

Construction of a High-density Genetic Map and Identification of Candidate Genes for Resistance to Fusarium Wilt based on the Resequencing of Recombinant Inbred Line Population in *Gossypium Barbadense*

Wanli Han

Xinjiang Agricultural University

Jieyin Zhao

Xinjiang Agricultural University

Xiaojuan Deng

Xinjiang Agricultural University

Aixing Gu

Xinjiang Agricultural University

Duolu Li

Xinjiang Agricultural University

Yuxiang Wang

Xinjiang Agricultural University

Xiaoshuang Lu

Xinjiang Agricultural University

Qianli Zu

Xinjiang Agricultural University

Qin Chen

Xinjiang Agricultural University

Quanjia Chen

Xinjiang Agricultural University

Yanying Qu (✉ xjyyq5322@126.com)

Xinjiang Agricultural University

Research Article

Keywords: single-nucleotide polymorphism (SNP), Quantitative trait locus(QTL), FW, Recombinant inbred lines

Posted Date: January 14th, 2021

DOI: <https://doi.org/10.21203/rs.3.rs-139530/v1>

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

[Read Full License](#)

Construction of a high-density genetic map and identification of candidate genes for resistance to Fusarium wilt based on the resequencing of recombinant inbred line population in *Gossypium barbadense*

Wanli Han⁺, Jieyin Zhao⁺, Xiaojuan Deng, Aixing Gu, Duolu Li, Yuxiang Wang, Xiaoshuang Lu, Qianli Zu, Qin Chen, Quanjia Chen and Yanying Qu^{*}

Engineering Research Centre of Cotton, Ministry of Education/College of Agriculture, Xinjiang Agricultural University, 311 Nongda East Road, Urumqi, 830052, China.

Wanli Han and Jieyin Zhao contributed equally to this work.

Correspondence: xjyq5322@126.com

Abstract

Background: Resistance to Fusarium wilt (FW) is of great significance for increasing the yield of *Gossypium barbadense*. Most published genetic studies on *G. barbadense* focus on yield and fiber quality traits, while there are few reports on resistance to FW.

Results: To understand the genetic basis of cotton resistance to FW, this study used 110 recombinant inbred lines (RILs) of *G. barbadense* obtained from the parental materials Xinhai 14 and 06-146, and Nannong was used to construct a high-density genetic linkage map. The high-density genetic map was based on the resequencing of 933,845 single-nucleotide polymorphism (SNP) markers, and 3627 bins covering 2483.17 cM were finally obtained. The collinearity matched the physical map. A total of 9 QTLs for FW resistance were identified, each QTL explained 4.27-14.92% of the observed phenotypic variation, and qFW-Dt3-1 was identified in at least two environments. According to gene annotation information from multiple databases, promoter homeopathic elements and transcriptome data, 10 candidate genes were screened in a stable QTL interval. qRT-PCR analysis showed that the *GOBAR_DD06292* gene was differentially expressed in the roots of the two parents under FW stress and exhibited the same expression trend in the *G. barbadense* resource materials.

Conclusions: These results indicate the importance of the *GOBAR_DD06292* gene in FW resistance in *G. barbadense* and lay a molecular foundation for the analysis of the molecular mechanism of FW in *G. barbadense*.

Key words: single-nucleotide polymorphism (SNP), Quantitative trait locus (QTL), FW, Recombinant inbred lines

Background:

In addition to being used in textile manufacturing, cotton is an important source of feed, food, petroleum and biofuel components [1]. However, various cotton diseases seriously threaten cotton production[2-4]. Fusarium wilt (FW) caused by the fungus *Fusarium oxysporum* is one of the most serious diseases of cotton[5]. *F. oxysporum* is a soil-borne plant fungal pathogen that causes vascular wilt through root infection in many plants, including important economic crops such as cabbage, banana, cotton, watermelon and tomato[6-12]. This pathogen exists in the soil and crop straw in the form of mycelia, chlamydia or small sclerotia and can survive in the soil for 5-10 years before causing infection[10]. FW usually develops after the seedling stage, and its obvious

symptoms include leaf yellowing, wilting, vascular tissue damage, and ultimately plant death[12].

With the increased yields of various crops, disease resistance is currently one of the most important biological characteristics in the breeding of global crop varieties and has always been an important component of the development of cultivation methods[13-16]. The demand for *Gossypium barbadense* has increased dramatically, making it necessary for cotton researchers to cultivate cotton plants that are highly resistant to FW. Therefore, understanding the genetic basis of resistance to FW in *G. barbadense* will contribute to the efficient breeding of resistant varieties. The control of FW is mainly achieved through the use of chemical fungicides[17]. However, their effects on the soil are often limited by the diversity of pathogens. Therefore, the development and application of disease-resistant varieties is an effective approach for reducing the threat of these pathogens. However, the molecular mechanism of FW infection and the determinants of disease resistance are still largely unclear.

The restriction of genetic recombination events and decreases in genetic diversity increase the difficulty of cross-breeding. With the completion and continuous improvement of cotton genome sequencing, molecular marker-assisted selection (MAS) has been increasingly widely used in cotton hybrid breeding to improve the accuracy and shorten the cycle of breeding[18-21]. For MAS, a high-density, high-quality genetic map is very important, which lays the foundation for locating QTLs of important agronomic traits and further identifying candidate genes in these QTL regions. With the construction of the first genetic map of allotetraploid cotton, much research has focused on the construction of genetic maps and QTL mapping[22,23]. To date, many genetic maps have been constructed using simple sequence repeat (SSR) markers. The first genetic map of cotton was constructed using restriction fragment length polymorphism (RFLP) markers, and many genetic linkage maps were developed for QTL mapping related to lint yield, the yield composition and fiber quality in cotton[24]. Other traits have also been mapped and identified using F₂, BC and recombinant inbred line (RIL) populations of *Gossypium hirsutum* × *Gossypium barbadense*[25-28]. Although the cotton population genetic map has been saturated, many constructed genetic maps are mainly used for QTL mapping within the population of *G. hirsutum*[29-31]. Zhang et al. constructed a multimarker linkage map of *G. hirsutum* from a three-parent compound population, with a total of 978 SSR markers covering 4184.4 cM and an average marker distance of 4.3 cM[32]. Wang et al. used chromosome fragment introduction lines (CSILs) to transfer fiber quality QTLs to *G. hirsutum* for the first time without affecting economic traits[33]. However, a low level of SSR marker polymorphism is a bottleneck in constructing a genetic map of the genome. With the rapid development of DNA sequencing technology and the decline in sequencing costs, the development of large numbers of genome-wide single-nucleotide polymorphism (SNP) markers at the cotton genome level has become inexpensive and reliable. Therefore, it has become possible to use SNP markers to construct a high-quality, high-density genetic linkage map (HDGM) in *G. barbadense*. HDGMs of multiple species have been successfully constructed using resequencing technology[34-37].

To provide more information for breeding programs, the genetic basis of resistance to FW in cotton needs to be further studied. In this study, we used a recombinant inbred line (RIL) population including 110 individual plants from a cross between *G. barbadense* Xinhai14 and 06-146. The purpose of this study was to construct a high-density genetic map using SNP markers by resequencing 110 RIL populations. A total of 3627 bin markers were identified in 26 linkage groups, and a genetic map covering a total map distance of 2483.17 cM was finally obtained. A

stable QTL for resistance to FW in *G. barbadense* was identified in various environments, which has also been reported in previous studies. Enrichment analysis combined with RNA-Seq was conducted to analyze the genes in the D03 region, and 10 candidate genes were finally selected. qRT-PCR was then used to evaluate different tissues and the results of different inoculation experiments in the two parents (Xinhai 14 and 06-146). According to a quantitative analysis of expression in different periods, the *GOBAR_DD06292* gene showed expression differences in the roots of the two parents under FW stress and showed the same expression trend in the *G. barbadense* resource materials. These results indicate that the *GOBAR_DD06292* gene may play an important role in resistance to FW in *G. barbadense*, which provides a molecular foundation for the analysis of the molecular mechanism of FW resistance in *G. barbadense*.

