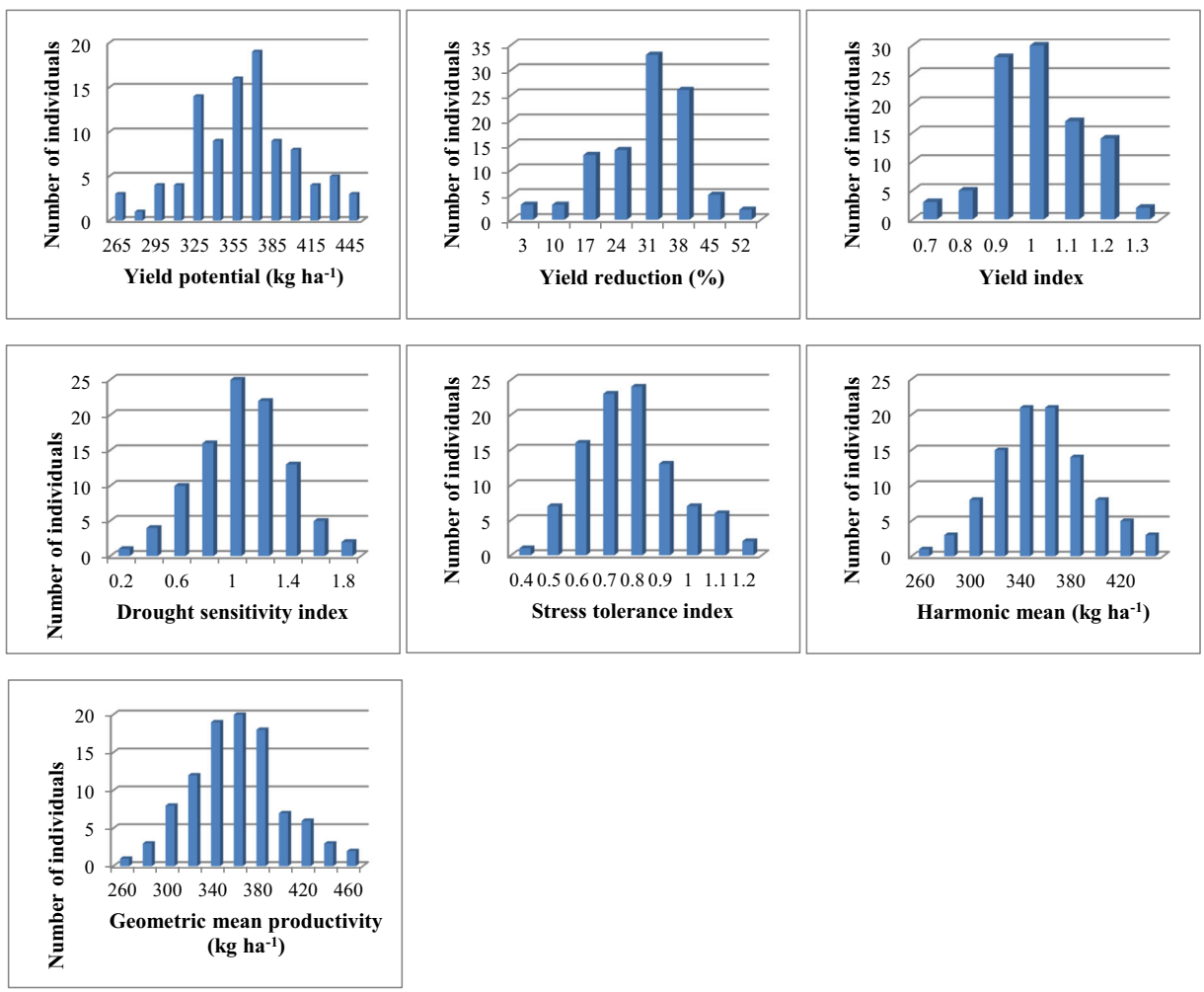


**Fig. 2** Distributions of drought parameters within the germplasm panel

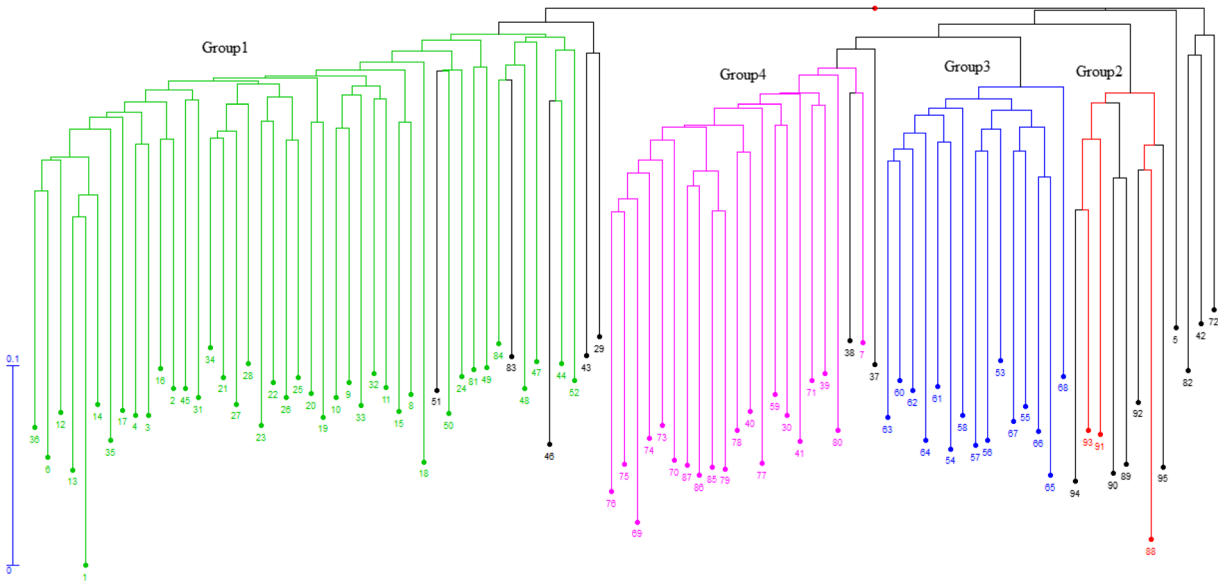


**Fig. 3** Genetic diversity of 95 upland cotton genotypes. Colors green, red, blue, and pink represent individuals of sub-groups A, B, C, and D, respectively. "Admixed" individuals are in black

Group 2, Group 3, and Group 4 were composed of 46, 8, 15, and 22 individuals, respectively. Four individuals were not grouped: Auburn M, Delcerro, Sicala 3/2, and SJ U 86. The pairwise dissimilarity values between cultivars ranged from a low of 22% between TKY 9309 and GC 555 to a high of 60% between Sealand 542 and PG 2018. The mean pairwise dissimilarity was 38%. A high correlation existed between the pairwise dissimilarities and distances as represented in the tree ($r = 0.92$).

