A large-scale genomic association analysis identifies a fragment in Dt11 chromosome conferring cotton Verticillium wilt resistance

The authors
Zhang Yan1, Chen Bin1, Sun Zhengwen, Liu Zhengwen, Cui Yanru, Ke Huifeng, Wang Zhicheng, Wu Liqiang, Zhang Guiyin, Wang Guoning, Li Zhikun, Yang Jun, Wu Jinhua, Shi Rongkang, Liu Song, Wang Xingfen*, Ma Zhiying*

State Key Laboratory of North China Crop Improvement and Regulation, Key Laboratory for Crop Germplasm Resources of Hebei, Hebei Agricultural University, Baoding, China.

1These authors are equally contributed to this work.

*Corresponding authors:
Dr. Zhiying Ma E-mail: mzhy@hebau.edu.cn; Dr. XingfenWang E-mail: cotton@hebau.edu.cn

Running head: A fragment in Dt11 chromosome conferred cotton Verticillium wilt resistance.

Keywords: Gossypium hirsutum; Verticillium wilt; GWAS, Dt11 chromosome; GhLecRKs gene cluster

Summary
Verticillium wilt (VW) is a destructive disease that results in great losses in cotton yield and quality. Identifying genetic variation that enhances crop disease resistance is a primary objective in plant breeding. Here we reported a GWAS of cotton VW resistance in a natural-variation population, challenged by different pathogenicity stains and different environments, and found 382 SNPs significantly associated with VW resistance. The associated signal repeatedly peaked in chromosome Dt11 (68,798,494-69,212,808) containing 13 core elite alleles undescribed previously. The core SNPs can make the disease reaction type from susceptible to tolerant or resistant in accessions with alternate genotype compared to reference genotype. Of the genes associated with the Dt11 signal, 25 genes

This article has been accepted for publication and undergone full peer review but has not been through the copyediting, typesetting, pagination and proofreading process, which may lead to differences between this version and the Version of Record. Please cite this article as doi: 10.1111/PBI.13650

This article is protected by copyright. All rights reserved.
differentially expressed upon *V. dahliae* stress, with 21 genes verified in VW resistance via gene knockdown and/or overexpression experiments. We firstly discovered that a gene cluster of L-type lectin-domain containing receptor kinase (*GhLecRKs-V.9*) played an important role in VW resistance. These results proved that the associated Dt11 region was a major genetic locus responsible for VW resistance. The frequency of the core elite alleles (FEA) in modern varieties was significantly higher than the early/middle varieties (12.55% vs 4.29%), indicating that the FEA increased during artificial selection breeding. The current developmental resistant cultivars, JND23 and JND24, had fixed these core elite alleles during breeding without yield penalty. These findings unprecedentedly provided genomic variations and promising alleles for promoting cotton VW resistance improvement.

**Introduction**

Verticillium wilt (VW), caused by *Verticillium dahliae*, brought great yield losses and quality decline in cotton production, despite the use of fungicide treatments estimated to cost farmers additional $1.2 billion (Bolek et al., 2005; Song et al., 2020). More recently, changes in climate and continuous cropping of cotton had made the VW more frequent and severe, even in cotton regions where had not been previously reported, such as Xinjiang located in the largest cotton region of China (Zhu et al., 2017). VW had even been thought as a “cancer” to many dicotyledons including cotton, owing to its complexity and destructiveness (Luo et al., 2014). The incorporation of resistance genes into commercial varieties was the ideal strategy to combat VW (Wang et al., 2016; Zhang et al., 2017). Previous studies had shown that VW resistance was a complexly inherited trait (Chen et al., 2020; Song et al., 2020). Therefore, it is crucial for researchers to understand the genetic basis of VW resistance in current breeding populations and continuously search for novel genes.

Cotton VW resistance had been proved to belong quantitative trait, which was based on multiple loci, each with a small effect (Said et al., 2015). In model plant tomato, it was reported that VW resistance was mainly controlled by *Ve* locus (Fradin et al., 2009; de Jonge et al., 2012). However, the genetic basis of quantitative variation for VW resistance in agriculturally important crops remained poorly understood. Several studies focused on cotton VW resistance inheritance and identified several related QTLs, whereas, none of which had been functional verification. Owing to lack of well-defined QTLs and the close-linked diagnostic markers and the crucial genes responsible for VW resistance, current resistance breeding strategies still primarily relied on phenotypic evaluation rather than targeted-genotyping-based selection, leading to extremely time-consuming and low efficiency. Thus, deployment of resistant genes or accurate diagnostic markers is highly expected by breeders.
With significant progresses in high-throughput sequencing and high-quality assembly of *G. hirsutum* genome, large numbers of molecular markers had been identified that facilitated the progress of more efficient analysis via associating mapping (Yang et al., 2020). Identifying new sources of genetic variation from diverse germplasm resources was an important strategy in crop improvement (Bevan et al., 2017). Genome-wide association studies (GWAS) were increasingly used for successfully identifying variation associated with complex traits in various crops (Ma et al., 2018), due to its powerfulness in simultaneously detecting numerous of natural allelic variations and a lot of historical chromosomal recombination events occurring over multiple generations of natural population (Dehghan, 2018). For example, Rice *bsr-d1* was discovered conferring broad-spectrum resistance to *Magnaporthe oryzae*, showing great utilization value because of its non-link to any gene known to confer undesirable grain quality or flavor (Li et al., 2017a). *ZmFDL41* modulated banded leaf and sheath blight resistance via targeting the host proteasome in maize (Li et al. 2019). Two studies had reported the association of DNA polymorphisms with cotton VW resistance (Fang et al., 2017; Li et al., 2017b), however, there was no or very few SNPs repeatedly detected under multiple environments, implying that the data set with great power for revealing VW resistance genomic variations was still needed. Also, the potential resistant germplasm containing genetic loci linked to broad-spectrum resistance genes are still lacking.

In this study, we used a cotton core collection with promising population structure for GWAS based on our previous research (Ma et al., 2018). We further resequenced this core collection up to 11.0-fold-coverage depth, phenotyped VW resistance across two *V. dahliae* stains and environments and uncovered genomic variations of the resistance. Combining transcriptome, gene expression and functional validation analyses, we newly identified a Dt11 genomic region conferring strong VW resistance, and firstly revealed a *GhLecRK-V.9* gene cluster playing crucial role in VW resistance. Our results unprecedentedly provided genomic variations and promising alleles for promoting cotton VW resistance improvement.