Results

Disease grade in parents and RIL populations

The disease resistance ratings of the two parents and the RIL population were measured in 8 environments. In these 8 environments, the disease grades of the RILs ranged from 0 to 4, with an average of 1.73; the disease grades of the parents Xinhai 14 and 06-146 ranged from 0 to 3 (Table 1). The disease grade of Xinhai 14 was significantly higher ($p < 0.01$) than that of 06-146. The skewness and kurtosis recorded in each environment showed that the disease grade followed a normal distribution in the RIL population (Figure 1, Table 1). In addition, compared with the parents Xinhai 14 and 06-146, superparental segregation of the disease grade was observed in the RIL population. An analysis of variance (ANOVA) showed that the generalized heritability estimate (H_2) of the disease resistance grade was greater than 90.6%, indicating that the disease grade was highly heritable.

Table 1 Disease resistance levels in Xinhai 14, 06-146 and RIL populations

Environment	Parent		RILs	sd	Skewness	Kurtosis	H_2			
	Xinhai 14	06-146						mean	min	max
11AC	2	0.6	**	2.08	0	4	2.09	-0.04	-0.87	94.90%
11AM	1.8	0.4	**	1.69	0	3.4	1.69	-0.25	-0.79	95.40%
11KC	1.6	0.2	*	1.64	0.2	3.2	1.64	-0.23	-0.37	90.60%
12AC	3	0.8	***	1.81	0.6	3.2	1.81	0.08	-0.48	90.70%
12KC	2	0.4	**	1.62	0	3.2	1.62	0.04	-0.22	95.30%
17SM	1.2	0	**	1.56	0	3.4	1.56	0.19	-0.33	95.20%
17AC	2	0.6	**	1.64	0	3.2	1.64	-0.30	-0.36	93.30%
18AC	1.8	0.4	**	1.74	0	3.8	1.74	0.10	-0.17	90.70%

11AC: Disease grade in the adult stage in Alar in 2011; 11KC: disease grade in the adult stage of Korla in 2011; 12AC: disease grade in the adult stage in Alar in 2012; 12KC: disease grade in Korla in 2012; 11AM: disease level in the seedling stage in Alar in 2011; 17SM: disease level in the indoor seedling stage in 2017. 17AC: disease grade in the adult stage in Alar in 2017. 18AC: disease level in adult plants in Aksu in 2018.

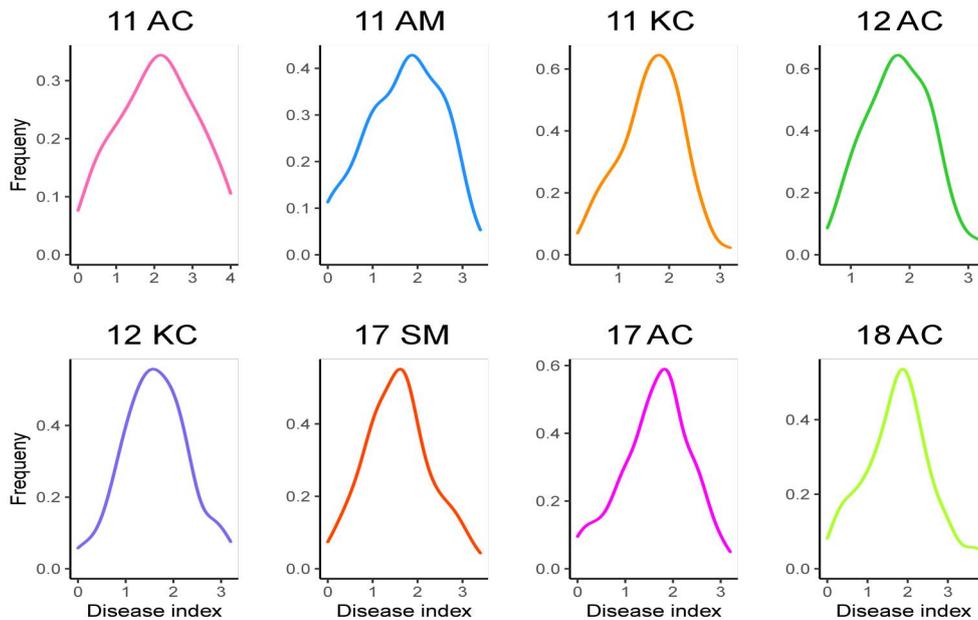


Figure 1 Frequency distribution of disease resistance levels in 110 RILs in different environments

Statistics of DNA-seq data

By resequencing the parents and the 110 RIL population, the sequencing depth for the parent Xinhai 14 was 13 \times , while that of 06-146 was 16 \times , and the average sequencing depth for the parents was 14.5 \times ; the coverage of the reference genome was 97%. The progeny sequencing depth was 7 \times , and the genome coverage was 95% (Table S1). For Xinhai 14, a total of 30.95 Gbp of clean data were obtained, while a total of 38.06 Gbp of clean data were obtained for 06-146. The total data volume for the 110 progeny was 1941.44 Gbp, and the total amount of data was 2010.45 Gbp. The Q30 value exceeded 90%, and the GC content was greater than 34.72% (Table S2).

Information marked by SNPs

Using the clean reads, a total of 18,446,57 SNPs were obtained in Xinhai 14; the conversion-type category included 1,257,644 SNPs, the number of conversion-type SNPs was 587,013, the number of heterozygous SNPs was 394,393, and the number of pure SNPs was 1,450,264. A total of 1,844,657 SNPs were obtained in 06-146, including 1,257,644 of the conversion type, 587,013 of the transversion type, 483,400 of the heterozygous type, and 1,361,257 of the pure type (Table S3).

Construction of a linkage genetic map

To ensure the quality of the map, the markers that were homozygous and discordant in the parents were filtered out, the depth of the parental markers was less than 10X, and the markers located on a chromosome were removed. The offspring were genotyped in relation to the parents and a high depth of parental sequencing ensured the correctness of the offspring typing results. For the 933,845 filtered SNPs, a window size of 15 SNPs was adopted, with 1 SNP as the step size, for a sliding window scan along the chromosome. When the number of SNP types in the sliding window was greater than or equal to 11, the type was recorded as aa. When the number of SNP types in the sliding window was greater than or equal to 11, the type was recorded as bb. In other cases, ab was used for genotype filling and correction. After completing marker filling and correction, bin division was performed according to the reorganization of the offspring. The

samples were arranged neatly according to the physical position of the chromosome. When a typing change was observed in any sample, a recombination breakpoint was considered to have occurred. The SNPs between recombination breakpoints were classified into bins. No recombination event was considered to have occurred in the bin. Finally, the bins were used as the mapped markers for the construction of a genetic map. In total, 3627 bins were finally obtained. The bins were used as a marker for subsequent analysis.

A total of 3627 bins were used to perform genotype analysis in 110 offspring (Figure 2). Several chromosomes in the genomes of most strains showed no recombination; that is, they were completely derived from a single parental genome. Some strains also exhibited heterozygous chromosomal segments, which may be related to incomplete or incorrect repair after chromosome exchange. The bins were divided into 26 linkage groups based on the available information, the linkage group was used as a unit, the linear arrangement of the markers in the linkage group was obtained via HighMap software analysis, the genetic distance between adjacent markers was estimated, and finally the total genetic map distance was 2483.17 cM (Table 2). The average number of bins in each linkage group was 138.5, among which D13 contained the fewest bins (53), and A06 contains the most (342 bins). The average distance between 26 linkage groups was 95.51 cM, D13 showed the shortest distance of 41.17 cM, and A01 showed the longest distance of 148.04 cM. The largest gap between A10 and D13 was 19.3 cM. The average proportion of gaps of less than 5 cM in the 26 linkage groups was 96.71%.

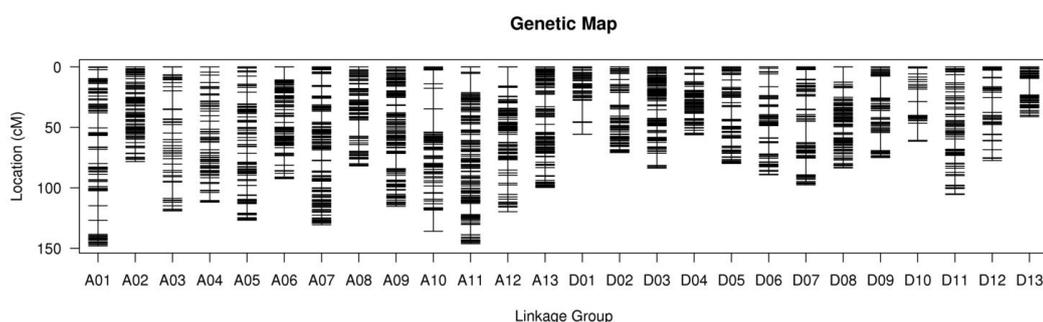


Fig. 2 The distribution of SNP markers in each of the 26 linkage groups/chromosomes.

