CrossMark

Molecular diversity and identification of alleles for *Verticillium* wilt resistance in elite cotton (*Gossypium hirsutum* L.) germplasm

Asena Akkose Baytar · Oktay Erdogan · Anne Frary · Amy Frary · Sami Doganlar

Abstract Cotton is an important crop in the textile, food and pharmaceutical industries. In the present study, a panel of 108 elite cotton (*Gossypium hirsutum* L.) lines was genotyped with 177 genome-wide SSR markers to assess genetic diversity, linkage disequilibrium, population structure and association analyses. A total of 967 loci were assayed and the lines fell into four main groups with a mean genetic distance of 39%. The linkage disequilibrium (LD) decay rate was estimated to be 20–30 cm ($r^2 \leq 0.5$). Association analyses were performed with both general linear model and mixed linear model methods to identify SSR marker loci linked to *Verticillium* wilt resistance. *Verticillium* wilt is a fungal disease that causes huge yield losses in cotton production throughout the world. A total of 26 marker loci distributed on 14 chromosomes were associated with resistance at $p \leq 0.05$. Eight of the 26 associated marker loci were highly significant ($p < 0.01$). The phenotypic variation explained (r^2) by individual markers ranged from 3.2% to 8.2%. Three of the 26 marker loci (JESPR153, JESPR274 and CIR218) were consistent with previous studies. Our results should be useful in improving *Verticillium* wilt resistance in cotton breeding lines.

Introduction

Cotton (*Gossypium* spp.) has been cultivated by humans for thousands of years. As an essential raw material in the pharmaceutical, food and textile industries, cotton is currently grown on over 30 million hectares and almost 105 million 480-lb. bales are produced worldwide (USDA-FAS 2016). The top three producers are India, China, and the United States. Turkey ranks seventh with 2.5% (2.6 million bales) of total worldwide cotton production (USDA-FAS 2016).

The genus *Gossypium*, in the family Malvaceae, consists of approximately 50 species including 45 diploids (n = x = 13) and five allotetraploids (n = 2x = 26). Only four of these species have been

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

O. Erdogan
Department of Biosystem Engineering, Nevsehir Haci Bektas Veli University, Nevsehir, Turkey

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

Springer

domesticated and are cultivated: *G. hirsutum* L. [n = 2x = 26, (AD)$_1$], *G. barbadense* L. [n = 2x = 26, (AD)$_2$], *G. herbaceum* L. (n = x = 13, A$_1$) and *G. arboreum* L. (n = x = 13, A$_2$). *G. barbadense* (Pima cotton) and *G. hirsutum* (upland cotton) are the dominant species providing most of the natural textile fiber in the world (Lee 1984). The chromosomes of diploid cotton are assigned to eight genome types from A to G and K depending on their meiotic size and pairing features (Beasley 1941; Endrizzi et al. 1985). In the cultivated tetraploid cotton species, A subgenome chromosomes are designated A1–A13 and the homeologous D subgenome chromosomes are designated D1–D13 (Wang et al. 2006).

Pathogens severely damage cotton fiber yield and quality resulting in massive costs to agriculture around the world. *Verticillium* wilt (VW) is an especially destructive disease of cotton that was first reported in Virginia, USA, in 1914 (Carpenter 1914). In Turkey, VW was first detected in Manisa in 1941 (Iyriboz 1941). Huge losses in cotton production due to VW have been reported: 75% yield loss in California (Bejarano-Alcazar et al. 1996), 60% in China (Wu and Wu 1997) and 4–12% in Turkey (Erdogan and Kurt 2013). Although there are several species in the genus *Verticillium* that cause disease, *Verticillium dahliae* is the only causative pathogen of *Verticillium* wilt in cotton (Pegg and Brady 2002).

V. dahliae Kleb. is a soil-borne fungal pathogen that has a broad host range of more than 200 dicotyledonous plant species (Meschke et al. 2012) and affects a variety of economically important crops in temperate areas. This fungus produces microsclerotia, resting structures that remain in dead plant tissue and survive in the soil for many years. Germination of microsclerotia is stimulated by root exudates from host species. The fungal hyphae penetrate the roots of the host and colonize the cortex. They then extend into the xylem and form conidia. Conidia spread to various tissues as water moves through the xylem (Meschke et al. 2012). Infection leads to blockage of the vascular system, resulting in characteristic symptoms: leaf wilting, chlorosis, necrosis, and an overall reduction in plant leaf surface area and plant dry weight (Pegg and Brady 2002). The severity of symptoms in cotton depends on genotype, inoculation density of the pathogen, degree of pathogenicity and environmental effects. However, *V. dahliae* Kleb. strains affecting cotton can be divided into two types: defoliating (D) strains and non-defoliating (ND) strains. Whereas D strains are highly virulent and cause complete leaf loss and death of the plant, ND strains are mildly virulent and lead to wilt and limited to no defoliation.

Controlling VW disease is not easy for several reasons: the wide range of host plants, the persistence of microsclerotia in the soil for decades, fungicides specific to *V. dahliae* Kleb. are ineffective once the vascular system is infected, the complexity of resistance mechanisms and the low level of genetic resistance in cotton (Zhang et al. 2014). Although conventional breeding methods have achieved some success in developing resistance in cotton lines, most commercially grown cultivars are susceptible to *V. dahliae* Kleb.. Improving resistance to this disease is crucial and, in recent years, molecular approaches toward this goal have been undertaken.