**Results**

**Extensive natural variations were detected in VW resistance among upland cotton**

To accurately explore the genetic resistance of each tested cotton genotype, we determined VW resistance under completely controllable environment (artificial climate room). A total of 401 cotton accessions including 56,140 plants, 35 plants for each genotype in an experiment, were characterized for phenotypic response to a strong pathogenicity strain LX2-1 (Wang et al., 2012) for four independent experiments. Extensive natural variation in VW
resistance was observed, the disease index (DI) ranged from 20.5 to 70.7 and 32.1 to 86.1 at 20 days post inoculation (dpi) and 25 dpi, with an average of 43.4 and 63.7, respectively (Fig. 1a, Table 1). Considering the disease resistance could be influenced by fungal stains and environmental variations (Li et al., 2017b), we further carried out VW resistance identification under another completely controllable environment (greenhouse) with a medium pathogenicity strain Vd991. A total of 296 accessions including 20,720 plants were characterized for phenotypic response to Vd991 with two respective experiments, and the DI ranged from 8.6 to 49.3 with an average of 23.5 and 11.79 to 71.07 with an average of 37.9 at 20 dpi and 25 dpi, respectively (Table 1). An analysis of variance of the DI in different environments and different stains revealed significant differences among the genotypes, investigating periods and environments (Table S1). Finally, we gained a DI data set with four types: LX-20dpi, LX-25dpi, Vd991-20dpi and Vd991-25dpi (Table S2). The correlation coefficients of the DI between 20 dpi and 25 dpi of the same strain (0.907 and 0.798) were obviously higher than those of different strains (Fig. S1). Thus, we performed GWAS analysis based on this DI data set.

A newly identified D11 genomic region was associated with cotton VW resistance

To greatly mining genomic mutation sites and enhancing large-scale genetic analytical potential, we performed deeply resequencing a core collection composed of 419 cotton accessions (Ma et al., 2018), and obtained a total of 10.66 Tb of sequence, with an average depth of 11.0× and coverage of 96.3% of the reference genome (Wang et al., 2018) (Table S3). After removing nucleotide polymorphisms with missing rates ≥0.25 and minor allele frequency <0.05, a total of 6,034,699 SNPs were identified when using the updated assemblies of TM-1 as a reference (https://cottonfgd.org/) (Table S4), with a density of 3.12 SNPs per kilobase, and variants number approximately 3 to 104 timed as in Fang et al. (2017) (0.97 SNPs per kilobase) and Li et al. (2017) (0.03 SNPs per kilobase) GWAS study, respectively.

We performed GWAS for VW resistance using above phenotypic data and SNP data. The $P < 10^{-6}$ was set as the significance threshold in each panel association, and a series of SNPs in a defined region with $P$-values lower than the threshold were detected. Based on the median distance of the linkage disequilibrium (LD) decay of this population (Ma et al., 2018), the significant SNPs within a 300-kb region were considered to represent a locus, and the SNP with the lowest $P$-value in a locus was defined as the SNP in the closest linkage to the causal gene. A total of 382 (excluding the repeated among different data types) SNPs signals were detected across four data types. In details, there were 17, 22, 6 and 354 significant SNPs were detected in LX-20dpi, LX-25dpi, Vd991-20dpi and Vd991-25dpi,

This article is protected by copyright. All rights reserved
respectively (Table S5). Manhattan plots showed that there were significantly associated signals in chromosome A02, A03, A05, A06, A08, A12, D05, D07, D10, D11 and D12 (Table S5). We found that the D11 peak could repeatedly detected in each data type, indicating that the D11 peak was a major associated signal, harboring 91.9% of the associated SNPs (351) (Fig. 1b and Table S6), which had not been reported previously. Four SNPs (D11:68877270, D11:68953719, D11:68966283, D11:69041458) could be identified in three types out of the data set, with $P$ value range from 6.09 to 10.74. Thirteen SNPs could be repeatedly detected in both pathogen strains (LX-20dpi and Vd991-25dpi), and we marked them as core SNPs which could make the disease reaction type from susceptible to tolerant or resistant in accessions with alternate genotype compared to reference genotype (Table S6). More important, the core SNPs sequentially located in a genomic region of D11 within 414.3Kb (68,798,494-69,212,808 bp) (Table S6).

We compared our GWAS results to previous studies and found that the significantly associated signal (A10_98704568) on A10 was overlapped to the reported SNP loci (Li et al., 2017b), supporting the reliability of our results. Although previously reported VW resistance QTLs or GWAS signals were not overlapped with the region of D11 that identified in the present study (Fang et al., 2017; Li et al., 2017b; Abdelraheem, et al., 2020), the SNPs identified in the region of D11 were repeatedly associated and exhibited clusters of resistance gene analogues (RGAs) (Holub, 2001) via bioinformatics analysis (Table S7). Thus, the significant SNPs that we identified on D11 were novel.

We focus on the 13 SNPs in the region of D11: 68,798,494-69,212,808 (Table S6), and extracted all genes within 300 kb of the most significant SNPs. We obtained 81 genes that represent plausible candidates for the causal gene of the resistance (Table S7) based on their consistent annotations in TM-1 assemblies (Wang et al., 2018 and https://cottonfgd.org/), functions of homologous genes and the indels generated by sequence blast between the resistance and susceptible varieties (Ma et al., 2018).

**A LecRK gene cluster was firstly identified to play crucial role in cotton VW resistance**

Among the 13 associated loci in the region of D11, there was a nonsynonymous SNP (D11:69088571) located within Ghir_D11G033410.1 (Gene ID: 2) (Fig. 2a, b), encoding a L-type lectin-domain containing receptor kinase V.9. The family of LecRKs including 77 and 173 members in Arabidopsis and rice, respectively, were plant specific and implicated in diverse biological processes (Bouwmeester and Govers 2009; Bouwmeester et al., 2011; Wang et al., 2011).
2017). In the present study, it is interesting that other three GhLecRKs (Ghir_D11G033420.1, Ghir_D11G033430.1 and Ghir_D11G033440.1) together with the GhLecRK-V.9 form a gene cluster at the locus. Thus, we designated it as GhLecRK-V.9 cluster. To date, the function of the GhLecRK-V.9 gene cluster is undescribed. The associated nonsynonymous SNP situated in the conserved protein kinase-like domain, and the mutation from A to G changed 251th amino acid from Isoleucine to Valine (Fig. 2b). The varieties with the alternate GG showed significantly lower DI compared to the reference AA (Fig. 2c). We found that the expression of Ghir_D11G033410.1 was suppressed upon V. dahliae infection in resistant cultivar ND601, but not in the susceptible (Fig. 2d, e), and Ghir_D11G033410.1 functioned on cytomembrane (Fig. 2f). Furtherly, two types of varieties containing the associated alleles were used to perform expression analysis. We discovered much higher levels of expression in the susceptible varieties (AA) than in the tolerant varieties (GG), although expression levels varied in different genetic backgrounds (Fig. 2g), indicating that Ghir_D11G033410.1 expression levels were strongly associated with VW resistance.

We further validated the function of Ghir_D11G033410.1 in cotton through virus-induced gene silencing (VIGS) technology and observed that knock down this gene both in tolerant (NDM 8) and susceptible (CCRI8) cultivars greatly improved VW resistance (Fig. 2h, j). The DI of silenced plants decreased from 27.9 to 12.9 and 53.7 to 13.6, respectively, reaching to resistant level (10 < DI ≤ 25). We then overexpressed GhLecRK-V.9. in Arabidopsis and found that the transgenic plants showed significantly increased DI (52.47) when compared to the wild-type (37.02) (Fig. 2i, k). According to above results, we concluded that Ghir_D11G033410.1 down regulated VW resistance in plants. Transcriptome analysis of the GhLecRK-V.9 cluster showed that Ghir_D11G033420.1 (ID: 26) could be significantly induced upon inoculation (Fig. S2), whereas Ghir_D11G033430.1 (ID: 27) displayed fluctuated expression pattern and Ghir_D11G033440.1 did not respond to V. dahliae stress (data not shown). Thus, we silenced Ghir_D11G033420.1 in cotton, and found that the silenced seedlings of Ghir_D11G033420.1 showed significantly more susceptible than the mock plants during the two repetitive experiments (Fig. 2). Taken together, these results proved the function of the GhLecRK-V.9 clusters in cotton immunity.