Sub-groups of the population were determined with the software STRUCTURE. According to the results, the $\Delta K$ value peaked at $K = 2$ with a smaller peak at $K = 4$ (Fig. S1). Hence, the optimum cluster number to avoid missing loci in the association analysis was determined to be two (Q matrix at $K = 2$) (Table S8). However, the secondary peak at $K = 4$ could support dividing the population into four sub-groups. At $K = 4$, the population was assigned to sub-groups A, B, C and D with 41 (41%), 3 (3%), 16 (16%), and 23 (23%) individuals in each group, respectively (Table S1). The individuals which failed to be assigned to a sub-group (16, 16%) were considered "admixed." The inclusion of population structure results in the diversity dendrogram (groups 1–4) revealed a high degree of correspondence between the two sets of results (Fig. 3). All individuals (100%) of sub-groups A and B were assigned to group 1 and group 2, respectively. Of sub-groups C and D, 94 and 96% of individuals were assigned to group 3 and group 4, respectively. Four of 16 "admixed" individuals could not be assigned to diversity groups (Auburn M, Delcerro, Sicala 3/2 and SJ U 86); the rest being distributed among group 1 (5), group 2 (5), and group 4 (2) (Fig. 3 and Table S1).

## Linkage disequilibrium analysis and LD decay

Site filtration of minor alleles brought the allelic data from 967 to 625 loci. Of 212,639 pairwise comparisons of the 625 SSR loci across 99 *G. hirsutum* L. cultivars, 9185 (4.3%) marker pairs showed linkage disequilibrium at a significance level of $p \leq 0.01$ and $r^2 \geq 0.01$, with 1.7% at $p \leq 0.001$. LD analysis of pairwise estimates for $r^2$ ranged from 0.06 to 1 for markers located within 0–170 cM. Most of the $r^2$ values were between 0.06 and 0.3. The average $r^2$ values (LD level) of global and unlinked SSR marker pairs were 0.16 and 0.15, respectively. For linked SSR marker pairs the average $r^2$ was 0.25. The LD decay plot shows how $r^2$ (LD) declined with genetic distance (cM) between marker pairs. The LD decayed relatively rapidly at 20–30 cM ($r^2 \geq 0.5$) (Fig. S2).

QTL analysis

GLM ($Q$) and MLM ($Q$ and $K$) models in TASSEL software were used to determine associated SSR marker loci under drought and irrigated conditions. Loci supported by both GLM and MLM analysis at a significance level $p \leq 0.01$ are reported here. Different sets of loci were discovered to be associated with the two watering regimes.

Well-watered conditions

Fifteen SSR markers distributed across 12 chromosomes were linked to four traits: seed cotton yield (SCY), lint yield (LY), lint percentage (LP), and water-use efficiency (WUE) under well-watered conditions (Table 1). Of these trait-associated markers, ten were distributed on D chromosomes and the remaining five were on A chromosomes. The total phenotypic variation explained (PVE) by individual markers ($r^2$) ranged from 6 to 18%. Seven of the trait-associated markers were relatively highly informative (GD values ≥ 0.3).

Seed cotton yield (SCY) was significantly associated with markers $CIR169_{180}$ (chromosome A07) and $DPL717_{434}$ (chromosome D11). Both markers had negative effects (− 30.4% and − 26.2% kg ha$^{-1}$, respectively) and explained 8–9% of the phenotypic variation in the trait (Table 1).

Three SSR markers, two on A02 and A08 and one on D01, were associated with lint yield (LY) with PVE values ranging from 6 to 10% (Table 1). $TMB0514_{191}$ had the largest positive effect (14 kg ha$^{-1}$) while $BNL3474_{170}$ had a negative effect (− 8 kg ha$^{-1}$) on lint yield with the largest PVE (10%).

Eleven different SSR marker loci were significantly associated with lint percentage (LP) with PVE values ranging between 7 and 18% (Table 1). These LP-associated marker loci ranged across nine chromosomes with eight loci on D chromosomes (D01, D03, D06, D07, D11, and D12) and three on A chromosomes

**Table 1** Yield and drought-associated SSR markers under well-watered conditions as determined by GLM and MLM analysis

| Trait | Marker loci | GLM | | MLM | | GD | Chromosome location |
|---|---|---|---|---|---|---|---|
| | | $p$ value | $r^2$ (%) | $p$ value | Marker effect[a] | | |
| SCY | $CIR169_{180}$ | 0.0039 | 8 | 0.0089 | − 30.4 | 0.21 | c07 (A07) |
| | $DPL717_{434}$ | 0.003 | 9 | 0.0046 | − 26.2 | 0.21 | c21 (D11) |
| LY | $BNL3474_{170}$ | 0.0035 | 10 | 0.009 | − 8.0 | 0.34 | c08 (A08) |
| | $DPL322_{196}$ | 0.01 | 6 | 0.01 | 10.4 | 0.28 | c15 (D01) |
| | $TMB0514_{191}$ | 0.0071 | 8 | 0.0057 | 14.0 | 0.32 | c02 (A02) |
| LP | $BNL1227_{228}$ | 0.006 | 9 | 0.01 | 1.2 | 0.22 | c26 (D12) |
| | $BNL3474_{170}$ | $5 \times 10^{-4}$ | 14 | 0.0019 | − 0.8 | 0.34 | c08 (A08) |
| | $DPL009_{207}$ | 0.01 | 7 | 0.0049 | 0.7 | 0.34 | c07 (A07) |
| | $DPL181_{176}$ | 0.003 | 10 | 0.0086 | − 1.0 | 0.3 | c21 (D11) |
| | $DPL223_{268}$ | 0.0027 | 12 | 0.0075 | 1.0 | 0.37 | c16 (D07) |
| | $DPL322_{191}$ | 0.0074 | 8 | 0.01 | − 0.6 | 0.28 | c15 (D01) |
| | $DPL520_{197}$ | $5.2 \times 10^{-4}$ | 12 | 0.0059 | − 0.8 | 0.19 | c25 (D06) |
| | $DPL520_{281}$ | 0.0012 | 11 | 0.0073 | − 0.8 | 0.19 | c25 (D06) |
| | $DPL717_{153}$ | 0.004 | 9 | 0.0079 | − 0.9 | 0.21 | c21 (D11) |
| | $TMB1356_{183}$ | 0.0027 | 9 | 0.008 | − 0.8 | 0.18 | c10 (A10) |
| | $TMB1910_{196}$ | 0.0075 | 8 | 0.01 | 0.9 | 0.18 | c15 (D01) |
| | $TMB2018_{240}$ | $5.6 \times 10^{-5}$ | 18 | $9.1 \times 10^{-4}$ | − 0.8 | 0.21 | c17 (D03) |
| WUE | $CIR169_{180}$ | 0.0029 | 9 | 0.0072 | − 0.1 | 0.21 | c07 (A07) |
| | $DPL247_{167}$ | $7.3 \times 10^{-4}$ | 11 | 0.01 | 0.1 | 0.36 | c19 (D05) |
| | $DPL307_{207}$ | 0.0011 | 10 | 0.01 | 0.1 | 0.37 | c23 (D09) |
| | $DPL717_{434}$ | 0.0041 | 8 | 0.0062 | 0.1 | 0.21 | c21 (D11) |