Table 2 Detailed information on the SNP markers in the genetic map

LG ID	Total BinMarkers	Total Distance(cM)	Average Distance(cM)	Max Gap(cM)	Gaps<5cM(%)
A01	213	148.04	0.7	12.41	96.70%
A02	136	78.32	0.58	4.42	100.00%
A03	75	119.06	1.61	14.97	90.54%
A04	81	110.86	1.4	6.67	93.75%
A05	141	126.76	0.91	9.8	95.71%
A06	342	92.44	0.27	10.76	99.41%
A07	198	130.55	0.66	9.8	96.45%
A08	178	81.89	0.46	13.38	98.87%
A09	195	115.37	0.59	12.61	98.45%
A10	105	135.91	1.31	19.3	95.19%
A11	236	146.22	0.62	15.9	98.72%
A12	145	119.78	0.83	15.83	95.83%

A13	192	99.73	0.52	10.24	97.91%
D01	105	55.68	0.54	17.76	98.08%
D02	110	70.88	0.65	10.48	98.17%
D03	163	83.75	0.52	10.08	98.77%
D04	153	56.17	0.37	6.09	99.34%
D05	89	79.87	0.91	10.48	96.59%
D06	88	89.21	1.03	10.57	94.25%
D07	96	97.56	1.03	12.62	95.79%
D08	138	83.58	0.61	18.12	98.54%
D09	110	74.99	0.69	14.17	97.25%
D10	70	61.44	0.89	8.5	95.65%
D11	131	105.48	0.81	12.22	96.92%
D12	58	77.46	1.36	13.72	89.47%
D13	53	41.17	0.79	19.3	98.08%
Total	3,601	2,483.17	0.79	19.3	96.71%

Quality assessment of genetic map

The results of collinearity analysis based on the positions of the markers in the genome and the genetic map in the above figure (Figure S1) showed that the SNP markers in the 26 linkage groups presented higher coverage in the cotton genome. The lowest Spearman correlation coefficient of the 26 linkage groups was 0.80 on chromosome D05, and the highest was 0.99 on chromosome D13, with an average of 0.96. However, with the exception of A12 and D05, showing a coefficient lower than 0.9, the Spearman correlation coefficients of the remaining 24 linkage groups are all higher than 0.9 (Table S4). These results indicated that the order of most of the markers in the linkage group was highly consistent with the sequence of the Nannong reference genome. These findings showed that the quality of the map was reliable.

Mapping of QTLs for resistance to FW in island cotton

According to the results of R/QTL mapping, the composite interval mapping method was employed to locate QTLs for FW resistance in *G. barbadense*, and such QTLs were detected on 7 chromosomes (A01, A05, A07, A09, D03, D05, D09) in 6 environments (Table 3). The explained phenotypic variation ranged from 1.037%-19.654%, and the LOD value ranged from 2-5.996. Among the different environments, the LOD values for the 11KC and 12AC environments were not located in the specified interval. qFW-Dt3-1, identified in 11AC, showed the widest range of phenotypic variation.

qFW-Dt3-1 was identified in 11AC and 17AC, and the corresponding chromosomes were the same as those reported by Wang[38]. Therefore, we used qFW-Dt3-1 as a stably inherited QTL for candidate gene mining.

Table 3 Summary of FW QTLs identified in different environments

QTL name	Environment	Left marker	Right marker	LOD	ADD	PVE
qFW-Dt3-1	AC11	Block6644	Block6655	4.836	-0.517	19.654
qFW-At7-1	AM11	Block2022	Block2027	2	-0.165	4.032
qFW-Dt5-1	AM11	Block7250	Block7246	2	0.154	3.491
qFW-At9-1	KC12	Block2986	Block2987	2.5	0.162	10.144
qFW-Dt3-2	KC12	Block6766	Block6767	2.5	-0.113	4.93

qFW-At1-1	SM17	Block89	Block91	3	0.225	8.248
qFW-At5-1	AC17	Block1451	Block1453	2	0.077	1.037
qFW-Dt3-1	AC17	Block6649	Block6654	2	-0.223	8.782
qFW-Dt9-1	AC18	Block8062	Block8065	2	-0.328	11.399

Screening of candidate genes for stable qtl interval

According to the previous mapping results, qFW-Dt3-1 was located on 11AC and 17AC, and the intersecting gene observed in these two environments was selected as the candidate gene for resistance to FW in *G. barbadense*. In the qFW-Dt3-1 interval, the total number of encoded genes was 166. First, 148 genes were compared and enriched in the NR, SwissProt, GO, COG, and KEGG databases (Table S5). Based on the annotation information, we directly excluded genes that were not related to plant resistance, and we finally selected 37 genes related to protein phosphorylation, carbohydrate metabolism processes, transferase activity, redox processes, phenylpropanoid biosynthesis, plant-pathogen interactions, starch and sucrose metabolism and the MAPK signaling pathway for the next step of the analysis. Soon after examining the transcriptome data of the parents and selecting highly FW-resistant and highly FW-susceptible materials in the 110 RIL population (Figure S2), it was found that the expression patterns of these 37 genes could be clearly divided into 2 types. There was no obvious difference in the first type of expression. Only the 8 genes of the second category showed significant differences in expression.

qRT-PCR

Adversity-related stress can cause transcriptome reprogramming events. Based on the analysis of the transcriptome expression profile combined with the identified promoter cis-acting elements (Table S6), it was speculated that 10 of the 37 genes (*GOBAR_DD02552*, *GOBAR_DD02562*, *GOBAR_DD06292*, *GOBAR_DD07063*, *GOBAR_DD07508*, *GOBAR_DD09955*, *GOBAR_DD23790*, *GOBAR_DD_34378*, *GOBAR_DDDD889*) may be involved in resistance to FW in *G. barbadense*. To further study the roles of these candidate genes in FW stress, qRT-PCR was used to detect the transcription levels of these 10 genes in the roots of the two parent seedlings under the conditions of FW stress. Under FW stress, compared with the results obtained at 0 h, genes other than *GOBAR_DD07508* exhibited obvious expression differences in the roots (Figure 3). This indicated that the transcription levels of these candidate genes were significantly increased after FW stress. However, only 7 genes (*GOBAR_DD02552*, *GOBAR_DD06292*, *GOBAR_DD07063*, *GOBAR_DD09955*, *GOBAR_DD28836*, *GOBAR_DD34889*, *GOBAR_DD37843*) showed significant differences in expression during the same period following inoculation in the two parents. In summary, these seven genes may play a role in the response to FW invasion in *G. barbadense*, rather than the expression of genes related to the growth and development of cotton.

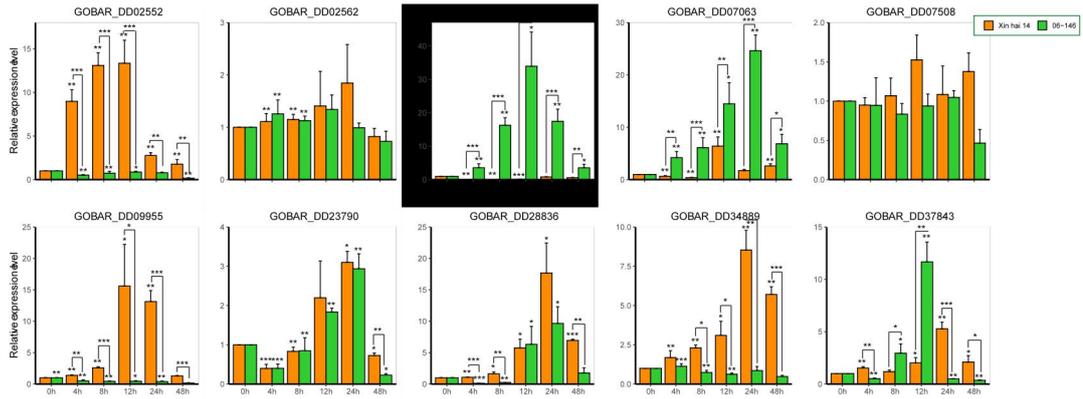


Fig. 3 Expression profiling of the 10 candidate genes related to FW. The error bars represent the means of three technical replicates \pm SEs. Statistically significant differences from the control group are indicated with *P < 0.05; **P < 0.01

Analyses of the expression of genes in different regions and the different molecular biological functions of the genes were performed. Regardless of whether a plant is under any stress, the roots are the initial sensory plant component. For this reason, we selected the most significant period before and after FW to further verify the tissue-specific expression of these 7 candidate genes in the two parents. Their expression was generally highest in the roots, both before and after stress (Figure 4). Interestingly, these 7 genes exhibited significant expression changes in different tissues before and after FW stress. Among these genes, three genes *GOBAR_DD06292*, *GOBAR_DD09955* and *GOBAR_DD34889*, showed significant changes in the roots but showed the opposite expression patterns in the two parents, which was consistent with the previous analysis results.