**The D11 region enriched many other functional genes in cotton VW resistance**

Except for three genes, Ghir_D11G033410.1, Ghir_D11G033420.1 and Ghir_D11G033430.1 in the GhLecRK-V.9 cluster, transcriptome analyses of the 81 associated genes in the Dt11 region (Table S7) also revealed that many other 22 genes differentially expressed upon V. dahliae stress (Fig. S2, Table S8). The 22 genes were functionally validated for their VW resistance via VIGS in cotton and/or overexpression in Arabidopsis. When
silenced each of 21 genes (excluding No.10 gene not successfully knock down) in tolerant and/or susceptible cotton cultivars, we found 18 genes functioned in VW resistance, with three downregulating genes and 15 upregulating genes. All the silenced plants of three downregulating genes, \textit{Ghir\_D11G033550.1} (ID:13), \textit{Ghir\_D11G033690.1} (ID:16) and \textit{Ghir\_D11G033820.1} (ID:32), showed greatly improved VW resistance, especially the resistance of ND601-silenced 13 and ND601-silenced 16 reaching to high resistance level (0 < DI≤10). In addition, the CCRI8-silenced 13, CCRI8-silenced 16 and CCRI8-silenced 32 also displayed more resistant than the susceptible CCRI8-mock plants, nearly achieving resistant level (10 < DI≤25) (Fig. 3). All the silenced plants of the 15 upregulating genes in NDM 8 exhibited more susceptible than the mock plants (Fig. 4), and the silenced plants displayed much more \textit{V. dahliae} biomass (Fig. 5).

The functional analyses of several genes by ectopic expression in \textit{Arabidopsis} showed that overexpressed three upregulating genes (ID: 5, 14 and 15) markedly increased the plant resistance. The DI for the wilt-type (WT) plants was 37.0, whereas DI for OE5, OE14 and OE15 transgenic plants declined to 27.6, 27.0 and 21.3, respectively. In contrast, the transgenic plants OE10, OE13 and OE16 of three downregulating genes (ID: 10, 13 and 16) displayed more seriously susceptibility than the WT, with increased DI of 55.3, 52.0 and 58.1, respectively (Fig.6).

Taken together, according to gene expression, VIGS and overexpression experiments, 21 candidate genes located in the region of Dt11 play important roles in regulating VW resistance, uncovering a crucial functional region for VW resistance.

**The associated SNPs in the D11 genomic fragment possibly functioned as a unit element**

To investigate how the associated genes to influence disease resistance of alterative and reference genotypes in GWAS panel accessions, we reuse two types of varieties as mentioned in Fig. 2 for detecting gene expression levels post \textit{V. dahliae} infection. qPCR analyses of 10 candidate genes (including gene 2, 5, 7, 10, 11, 12, 13, 14, 15 and 16) showed that the average expression level of all the up-regulated genes (ID: 5, 7, 11, 12, 14 and 15) were significantly higher in resistant accessions (alternative genotype) than in susceptible accessions (reference genotype), with negative correlation ($r = -0.68$) between expression level and DI (Fig. S3a). The opposite trend was observed in downregulated genes (ID: 2, 10, 13 and 16) ($r = 0.88$) (Fig. S3b). Besides, the SNP number investigation revealed that all the resistant cultivars had at least 84.6% core elite alleles, and 50.0% resistant cultivars had all the elite alleles (Fig. S4). Thus, we deduced that the associated SNPs locating on D11 genomic fragment possibly functioned as a unit element.

**Elite-allele frequency increased during modern cotton breeding**
The frequency of the elite alleles (FEA), i.e. the core SNPS that clearly enhance VW resistance as identified above, was evaluated in early/middle and modern varieties (Ma et al., 2018). The average FEAs of modern varieties was 12.55%, significantly higher than that of early/middle varieties (4.29%) (Table S9), indicating that the FEAs was increased during the artificial selection. Moreover, we also investigated the FEA of modern varieties from different cotton-planting areas and found that most of the varieties with elite alleles gathered in Yellow River cotton-growing area (Table S10). This finding is accordant with the actual situation that Yellow River cotton-growing area, especially in Hebei and Henan provinces, is seriously and frequently occurring region of VW in cotton production (Zhang et al., 2006).

To further validate the effect of variation in the D11 region in improving resistance in the field, a part of varieties in GWAS panel with alternate or reference genotypes were planted in field disease nursery. Twenty varieties with reference genotype displayed seriously susceptible, with an average DI 66.1, significantly higher than that of seven varieties with alternate genotype (36.2). These results suggested that the SNPs should greatly improve cotton resistance. To assess if the variation in the D11 fragment has been selected during the breeding of commercial varieties, the DI and corresponding genotypes in the D11 fragment were investigated across five current cotton cultivars, including NDM13, ND601, JND23, JND24 and JND25 (Fig. S5). The five cultivars exhibit good resistance compared to TM-1, and the resequencing results showed that JND23 and JND24 have the same SNP as the alternate genotype, the others were the same as the reference genotype. These results indicated that these core elite alleles had been selected during the process of breeding in some areas under unknowing the alleles. However, because these elite SNPs were present in only approximate 8.8% of the GWAS panel varieties (Table S6), there was a great potential in introducing these alleles into other cottons in the future efforts.

**Discussion**

Cotton VW is the most devastating disease and greatly reduces yield and fiber quality (Xu et al., 2018). Despite the popularity of the association study approach to explore elite natural alleles associated with VW resistance, there were few reports of successfully used in practical application (Li et al., 2017b; Fang et al., 2017). It was reported that VW resistance always influenced by different *V. dahliae* strains and various environments (Li et al., 2017b). Given this previous result, we performed disease resistance investigation under two different *V. dahliae* strains and two environments to identify the overlapping QTLs associated with VW resistance in cotton. Besides, the interaction between *V. dahliae* and cotton was a complex process (Song et al., 2020). Inadequate interaction between them
greatly affected the displaying of phenotypic symptoms. An example reflected this phenomenon in this study, the GWAS results based on DI of Vd991-20dpi and LX-25dpi showed poor detection power. On the other hand, higher number of genetic markers may provide adequate coverage to detect linkage between markers and causal loci controlling complex traits (Ma et al., 2018). In previous studies, Fang et al. (2017) carried out GWAS of VW by using 258 diverse accessions and 1,871,401 SNPs (with 0.97 SNPs per kilobase), and only three unrepeated SNPs were detected. Another study reported by Li et al. (2017b), performing association analysis based on 299 accessions and 85,630 SNPs (0.03 SNPs per kilobase), only found two SNPs (A10_99071906 and A10_99110423) exceeding significance threshold. It was reported that greater power in association study could be achieved by increasing the sample size and the number of polymorphisms towards a complex trait (Long et al., 1999; Korte et al., 2013; Ma et al., 2018). To date, our GWAS using a largest-sample-size cotton core collection and deepest resequencing data (average 11×) could significantly enhanced the detection power. Using the same analytical approaches as in a previous report, we detected previously identified SNP loci on A10 (Li et al., 2017b). More importantly, we discovered more previously undescribed loci significantly associated with VW resistance. And many more loci (single SNP) could be repeatedly detected in at least three datatypes out from four. Furthermore, we identified 13 novel SNPs in the D11 region of Dt11:68,798,494-69,212,808 significantly associated with VW resistance.