[a] The additive effects of the allele on the phenotype, either positive or negative

(A07, A08, and A10). Six loci ($DPL181_{176}$, $DPL520_{281}$, $DPL520_{197}$, $DPL223_{268}$, $BNL3474_{170}$, and $TMB2018_{240}$) had PVE values higher than 10%. The additive effects of the individual alleles were less than ± 1.2%.

Four SSR markers, three on D chromosomes: D05, D09, and D11 and one on A07, were associated with water-use efficiency (WUE) with PVE values ranging from 8 to 11% (Table 1). Among them, $DPL247_{167}$ and $DPL307_{207}$ had the highest PVE values of 11 and 10%, respectively (Table 1). The allelic effects of all four markers were less than ± 0.1 kg $ha^{-1}$.

## Water-limited conditions

Twenty-three different SSR markers distributed across 17 chromosomes were linked to ten yield traits and drought parameters under water-limited conditions (Table 2). Of these loci, 14 were distributed on D chromosomes and the remaining nine markers were on A chromosomes. The total phenotypic variation explained by the markers ($r^2$) ranged from 7 to 14% under water-limited conditions. Thirteen of the associated marker loci were highly informative with GD value higher than 0.34.

Six SSR markers distributed across six chromosomes were significantly associated with seed cotton yield (SCY) with PVE values ranging from 9 to 13% (Table 2). Of them, three were on D chromosomes (D07, D08, and D09) and three were on A chromosomes (A01, A08, and A11). Two allelic loci of TMB2068 ($TMB2068_{140}$ and $TMB2068_{146}$) had opposite effects on SCY with negative and positive allelic effects of − 16.9 and 16 kg $ha^{-1}$, respectively. $BNL1034_{315}$, $BNL1667_{148}$, $JESPR274_{137}$, $JESPR157_{233}$, and $TMB2068_{140}$ had PVE values higher than 10%.

Eight SSR markers were associated with lint yield (LY) with PVE values ranging from 7 to 12% (Table 2). They ranged across six D chromosomes (D02, D05, D07, D08, D09, and D13) and two A chromosomes (A07 and A08). Two allelic loci of TMB2068 ($TMB2068_{140}$ and $TMB2068_{146}$) were negatively associated with LY. The allelic effects of the marker loci associated with LY ranged from − 7.1 to 8.7 kg $ha^{-1}$. Three markers $DPL405_{265}$, $JESPR157_{233}$, and $TMB2068_{140}$ had PVE values higher than 10%.

Eleven SSR markers distributed across 11 chromosomes were associated with lint percentage (LP) with PVE values ranging from 7 to 14% (Table 2). Eight of these markers were on D chromosomes (D01, D02, D03, D05, D06, D07, D12, and D13) and the remaining three were on A chromosomes (A07, A08, and A11). Eight of the marker loci ($BNL1227_{228}$, $BNL3474_{170}$, $DOW006_{264}$, $DPL223_{268}$, $DPL322_{196}$, $GH537_{161}$, $TMB1295_{271}$, and $TMB2018_{240}$) had PVE values higher than 10%. All LP-associated marker loci had allelic effects less than ± 1.4%.

Five SSR markers were significantly associated with water-use efficiency (WUE) explaining between 10 and 13% of the phenotypic variation in the trait (Table 2). The markers were distributed on three D chromosomes (D07, D08, and D09) and two A chromosomes (A01 and A11). Two allelic loci of TMB2068 ($TMB2068_{140}$ and $TMB2068_{146}$) were associated with WUE with opposite but very low allelic effects. None of WUE-associated markers' allelic effects were higher than 0.1 kg $ha^{-1}$. However, all marker loci had PVE values higher than 10%.

Five SSR markers distributed on four chromosomes were associated with yield reduction (YR) with PVE values ranging from 7 to 12% (Table 2). Three of the markers were on A chromosomes (A07 and A12) and the remaining two were on D chromosomes (D07 and D11). Only $CIR169_{180}$ had a PVE value higher than 10% (12%). However, two allelic loci of the marker DPL100 ($DPL100_{160}$ and $DPL100_{199}$) had negative (− 3.7%) and positive (5%) allelic effects on YR, respectively.

Four SSR markers were associated with yield index (YI) with PVE ranging from 8 to 13% (Table 2). They were located on three D chromosomes (D07, D08, and D09) and one A chromosome (A11). Two allelic loci of the marker JESPR274 ($JESPR274_{137}$ and $JESPR274_{230}$) had positive and negative effects on YI, respectively. Moreover, two allelic loci of TMB2068 ($TMB2068_{140}$ and $TMB2068_{146}$) also had opposite but very low allelic effects. None of the YI-associated marker effects were higher than 0.1%. However, $BNL1034_{315}$, $JESPR274_{137}$, $JESPR157_{233}$, and $TMB2068_{140}$ had PVE values higher than 11%.

Five SSR markers distributed across five chromosomes were associated with drought sensitivity index (DSI) with PVE values ranging from 7 to 12% (Table 2). Three of the markers were on A chromosomes (A07, A12, and A13) and two were on D chromosomes (D05 and D07). Three allelic loci of the marker DPL100 ($DPL100_{160}$, $DPL100_{175}$, $DPL100_{199}$) were significantly associated with DSI with PVE values of 12, 8, and 10%, respectively. All of the DSI-associated markers' allelic effects were less than 0.2%.