A number of QTL analyses have identified markers associated with VW resistance. Wang et al. (2014) inoculated interspecific introgression lines with three different *V. dahliae* Kleb. isolates and detected 42 QTLs located on 18 chromosomes. Of these QTLs, 23 increased resistance and 19 decreased resistance to the fungus. In another study, Du et al. (2004) used an intercross population (*G. barbadense* α15-3493 X *G. hirsutum* Shihezi 875) to calculate a genetic distance of 13.1 cm between VW resistance and the SSR marker BNL3556 on chromosome A8. Using a RIL population derived from sensitive upland TM-1 and resistant upland NM 24016, Fang et al. (2014) detected 21 QTLs for VW resistance on 11 separate chromosomes, mostly D chromosomes. In an F$_2$ cross population between *G. hirsutum* CCRI 8 and *G. barbadense* Pima 90 and 53, VW resistance was mapped 13.7 cm away from the chromosome A5 marker BNL3255$_{208}$ (Zhen et al. 2006). The reliability of linked marker BNL3255$_{208}$ was evaluated by inoculating two F$_2$ cross populations derived from susceptible CCRI8 or Han 208 and resistant Pima 90–53 (Wang et al. 2007a, b). Over 80% of the plants that demonstrated resistance carried BNL3255$_{208}$. The aforementioned results demonstrate that QTL mapping is an effective means of identifying markers linked to VW resistance. Unfortunately, none of these linked markers has been reliable enough to perform efficient MAS (marker assisted selection) for VW resistance. Thus, researchers continue to search for molecular marker(s) for this purpose.

In the present study, association analysis was used in a panel consisting of 108 upland cotton cultivars in order to identify QTL regions with roles in VW resistance. It is hoped that the identification of markers linked to VW resistance in cotton will facilitate breeding efforts to improve resistance in the crop. Moreover, diversity and LD (linkage disequilibrium) analyses were performed and the results should help in the selection of suitable parental combinations to meet specific breeding goals.

Materials and methods

Plant materials

A panel consisting of 108 elite cotton materials (*G. hirsutum L.*) frequently grown in Turkey was used in this study (Table S1 and Table S2). The material represents cultivars bred in Turkey, those introduced into the country by seed companies and breeding/germplasm lines. Seeds were provided by Nazilli Cotton Research Institute (NCRI, Nazilli, Turkey).

Verticillium wilt resistance assay

Cotton cultivars were planted individually in pots (5 cm in diameter) with five replicates of each variety. The plants were grown in the growth chamber at 24 ± 1 °C under 12 h daylight period and inoculated with *V. dahliae* Kleb. fungal pathogen at the 4 to 5 leaf stage. ND type pathogen was used because the genotypes were *G. hirsutum* cultivars which do not have resistance to D type pathogen. The highly virulent fungal isolate Vd11 (ND), which is the most common type in cotton-growing areas of Turkey, was grown in Potato Dextrose Agar (PDA) medium. After 14 days of growth, a conidial suspension containing Tween-80 was prepared and used to inoculate the roots of cotton plants by carefully unpotting them and adding 5 ml suspension to the bottom of the pots (Erdogan and Kurt 2013). Five weeks later the inoculated plants were observed for their disease phenotypes. Leaves of each individual were examined and scored from 0 to 4 based on the extent of pathogen symptoms (Bejarano-Alcazar et al. 1996). While a score of 0 represented resistant individuals with no visible symptoms, values up to 1 indicated tolerant individuals (1–33% of the plant showing symptoms),

1–2 indicated moderate tolerance (34–66% symptoms), 2–3 indicated sensitive individuals (67–97% symptoms), 3–4 highly sensitive individuals and 4 represented dead individuals. Disease rate was calculated by averaging the disease scores across replicates (Erdogan et al. 2014). The disease scores were evaluated to calculate CV (coefficient of variance) and LSD (least significant difference) with the JMP 5.0 (SAS Institute, Cary, NC, 5.0 PC version) statistical program at 95% confidence level. Before association analysis was performed the data were transformed whereby each disease score was subtracted from 4 so that higher values represented more tolerance.

DNA extraction

Genomic DNA was extracted from the youngest leaves using a CTAB DNA isolation procedure (Doyle and Doyle 1987). All DNA samples were resuspended in TE buffer before storage at −20 °C. The quantity and quality of DNA were measured using a Nanodrop ND-1000 spectrophotometer.

SSR analysis

A total of 177 pairs of simple sequence repeat (SSR) primers was used to detect polymorphic loci within the population (Table S3). Half of them were selected from previous research in which a core set of SSR markers was developed (Yu et al. 2012). The rest of them were selected to analyze all chromosomes with at least three markers per chromosome. Primer information was obtained from the Cotton Marker Database (www.cottonmarker.org). PCR was conducted in a total volume of 25 µl, containing 10X PCR buffer (50 mM KCl 10 mM, Tris–HCl 1.5 mM MgCl$_2$, pH 8.3), 25 mM MgCl$_2$, 0.2 mM dNTP, 10 pmol of each forward and reverse primer, 0.25 U Taq DNA polymerase, sterile ultrapure water, and $\sim$25 ng/µl DNA. Reactions were performed with 3 min at 94 °C for denaturation, followed by 36 amplification cycles (1 min at 94 °C, 45 s at 55–60 °C, and 1 min at 72 °C) and 10 min at 72 °C for final extension in BIO-RAD Thermal Cycler^{TM}. A Fragment Analyzer^{TM} 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 determination was performed within ± 2 base pair. Polymorphism information content (PIC) values were calculated using Gene Diversity software (GD-dom) (http://plantmolgen.iyte.edu.tr/programs/).