Based on the geographical distribution and breeding time of 32 cotton accessions carrying elite allele on the D11 region, 68.8% (22 accessions) came from Yellow River cotton-growing area and belonged to modern varieties, indicating that the resistant mutations predominantly occurring in the area. In China, the Yellow River cotton area is one of main cotton planting-areas, and it is also deeply remembered because of VW occurring seriously and frequently compared to the others (Sun et al., 2013). Thus, we deduced that the novel identified D11 region probably originated from mutation(s) of cotton cultivar(s) grown in Yellow River area, especially in Hebei and Henan provinces, where cotton production constantly faces heavy VW pressure due to favorite environmental factors (25~28°C and relative higher humid climate) for V. dahliae proliferation (Wang et al., 2012).

Crop genetic improvement for resistance was the most economical and environmentally friendly approach to guarding against disease outbreaks (Zhou et al., 2018). Many resistant (R) genes conferring race-specific resistance had been used in developing resistant cultivars, however, the extremely time-consuming process of pyramiding R genes and non-durable resistance limited its utilization in practice (Li et al., 2017a). In contrast, the broad-spectrum resistance controlled by multiple genes or quantitative trait loci was often durable and therefore was more effective.
This article is protected by copyright. All rights reserved
resistant cultivars. Functional analysis demonstrated that down-regulation of Ghir_D11G033410.1 expression greatly enhanced cotton resistance, addressing its important role firstly.

Furthermore, we proved the GhLecRK-V.9 gene cluster functioned in cotton immunity. Gene clustering was rare in eukaryotes except nucleotide-binding leucine-rich repeat (NLR)-encoding disease resistance (R) genes constantly clustering in plant genomes (Wersch and Li, 2019). And the arrangement of the paired or larger clusters of NLRs acting together could trigger immunity stronger than other station. These patterns could help host conferring broad resistance despite their finite number of R genes (Wersch and Li, 2019). According to previous report, gene cluster general fall into two categories: homologous gene duplications and functionally linked genes with little sequence similarity (Osbourn and Field, 2009). In details, the later tended to contain the genes required for metabolic or signaling pathways (Michalak, 2008; Lee and Sonnhammer, 2003). In plant, LecRLKs mediated resistance to fungal pathogens (Meyers et al., 2003; Steuernagel et al., 2018) or facilitated symbiosis (Seo et al., 2016), which typically comprised of an extracellular N-terminal lectin domain, a transmembrane region, and a cytosolic kinase domain (Muchero et al., 2018). Lectins were proteins that possessed at least one noncatalytic domain that could selectively recognize and reversibly bind to specific carbohydrate structures (Gao et al., 2018). We hypothesized that the functional receptors of our GhLecRKs in alternate genotypes were necessary to perceive V. dahliae derived ligand(s) and mediated the interactions between the host and fungal.

In summary, we identified a novel chromosomal region in Dt11, and the haplotype and candidate gene analyses reveal promising alleles that function in VW resistance, which will accelerate future efforts aiming at cotton VW resistance improvement through the approaches of transgene, marker-assisted selection and genome editing.

Experimental procedures

Deepening resequencing, sequence quality checking and population SNP detection

The DNA samples of a cotton core collection used previously (Ma et al., 2018) were chosen for further sequencing. A 350-bp whole-genome library for each accession was constructed according to the manufacturer’s instructions (Illumina). Then we used the Illumina HiSeq platform to generate raw sequences with 150-bp read length. The sequence quality checking and filtering, sequence alignment and SNP calling on a population scale were carried out according to the process of our previously reported (Ma et al., 2018).

Phenotypic evaluation

Cotton seeds were surface-disinfected in 0.5% sodium hypochlorite (NaOCl) for 5 min and then were washed five
times with distilled water. After they were transferred to wet, sterile towels in petri dishes for promoting germination
in a biochemical incubator at room temperature (RT) for 36h. Any seed presenting internal fungal contamination were
discarded. Each seed with similar size was selected and cultivated in a pot containing sterile vermiculite under the
following growth chamber conditions: 16-h photoperiod, 28±2°C (day) and 25±2°C (night), and 75% relative humidity.
Thirty-five seedlings comprised a replication for each accession. Hoagland's nutrient solution was added to the plots
every 7 days.

*V. dahliae* isolates, strong pathogenicity strain LX2-1 and medium pathogenicity strain Vd991, were used for
inoculation. Both isolates were grown in a growth chamber at 25°C in a 12 hours light /12 hours dark photoperiod
before used for inoculation. To initiate conidia production, we subculture the pathogen strain from PDA plates into
Czapek's medium (Zhang et al., 2017). After incubation at 25°C for 5 to 7 days, the concentration of conidia was
adjusted with deionized water to approximately 10^7 ml^-1. For Verticillium inoculation, a total of 10 ml conidia (10^7
ml^-1) was well-distributed injected into seedling pot. Symptom development was recorded at 20pi and 25dpi and
categorized into five grades: 0 = healthy plant, no symptoms on leaves; 1 = one or two cotyledons showing symptoms
but no symptoms on true leaves; 2 = both cotyledons and one true leaf showing symptoms; 3 = both cotyledons and
two true leaves showing symptoms; and 4 = all leaves showing symptoms, symptomatic leaves dropped, apical
meristem necrotic or plant death. The DI was calculated according to the same method used in the Li et al., 2017.

During the screening in the artificial climate room, the resistant cultivar ND601 and the susceptible cultivar
Jimian11 were used as the resistant and susceptible controls, respectively, which were randomized placed on each
floor of shelves. We could finish the 401 samples screening in three identical rooms at the same batch, and the results
from different artificial climate room were corrected via the resistant and susceptible control. The mean value of the
resistant and susceptible control in the same batch was used for correcting data from different replications. For disease
resistance screening in the greenhouse, all the samples could be planted in two identical greenhouses. Five resistant-
and five-susceptible controls were randomized placed in each greenhouse. The results from the two greenhouses were
corrected based on the controls. The mean value of the resistant and susceptible control in the same batch was used for
correcting data from different replications.