**Table 2** Yield and drought associated SSR markers under water-limited conditions as determined by GLM and MLM analysis

| Trait | Marker loci | GLM | | MLM | | GD | Chromosome location (c) |
|---|---|---|---|---|---|---|---|
| | | *p* value | $r^2$ (%) | *p* value | Marker effect[a] | | |
| SCY | $BNL1034_{315}$ | $4 \times 10^{-4}$ | 13 | 0.0072 | − 12.4 | 0.37 | Not certain |
| | $BNL1667_{148}$ | 0.0015 | 11 | 0.01 | − 10.8 | 0.38 | Not certain |
| | $DPL176_{274}$ | 0.0028 | 9 | 0.01 | − 13.8 | 0.37 | c08 (A08) |
| | $JESPR274_{137}$ | $8.1 \times 10^{-4}$ | 12 | 0.0028 | 13.8 | 0.25 | c23 (D09) |
| | $JESPR157_{233}$ | $1.8 \times 10^{-4}$ | 13 | 0.0067 | 13.8 | 0.19 | Not certain |
| | $TMB2068_{140}$ | $3.3 \times 10^{-4}$ | 13 | $9.9 \times 10^{-4}$ | − 16.9 | 0.18 | c16 (D07) |
| | $TMB2068_{146}$ | 0.0028 | 9 | 0.0069 | 16.0 | 0.18 | c16 (D07) |
| LY | $DOW006_{264}$ | 0.0093 | 8 | 0.0055 | 8.3 | 0.38 | c18 (D13) |
| | $DPL009_{207}$ | 0.0071 | 7 | 0.0044 | 6.2 | 0.34 | c07 (A07) |
| | $DPL140_{240}$ | 0.01 | 9 | 0.01 | 6.5 | 0.45 | c19 (D05) |
| | $DPL176_{275}$ | 0.0057 | 8 | 0.0068 | 8.7 | 0.37 | c08 (A08) |
| | $DPL405_{265}$ | $7.1 \times 10^{-4}$ | 12 | 0.0048 | − 5.6 | 0.37 | c14 (D02) |
| | $JESPR274_{137}$ | 0.0046 | 9 | 0.0057 | 5.7 | 0.25 | c23 (D09) |
| | $JESPR157_{233}$ | 0.0016 | 10 | 0.0059 | 6.0 | 0.19 | Not certain |
| | $TMB2068_{140}$ | $7.8 \times 10^{-4}$ | 12 | 0.0016 | − 7.1 | 0.18 | c16 (D07) |
| | $TMB2068_{146}$ | 0.0048 | 8 | 0.0097 | − 5.7 | 0.18 | c16 (D07) |
| LP | $BNL1034_{205}$ | 0.01 | 7 | 0.01 | 0.9 | 0.37 | Not certain |
| | $BNL1227_{228}$ | 0.0022 | 11 | 0.0061 | 1.4 | 0.22 | c26 (D12) |
| | $BNL3474_{170}$ | 0.0028 | 11 | 0.0068 | − 0.7 | 0.34 | c08 (A08) |
| | $DOW006_{264}$ | 0.0007 | 14 | 0.001 | 1.4 | 0.38 | c18 (D13) |
| | $DPL119_{127}$ | 0.0053 | 9 | 0.0054 | − 0.9 | 0.4 | c07 (A07) |
| | $DPL223_{268}$ | 0.0019 | 13 | 0.0053 | 1.2 | 0.37 | c16 (D07) |
| | $DPL322_{196}$ | 0.0025 | 10 | 0.0015 | 1.2 | 0.28 | c15 (D01) |
| | $DPL405_{265}$ | 0.0047 | 9 | 0.01 | − 0.7 | 0.37 | c14 (D02) |
| | $GH537_{161}$ | 0.0014 | 13 | $8.1 \times 10^{-4}$ | −1.2 | 0.28 | c25 (D06) |
| | $TMB1295_{271}$ | $4.8 \times 10^{-4}$ | 13 | 0.0021 | − 0.9 | 0.28 | c19 (D05) |
| | $TMB2018_{240}$ | $6.6 \times 10^{-4}$ | 13 | 0.0055 | − 0.7 | 0.21 | c17 (D03) |
| WUE | $BNL1034_{315}$ | $7 \times 10^{-4}$ | 12 | 0.0089 | 0.1 | 0.37 | Not certain |
| | $BNL1667_{148}$ | 0.0011 | 12 | 0.0098 | 0.1 | 0.38 | Not certain |
| | $JESPR274_{137}$ | $7.9 \times 10^{-4}$ | 12 | 0.0025 | 0.1 | 0.25 | c23 (D09) |
| | $JESPR157_{233}$ | $4.4 \times 10^{-4}$ | 12 | 0.01 | 0.1 | 0.19 | Not certain |
| | $TMB2068_{140}$ | $2.8 \times 10^{-4}$ | 13 | $9.3 \times 10^{-4}$ | − 0.1 | 0.18 | c16 (D07) |
| | $TMB2068_{146}$ | 0.0023 | 10 | 0.0058 | 0.1 | 0.18 | c16 (D07) |
| YR | $CIR169_{180}$ | $7.3 \times 10^{-4}$ | 12 | $7.4 \times 10^{-4}$ | 6.3 | 0.21 | c07 (A07) |
| | $DPL009_{207}$ | 0.0085 | 7 | 0.0069 | 3.6 | 0.34 | c07 (A07) |
| | $DPL100_{160}$ | 0.0073 | 8 | 0.0079 | − 3.7 | 0.44 | c12 (A12) |
| | $DPL100_{199}$ | 0.0049 | 9 | 0.0054 | 5.0 | 0.44 | c12 (A12) |
| | $DPL717_{277}$ | 0.0055 | 8 | 0.0057 | 2.9 | 0.21 | c21 (D11) |
| | $TMB2068_{140}$ | 0.0089 | 8 | 0.0083 | − 3.5 | 0.18 | c16 (D07) |
| YI | $BNL1034_{315}$ | $4.2 \times 10^{-4}$ | 13 | 0.0082 | 0.1 | 0.37 | Not certain |
| | $JESPR274_{137}$ | 0.0015 | 11 | 0.0043 | 0.1 | 0.25 | c23 (D09) |
| | $JESPR274_{230}$ | 0.0094 | 8 | 0.0034 | − 0.1 | 0.25 | c23 (D09) |
| | $JESPR157_{233}$ | $2 \times 10^{-4}$ | 13 | 0.0094 | 0.1 | 0.19 | Not certain |