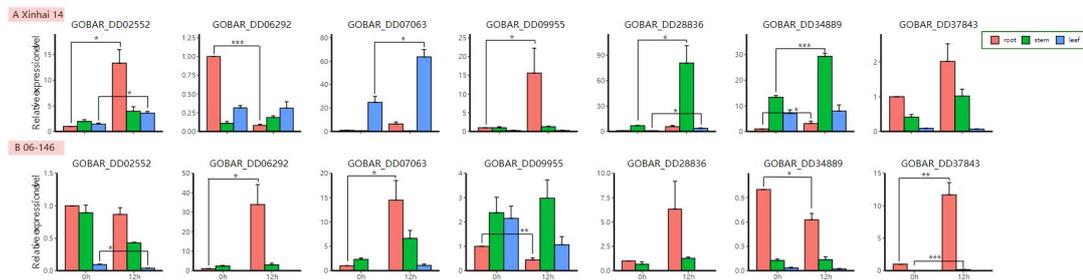


Fig. 4 Tissue-specific expression analysis of the 7 candidate genes related to FW. The error bars represent the means of three technical replicates \pm SEs. Statistically significant differences from the control group are indicated as *P < 0.05; **P < 0.01

To further verify the functions of these 3 genes after *G. barbadense* was subjected to FW stress, based on results from our research group, 4 disease-resistant *G. barbadense* materials and 4 disease-susceptible *G. barbadense* materials (Figure 5) were used to study prevalent gene functions. It was found that only *GOBAR_DD06292* expression was significantly higher in resistant materials than in susceptible materials. The linear regression equation between the disease resistance grade (X) and expression levels (Y) in 8 materials was as follows: $y = 36.9 - 15.5x$, $R^2_{adj} = 0.76$. In summary, we speculate that the expression level of *GOBAR_DD06292* is positively correlated with resistance to FW of *G. barbadense*. The results further revealed that the *GOBAR_DD06292* gene exhibited a significant difference in expression after the two parents were subjected to FW stress and showed the same trend of gene expression and the same relationship with disease resistance in *G. barbadense* resource materials. It is very likely that when *G. barbadense* is subjected to FW stress, the significant change in its expression in the roots will enhance resistance to FW in *G. barbadense*. These results further indicate the importance of the *GOBAR_DD06292* gene in resistance against FW in *G. barbadense* and lay a molecular foundation for the analysis of the molecular mechanism of resistance to FW in *G. barbadense*.

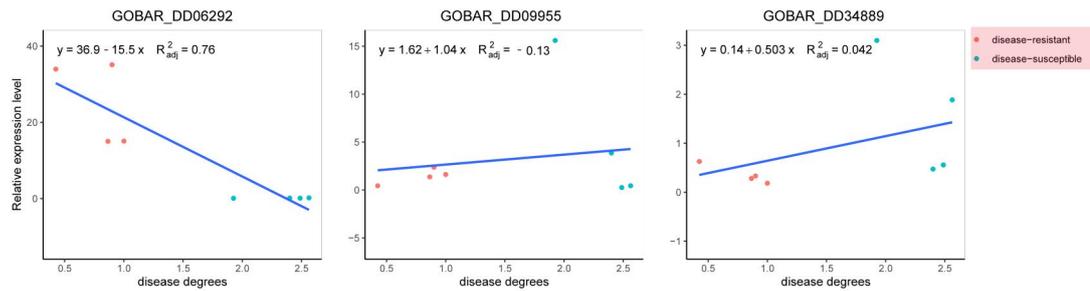


Fig. 5 Expression patterns of 3 candidate genes in mixed apical organs of eight varieties of natural population in FW stress, detected by qRT-PCR; red and blue dots indicate four disease-resistant cotton and disease-susceptible cotton, respectively.

Discussion

A detailed genetic map plays a key role in the identification of QTLs. Recently, many genetic maps have been constructed by using PCR-based markers in cotton. For example, Sun et al. constructed a genetic map with a total length of 959.4 cM containing 155 SSR markers[39]. Shang et al. used 581 SSR markers to construct a genetic map, and the average marker distance between two adjacent genes was approximately 6.39 cM[40]. Liu et al. used 1675 SSR markers to construct a genetic map with a total genetic distance of 3338.2 cM[41]. However, the number of SSR markers was not sufficient to construct a genetic map of high-density cotton. Through the use of SNP markers, the marker density in the cotton genetic map has been greatly increased. Jia et al. used 6295 SNP markers and 139 SSR markers to construct a genetic map with a total length of 4,07.98 cM[42]. Zhang et al. used 5521 SNP markers to construct a genetic map covering 3259.37 cM, with an average of 0.78 cM between adjacent markers[43]. Sun et al. used 3978 SNP marker to cover a total genetic distance of 2483.17 cM[44]. Therefore, SNP markers are more effective for constructing genetic maps. Many of the genetic maps constructed for cotton come from *G. hirsutum*.

In this study, we resequenced 110 *G. barbadense* RIL populations for the first time and constructed the first high-density genetic map of a *G. barbadense* RIL population using 933,845 high-quality SNP markers. There were a few chromosome in the genomes of most strains that showed no recombination; that is, they were completely derived from a single parental genome. Some strains also exhibited heterozygous chromosomal segments, which may be related to incomplete or incorrect repair after chromosome exchange. Although the number of RILs in the population was small, the RIL population showed a higher recombination rate and higher position resolution than the F2 population. The population traits were normally distributed, indicating that the population was suitable for constructing a genetic map and identifying QTLs for FW resistance in *G. barbadense*.

To date, most of the published cotton genetic research has focused on characteristics such as yield and fiber, while few studies have focused on disease. Bolek et al. mapped F₂ populations showing high resistance to FW (Pima S-7) and susceptibility to FW (Acala 44) and used SSR to reveal polymorphism the between susceptible and highly resistant parents. A total of 225 SSR marker pairs were screened in a bulk composed of 10 resistant progeny and 10 susceptible progeny. Sixty markers were utilized to break down QTLs, and 11 linkage groups were built using 35 markers covering 531 cM, with a mean distance of 15.17 cM[45]. Wang et al. (2009) used molecular mapping technology to identify a gene for FW resistance, which was closely linked to

the SSR marker "JESPR304-", and designated the gene FWR. Simultaneously, 4 FW resistance QTL were analyzed: A7, D3, D9 (c23) and D1 (c15)[38]. In this study, a total of 8 environmental phenotypic values were used to locate 8 QTLs for resistance to FW in *G. barbadense*. Combining the functional annotations of genes within stable QTLs and orthologous genes in *Arabidopsis* with transcriptome data is an effective method for mining candidate genes. The candidate genes were further verified via transcriptome analysis combined with qRT-PCR. Finally, the *GOBAR_DD06292* gene was found to be differentially expressed after the two parents were stressed by FW invasion and showed the same expression trend in *G. barbadense* resource materials. We found that the homologous gene of *GOBAR_DD06292* in *Arabidopsis thaliana* was *AT3G10300*, and the homology was 80.50. SA, JA and ET treatments applied to *Arabidopsis* plants induced the expression of this gene. These results further indicated the importance of the *GOBAR_DD06292* gene in resistance to FW in *G. barbadense* and lay a molecular foundation for the analysis of the molecular mechanism of resistance to FW in *G. barbadense*.

Materials and Methods

Plant materials

In this study, single-grain F₂ plants were obtained by crossing the FW-susceptible *G. barbadense* material Xinhai 14 with the FW-resistant material 06-146. The obtained plants were then propagated to the 7th generation to form an RIL population containing 110 lines.