**GWAS analysis**

In our association panel containing 401 samples, a total of 3766828 SNPs (MAF ≥ 0.05; missing rate ≤ 0.2, depth ≥ 3)
were used in GWAS for different traits. Association analysis was conducted with the genome-wide efficient
mixed-model association (GEMMA) software package (Zhou et al., 2012). For mixed-linear-model analysis, we used the following equation: $y = X\alpha + S\beta + K\mu + e$, where $y$ represents phenotype; $\alpha$ and $\beta$ are fixed effects representing marker effects and nonmarker effects, respectively; and $\mu$ represents unknown random effects. $X$, $S$, and $K$ are the incidence matrices for $\alpha$, $\beta$, and $\mu$, respectively, and $e$ is a vector of random residual effects. The top three PCs were used to build up the $S$ matrix for population-structure correction. The matrix of simple matching coefficients was used to build up the $K$ matrix. The analyses were performed in the GEMMA software package.

**RNA isolation and qPCR**

Total RNA was extracted using EASYspin Plus Plant RNA Kit (Aidlab Biotech, China) according to the manufacturer’s protocols. cDNA was synthesized using an RNA reverse transcription kit (Takara, Dalian, China). The qPCR was conducted using a Bio-Rad CFX96 Real-Time System coupled to a C1000 Thermal Cycler (Bio-Rad, Hercules, CA, USA). The $2^{-\Delta\Delta CT}$ method was used to calculate the relative expression levels in a variety across different time points (Livak and Schmittgen, 2000), and the $\Delta CT$ method was used to calculate the relative expression levels in different varieties at the same time point (Pfaffl et al., 2001; Ma et al., 2018). The primers used in this study are listed in Table S11.

**Gene cloning and plant transformation.**

The full-length open reading frames of the related genes were amplified through PCR using cDNAs synthesized from RNA that were isolated from seedlings of varieties with the corresponding SNP alleles. The amplified products were further cloned into the pGreen vector driven by the cauliflower mosaic virus (CaMV) 35S promoter. The resulting constructs were further transformed into A. thaliana Columbia type by Agrobacterium tumefaciens GV3101 and selected with Basta (Ma et al., 2018). The assay for disease resistance of transgenic Arabidopsis was performed according to Zhang et al. (2011).

The primers used for gene cloning are listed in Supplementary Table S11.

**Subcellular localization**

For subcellular localization in Arabidopsis protoplasts, we amplified the coding region of $GhLecRK$ and cloned it into the PBI211 vector to generate fusion construct with GFP (p35S:$GhLecRK$:GFP). The fusion construct was transformed or co-transformed into Arabidopsis leaf protoplasts and onion epidermal cells following the method as described previously (Chen et al., 2020; Zhang et al., 2016). For controls, protoplasts or onion epidermal cells were transformed with empty vector (p35S:GFP). Fluorescence was examined under a confocal microscopy (Nikon A1

This article is protected by copyright. All rights reserved
Virus-induced gene silencing in cotton

The gene-specific region for target gene was amplified as a template and cloned into the pTRV2 vector. The resulting pTRV2 construct was co-infiltrated with pTRV1 via A. tumefaciens GV3101 into cotton seedlings of NDM 8 through syringe inoculation when the cotyledons had spread (Zhang et al., 2019). Plants co-inoculated with empty pTRV2 and pTRV1 were used as the control. The primers used for construction of the VIGS vector are listed in Supplementary Table S11.

Verticillium dahliae recovery assay

At 7dpi, a cotton stem fragment of 1 cm was sectioned from the first node of the stem base to the above. Their surfaces were sterilized for 20-30 s in 70% ethanol, then for 1 min in 0.1% corrosive sublimate. They were then rinsed once in 0.1% HgCl₂ solution, five times with sterilized water and sliced into 3-4 mm of each cross-sectional cutting. From each plant, six to seven stem slices were obtained and plated on PDA before being incubated at 25°C.

Statistical analysis

SPSS22 was used for the statistical analysis of phenotypic traits. Statistically significant differences between control and experimental groups were determined by one-way ANOVA. The significance level was set at $P = 0.05$ or $0.01$. In the transcriptome analysis, the FPKM values of genes from inoculated and control cottons were calculated with Cufflinks (version 2.1.1). For genes differentially expressed between inoculated and control samples, the expression changes were set as fold change $> 1$.

Acknowledgments

This work was supported by the National Natural Science Foundation of China (Grant 31871672), the Outstanding Youth Fund of Hebei Province (Grant C2019204365), the China Agriculture Research System (CARS-15-03), and the Talents Support Program of Hebei Province.

Competing Interest Statement: The authors declare no conflicts of interest.

Author contributions

YZ and BC conducted the experiments, analyzed the data, and YZ wrote the manuscript. ZWS and YRC performed GWAS analysis. ZWL and GNW participated in bioinformatics analysis and gene expression analysis, respectively, and WZC conducted gene subcellular localization assays. KHF, WLQ, ZGY, LZK, YJ, WJH, WZC, SRK and LS participated in VW resistance detection experiments. ZYM and XFW conceived and directed the project and revised

This article is protected by copyright. All rights reserved
the manuscript.

References

1. Abdelraheem, A., Elassbli, H., Zhu, Y., Kuraparthy, V., Hinze, L., Stelly, D., Wedegaertner, T. and Zhang, J.F. (2020) A genome-wide association study uncovers consistent quantitative trait loci for resistance to Verticillium wilt and Fusarium wilt race 4 in the US Upland cotton. *Theor. Appl. Genet.* **133**, 563–577.

2. Arnaud, D., Desclos-Theveniau, M. and Zimmerli, L. (2012) Disease resistance to *Pectobacterium carotovorum* is negatively modulated by the Arabidopsis lectin receptor kinase LecRK-V.5. *Plant Signal Behav.* **7**, 1070–1072.

3. Balague, C., Gouget, A., Bouchez, O., Souriac, C., Haget, N., Boutet-Mercey, S. *et al.* (2016) The Arabidopsis thaliana lectin receptor kinase LecRK-I.9 is required for full resistance to *Pseudomonas syringae* and affects jasmonate signalling. *Mol Plant Pathol.*, **18**, 937–948.

4. Bevan, M.W., Uauy, C., Wulff, B.B.H., Zhou, J., Krasileva, K. and Clark, M.D. (2017) Genomic innovation for crop improvement. *Nature*, **543**, 346–354.

5. Bolek, Y., El-Zik, K.M., Pepper, A.E., Bell, A.A., Magill, C.W., Thaxton, P.M. and Reddy, O.U.K. (2005) Mapping of Verticillium wilt resistance genes in cotton. *Plant Sci.* **168**, 1581–1590.

6. Bouwmeester, K. and Govers, F. (2009) Arabidopsis L-type lectin receptor kinases: phylogeny, classification, and expression profiles. *J. Exp. Bot.* **60**, 4383–4396.

7. Bouwmeester, K., de Sain, M., Weide, R., Gouget, A., Klamer, S., Canut, H., *et al.* (2011) The lectin receptor kinase LecRK-I.9 is a novel Phytophthora resistance component and a potential host target for a RXLR effector. *PLoS Pathog.*, **7**, e1001327.

8. Chen, H., Iqbal, M., Yang, R.C. and Spaner, D. (2016) Effect of Lr34/Yr18 on agronomic and quality traits in a spring wheat mapping population and implications for breeding. *Mol. Breed.* **36**, 53.