**Table 2** (continued)

| Trait | Marker loci | GLM | | MLM | | GD | Chromosome location (c) |
|---|---|---|---|---|---|---|---|
| | | *p* value | $r^2$ (%) | *p* value | Marker effect[a] | | |
| | $TMB2068_{140}$ | $3.5 \times 10^{-4}$ | 13 | 0.0012 | − 0.1 | 0.18 | c16 (D07) |
| | $TMB2068_{146}$ | 0.0028 | 9 | 0.0068 | 0.1 | 0.18 | c16 (D07) |
| DSI | $CIR169_{180}$ | 0.0025 | 10 | 0.0027 | 0.2 | 0.21 | c07 (A07) |
| | $DOW053_{367}$ | 0.0083 | 8 | 0.0097 | 0.1 | 0.45 | c19 (D05) |
| | $DPL100_{160}$ | $7.2 \times 10^{-4}$ | 12 | $9.5 \times 10^{-4}$ | − 0.2 | 0.44 | c12 (A12) |
| | $DPL100_{175}$ | 0.0059 | 8 | 0.0076 | 0.2 | 0.44 | c12 (A12) |
| | $DPL100_{199}$ | 0.0022 | 10 | 0.0028 | 0.2 | 0.44 | c12 (A12) |
| | $JESPR153_{127}$ | 0.008 | 8 | 0.0095 | 0.1 | 0.37 | c13 (A13) |
| | $TMB2068_{160}$ | 0.01 | 7 | 0.01 | 0.2 | 0.18 | c16 (D07) |
| STI | $BNL1034_{315}$ | $8.1 \times 10^{-4}$ | 12 | 0.01 | − 0.1 | 0.37 | Not certain |
| | $JESPR274_{137}$ | 0.0024 | 10 | 0.01 | 0.1 | 0.25 | c23 (D09) |
| HM | $BNL1034_{315}$ | $6.4 \times 10^{-4}$ | 12 | 0.01 | − 12.7 | 0.37 | Not certain |
| | $JESPR274_{137}$ | 0.0031 | 10 | 0.01 | 13.2 | 0.25 | c23 (D09) |
| GMP | $BNL1034_{315}$ | 0.001 | 11 | 0.01 | − 12.6 | 0.37 | Not certain |
| | $JESPR274_{137}$ | 0.0036 | 9 | 0.01 | 13.2 | 0.25 | c23 (D09) |

[a] The additive effects of the allele on the phenotype, either positive or negative

Stress tolerance index (STI), harmonic mean (HM), and geometric mean productivity (GMP) were each significantly associated with two markers: $BNL1034_{315}$ (location not certain; A11, D03, or D11; Blenda et al. 2012) (PVE ≥ 11%) negatively affected the traits (12.6, 12.7, and 0.1, respectively), while $JESPR274_{137}$ (PVE ≥ 9%) positively affected these three traits (13.2, 13.2, and 0.1, respectively) (Table 2).

## Discussion

In the present study, a germplasm panel of 99 upland cotton genotypes was evaluated with 177 genome-wide SSR markers to assess genetic diversity and perform association analysis of yield and drought parameters under two watering regimes. This analysis also revealed which of the lines showed the greatest phenotypic stability under drought stress conditions.

The genotypes showing the least change in seed cotton yield under drought stress were Zeta 2 (0%), Delcerro (− 2%), Nazilli 87 (− 3%), and DAK 66/3 (− 4%). Lint yield was fairly stable under drought stress in three of the same genotypes: Zeta 2 (+2%), Delcerro (−1%), DAK 66/3 (−3%), and also Vulcano (− 3%). Lint percentage (LP) was not altered much by watering regime. In Stoneville 213 and Vulcano LP increased 8 to 9% under drought stress indicating the potential of these cultivars to adapt to drought conditions. Water-use efficiency increased in all genotypes under drought stress. The top-performing genotypes were, of course, those which had little change in yield under drought: Zeta 2 (100%), Delcerro (100%), Nazilli 87 (91%), and DAK 66/3 (88%).

Under drought conditions, Np Ege 2009 (525 kg $ha^{-1}$), Nazilli M39 (452 kg $ha^{-1}$), Sj U 86 (434 kg $ha^{-1}$), Barut 2005 (430 kg $ha^{-1}$), Nazilli 143 (429 kg $ha^{-1}$), and Np Ozbek 100 (426 kg $ha^{-1}$) had the highest yield potential. In contrast, yield reduction ($\bar{x}$ = 26%) was lowest in Zeta 2 (0%), Delcerro (2%), Nazilli 87 (3%), and DAK 66/3 (4%). Yield index was higher than the average value of 1 in 63 genotypes. The best genotypes were Np Ege 2009 (1.3) and DAK 66/3 (1.3).

High (≥ 1) and low (≤ 1) drought sensitivity index ($\bar{x}$ = 0.95) indicates susceptibility and tolerance against drought stress, respectively. Fifty-six genotypes showed some level of drought tolerance. The top five genotypes were Zeta 2 (0.1), Nazilli 87 (0.2), DAK 66/3 (0.3), Niab 999 (0.3), and Delcerro (0.3). Forty-three genotypes did not show significant level of drought tolerance. The most sensitive cultivars were Taskent 1 (1.9), Tamcot 22 (1.7), Taskent Uzbek (1.7), and Coker 208 (1.5).

Stress tolerance index ($\bar{x} = 0.7$) was highest in Np Ege 2009 (1.6), Nazilli M39 (1.2), Barut 2005 (1.1), and Nazilli 143 (1.1). Harmonic mean ($\bar{x} = 343$) and geometric mean productivity were highest in the same genotypes: Np Ege 2009 (495 and 510 kg ha$^{-1}$, respectively), Nazilli M39 (437 and 444), Barut 2005 (421 and 425).

Stoneville 453, Caroline Queen, Sayar 314, Cukurova 1453, Nazilli 84, Nazilli 87, Ersan 92, and Ege 7913 are widely grown in Turkey (Cukobirlik 2017). Two of these cultivars (Ersan 92 and Nazilli 87) performed well in our study under drought conditions. Our study identified several other genotypes with good drought tolerance. For example, yield component traits were fairly stable in DAK 66/3, Ms. 30/1, Zeta 2, Delcerro, Delcerro Ms., Niab 999, and Vulcano under water-limited conditions. Changes in climate (temperature, precipitation) can have profound impacts on agricultural production including cotton yield (ITC 2011). Cotton genotypes that show little change in yield-based traits between well-watered and water-stress conditions may be more adaptive and less susceptible to unforeseen changes in climate.