Planting method and character inspection

The two parents and 110 RIL lines were grown in 8 environments at 3 locations. The experiments were carried out at the experimental base of the Xinjiang Academy of Agricultural Sciences in the 16th Regiment of the First Agricultural Division of Alar City, Xinjiang, and the experimental site of the Xinjiang Korla Academy of Agricultural Sciences from 2011 to 2012. The experiment aimed at disease resistance identification was conducted in the Sixteenth Regiment of the First Agricultural Division of Alar City over 3 consecutive years from 2016 to 2018. The planting plot length was 3 m, the row spacing was 0.35 m, and the plant distance was 0.10 m. IN each setting, three repeats were conducted, and field management was routine. In 2018, a duplicate control was set up in the cotton cultivation room of Xinjiang Agricultural University to conduct an indoor identification experiment. The tested cotton seeds identified in the laboratory were broadcast live in a nutrient bowl, and the nutrient bowl was filled with vermiculite and sterilized nutrient soil at 1:2, 10 holes per bowl, 2 seeds per hole, and 3 groups of repeated controls were set up. When the seedlings grew to the 2-true-leaf stage, the redundant seedlings were removed. When the cotyledons of the seedlings were fully expanded, the seedlings were inoculated via the root-cutting method. The noninoculated material was used as the CK control, and disease was observed at the 2-leaf and 1-heart stages.

In all environments, the foliar disease area was used as an index to classify the disease. Five normally growing plants were randomly selected in each replication zone for the leaf surface disease area test. SPSS 23.0 was used to perform statistical analysis of the RIL population, including an ANOVA of disease value or the relative disease index in different environments[46]. The R software "qtl" package was used for QTL mapping[47].

DNA library preparation

In July 2017, the young leaves of two pairs of the parents and the 110 RIL populations were collected, and the young leaves were stored in a refrigerator at -70°C. Using the miniprep method,

genomic DNA was extracted from the two parents and each individual RIL.

Development of filtering quality control and labeling of sequencing data

The original sequences (raw reads) obtained by sequencing were filtered to obtain clean reads. The main data filtering quality control steps were as follows: remove adapter sequences; if the proportion of N bases in a read (the specific base type cannot be determined) is greater than 10%, then filter out paired-end reads; remove low-quality reads (in which the proportion of bases with a Q quality value of less than or equal to 10 accounts for more than 50% of the entire read). The positions of the clean reads in the reference genome of *G. barbadense* Nannong were compared using Burrows–Wheeler Alignment (BWA) software, the sequencing depth and genome coverage of each sample were calculated, and mutation detection was performed[48]. According to the results of clean read localization in the reference genome, Picard's Mark Duplicate tool was used to remove duplications and shield the impact of PCR duplications, and GATK was used for indel realignment (i.e., to locally recompare of the sites located near the indel comparison results)[49]. The error in the comparative results caused by indels was corrected to ensure the accuracy of the detected SNPs. When using GATK for variant calling, strict filtering was performed with the SNP: SNP cluster filter (if there were 2 SNPs within 5 bp, then the sequence was filtered out). Thus, the final SNP site set was produced.

Construction and quality assessment of the genetic map

The linkage map of the RIL population was constructed using the *G. barbadense* Nannong reference genome database. HighMap software was used to correct genotyping errors and sort sequences along chromosomes[50]. In addition, SMOOTH was used to correct errors based on parental genotype contributions, and the k-nearest neighbor algorithm was used to correct incorrect genotypes[51]. The multipoint maximum likelihood (MML) method was used to add the following distortion markers to the linkage map. Map distance was calculated with the Kosambi mapping function. The chi-square test was used to identify separation distortion marks. As mentioned in previous studies, the separation of distortion markers (SDMs) in the HDGM was acceptable[43]. The SNP markers in the linkage map were aligned with the genome using the local BLAST method.

To ensure the quality of the map, markers that were homozygous and showed inconsistency in relation to the parents were selected. The marker depth for the parents was at least 10X; markers that were not located on chromosomes were removed. The offspring were genotyped in relation to the parents, where the high depth of parental sequencing ensured the correctness of the offspring typing. Each sample was neatly arranged according to the physical location of the chromosome. When there was a typing change in any sample, a recombination breakpoint was considered to exist. A SNP between recombination breakpoints was classified as a bin. No recombination events were considered to occur within a bin. Finally, the bins were used as markers for mapping to construct a genetic map. Bins with a length of less than 10 kb were screened out. The polymorphic markers with severe partial separation (chi-square test $P < 0.01$) were filtered.

QTL positioning

The phenotypic values recorded in 8 different environments were analyzed with the R-3.62 software package "qtl". A strict LOD threshold was calculated through permutation experiments. The parameters for QTL positioning were set as follows: calculation time of 1000, I type error p value of 0.05, PIN of 0.001, and mapping step length of 1.0 cm. If the same QTL was identified in two or more environments where the LOD threshold was greater than 2, it was considered a significant QTL. The QTL confidence interval (95%) was set as the location distance interval corresponding to a 1 LOD drop on both sides of the peak. When the confidence intervals overlapped, the QTLs detected in two or more environments were considered stable.

Screening of candidate genes

To obtain potential candidate genes, BLAST software was used to extract gene sequences within stable QTLs, and the genes in the associated regions were compared with the NR, SwissProt, GO, COG, and KEGG databases[52]. To reveal the general expression pattern of the candidate genes, the transcriptome sequencing data from the parents (Xinhai14×06-146) were used as a reference. The roots and young leaves of the two parents were sampled at the three-true leaf-stage, and the expression levels of candidate genes were detected by qRT-PCR.

qRT-PCR

According to the cDNA information for the candidate genes, Primer 5.0 software was used to design primers in a specific regions at the 5' end or 3' end of the gene sequence. Root tissue cDNA was used as a template, and the expression of candidate genes was detected by qRT-PCR. Each sample was set to 3 replicates, and the internal reference gene was GbUBQ7. A total RNA extraction kit (Tiangen, China) was used to extract RNA, and 5×All-In-One Mastermix (Abm, Canada) was used for reverse transcription. Real-time PCR amplification was performed on an ABI 7500 Fast system. A BrightGreen 2×Qpcr Mastermix (Abm, Canada) kit was used according to the provided method, and the total amplification volume was 20 μL. The reaction program was as follows: predenaturation at 94°C for 30 s, followed by 40 cycles of denaturation at 95°C for 5 s, annealing at 57°C for 5 s, and extension at 72°C for 34 s. The results were used for relative quantitative analysis via the 2- $\Delta\Delta$ Ct method[53].

Supplementary information

Additional file 1: Table S1. sequencing depth of the parents and RIL populations

Additional file 2: Table S2. Sequencing data statistics for each sample

Additional file 3: Table S3. SNP statistics detected between parents

Additional file 4: Table S1. Genetic linkage map and genome collinearity map

Additional file 5: Table S4. Correlation coefficient of each linkage group

Additional file 6: Table S5. Annotation information

Additional file 7: Table S6. Cis-acting element analysis of 37 candidate genes

Additional file 8: Figure S2. Expression analysis of the 37 genes from the parent and RIL populations under Fusarium wilt stress based on transcriptome data

Additional file 9: Table S7. All primers used in this study

Abbreviations

FW:Fusarium wilt;RIL:Recombinant inbred line;MAS:marker-assisted selection;SSR:simple sequence repeat;RFLP:restriction fragment length polymorphism;SNP:single-nucleotide polymorphism;HDGM:high-density genetic linkage map;ANOVA:An analysis of

variance;H₂:heritability estimate;BWA:Burrows–Wheeler Alignment;MML:multipoint maximum likelihood;SDM:separation of distortion marker

Declarations

Ethics approval and consent to participate

Not applicable.

Consent for publication

Not applicable.

Availability of data and materials

The genome databases were downloaded from Nannong (<http://database.chgc.sh.cn/cotton/index.html>).

Competing interests

The authors declare that they have no competing interests.

Funding

This work was supported by the Major Special Project of the Xinjiang Uygur Autonomous Region (2020A01002-2) and New Cultivar Breeding of Transgenic Herbicide Resistance in Cotton (2020ZX08005-005). The funding agencies played no role in the design of the study, data collection or analysis, or interpretation or writing of the manuscript.

Authors' contributions

Conceived and designed the experiments: WLH, QJC and YYQ. Collected public datasets and performed experiments: HWL, JYZ, DLL, YXW, QC, QLZ, XSL and XJD. Analyzed the data: WLH, JYZ, XJD, AXG, QJC and YYQ. Wrote the manuscript: HWL and ZJY. Revised the manuscript: HWL, JYZ, QJC and YYQ. All authors read and approved the final manuscript

Acknowledgments

Not applicable.