9. Chen, B., Zhang, Y., Yang, J., Zhang, M., Ma Q.M., Wang X.F., Ma Z.Y. (2020) The G-protein a subunit GhGPA positively regulates *Gossypium hirsutum* resistance to *Verticillium dahliae* via induction of SA and JA signaling pathways and ROS accumulation. *Crop J.* https://doi.org/10.1016/j.cj.2020.09.008.

10. de Jonge, R., van Esse, H.P., Maruthachalam, K., Bolton, M.D., Santhanam, P., Saber, M.K., Zhang, Z. *et al.* (2012) Tomato immune receptor Ve1 recognizes effector of multiple fungal pathogens uncovered by genome and RNA sequencing. *Proc. Natl Acad Sci. USA.* **109**, 5110–5115.

11. Dehghan, A. (2018) Genome-wide association studies. *Methods Mol Biol.* **1793**, 37–49.
12. Desclos-Theveniau, M., Arnaud, D., Huang, T.Y., Lin, G.J.C., Chen, W.Y., Lin, Y.C. et al. (2012) The Arabidopsis lectin receptor kinase LecRK-V.5 represses stomatal immunity induced by Pseudomonas syringae pv. tomato DC3000. *PLoS Pathog.* **8**, e1002513.
13. Fang, L., Wang, Q., Hu, Y., Jia, Y., Chen, J., Liu B., Zhang, Z. et al. (2017) Genomic analyses in cotton identify signatures of selection and loci associated with fiber quality and yield traits. *Nat. Genet.* **49**, 1089–1098.
14. Fradin, E.F., Zhang, Z., Juarez Ayala, J.C., Castroverde, C.D.M., Nazar, R.N., Robb, J., Liu, C.M. et al. (2009) Genetic dissection of Verticillium wilt resistance mediated by tomato Ve1. *Plant Physiol.* **150**, 320–332.
15. Fukuoka, S., Saka, N., Koga, H., Ono, K., Shimizu, T., Ebana, K. and Hayashi N. et al. (2009) Loss of function of a proline-containing protein confers durable disease resistance in rice. *Science*, **325**, 998–1001.
16. Gao, Y., Wang, W., Zhang, T., Gong, Z., Zhao, H., Han, G. et al. (2018) Out of water: the origin and early diversification of plant R-genes. *Plant Physiol.* **177**, 82–89.
17. Gouhier-Darimont, C., Schmiesing, A., Bonnet, C., Lassueur, S. and Reymond, P. (2013) Signalling of *Arabidopsis thaliana* response to Pieris brassicae eggs shares similarities with PAMP-triggered immunity. *J. Exp. Bot.* **64**, 665–674.
18. Holub, E.B. (2001) The arms race is ancient history in Arabidopsis, the wildflower. *Nat. Rev. Genet.* **2**, 516–527.
19. Huang, P., Ju, H.W., Min, J.H., Zhang, X., Kim, S.H., Yang, K.Y., et al. (2013) Overexpression of L-type lectin-like protein kinase 1 confers pathogen resistance and regulates salinity response in Arabidopsis thaliana. *Plant Sci.* **203**, 98–106.
20. Huang, P.Y., Yeh, Y.H., Liu, A.C., Chien, C.P. and Zimmerli, L. (2014) The Arabidopsis LecRK-VI.2 associates with the pattern-recognition receptor FLS2 and primes *Nicotiana benthamiana* pattern-triggered immunity. *Plant J.* **79**, 243–255.
21. Korte, A. and Farlow, A. (2013) The advantages and limitations of trait analysis with GWAS: a review. *Plant Methods*, **9**, 29.
22. Kou, Y. and Wang, S. (2010) Broad-spectrum and durability: Understanding of quantitative disease resistance. *Curr. Opin. Plant Biol.* **13**, 181–185.
23. Kou, Y. and Wang, S. (2012) Toward an understanding of the molecular basis of quantitative disease resistance in rice. *J. Biotechnol.* **159**, 283–290.
24. Lee, J.M. and Sonnhammer, E.L.L. (2003) Genomic gene clustering analysis of pathways in eukaryotes. *Genome*
25. Li, W.T., Zhu, Z.W., Chern, M.S., Yin, J.J., Yang, C., Ran, L., Cheng, M.P. et al. (2017a) A natural allele of a transcription factor in rice confers broad-spectrum blast resistance. *Cell* **170**, 114–126.

26. Li, T.G., Ma, X.F., Li, N.Y., Zhou, L., Liu, Z., Han, H.Y., Gui, Y.J. et al. (2017b) Genome-wide association study discovered candidate genes of Verticillium wilt resistance in upland cotton (*Gossypium hirsutum* L.). *Plant Biotechnol. J.* **15**, 1520–1532.

27. Li, N., Lin, B., Wang, H., Li, X.M., Yang, F.F., Ding, X.H., Yan, J.B. et al. (2019) Natural variation in ZmFBL41 confers banded leaf and sheath blight resistance in maize. *Nat Genet.* **51**, 1540–1548.

28. Livak, K.J. and Schmittgen, T.D. (2000) Analysis of relative gene expression data using realtime quantitative PCR and the $2^{-\Delta\Delta Ct}$ method. *Methods*, **25**, 402–408.

29. Long, A. D. and Langley, C. H. (1999) The power of association studies to detect the contribution of candidate genetic loci to variation in complex traits. *Genome Res.* **9**, 720-731.

30. Luo, X., Xie, C., Dong, J., Yang, X., Sui, A. (2014) Interactions between *Verticillium dahliae* and its host: vegetative growth, pathogenicity, plant immunity. *Appl. Microbiol. Biotechnol.* **98**, 6921–6932.

31. Ma, Z., He, S., Wang, X., Sun, J., Zhang, Y., Zhang, G., Wu, L., et al. (2018) Resequencing a core collection of upland cotton identifies genomic variation and loci influencing fiber quality and yield. *Nat. Genet.* **50**, 803–813.

32. Meyers, B.C., Kozik, A., Griego, A., Kuang, H., Michelmore, R.W. (2003) Genome-wide analysis of NBS-LRR encoding genes in Arabidopsis. *Plant Cell* **15**, 809–834.

33. Michalak, P. (2008) Coexpression, coregulation, and cofunctionality of neighboring genes in eukaryotic genomes. *Genomics*, **91**, 243–248.

34. Miller, C., Wells, R., McKenzie, N., Trick, M., Ball, J., Fatihi, A., Dubreucq, B. et al. (2019) Variation in expression of the HECT E3 ligase UPL3 modulates LEC2 levels, seed size and crop yields in Brassica napus. *Plant Cell*, **31**, 2370–2385.

35. Muchero, W., Sondreli, K., Chen, J., Urbanowicz, B., Zhang, J., Singan, V., Yang, Y. et al. (2018) Association mapping, transcriptomics, and transient expression identify candidate genes mediating plant-pathogen interactions in a tree. *Proc. Natl Acad. Sci. USA* **115**, 11573–11578.

36. Osbourn, A.E. and Field, B. (2009) Operons. *Cell. Mol. Life Sci.* **66**, 3755-3775.

37. Pfaffl, M.W. (2002) A new mathematical model for relative quantification in real-time RT-PCR. *Nucleic Acids Res.* **30**, 3, e11.
38. Piffanelli, P., Zhou, F., Casais, C., Orme, J., Jarosch, B., Schaffrath, U., Collins, N. et al. (2002) The barley MLO modulator of defense and cell death is responsive to biotic and abiotic stress stimuli. Plant Physiol. 129, 1076–1085.