Diversity and population structure analysis

Allelic data were scored dominantly with "1" for presence, "0" for absence and "9" for missing data. To identify genetic distances between cultivars, the DARwin5 (Dissimilarity Analysis and Representation for Windows) computer program (Perrier and Jacquemoud-Collet 2006) was used with the Dice coefficient and the unweighted neighbor-joining algorithm. A Mantel test was used to calculate the correlation between the dissimilarity matrix and the dendrogram. To evaluate discrimination of the individuals between dendrogram groups, PAWS statistics software with Wilk's lambda method was employed (SPSS Inc. Released 2009, PASW Statistics for Windows, Version 18.0, Chicago: SPSS Inc). To detect the structure of the population, STRUCTURE 2.3.4 software, a model-based Bayesian approach, was used (Pritchard et al. 2000). This program employs a 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 cluster determination, cluster numbers (K) from 1 to 10 were tested with 20 iterations each. The Q matrix showing the proportion of assignment to each cluster for each individual was obtained from the STRUCTURE analysis. These data were processed with the STRUCTURE HARVESTER program (Earl and vonHoldt 2012) to visualize the 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 70%. Individuals with a probability lower than 70% 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 SSR markers using TASSEL 2.1 (Trait Analysis by aSSociation, Evolution and Linkage; Bradbury et al. 2007). Before conducting LD analysis, it is important to filter

genotype alignment data to remove minor alleles that can bias LD estimations. Thus, SSR alleles with frequencies below 0.05 were removed using the site filtration function of TASSEL. LD analysis was then performed on the filtered dataset based on the squared allele-frequency correlations (r^2) between marker pairs using the rapid permutation test with 10,000 shuffles ($p \leq 0.01$). 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). LD levels (average r^2) were calculated for all significant SSR pairs (global), linked (on the same chromosomes) and unlinked (on different chromosomes) SSR marker pairs in LD. The LD decay pattern of global marker pairs was generated for pairs with significant LD ($p \leq 0.01$ and $r^2 \geq 0.01$). Chromosomal positions of molecular markers were based on Blenda et al. (2012) and Yu et al. (2012). Assignments of chromosomes to A and D subgenomes were based on Wang et al. (2006).

Association analysis

Linkage analysis between morphological data and SSR allelic data was performed with TASSEL 2.1 using the GLM and MLM models. Different models were evaluated to detect the one with the best fit for association analysis. The Q matrix which shows the quantities of individuals for the best cluster number was calculated with STRUCTURE software. Principal components were calculated with DARwin5. In the first attempt, the GLM model was corrected with Q matrix and principal components (PC), separately. Subsequently, the GLM model was corrected with both Q and PC. For the MLM model, kinship matrix (K) was calculated with TASSEL software. MLM with K was corrected with Q matrix and PC, separately. Lastly, the MLM was corrected with both Q and PC. The p values from seven different models were evaluated with QVALUE software (Storey 2002) to determine the best model. To calculate the estimation of $\pi 0$ values (proportions of null p-values), bootstrapping with a false discovery rate (FDR) of 0.05 was performed. Comparison of $\pi 1$ values ($1 - \pi 0$) was used to determine the proportion of significant p values and thus the best model. The results were assessed according to significance levels of $p < 0.05$.

Results

Verticillium wilt resistance assay

Based on the intensity of pathogen symptoms, tolerant, moderately tolerant and susceptible individuals were identified (Table S1; Fig. 1). Completely dead (disease score of 4) or completely resistant (0) individuals were not observed. TKY9409 (3.44) showed the greatest sensitivity. Carmen and Prestige had disease scores of 0.68 and were the most tolerant. Other tolerant lines were Nazilli 87, Nazilli M503/2, DPL SR 383, BA 308 and PG 2018 with scores ranging from 0.80 to 1.04. CV and LSD values were calculated as 3.44 and 0.55 at a 95% confidence level. Thus, significant variation in disease response existed within the population.

Diversity and population structure analysis

Within the panel of 108 cotton cultivars, 177 SSR markers generated 967 loci with an average of 5.5 alleles per marker. The length of the fragments varied from 76 to 434 bp. Overall average PIC values of markers ranged from 0.01 to 0.50 with an average of 0.25. DOW038 and DPL039 were the most and least informative markers, respectively.