References

1. Xiongming D , Shouye L , Junling S , et al. Dissection of complicate genetic architecture and breeding perspective of cottonseed traits by genome-wide association study[J]. *Bmc Genomics*, 2018, 19(1):451-.
2. Tian, J., et al., Expression of baculovirus anti-apoptotic genes p35 and op-iap in cotton (*Gossypium hirsutum* L.) enhances tolerance to verticillium wilt. *PLoS One*, 2010. 5(12): p. e14218.
3. Syed Shan [∗] | li Zaidi, Naqvi R Z , Asif M , et al. Molecular insight into cotton leaf curl geminivirus disease resistance in cultivated cotton (*Gossypium hirsutum*)[J]. *Plant Biotechnology Journal*, 2020, 18(3).
4. Xiao-Kun W , Chun-Yan W , Cheng-Jian X , et al. Advances in Molecular Mechanisms of Verticillium Pathogenicity and Plant Resistance to Verticillium Wilt[J]. *Journal of Henan Agricultural Sciences*, 2014.
5. Yao Z , Chen Q , Chen D , et al. The susceptibility of sea-island cotton recombinant inbred lines to *Fusarium oxysporum* f. sp. *vasinfectum* infection is characterized by altered expression of long noncoding RNAs[J]. *Scientific Reports*, 2019, 9(1).
6. Miaomiao, Xing, Honghao, et al. Transcriptome Profiling of Resistance to *Fusarium oxysporum* f. sp. *conglutinans* in Cabbage (*Brassica oleracea*) Roots.[J]. *PloS one*, 2016.
7. Li X , Zhang, Ya'nan, Ding C , et al. Temporal patterns of cotton *Fusarium* and *Verticillium* wilt in Jiangsu coastal areas of China[J]. *Scientific Reports*, 2017, 7(1):12581.
8. De, Cal, Garcia-Lepe, et al. Induced Resistance by *Penicillium oxalicum* Against *Fusarium oxysporum* f. sp. *lycopersici*: Histological Studies of Infected and Induced Tomato Stems.[J].

- Phytopathology, 2000, 90(3):260-8.
9. Dale J , James A , Pau J Y , et al. Transgenic Cavendish bananas with resistance to Fusarium wilt tropical race 4[J]. Nature Communications, 2017.
 10. Galindo-González Leonardo, Deyholos M K . RNA-seq Transcriptome Response of Flax (*Linum usitatissimum* L.) to the Pathogenic Fungus *Fusarium oxysporum* f. sp. lini[J]. Frontiers in Plant Science, 2016, 7:1766-.
 11. Upasani M L , Limaye B M , Gurjar G S , et al. Chickpea-Fusarium oxysporum interaction transcriptome reveals differential modulation of plant defense strategies[J]. Rep, 2017, 7(1):7746.
 12. Rana A , Sahgal M , Johri B N . Fusarium oxysporum: Genomics, Diversity and Plant–Host Interaction[M]// Developments in Fungal Biology and Applied Mycology. 2017.
 13. R. W. Michelmore, I. Paran, R. V. Kesseli. Identification of Markers Linked to Disease-Resistance Genes by Bulk Segregant Analysis: A Rapid Method to Detect Markers in Specific Genomic Regions by Using Segregating Populations. 1991, 88(21):9828-9832.
 14. Delong J A , Saito S , Xiao C L , et al. Population Genetics and Fungicide Resistance of *Botrytis cinerea* on *Vitis* and *Prunus* spp. in California[J]. Phytopathology, 2020, 110(3):PHYTO-09-19-036.
 15. Dodds P N , Rathjen J P . Plant immunity: towards an integrated view of plant-pathogen interactions.[J]. Nature Reviews Genetics, 2010, 11(8):539-548.
 16. Zhou Z , Luo D , Shao J , et al. Immune genetic algorithm based multi-UAV cooperative target search with event-triggered mechanism[J]. Physical Communication, 2020, 41:101103.
 17. Swarupa V , Ravishankar K V , Rekha A . Plant defense response against *Fusarium oxysporum* and strategies to develop tolerant genotypes in banana[J]. Planta, 2014, 239(4):735.
 18. Cao Z , Wang P , Zhu X , et al. SSR marker-assisted improvement of fiber qualities in *Gossypium hirsutum* using *G. barbadense* introgression lines[J]. Theoretical and Applied Genetics, 2014, 127(3):587-594.
 19. Huang C , Shen C , Wen T , et al. SSR-based association mapping of fiber quality in upland cotton using an eight-way MAGIC population[J]. Molecular Genetics & Genomics Mgg, 2018, 293(8):1-13.
 20. Xiaoyun J , Chaoyou P , Hengling W , et al. High-density linkage map construction and QTL analysis for earliness-related traits in *Gossypium hirsutum* L[J]. BMC Genomics, 2016, 17(1):909.
 21. Linkage and association mapping reveals the genetic basis of brown fibre (*Gossypium hirsutum*)[J]. Plant Biotechnology Journal, 2018, 16(9).
 22. Shang, LG, Wang, et al. Genetic analysis of Upland cotton dynamic heterosis for boll number per plant at multiple developmental stages[J]. Sci Rep Uk, 2016.
 23. Zhu, Qian-Hao, Yuan, et al. Genetic dissection of the fuzzless seed trait in *Gossypium barbadense*[J]. Journal of Experimental Botany, 2018.
 24. Myers G O , Jiang B , Akash M W , et al. Chromosomal assignment of AFLP markers in upland cotton (*Gossypium hirsutum* L.)[J]. Euphytica, 2009, 165(2):391-399.
 25. Shi Y , Li W , Li A , et al. Constructing a high-density linkage map for *Gossypium hirsutum* × *Gossypium barbadense* and identifying QTLs for lint percentage: A cotton BC 1 F 1 high-density SSR genetic linkage map[J]. Journal of Integrative Plant Biology, 2015(05):450-467.
 26. Shi Y . Constructing a high-density linkage map for *Gossypium hirsutum* × *Gossypium barbadense* and identifying QTLs for lint percentage[J]. Journal of Integrative Plant Biology, 2015(5).

27. Jiwen, Shuxun, Gore, et al. Identification of quantitative trait loci across interspecific F-2, F-2:3 and testcross populations for agronomic and fiber traits in tetraploid cotton[J]. *Euphytica : International journal of plant Breeding*, 2013, 191(3):375-389.
28. Mapping quantitative trait loci for lint yield and fiber quality across environments in a *Gossypium hirsutum* × *Gossypium barbadense* backcross inbred line population[J]. *Theoretical and Applied Genetics*, 2013, 126(1):275-287.
29. Identification of candidate genes for key fibre-related QTLs and derivation of favourable alleles in *Gossypium hirsutum* recombinant inbred lines with *G.barbadense* introgressions[J]. *Plant Biotechnology Journal*, 2020, 18(3).
30. Guoyuan Liu, Wenfeng Pei, Dan Li, et al. A targeted QTL analysis for fiber length using a genetic population between two introgressed backcrossed inbred lines in upland cotton (*Gossypium hirsutum*). 2018,
31. Liu X , Teng Z , Wang J , et al. Enriching an intraspecific genetic map and identifying QTL for fiber quality and yield component traits across multiple environments in Upland cotton (*Gossypium hirsutum* L.)[J]. *Molecular Genetics & Genomics*, 2017, 292(1):1-26.
32. Zhang K , Zhang J , Ma J , et al. Genetic mapping and quantitative trait locus analysis of fiber quality traits using a three-parent composite population in upland cotton (*Gossypium hirsutum* L.)[J]. *Molecular Breeding*, 2012, 29(2):335-348.
33. Wang P , Ning Z , Lin L , et al. Genetic dissection of tetraploid cotton resistant to *Verticillium* wilt using interspecific chromosome segment introgression lines[J]. *The Crop Journal*, 2014, 2(5):278-288.
34. Ji G , Zhang Q , Du R , et al. Construction of a high-density genetic map using specific-locus amplified fragments in sorghum[J]. *Bmc Genomics*, 2017, 18(1):51.
35. Keerio A A , Shen C , Nie Y , et al. QTL Mapping for Fiber Quality and Yield Traits Based on Introgression Lines Derived from *Gossypium hirsutum* × *G. tomentosum*[J]. *International Journal of Molecular Sciences*, 2018, 19(1):243-.
36. Tao A , Huang L , Wu G , et al. High-density genetic map construction and QTLs identification for plant height in white jute (*Corchorus capsularis* L.) using specific locus amplified fragment (SLAF) sequencing[J]. *Bmc Genomics*, 2017, 18(1):355.
37. Zhang S , Zhang Z , Bales C , et al. Mapping novel aphid resistance QTL from wild soybean, *Glycine soja* 85-32[J]. *Theoretical and Applied Genetics*, 2017.
38. Xingfen W , Jun M A , Shuo Y , et al. Assessment of genetic diversity among Chinese upland cottons with *Fusarium* and/or *Verticillium* wilts resistance by AFLP and SSR markers[J]. *Academic Abstracts of Chinese Universities*, 2007, 001(002):129-135.
39. Sun F D , Zhang J H , Wang S F , et al. QTL mapping for fiber quality traits across multiple generations and environments in upland cotton[J]. *Molecular Breeding*, 2012, 30(1):569-582.
40. Shang L , Liu F , Wang Y , et al. Dynamic QTL mapping for plant height in Upland cotton (*Gossypium hirsutum*) [J]. *Plant Breeding*, 2015, 134(6):703-712.
41. Liu D , Liu F , Shan X , et al. Construction of a high-density genetic map and lint percentage and cottonseed nutrient trait QTL identification in upland cotton (*Gossypium hirsutum* L.)[J]. *MGG Molecular & General Genetics*, 2015, 290(5).
42. Xiaoyun J , Chaoyou P , Hengling W , et al. High-density linkage map construction and QTL analysis for earliness-related traits in *Gossypium hirsutum* L [J]. *Bmc Genomics*, 2016, 17(1):909.
43. Zhen Zhang, Haihong Shang, Yuzhen Shi, et al. Construction of a high-density genetic map by