A total of 967 marker loci were generated from 177 SSRs with an average of 5.5 alleles per marker. Previous analyses of cotton SSR markers (Qin et al. 2015; Nie et al. 2016; Cai et al. 2014; Du et al. 2016; Zhang et al. 2011) have revealed somewhat lower levels of polymorphism, with averages ranging from 2.2 to 5.1 alleles per marker. Diversity analysis of the germplasm panel revealed an average genetic diversity (GD) of 38%. This is consistent with previous molecular marker studies in which different panels of upland cotton cultivars were analyzed with SSR markers that revealed average genetic diversity of 36 and 38% (Du et al. 2016; Nie et al. 2016). However, the genetic diversity of our germplasm panel is much higher than reported (13%) in an analysis of 335, mostly Uzbek, cotton accessions (Abdurakhmonov et al. 2008).

Cotton is one of the earliest domesticated fiber crops. The first trace of domesticated *G. hirsutum* L. (upland cotton) dates to 3400–2300 B.C. (Rajpal et al. 2016). Domestication and subsequent breeding has restricted the gene pool of cotton resulting in the low genetic diversity of modern cultivated cotton lines. Our diversity results (Fig. 3) revealed that individuals from the same breeding program exhibited close similarity to each other and clustered in the same diversity groups. For example: DPL 6 (12) and DPL 882 (14) had 76% similarity; BA 525 (90) and BA 308 (89) had 71% similarity; and Taskent 1 (66) and Taskent 6 (67) had 72% similarity. Our results clearly show the consequences of intensive breeding on genetic diversity in cotton. Our findings could provide useful information for breeders looking to enhance genetic diversity in their programs by selecting elite cotton genotypes based on their dissimilarity.

Linkage disequilibrium is defined as the non-random co-segregation of loci through generations. Association mapping uses this property to predict the association of marker loci/chromosomal regions with a trait/phenotype (Ersoz et al. 2007). In our study, only 4.3% of linked (on the same chromosome) and unlinked (on different chromosomes) SSR locus pairs were in LD ($p \leq 0.01$) which is considerably lower than in previous studies (9.4, 17.3, and 21% as reported by Qin et al. 2015; Nie et al. 2016; Mei et al. 2013, respectively). As the coefficient of determination ($r^2$) approaches 1 for two loci, those loci co-occur more frequently in the population. Analysis of pairwise LD based on average $r^2$ (LD level) revealed that linked marker pairs were higher (an average $r^2 = 0.25$) than all entire (an average $r^2 = 0.16$) and unlinked marker pairs (an average $r^2 = 0.15$). This is similar to the results of previous studies that clearly showed that physical linkage affects the detection of LD (Mei et al. 2013; Zhao et al. 2014).

In our study, genome-wide LD extended to 74.2 cM at the level of $r^2 = 0.1$ and it rapidly decayed to 22.45 cM at $r^2 = 0.5$, much higher values than reported by others (25 cM ($r^2 \geq 0.1$), 12–13 cM ($r^2 = 0.1$) and 8.6 cM ($r^2 > 0.1$) as reported by Abdurakhmonov et al. 2009; Mei et al. 2013; Qin et al. 2015, respectively). LD is affected by many factors: genetic drift, natural selection, and especially recombination rate. LD tends to be high (low decay) in self-pollinated crops because of their low effective recombination rate. While cotton is naturally cross-pollinating, it has been bred to be self-pollinating as a means of maintaining genomic purity in the crop (Simpson 1954) Therefore, we expect the LD in upland cotton to be relatively high, as demonstrated in our study.

The extent of decay in LD indicates how many markers (marker density) are required for association analysis. The genome of tetraploid cotton spans 5200 cM (Paterson and Smith 1999). Hence, our LD decay rate suggests that nearly 230 polymorphic markers are required to implement association analysis correctly. As previously stated, we used 967 marker loci

dispersed across the tetraploid cotton's 26 chromosomes, thus providing sufficient theoretical coverage.

Many QTL analyses related to fiber and yield traits under different environments have been published for cotton (Wang et al. 2007; Zhang et al. 2013; Qin et al. 2015; Wang et al. 2015; Jamshed et al. 2016) while drought tolerance in cotton has been considered in only a few reports (Saranga et al. 2001; Saranga et al. 2004; Saeed et al. 2011; Zheng et al. 2016). In our study, an association analysis combining drought tolerance and yield parameters was conducted. We identified 30 different QTLs for all yield and drought parameters in *G. hirsutum* under both watering regimes. Among them, 15 QTLs were identified under well-watered and 23 QTLs under water-limited conditions. These 30 QTLs were widely distributed on 19 chromosomes. Chromosomes A03, A04, A05, A06, A07, D04, and D10 were not associated with any of the traits we analyzed. The majority of SSR markers (63%; 19 loci) mapped within the D sub-genome and the remainder to the A sub-genome. This finding is consistent with the greater diversity of the D sub-genome (Paterson et al. 2000). A high degree of polymorphism in the D sub-genome has also been reported in QTL analyses of fiber quality (Jiang et al. 1998), plant structure (Jiang et al. 2000), disease tolerance (Wright et al. 1998), and drought tolerance traits (Paterson et al. 2000; Saranga et al. 2001).

Several marker loci were associated with more than one trait which was expected given the related nature of most of the traits. Under water-stress conditions, BNL1034 (location not certain, A11, D03 or D11; Blenda et al. 2012) was associated with seven traits (SCY, LP, WUE, YI, STI, HM, and GMP), TMB2068 on D07 was associated with six traits (SCY, LY, WUE, YR, YI, and DSI), and JESPR274 on D09 was associated with seven traits (SCY, LY, WUE, YI, STI, HM, and GMP). These markers could potentially lie within genomic regions controlling drought tolerance.

Interestingly, CIR169 on A07 was associated with different traits under the two different regimes: SCY and WUE under well-watered conditions and YR and DSI under water-stress conditions (Table 2). Furthermore, we identified completely different sets of marker loci for the traits (except LP) under the two watering regimes suggesting that different alleles may be activated in response to drought conditions. *G. hirsutum* is an allopolyploid species ($n = 2\times = 26$, AADD) that originated from a hybridization event between two different diploid genomes (an African or Asian species with an American species) (Wendel and Cronn 2001). Subsequent genome doubling has resulted in the multiplication of genes in each sub-genome (Reinisch et al. 1994; Saranga et al. 2001) and the possibility of genetic redundancy (Gottlieb 2003) as well as functional divergence of duplicate genes.