39. Said, J.I., Song, M.Z., Wang, H.T., Lin, Z.X., Zhang, X.L., Fang, D.D. and Zhang, J.F. (2015) A comparative meta-analysis of QTL between intraspecific Gossypium hirsutum and interspecific G. hirsutum × G. barbadense populations. Mol. Genet. Genomics, 290, 1003–1025.

40. Seo, E., Kim, S., Yeom, S. and Choi, D. et al. (2016) Genome-wide comparative analyses reveal the dynamic evolution of nucleotide-binding leucine-rich repeat gene family among Solanaceae plants. Front. Plant Sci. 7, 1205.

41. Singh, P., Chien, C.C., Mishra, S., Tsai, C.H. and Zimmerli, L. (2013) The Arabidopsis LECTIN RECEPTOR KINASE-VI.2 is a functional protein kinase and is dispensable for basal resistance to Botrytis cinerea. Plant Signal Behav. 8, 1–3.

42. Singh, P., Kuo, Y.C., Mishra, S., Tsai, C.H., Chien, C.C., Desclos-Theveniau, M., et al. (2012) The lectin receptor kinase-VI.2 is required for priming and positively regulates Arabidopsis pattern-triggered immunity. Plant Cell, 24, 1256–1270.

43. Song, R.R., Li J.P., Xie C.J., Jian W. and Yang X.Y. (2020) An overview of the molecular genetics of plant resistance to the Verticillium wilt pathogen Verticillium dahliae. Int. J. Mol. Sci. 21, 1120.

44. Steuernagel, B., Witek, K., Krattinger, S., Ramirez-Gonzalez, R., Schoonbeek, H., Yu, G., Baggs, E. et al. (2018) Physical and transcriptional organisation of the bread wheat intracellular immune receptor repertoire. bioRxiv 339424.

45. Sun X., Wang J., Huang X., Cai Y. and Ji, Z. (2013) Analysis on the relationship between defoliating type of Verticillium wilt of cotton and meteorological factors in the Yellow River Valley. Journal of Shanxi Agricultural Sciences, 41, 167–171.

46. Wang, G.N., Zhang, G.Y., Yue, X.W., Li, Z.K., Zhang, Y., Zhang, G.Y., Wu, L.Q. et al. (2012) Pathogenicity and ISSR genetic differentiation of Verticillium dahliae isolates from cotton growing areas of Hebei province. Cotton Sci. 24, 348–357.

47. Wang, M., Tu, L., Yuan, D., Zhu, D., Shen, C., Li, J., Liu, F. et al. (2019) Reference genome sequences of two
cultivated allotetraploid cottons, *Gossypium hirsutum* and *Gossypium barbadense*. Nat. Genet. **51**, 224–229.

48. Wang, Y. and Bouwmeester, Klaas. (2017) L-type lectin receptor kinases: New forces in plant immunity. *PLoS Pathog.*, **13**, e1006433.

49. Wang, Y., Bouwmeester, K., Beseh, P., Shan, W. and Govers, F. (2014) Phenotypic analyses of Arabidopsis T-DNA insertion lines and expression profiling reveal that multiple L-type lectin receptor kinases are involved in plant immunity. *Mol. Plant Microbe Interact.* **27**, 1390–1402.

50. Wang, Y., Liang, C., Wu, S., Zhang, X., Tang, J., Jian, G., Jiao, G. *et al*. (2016) Significant improvement of cotton Verticillium wilt resistance by manipulating the expression of Gastrodia antifungal proteins. *Mol. Plant*, **9**, 1436–1439.

51. Wang, Y., Weide, R., Govers, F. and Bouwmeester, K. (2015) L-type lectin receptor kinases in Nicotiana benthamiana and tomato and their role in Phytophthora resistance. *J. Exp. Bot.* **66**, 6731–6743.

52. Wersch, S. and Li, X. (2019) Stronger when together: Clustering of plant NLR disease resistance genes. *Trends Plant Sci.* **24**, 8–19.

53. Woo, J.Y., Jeong, K.J., Kim, Y.J. and Paek, K.H. (2016) CaLecRK-S.5, a pepper L-type lectin receptor kinase gene, confers broad-spectrum resistance by activating priming. *J. Exp. Bot.* **67**, 5725–5741.

54. Xu, J., Wang, X., Li, Y., Zeng, J., Wang, G., Deng, C. and Guo, W. (2018) Host-induced gene silencing of a regulator of G protein signaling gene (*VdRGS1*) confers resistance to Verticillium wilt in cotton. *Plant Biotechnol. J.* **16**, 1629–1643.

55. Yang, Z.R., Qammar, G., Wang, Z., Yang, Z.E. and Li, F.G. (2020) *Gossypium* genomics: trends, scope, and utilization for cotton improvement. *Trends Plant Sci.* **25**, 488–500.

56. Zhang X.H. and Li J. (2006) Advance in occurrence and researches on cotton Verticillium wilt. *Acta Agriculturae Jiangxi*, **18**, 99–104.

57. Zhang, Y., Wang, X.F., Yang, S., Chi, J.N., Zhang, G.Y., Ma, Z.Y. (2011) Cloning and characterization of a Verticillium wilt resistance gene from Gossypium barbadense and functional analysis in *Arabidopsis thaliana*. Plant Cell Rep. **11**, 2085-2096.

58. Zhang, Y., Wang, X.F., Rong, W., Yang, J., Ma, Z.Y. (2016) Island cotton enhanced disease susceptibility I gene encoding a lipase-like protein plays a crucial role in response to *Verticillium dahliae* by regulating the SA level and H₂O₂ accumulation. *Front Plant Sci.* **7**, 183.
59. Zhang Y., Wang X., Rong W., Yang J., Li Z., Wu L., Zhang G. and Ma Z. (2017) Histochemical analyses reveal that stronger intrinsic defenses in *Gossypium barbadense* than in *G. hirsutum* are associated with resistance to *Verticillium dahliae*. *Mol. Plant Microbe Interact.* **30**, 984–996.

60. Zhang, Y., Wu, L.Z., Wang, X.F., Chen, B., Zhao, J., Cui, J., Li, ZK. et al. (2019) The cotton laccase gene *GhLAC15* enhances *Verticillium* wilt resistance via an increase in defence-induced lignification and lignin components in the cell walls of plants. *Mol. Plant Pathol.* **20**, 309–322.

61. Zhou, X. and Stephens, M. (2012) Genome-wide efficient mixed-model analysis for association studies. *Nat. Genet.* **44**, 821–824.

62. Zhou, X., Liao, H., Chern, M., Yin, J., Chen, Y., Wang, J., Zhu, X. et al. (2018) Loss of function of a rice TPR-domain RNA-binding protein confers broad-spectrum disease resistance. *Proc. Natl Acad Sci. USA* **115**, 3174–3179.