Six individuals were excluded from the diversity analysis because of missing data: Cukurova 1518, Delta Diamond, Gloria, Niab 111, Nazilli 143 and Tomcot 22. Analysis of the remaining 102 cultivars was performed with Unweighted Neighbor-joining analysis. The fit criterion (Mantel test) between the dissimilarity matrix and the tree showed a high
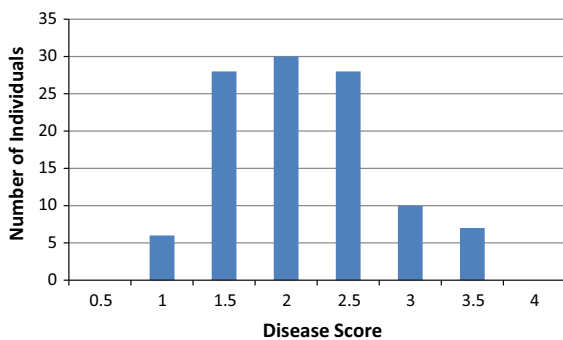

Fig. 1 Frequency of *Verticillium* wilt disease scores in the association mapping population showing a normal distribution of phenotypic variation for the trait

correlation (r = 0.92). The cultivars fell into four main groups: Group 1 (G1), Group 2 (G2), Group 3 (G3) and Group 4 (G4) with 20, 16, 15 and 44 individuals, respectively (Fig. 2). Certain individuals were not assigned to the four main groups and are designated "non" in Table S2. Within the population, the least genetic diversity (23%) existed between TKY 9309 and GC 555. The greatest genetic diversity (60%) was detected between Famosa and Sealand 542. The mean genetic diversity within the population was 39%.

In terms of origin, Australian, Turkish and American material tended to cluster away from each other. To wit, 30% of G1, 73% of G3 and 73% G4 originated from Australia, Turkey and USA, respectively. As expected, cultivars generated from the same breeding program tended to be genetically more similar. For instance, Tashkent cultivars (Uzbekistan cultivars) and Nazilli cultivars (NCRI) clustered together in G3, and all DPL cultivars (Delta and Pine Land Company) clustered in G4.

To see if there was a correlation between disease response and genetically similar cultivars, we combined the results of the *Verticillium* wilt assay and diversity analysis (Fig. S1, Fig. S2). Whereas 65% of G1, 75% of G2 and 80% of G3 individuals were highly to moderately tolerant, 63% of G4 was composed of highly and moderately sensitive individuals. Thus, sensitive individuals tended to fall in G4. Most of the individuals in this group originated from the USA including four of the six highly sensitive individuals. Four of the seven highly tolerant individuals were Turkish cultivars (Nazilli 87, Nazilli M503/2, BA 308 and Pg 2018) and were found in G2 and G3 (Fig. S2). Discriminant analysis between groups and VW resistance yielded a significant r^2 of 0.23, meaning that 23% of the VW resistance data was correlated with DARwin clustering results.

To obtain additional insight into the extent of genetic diversity, population structure was analyzed. The highest delta K (ΔK) was obtained for K = 2 (Fig. S3). Therefore K = 2 was determined as the best cluster number to represent population structure and later used in TASSEL for association analysis. We also obtained a relatively high ΔK value at K = 4 from the Q matrix which coincided with the diversity results from dendrogram analysis. Using the 70% cut-off value, four subgroups were determined: A (38 individuals), B (7 individuals), C (16 individuals) and
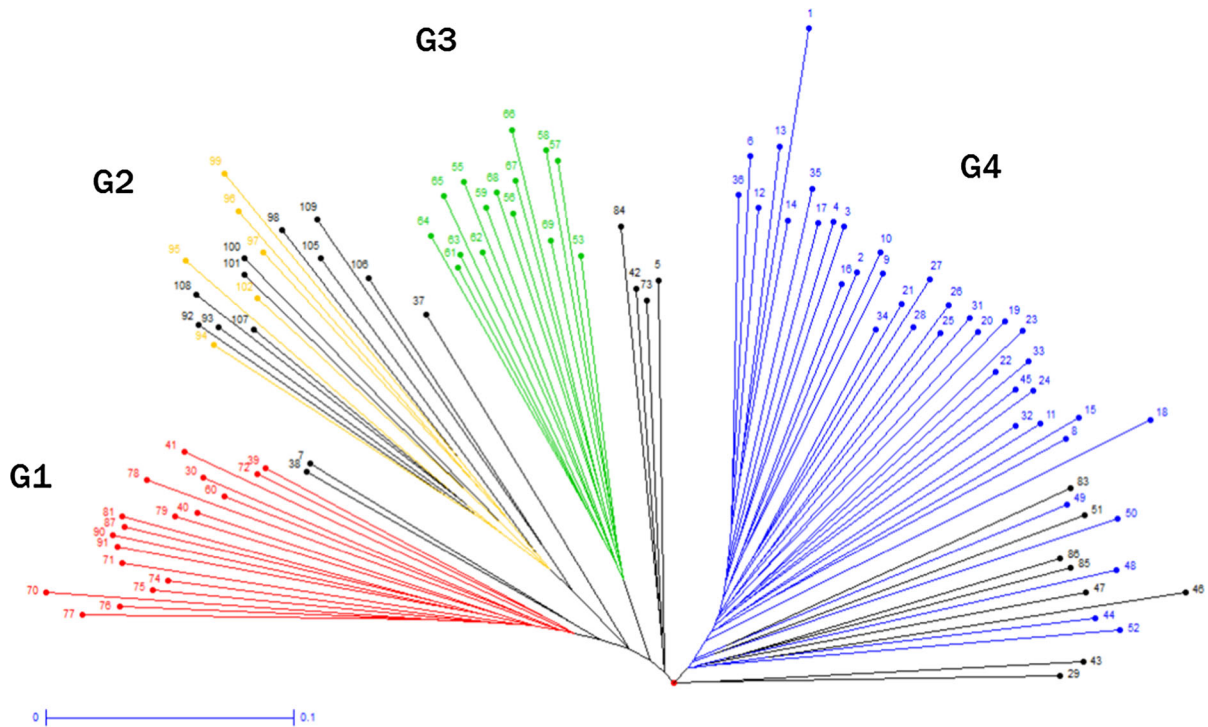
Fig. 2 Dendrogram showing genetic diversity of the population. *G1*, *G2*, *G3* and *G4* show groups 1, 2, 3 and 4, respectively. Demonstration of consistency between genetic diversity and population structure results is shown with *colors*. Subpopulation A is shown in *blue*, Subpopulation B is shown in *yellow*, Subpopulation C is shown in *green*, Subpopulation D is shown in *red* and admixed individuals are shown in *black* on neighbor joining dendrogram. Names corresponding to the numbers are given in Table S2