Marker loci with high PVE and positive effects could be useful for marker-assisted selection of yield and drought tolerance traits under water-limited conditions. For example, $TMB2068_{146}$ had a relatively strong positive effect (16%; PVE = 9%) on seed cotton yield (SCY) (Table 2) suggesting that this marker could be useful for increasing SCY under water stress conditions. However, $TMB2068_{140}$ was also associated with SCY with relatively high negative allelic effect of 16.9% (PVE = 13%) (Table 2). Therefore, selection against this negative allele would be just as important as selection for the positive allele $TMB2068_{146}$ when using this marker to breed for improved seed cotton yield.

One way of targeting potentially useful loci for marker-assisted selection is to compare our results with those of previous QTL analyses using these SSR markers. BNL1227 was also associated with lint percentage in a study by An et al. (2010). The markers JESPR153 and JESPR274, associated with the yield and drought components in our study, were identified with fiber traits in previous studies: JESPR153 with fiber elongation and fiber length (Shen et al. 2005) and with fiber strength (Cai et al. 2014; Wang et al. 2015; Qin et al. 2015); JESPR274 with fiber micronaire (Wang et al. 2012; Qin et al. 2015). Thus, these markers are important targets for marker-assisted selection.

In conclusion, we identified 30 different SSR marker loci associated with drought and yield components under two watering regimes. Our study is unique in looking at drought and yield traits under both well-watered and water-limited conditions. To our knowledge, most of the loci associated with the aforementioned traits were newly identified. The genetic diversity and association mapping results should facilitate the introgression of quantitative trait loci and the development of drought-tolerant cotton lines with high yield.

**Acknowledgments** We are grateful to Nazilli Cotton Research Institute (NCRI, Nazilli, Aydın, TURKEY) for providing cotton seeds.

**Author contributions** AAB: molecular characterization, data analysis, interpretation of data, manuscript drafting and revision; CP, VS, HB: field experiments and drought tests; AmF: interpretation of data, manuscript revision; AnF: conception and design, interpretation of data, manuscript revision; SD: conception and design, manuscript revision; All: final approval of the version to be published.Funding informationThis study was supported by the Scientific and Technological Research Council of Turkey–TUBİTAK, Project No. 3110087.

### Compliance with ethical standards

**Conflict of interest** The authors declare that they have no conflict of interest.

## References

Abdurakhmonov IY, Kohel RJ, Yu JZ, Pepper AE, Abdullaev AA, Kushanov FN, Salakhutdinov IB, Buriev ZT, Saha S, Scheffler BE, Jenkins JN, Abdukarimov A (2008) Molecular diversity and association mapping of fiber quality traits in exotic *G. hirsutum* L. germplasm. Genomics 92(6): 478–487

Abdurakhmonov IY, Saha S, Jenkins JN, Buriev ZT, Shermatov SE, Scheffler BE, Pepper AE, Yu JZ, Kohel RJ, Abdukarimov A (2009) Linkage disequilibrium based association mapping of fiber quality traits in *G. hirsutum* L. variety germplasm. Genetica 136(3):401–417

Abuzayed M, El-Dabba N, Frary A, Doganlar S (2016) GDdom: an online tool for calculation of dominant marker gene diversity. Biochem Genet 55:155–157. https://doi.org/10.1007/s10528-016-9779-0

An C, Jenkins JN, Wu J, Guo Y, McCarty JC (2010) Use of fiber and fuzz mutants to detect QTL for yield components, seed, and fiber traits of upland cotton. Euphytica 172:21–34

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

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. PLoSONE7 (9), e45739

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

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

Cukobirlik (Çukurova Pamuk Yerfistiği Ve Yağli Tohumlar Tarim Satiş Kooperatifleri Birliği), http://www.cukobirlik.com.tr/?tekd=777&ikid=1&syf=*PAMUK* Accessed 15 February 2017

Doyle JJ, Doyle JL (1987) A rapid DNA isolation procedure for small quantities of fresh leaf tissue. Phytochem Bull 19:11–15

Du L, Cai C, Wu S, Zhang F, Hou S, Guo W (2016) Evaluation and Exploration of favorable QTL alleles for salt stress related traits in cotton cultivars (*G. hirsutum* L.). PLoS One 11(3), e0151076. https://doi.org/10.1371/journal.pone.0151076

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) Chapter 5 Applications of linkage disequilibrium and association mapping in crop plants. Genomics 1:97–119. https://doi.org/10.1007/978-1-4020-6295-7_5

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

Fernandez GCJ (1992) Effective selection criteria for assessing plant stress tolerance. In: Kuo CG (ed.) Adaptation of food crops to temperature and water stress. Proceeding of an International Symposium, Aug13–18 Taiwan 257–270

Fischer RA, Maurer R (1978) Drought resistance in spring wheat cultivars. I. Grain yield responses. Aust J Agric Res 29:897–912

Gavuzzi P, Rizza F, Palumbo M, Campaline RG, Ricciardi GL, Borghi B (1997) Evaluation of field and laboratory predictors of drought and heat tolerance in winter cereals. Can J Plant Sci 77:523–531

Gottlieb LD (2003) Plant polyploidy: gene expression and genetic redundancy. Heredity 91:91–92. https://doi.org/10.1038/sj.hdy.6800317

Howell TA, Hiler EA (1975) Optimization of water use efficiency under high frequency irrigation: evapotranspiration and yield relationship. Trans ASAE 18:873–878

Iqbal M, Khan MA, Naeem M, Aziz U, Afzal J, Latif M (2013) Inducing drought tolerance in upland cotton (*Gossypium hirsutum* L.), accomplishments and future prospects. World Appl Sci J 21:1062–1069

ITC (International Trade Centre) (2011) Cotton and climate change: impacts and options to mitigate and adapt. Geneva: ITC xii, 32 p. (Technical paper) Doc. No. MAR-11-200.E

Jamshed M, Jia F, Gong J, Palanga KK, Shi Y, Li J, Shang H, Liu A, Chen T, Zhang Z, Cai J, Ge Q, Liu Z, Lu Q, Deng X, Tan Y, or Rashid H, Sarfraz Z, Hassan M, Gong W, Yuan Y (2016) Identification of stable quantitative trait loci (QTLs) for fiber quality traits across multiple environments in *Gossypium hirsutum* recombinant inbred line population. BMC Genomics 17:197