- specific locus amplified fragment sequencing (SLAF-seq) and its application to Quantitative Trait Loci (QTL) analysis for boll weight in upland cotton (*Gossypium hirsutum* L.). 2016, 16(1)
44. Sun Zhengwen, Wang Xingfen, Liu Zhengwen, et al. Genome-wide association study discovered genetic variation and candidate genes of fibre quality traits in *Gossypium hirsutum* L.. 2017, 15(8):982-996.
 45. Bolek Y , El-Zik K M , Pepper A E , et al. Mapping of verticillium wilt resistance genes in cotton[J]. *Plant Science*, 2005, 168(6):1581-1590.
 46. Cong L , Yating D , Tianlun Z , et al. Genome-Wide SNP Linkage Mapping and QTL Analysis for Fiber Quality and Yield Traits in the Upland Cotton Recombinant Inbred Lines Population[J]. *Frontiers in Plant Science*, 2016, 7(218).
 47. Lei Meng Huihui Li Luyan Zhang Jiankang Wang. QTL IciMapping: Integrated software for genetic linkage map construction and quantitative trait locus mapping in biparental populations[J]. *The Crop Journal*, 2015.
 48. Li H , Durbin R . Fast and accurate short read alignment with Burrows-Wheeler transform[J]. *Bioinformatics*, 2009.
 49. De Summa S , Malerba G , Pinto R , et al. GATK hard filtering: tunable parameters to improve variant calling for next generation sequencing targeted gene panel data[J]. *Bmc Bioinformatics*, 2017, 18(5 Supplement):119.
 50. Dongyuan L , Chouxian M , Weiguo H , et al. Construction and Analysis of High-Density Linkage Map Using High-Throughput Sequencing Data[J]. *Plos One*, 2014, 9(6):e98855.
 51. Os H V , Stam P , Visser R G F , et al. SMOOTH: a statistical method for successful removal of genotyping errors from high-density genetic linkage data[J]. *Theoretical & Applied Genetics*, 2005, 112(1):187-194.
 52. Zhang D , Hu Q , Liu X , et al. AllEnricher: a comprehensive gene set function enrichment tool for both model and non-model species[J]. *BMC Bioinformatics*, 2020, 21.
 53. Tanino Y , Kodama M , Daicho H , et al. Selection of Laboratory Procedures to Detect Toxigenic by the 2-Step Method[J]. *Rinsho Biseibutsu Jinsoku Shindan Kenkyukai shi = JARMAM: Journal of the Association for Rapid Method and Automation in Microbiology*, 2017, 27(1):9-14.

Figures

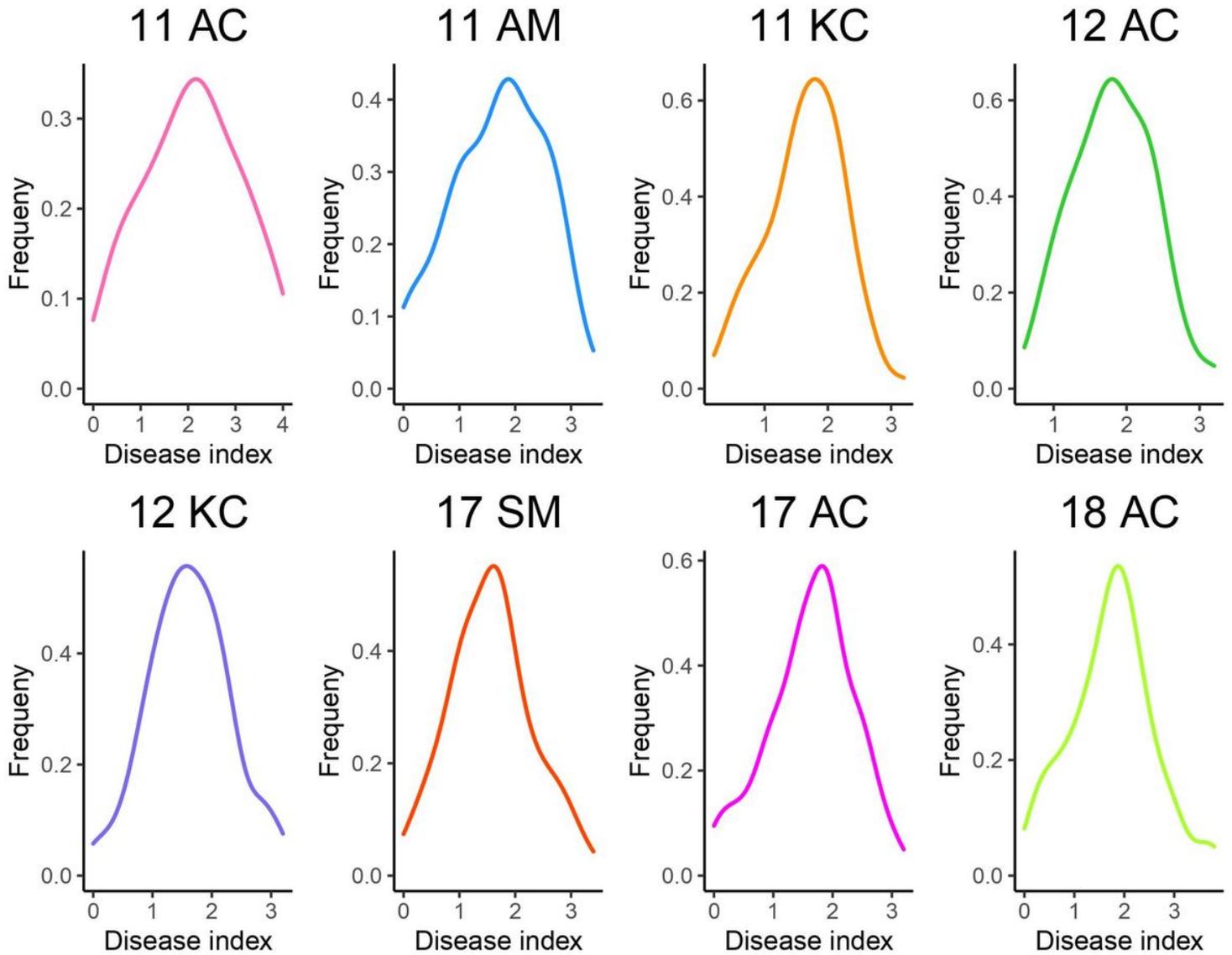


Figure 1

Frequency distribution of disease resistance levels in 110 RILs in different environments