D (21 individuals). The 26 individuals below the cut-off value were not clearly assigned to any subgroups and named "admixed" (Table S2). Subpopulation assignment was generally consistent with diversity analysis clustering. Thus, 90% of G1, 37% of G2, 100% of G3 and 86% G4 were found in the corresponding subpopulations (Table S2).

Linkage disequilibrium analysis and LD decay

Of a total of 212,639 pairwise comparisons of 967 SSR loci across 108 *G. hirsutum* 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% of marker pairs significant at $p \leq 0.001$. Pairwise estimates of r^2 ranged from 0.06 to 1 for markers located within 0–170 cm of each other. Most 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. However, 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 (Fig. S4). The LD decay rate was determined to be 20–30 cm ($r^2 \leq 0.5$) using the half-length method (i.e. the r^2 value at which r^2 fell below half its maximum value) (Singh and Singh 2015).

Association analysis

Different models were tested to determine the best model for association analysis. GLM (Q) and MLM (K + Q) had the highest proportion of significant p values, 18% (as determined from π1) (Table S4) and generated the exact same results (Table 1). A total of 26 marker loci were found to be associated with VW resistance at $p \leq 0.05$ (Table 1). Of these, 16 marker loci were relatively highly informative (PIC value > 0.3 with an average PIC of 0.4). The associated marker loci were widely distributed on 14 chromosomes; 6 and 8 of the chromosomes belonging to the A and D subgenomes, respectively. Three

Table 1 Markers associated with *Verticillium* wilt resistance in the association mapping panel

Association markers[a]	Marker effect[b]	p values ≤0.05	r^2 marker (%)	F marker	Chromosome location (C)	PIC value
TMB1295$_{225}$	0.2	0.0017	8.1	10.5	C19 (D5)	0.41
JESPR274$_{150}$	0.24	0.0027	8.2	9.5	C23 (D9)	0.28
TMB1295$_{220}$	0.19	0.0029	7.3	9.4	C19 (D5)	0.38
DOW038$_{324}$	−0.15	0.0042	6.6	8.6	C04 (A4)	0.50
DOW038$_{328}$	0.15	0.0042	6.6	8.6	C04 (A4)	0.50
TMB1295$_{259}$	−0.29	0.0069	6.1	7.7	C19 (D5)	0.13
JESPR153$_{150}$	0.16	0.0088	6.0	7.2	C13 (A13)	0.45
DOW059$_{338}$	0.18	0.0099	5.2	6.9	C20 (D10)	0.35
DPL176$_{274}$	−0.19	0.0133	5.2	6.5	C08 (A8)	0.31
DPL717$_{153}$	−0.25	0.0150	5.0	6.1	C21 (D11)	0.16
DPL188$_{156}$	−0.17	0.0150	4.8	6.1	Unknown	0.47
DPL188$_{158}$	0.17	0.0150	4.8	6.1	Unknown	0.47
DPL216$_{233}$	−0.2	0.0155	5.3	6.1	C02 (A2)	0.34
DPL204$_{174}$	−0.18	0.0183	4.9	5.8	C12 (A12)	0.32
DPL204$_{177}$	0.18	0.0183	4.9	5.8	C12 (A12)	0.32
DPL890$_{204}$	0.23	0.0185	4.5	5.7	C26 (D12)	0.16
TMB1295$_{230}$	0.14	0.0216	4.5	5.5	C19 (D5)	0.46
JESPR274$_{280}$	−0.28	0.0263	4.6	5.1	C23 (D9)	0.12
CIR218$_{175}$	0.13	0.0294	4.7	4.9	C22 (D4)	0.47
TMB0514$_{271}$	−0.15	0.0325	4.0	4.7	C02 (A2)	0.33
TMB1910$_{206}$	0.16	0.0366	3.9	4.5	C15 (D1)	0.31
DOW055$_{276}$	0.24	0.0407	3.4	4.3	C14 (D2)	0.11
TMB0799$_{175}$	0.19	0.0409	3.7	4.3	C12 (A12)	0.18
DPL717$_{434}$	−0.18	0.0461	3.4	4.1	C21 (D11)	0.20
DOW003$_{426}$	−0.22	0.0468	3.2	4.0	Unknown	0.14
DPL199$_{244}$	−0.19	0.0500	4.0	3.9	C11 (A11)	0.22

Positive additive effect indicates that the presence of the allele increases the phenotypic value of the trait. Negative value indicates that the presence of the allele decreases the phenotypic value of the trait