63. Zhu, H., Li, F., Feng, H., Zhao, L., Shi, Y., Liu, S. and Zhou, J. (2017) Overview of cotton *Verticillium* wilt research over the past decade in China and its prospect in future. *Cotton Sci.* **29**, 37–50.

**Figure legends**

**Figure 1** The cotton VW disease resistance identification and GWAS for VW resistance in cotton. (a) The disease symptoms of representative accessions under diverse pathogenicity *V. dahliae* in green house. The comprehensive performance reflected by disease index value of each plot containing 35 seedlings in one repetitive experiment. (b) Presentation of significant Manhattan and Q-Q Plots of GWAS results in four datatypes.

**Figure 2** Natural variation in *Ghir_D11G033410.1* was significantly associated with cotton resistance to *V. dahliae*. (a) Local Manhattan plot (top) and LD heat map (bottom) surrounding the peak on Dt11. The statistical analysis was performed with two-tailed Wald test. Red arrow indicates the position of the nonsynonymous (D11:69088571) SNP. (b) Gene structure of *Ghir_D11G033410.1*. Ref., reference; alt., alternate. (c) Box plots for disease index, based on the haplotypes of the nonsynonymous SNP. In the box plots, the center line denotes the median, box limits are the upper and lower quartiles, and whiskers mark the range of the data. n indicates the number of accessions with the same genotype. The significance of difference was analyzed with two-tailed t test. (d) The expression of *Ghir_D11G033410.1* in NDM8 under normal growth condition, indicating that *Ghir_D11G033410.1* expression level did not significant change during the plant growth. (e) Different expression patterns of *Ghir_D11G033410.1* during 0 to 48 hours post *V. dahliae* infection in two types cottons via qPCR. (f) Comparison the *Ghir_D11G033410.1*
expression level between the resistant (GG) and the susceptible (AA) varieties at 12 hpi, detected by qPCR. 
Ghhistone3b was used as an internal control. Data are represented as average values with s.d. (n=3 technical replicates). Data are represented as average values with s.d. (n=3 technical replicates). (g) Ghir_D11G033410.1 localized to the cytomembrane of Arabidopsis protoplast cells and onion epidermal cells. Scale bars, 50 μm. (h, j) The disease resistant performance of silencing Ghir_D11G033410.1 in tolerant (NDM8) and susceptible (CCRI8) cotton cultivars, respectively. (i, k) the disease resistance performance of transgenic Arabidopsis and wild type. The picture photographed at 20 dpi and representative of two independent experiments.

**Figure 3** Function validation of three downregulating genes (Ghir_D11G033550.1, ID13; Ghir_D11G033690.1, ID16; and Ghir_D11G033820.1, ID 32) in cotton VW resistance. Silencing of each gene in the tolerant cultivar NDM8 (a) and susceptible cultivar CCRI8 (b), respectively. Ten-day-old seedlings were hand-infiltrated with Agrobacterium carrying target gene in the VIGS vector. Ten days after infiltration, the seedlings were inoculated with LX2-1. The DI was the mean value of two independent experiments and the picture photographed at 20-day post inoculation.

**Figure 4** Function validation of the upregulating genes in the associated D11 region. Each gene was successfully silenced in the tolerant cultivar NDM8. Ten days after infiltration, all the silenced plants were inoculated with LX2-1. The DI was the mean value of two independent experiments and the picture photographed at 20 dpi. The image of mock plants (Mock-experiment) was the same as used in the Figure 3 (NDM8-mock) due to all the function validation assay was performed in an experiment panel and shared the mock plants.

**Figure 5** V. dahliae recovery assay and fungal biomass detection. (a) Mycelia growth in vascular tissue of surface-sterilized hypocotyl section prepared from cotton seedling at 7 dpi, then incubated for 7 d on 25% potato dextrose agar. (b) Detection the biomass of the V. dahliae in the silenced plants with qPCR. The relative average fungal biomass is shown with standard errors. The average fungal biomass was determined using at least three inoculated plants for each gene.

**Figure 6** Validation the function of six genes by ectopic expression in Arabidopsis. (a) Three-week-old plants were inoculated with LX2-1 (5×10⁷ conidia mL⁻¹). Transgenic lines of gene 5, 14 and 15 showed obvious improved resistance, and transgenic lines of gene 10, 13 and 16 displayed seriously susceptible than wild type plants. The images are representative of three independent transgenic lines at 20 dpi. (b) The average DI of each transgenic Arabidopsis at 20 dpi based on three independent experiments.

**Supplementary files**

This article is protected by copyright. All rights reserved
Figure S1 Correlation analysis of Verticillium wilt resistance in different conditions. ** indicates extremely significant difference at $P = 0.01$.

Figure S2 Expression analysis of the candidate genes upon *V. dahliae* based on transcriptome data.

Figure S3 Correlation analysis between 10 genes (upregulating genes ID: 5, 7, 11, 12, 14 and 15; downregulating genes ID: 2, 10, 13 and 16) expression and disease index from 22 accessions inoculated with *V. dahliae* at 12 hpi. We investigated 10 candidate genes expression in 22 accessions via qPCR. The expressed candidate gene number and corresponding expression level were regarded as multi-gene expression. (a) The correlation analysis between multi-gene expression and disease index based on upregulating gene. (b) The correlation analysis between multi-gene expression and disease index based on downregulating genes. *Ghistone3b* was used as an internal control. Data are represented as average values with s.d. (n = 3 technical replicates).

Figure S4 The analysis of core SNPs and disease index based on two types of cotton accessions.

Figure S5 The newly development cotton varieties with elite alleles displayed good VW resistance. (a) The newly developmental cotton varieties NDM23 and NDM24 displayed better VW resistance than TM-1 in the field nursery. (b) The information of 13 core SNPs in five newly development cotton varieties based on resequencing.

Table 1 Summary of the phenotypic variation of 401 cotton accessions.

Table S1 Analysis of variance (ANOVA) results of VW resistance.

Table S2 The disease index of 401 accessions upon *V. dahliae* stress.

Table S3 Information of the resequencing data of the accessions.

Table S4 Information of SNPs used in this study.

Table S5 List of the associated SNPs identified through GWAS for VW resistance in cotton.

Table S6 Information of the core SNPs identified in this study.

Table S7 The causal genes associated with 13 core SNPs.

Table S8 The information of 25 genes differentially expressed under *V. dahliae* stress in ND601.

Table S9 Elite-allele frequency in early & middle and modern varieties.

Table S10 The information of varieties with alternative genotype in D11 target region.

Table S11 A list for all the primers in this study.

Table 1 Summary of the phenotypic variation of 401 cotton accessions

This article is protected by copyright. All rights reserved
| Environment    | Mean | SD  | Min  | Max  | CV (%) |
|---------------|------|-----|------|------|--------|
| LX2-1 20 dpi  | 43.47| 9.71| 20.53| 70.71| 22.35  |
| LX2-1 25 dpi  | 63.71| 9.31| 32.14| 86.07| 14.61  |
| Vd991 20 dpi  | 23.42| 7.04| 8.57 | 49.29| 30.06  |
| Vd991 25 dpi  | 37.90| 10.74| 11.79| 71.07| 28.34  |

SD, standard deviation; CV, coefficient of variation.