Genetic Map

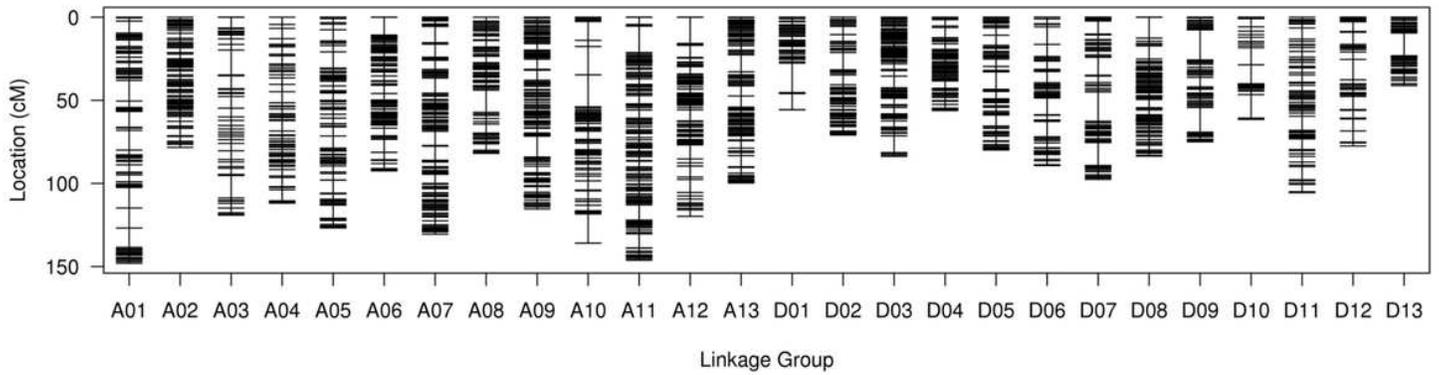


Figure 2

The distribution of SNP markers in each of the 26 linkage groups/chromosomes.

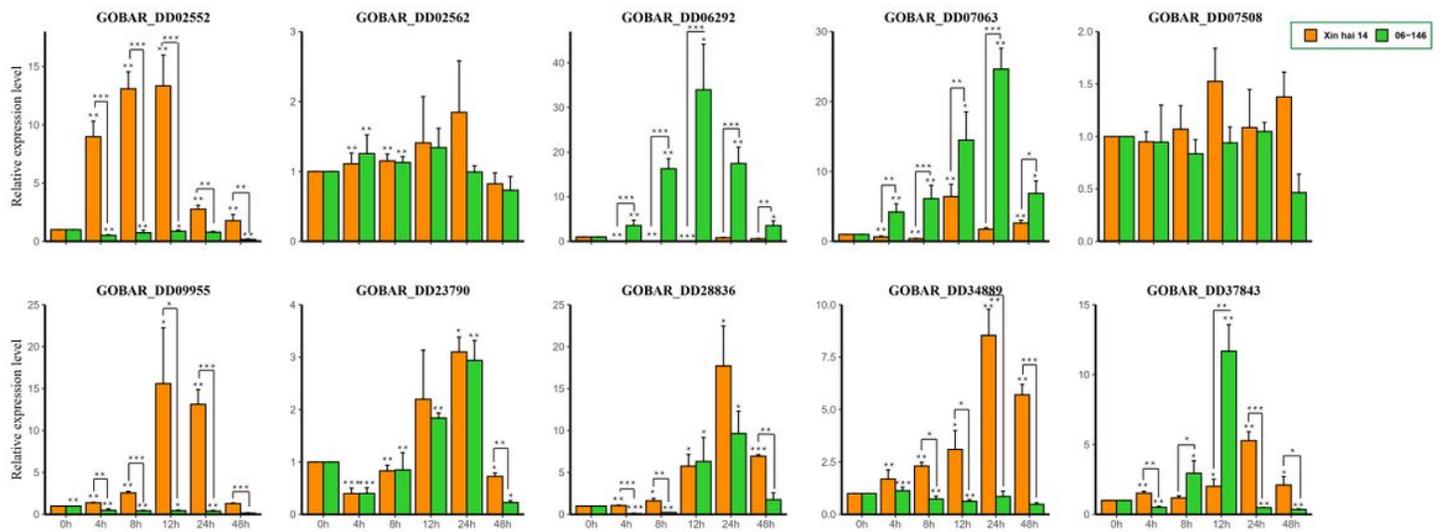


Figure 3

Expression profiling of the 10 candidate genes related to FW. The error bars represent the means of three technical replicates \pm SEs. Statistically significant differences from the control group are indicated with * $P < 0.05$; ** $P < 0.01$

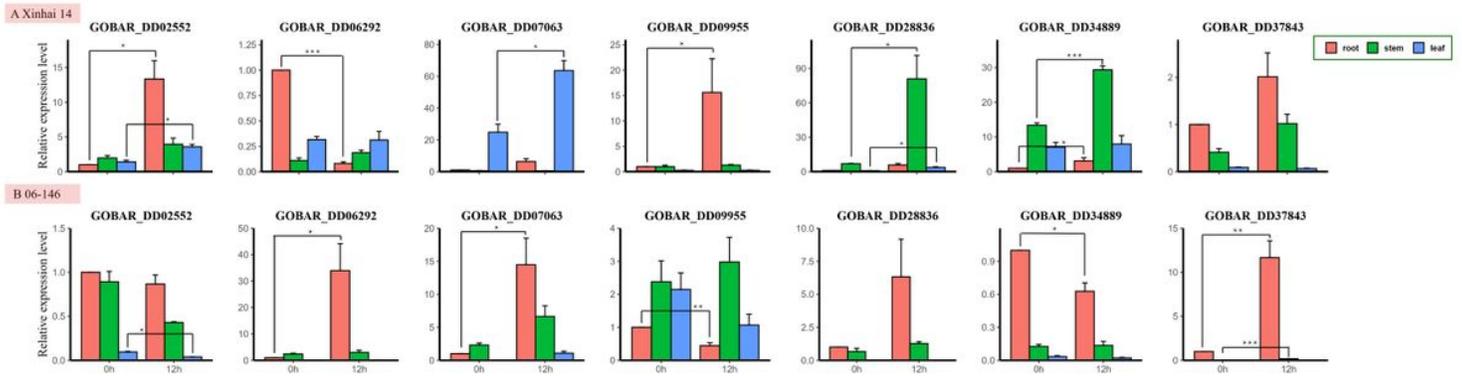


Figure 4

Tissue-specific expression analysis of the 7 candidate genes related to FW. The error bars represent the means of three technical replicates \pm SEs. Statistically significant differences from the control group are indicated as * $P < 0.05$; ** $P < 0.01$

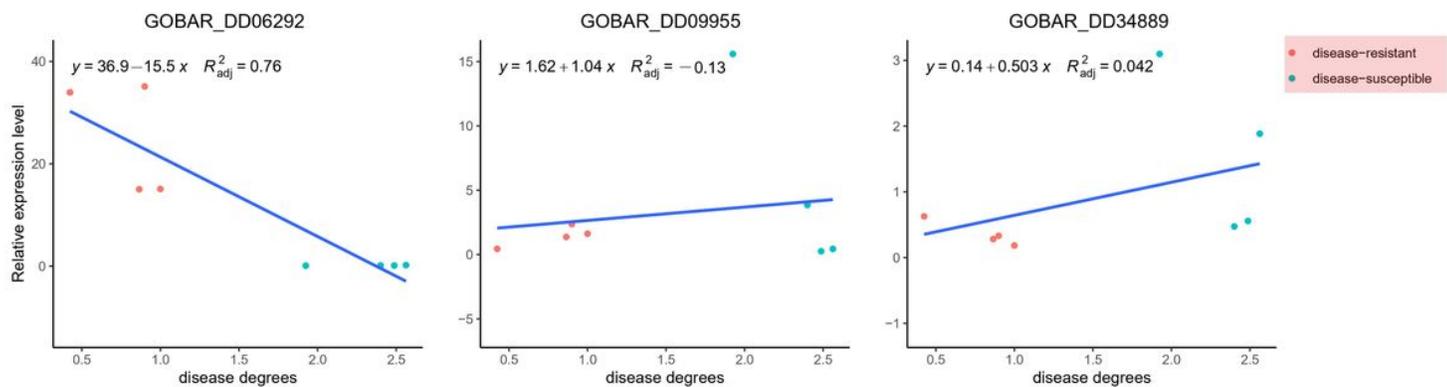


Figure 5

Expression patterns of 3 candidate genes in mixed apical organs of eight varieties of natural population in FW stress, detected by qRT-PCR; red and blue dots indicate four disease-resistant cotton and disease-susceptible cotton, respectively.

Supplementary Files

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

- [FigS1.png](#)
- [FigS2.pdf](#)
- [TableS1.xlsx](#)
- [TableS2.xlsx](#)

- [TableS3.xlsx](#)
- [TableS4.xlsx](#)
- [TableS5.xlsx](#)
- [TableS6.xlsx](#)
- [TableS7.xlsx](#)