[a] The subscripts indicate the fragment length amplified by the marker

[b] Marker effect indicates additive effect due to presence of the allele

resistance markers were located on chromosomes A12 and D5 and two each on A2, A4, D11, and D9. Single markers were detected on A8, A11, A13, D1, D2, D4, D10 and D12. Positions of two of the trait-linked markers are unknown. Fourteen loci showed positive effects and increased VW resistance, ranging from 0.13 to 0.24 and 12 loci showed negative effects on VW resistance, ranging from 0.15 to 0.29. The phenotypic variation explained (r^2) by individual markers ranged from 3.2 to 8.2% (Table 1). Eight of the 26 associated marker loci were determined to be highly significant ($p < 0.01$). TMB1295$_{225}$ on D5 was identified as the most significant marker for increased resistance with a p value of 0.0017 and a PVE value of 8.1%. Of the significantly associated markers, one marker pair was found linked to VW resistance on chromosome A2. On A2, markers TMB0514 and DPL216 are positioned at 56.8 and 62.1 cm, respectively (Blenda et al. 2012; Yu et al. 2012). Both markers resulted in negative additive effects on VW resistance of 0.15 and 0.20, respectively.

Discussion

To identify genetic markers for VW resistance, 108 upland cotton cultivars were subjected to VW inoculation. Disease scores showed a normal distribution

within the panel, indicating that VW resistance is indeed a quantitative trait. Only 6% (7) of the individuals were highly tolerant, with Turkish cultivars making up the highest proportion (57%) in terms of origin.

A total of 177 SSR markers were assayed on the cotton material and produced 967 loci that were used for analysis of genetic diversity. The diversity analysis provided information that will be helpful for breeders. The elite lines fell into four clusters/subpopulations with material from Turkey, the USA and Australia found in distinct clusters suggestive of their different breeding histories. The results also show that most individuals from the same breeding program were genetically close to each other such as DPL 20/DPL90, Acala 5/Acala 1517 and BA 151/BA 308. Therefore introducing greater genetic diversity into breeding programs might be desirable. Overall, the results allow selection of lines for further breeding based on both genetic relationships and VW disease results.

The number of alleles per marker ranged from 2 to 26 and averaged 5.5. Similar numbers of alleles per marker (5.1 and 5.6) were reported in two previous studies of cotton genetic diversity (Zhang et al. 2011; Lacape et al. 2007, respectively). The observed number of alleles per marker is a function not only of germplasm diversity but also of marker type and the method used for fragment separation and resolution (Lacape et al. 2007). Based on analysis using these SSR markers, the cotton panel showed relatively low levels of genetic diversity with an average value of 0.39. This value is comparable to the diversity found in analyses of 158 *G. hirsutum* accessions (0.34; Zhao et al. 2014) and a panel of 53 cotton cultivars (0.40; de Magalhães Bertini et al. 2006) using SSR markers. However, these values are considerably higher than reported elsewhere for SSRs. For example, 285 exotic *G. hirsutum* accessions yielded a much lower value of 0.13 (Abdurakhmonov et al. 2008). Similarly, 378 upland and 3 *G. barbadense* accessions averaged 0.19 similarity (Tyagi et al. 2014). Despite differences across studies, the germplasm of cultivated material generally shows the reduced diversity indicative of a genetic bottleneck. The high degree of similarity is likely due to decades of intensive breeding among cultivars within breeding programs. One consequence of the limited diversity within cotton is that it is difficult to identify associations between molecular markers and agronomically important traits. However,

compared with other studies, our panel was found to encompass a relatively high amount of diversity for cultivated cotton and, thus, was suitable for association mapping.

LD is non-random association of two loci/traits. When two loci are in linkage disequilibrium it means that they are inherited together, non-randomly each generation. In our study, 4.3% of the cotton SSR pairs were in LD. This proportion was lower than other reports in cotton (10.9%: Jia et al. 2014; 11–12%: Abdurakhmonov et al. 2008), opium poppy (6.5%: Celik et al. 2016) and maize (10%: Remington et al. 2001). However, a much higher percentage of LD markers has been reported in durum wheat (52–86%) (Maccaferri et al. 2005) and cultivated barley (45%) (Kraakman et al. 2004). A range of factors affect linkage disequilibrium: genetic drift, natural selection, high mutation rates, migration into the population and gene conversion (Singh and Singh 2015). The high recombination rate in allopolyploid cotton (Brubaker et al. 1999; Leflon et al. 2010) could be a primary contributor to the crop's relatively low LD level. The reasons underlying enhanced recombination in allopolyploid cotton have not yet been clarified.

The LD levels of global and unlinked SSR marker pairs (0.16 and 0.15, respectively) were lower than that of linked markers (0.25) which suggests that physical linkage highly affects LD determination in this population. These findings indicate that the germplasm used in this study was suitable for association analysis. Similar results in cotton were observed by Zhao et al. (2014) using 158 elite cotton germplasm lines which were genotyped with SSR markers.