Jiang C, Wright R, Woo SS, Delmonte TA, Paterson AH (2000) QTL analysis of leaf morphology in tetraploid *Gossypium* (cotton). Theor Appl Genet 100:409–418. https://doi.org/10.1007/s001220050054

Jiang CX, Wright RJ, El-Zik KM, Paterson AW (1998) Polyploid formation created unique avenues for response to election in *Gossypium* (cotton). Proc Natl Acad Sci U S A 95:4419–4424

Kristin AS, Senra RR, Perez FI, Enriquez BC, Gallegos JAA, Vallego PR, Wassimi N, Kelley JD (1997) Improving common bean performance under drought stress. Crop Sci 37:43–50

Kruglyak L (1999) Prospects for whole-genome linkage disequilibrium mapping of common disease genes. Nat Genet 22: 139–144

Levi A, Paterson AH, Barak V, Yakir D, Wang B, Chee PW, Saranga Y (2009) Field evaluation of cotton near-isogenic lines introgressed with QTLs for productivity and drought related traits. Mol Breed 23:179–195

Liang Q, Li P, Hu C, Hua H, Li Z, Rong Y, Wang K, Hua J (2014) Dynamic QTL and epistasis analysis on seedling root traits in upland cotton. J Genet 93:63–78

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

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. https://doi.org/10.1371/journal.pone.0082193

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

Paterson AH, Bowers JE, Burow MD, Draye X, Elsik CG, al JC-X (2000) Comparative genomics of plant chromosomes. Plant Cell 12(9):1523–1540. https://doi.org/10.1105/tpc.12.9.1523

Paterson AH, Smith RH (1999) Future horizons: biotechnology for cotton improvement. In: Smith CW, Cothren JT (eds) Cotton: origin, history, technology, and production. Wiley, Inc, New York pp 415–432

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

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

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

Quisenberry JE, Roark B, Fryer DW, Kohel RJ (1980) Effectiveness of selection in upland cotton in stress environments. Crop Sci 20:450–453

Rajpal VR, Rao S, Raina, S.N. (eds.) (2016) Gene pool diversity and crop improvement. Sustain Dev Biodivers 10 p: 117–122 doi https://doi.org/10.10007/978-3-319-27096-8_1

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

Rosielle AA, Hamblin J (1981) Theoretical aspects of selection for yield in stress and non-stress environment. Crop Sci 21:943–946

Saeed M, Guo W, Ullah I, Tabbasam N, Zafar Y, Mehboob-ur-Rahman, Zhang T (2011) QTL mapping for physiology, yield and plant architecture traits in cotton (*Gossypium hirsutum* L.) grown under well-watered versus water-stress conditions. Electron J Biotechnol 14(3)

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

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

Sezener V, Basal H, Peynircioglu C, Gurbuz T, Kizilkaya K (2015) Screening of cotton cultivars for drought tolerance under field conditions. Turk J Field Crops 20(2):223–232 doi: https://doi.org/10.17557/tjfc.57032

Shen X, Guo W, Zhu X, Yuan Y, Yu JZ, Kohel RJ, Zhang T (2005) Molecular mapping of QTLs for fiber qualities in three diverse lines in Upland cotton using SSR markers. Mol Breed 15:169–181. https://doi.org/10.1007/s11032-004-4731-0

Shen X, Guo W, Lu Q, Zhu X, Yuan Y, Zhang T (2006) Genetic mapping of quantitative trait loci for fiber quality and yield trait by RIL approach in Upland cotton. Euphytica 155:371–380

Simpson DM (1954) Natural cross-pollination in cotton. United States Department of Agriculture, Technical Bulletin No. 1094

Sun FD, Zhang JH, Wang SF, Gong WK, Shi YZ, Liu AY, Li JW, Gong JW, Shang HH, Yuan YL (2012) QTL mapping for fiber quality traits across multiple generations and environments in upland cotton. Mol Breed 30:569–582

Terwilliger JD, Haghighi F, Hiekkalinna TS, Göring HHH (2002) A bias-ed assessment of the use of SNPs in human complex traits. Curr Opin Genet Dev 12:726–734

USDA-FAS (2016) United States Department of Agriculture, Foreign Agricultural Service. https://apps.fas.usda.gov/psdonline/app/index.html#/app/advQuery (accessed 28 December 2016)

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

Wang H, Huang C, Guo H, Li X, Zhao W, Dai B, Yan Z, Lin Z (2015) QTL mapping for fiber and yield traits in upland cotton under multiple environments. PLoS One 10(6): e0130742. https://doi.org/10.1371/journal.pone.0130742

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

Wang P, Zhu YJ, Song XL, Cao ZB, Ding YZ, Liu B, Zhu X, Wang S, Guo W, Zhang T (2012) Inheritance of long staple fiber quality traits of *Gossypium barbadense* in *G.hirsutum* background using CSILs. Theor Appl Genet 124(8):1415–1428. https://doi.org/10.1007/s00122-012-1797-7

Wendel JF, Cronn RC (2001) Polyploidy and the evolutionary history of cotton. Adv Agron 78:139–186

Wright R, Thaxton P, Paterson AH, El-Zik K (1998) Polyploid formation in *Gossypium* has created novel avenues for response to selection for disease resistance. Genetics 149: 1987–1996

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

Yu JZ, Fang DD, Kohel RJ, Ulloa M, Hinze LL, Percy RG, Zhang J, Chee P, Scheffler BE, Jones DC (2012) Development of a core set of SSR markers for the characterization of

*Gossypium* germplasm. Euphytica 187:203–213. https://doi.org/10.1007/s10681-012-0643-y

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. https://doi.org/10.1371/journal.pone.0057220

Zhang Y, Wang XF, Li ZK, Zhang GY, Ma ZY (2011) Assessing genetic diversity of cotton cultivars using genomic and newly developed expressed sequence tag-derived microsatellite markers. Genet Mol Res 10:1462–1470

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

Zheng JY, Oluoch G, Riaz Khan MK, Wang XX, Cai XY, Zhou ZL, Wang CY, Wang YH, Li XY, Liu F, Wang KB (2016) Mapping QTLs for drought tolerance in an F2:3 population from an inter-specific cross between *Gossypium tomentosum* and *Gossypium hirsutum*. Genet Mol Res 15(3). https://doi.org/10.4238/gmr.15038477