LD decays at a rate determined by recombination frequencies (cM). LD decay with increasing genetic distance demonstrates that the SSR loci used for association analysis are indeed linked. The decay rate of LD is therefore a useful parameter for assessing marker density and, thus, the quality of QTL detection in association mapping studies. Recombination rates in outcrossing species like cotton are high because of their high level of heterozygosity. Thus, LD generally decays faster in cross-pollinated crops or diverse populations compared to self-pollinated or genetically narrow populations. Because LD extends over longer distances in self-pollinated crops, fewer markers are required for genome coverage. However in cross-pollinated species like cotton, LD extends over shorter distances and consequently higher numbers of markers

are required (Zhou et al. 2012). The total genetic length of the genome of tetraploid cotton is 5200 cM with an average of 400 kb per cM (Paterson and Smith, 1999). In our study the LD decay rate was estimated as 25–30 cM indicating that approximately 175–200 polymorphic markers are required to carry out association analysis properly. We used 177 SSR markers in this study, consistent with the required marker number. To increase LD resolution, the diversity and size of the population could be increased and/or the number of the markers could be augmented.

We identified 26 SSR loci linked to VW resistance. These loci were widely distributed on six A and eight D chromosomes, that is, 57% of the associated markers were on the D subgenome. We did not detect any significantly linked loci on chromosomes 10 and 18. Similarly, of the 193 QTLs identified by Zhang et al. (2015) in a meta-analysis, none were specific to chromosomes 10 or 18. Moreover, the majority (57%) of their QTLs were also mapped to the D subgenome (Zhang et al. 2015). It is reasonable that the QTLs linked to VW reside mostly within the D genome because a higher number of NBS domain-encoded genes with roles in disease resistance have been found in the D subgenome than the A subgenome. In addition, the D subgenome is thought to originate from *G. raimondii* which is immune to a range of pathogens (Li et al. 2014).

It has been difficult to produce reliable and reproducible data from QTL analyses of VW resistance. The highly complex inheritance of the trait is one factor. In addition, differences in the composition and size of mapping populations, inoculation methods and disease scoring as well as the possibility of environmental impact on disease symptoms all contribute to a lack of correspondence in findings from these analyses (Zhang et al. 2014). Nevertheless we have identified some overlap between our results and those of others. Li et al. (2013) compiled a total of 39 cotton markers (34 SSR, 2 DDRT, 2 RGAP and 1 DGAP marker), which were reported to be linked to VW and tested them in a population derived from mass-selection-mass crossing of upland cotton. While only 12 of the markers were polymorphic in their population, they confirmed the usefulness of five SSR markers for marker-assisted selection for VW resistance. One of these markers, JESPR153, was significantly linked (p = 0.0088) to VW resistance in our study. JESPR153 has also been

linked to VW by Jiang et al. (2009). CIR218 was also reported to be associated with VW (Zhang et al. 2015). Similarly we found a positive association between CIR218 and VW resistance. Zhao et al. (2014) performed association mapping analysis in elite cotton germplasm and reported 42 loci associated with VW. One of these, JESPR274, was also linked (p = 0.0027) to VW in our study. We found one SSR marker pair significantly linked to VW: TMB0514-DPL216 on A2. This region could be surveyed with more markers to saturate the area and more narrowly define the genomic region carrying the significant QTL(s). All of the mentioned markers should be considered as candidates for marker-assisted selection studies in the future.

To summarize, we used a diverse collection of elite upland cotton germplasm to identify 26 SSR marker loci linked to VW resistance with positive and negative effects. Of the identified loci, three were consisted with previous work: JESPR153, JESPR274 and CIR218. These marker loci should be useful in efforts to improve *Verticillium* resistance in cotton via marker-assisted selection.

Author contributions   AAB: molecular characterization, data analysis, interpretation of data, manuscript drafting and revision; OE: experimental design, pathogen assays, manuscript revision; AF: conception and design, interpretation of data, manuscript revision; AF: interpretation of data, manuscript revision; SD: conception and design, manuscript revision; All: final approval of the version to be published.

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 92:478–487

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

Beasley JO (1941) Hybridization, cytology, and polyploidy of *Gossypium*. Chron Bot 6:394–395

Bejarano-Alcazar J, Blanco-Lopez MA, Melero-Vara JM, Jimenez-Diaz RM (1996) Etiology, importance, and distribution of *Verticillium* wilt of cotton in southern Spain. Plant Dis 80:1233–1238

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

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

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

Carpenter CW (1914) The *Verticillium* wilt problem. Phytopathology 4:393

Celik I, Camci H, Kose A, Kosar FC, Doganlar S, Frary A (2016) Molecular genetic diversity and association mapping of morphine content and agronomic traits in Turkish opium poppy (*Papaver somniferum*) germplasm. Mol Breed 36:46

de Magalhães Bertini CHC, Schuster I, Sediyama T, de Barros EG, Moreira MA (2006) Characterization and genetic diversity analysis of cotton cultivars using microsatellites. Genet Mol Biol 29:321–329

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

Du WS, Du XM, Ma ZY (2004) Studies on SSR markers of resistance gene of *Verticillium* wilt in cotton. J Northwest Sci Tech Univ Agric For (Natural Science Edition) 32:20–24 **(in Chinese with an English abstract)**

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

Endrizzi JD, Turcotte EL, Kohel RJ (1985) Genetics, cytology, and evolution of *Gossypium*. Adv Genet 23:271–375

Erdogan O, Kurt S (2013) Determination of susceptibility against two different pathotype of *Verticillium dahliae* kleb. on some cotton varieties. Derim 30:36–47

Erdogan O, Kurt S, Gore ME (2014) Studies on the different inoculation methods of *Verticillium* wilt disease caused by *Verticillium dahliae* Kleb. in Cotton. Turk J Agric Nat Sci 2:188–193

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 H, Zhou H, Sanogo S, Lipka AE, Fang DD, Percy RG, Hughs SE, Jones DC, Gore MA, Zhang J (2014) Quantitative trait locus analysis of *Verticillium* wilt resistance in an introgressed recombinant inbred population of Upland cotton. Mol Breed 33:709–720

Iyriboz N (1941) Mahsul Hastalıkları, Ziraat Vekaleti Neşriyatı Umum No: 237

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

Jiang F, Zhao J, Zhou L, Guo WZ, Zhang TZ (2009) Molecular mapping of *Verticillium* wilt resistance QTL clustered on chromosomes D7 and D9 in upland cotton. Sci China C Life Sci 52:872–884

Kraakman ATW, Niks RE, Van Den Berg PMMM, Stam P, Van Eeuwijk FA (2004) Linkage disequilibrium mapping of yield and yield stability in modern spring barley cultivars. Genetics 168:435–446

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

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

Leflon M, Grandont L, Eber F, Huteau V, Coriton O, Chelysheva L, Jenczewski E, Chèvre A-M (2010) Crossovers get a boost in Brassica allotriploid and allotetraploid hybrids. Plant cell 22:2253–2264

Li CQ, Liu GS, Zhao HH, Wang LJ, Zhang XF, Liu Y, Zhou WY, Yang LL, Li PB, Wang QL (2013) Marker-assisted selection of *Verticillium* wilt resistance in progeny populations of upland cotton derived from mass selection-mass crossing. Euphytica 191:469–480

Li F, Fan G, Wang K, Sun F, Yuan Y, Song G, Li Q, Ma Z, Lu C, Zou C et al (2014) Genome sequence of the cultivated cotton *Gossypium arboreum*. Nat Genet 46:567–572

Maccaferri M, Sanguineti MC, Noli E, Tuberosa R (2005) Population structure and long-range linkage disequilibrium in a durum wheat elite collection. Mol Breed 15:271–289

Meschke H, Walter S, Schrempf H (2012) Characterization and localization of prodiginines from Streptomyces lividans suppressing *Verticillium* dahliae in the absence or presence of *Arabidopsis thaliana*. Environ Microbiol 14:940–952

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

Pegg GF, Brady BL (2002) *Verticillium* wilts. CABI Publishing, New York

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

Remington DL, Thornsberry JM, Matsuoka Y, Wilson LM, Whitt SR, Doebley J, Kresovich S, Goodman MM, Buckler ES (2001) Structure of linkage disequilibrium and phenotypic associations in the maize genome. Proc Natl Acad Sci USA 98:11479–11484

Singh BD, Singh AK (2015) Marker-assisted plant breeding: principles and practices. Springer, Varanasi

Storey JD (2002) A direct approach to false discovery rates. J R Stat Soc Series B Stat Methodol 64:479–498

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

Tyagi P, Gore MA, Bowman DT, Campbell BT, Udall JA, Kuraparthy V (2014) Genetic diversity and population structure in the US Upland cotton (*Gossypium hirsutum* L.). TAG 127:283–295

USDA-FAS (2016) United States Department of Agriculture, Foreign Agricultural Service. http://apps.fas.usda.gov/psd online/psdReport.aspx?hidReportRetrievalName=Table+ 04+Cotton+Area%2c+Yield%2c+and+Production&hid ReportRetrievalID=851&hidReportRetrievalTemplateID= 1. Accessed 01 June 2016

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 X, Ma J, Yang S, Zhang G, Ma Z (2007a) Assessment of genetic diversity among Chinese upland cottons with *Fusarium* and/or *Verticillium* wilts resistance by AFLP and SSR markers. Front Agric China 1:129–135

Wang XF, Zhen R, Ma ZY, Zhang GY, Zhang Y, Wang X (2007b) Verification and cloning of SSR marker linked with the gene of *Verticillium* wilt resistance in *Gossypium barbadense* L. J Plant Genet Resour 8:149–152

Wang P, Ning Z, Lin L, Chen H, Mei H, Zhao J, Liu B, Zhang X, Guo W, Zhang T (2014) Genetic dissection of tetraploid cotton resistant to *Verticillium* wilt using interspecific chromosome segment introgression lines. Crop J 2:278–288

Wu F, Wu FA (1997) Resistant response of the new upland cotton varieties to the defoliating strain of *V. dahliae* Kleb. China Cotton 24:11–13

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

Zhang 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

Zhang J, Fang H, Zhou H, Sanogo S, Ma Z (2014) Genetics, breeding, and marker-assisted selection for *Verticillium* wilt resistance in cotton. Crop Sci 54:1289–1303

Zhang J, Yu J, Pei W, Li X, Said J, Song M, Sanogo S (2015) Genetic analysis of *Verticillium* wilt resistance in a back-cross inbred line population and a meta-analysis of quantitative trait loci for disease resistance in cotton. BMC Genom 16:577

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

Zhen R, Wang XF, Ma ZY, Zhang GY, Wang X (2006) A SSR marker linked with the gene of *Verticillium* wilt resistance in *Gossypium barbadense*. Cotton Sci 5:269–272

Zhou H, Muehlbauer G, Steffenson B (2012) Population structure and linkage disequilibrium in elite barley breeding germplasm from the United States. J Zhejiang Univ Sci B 13:438–451