PROF. XIANLONG   ZHANG (Orcid ID : 0000-0002-7703-524X)

Article type   : Research Article

Long non-coding RNAs involve in resistance to *Verticillium dahliae*, a fungal disease in cotton

Lin Zhang[1], Maojun Wang[1], Nannan Li[1], Honglei Wang[1], Ping Qiu[1], Liuling Pei[1], Zheng Xu[1], Tianyi Wang[1], Erlin Gao[1], Junxia Liu[1], Shiming Liu[1], Qin Hu[1], Yuhuan Miao[1], Keith Lindsey[2], Lili Tu[1], Longfu Zhu[1] and Xianlong Zhang[1]*

[1]National Key Laboratory of Crop Genetic Improvement, Huazhong Agricultural University, Wuhan 430070, Hubei, China

[2]Integrative Cell Biology Laboratory, School of Biological and Biomedical Sciences, Durham University, South Road, Durham, DH1 3LE, UK

*Corresponding author: Xianlong Zhang

E-mail: xlzhang@mail.hzau.edu.cn

Tel: +86-27-87280510

Fax: +86-27-87280196

Accession number: PRJNA360482

SUMMARY

Long non-coding RNAs (lncRNAs) have several known functions in plant development, but their possible roles in responding to plant disease remain largely unresolved. In this study, we described a comprehensive disease-responding lncRNA profiles in defense against a cotton fungal disease *Verticillium dahliae*. We further revealed the conserved and specific characters of disease responding process between two cotton species. Conservatively for two cotton species, we found the expression dominance of induced lncRNAs in the Dt subgenome, indicating a biased induction pattern in the co-existing subgenomes of allotetraploid cotton. Comparative analysis of lncRNA expression and their proposed functions in resistant *Gossypium barbadense* cv. '7124' versus susceptible *G. hirsutum* cv. 'YZ1' revealed their distinct disease response mechanisms. Species-specific (LS) lncRNAs containing more SNPs displayed a fiercer inducing level post infection than the species-conserved (core) lncRNAs. Gene Ontology enrichment of LS lncRNAs and core lncRNAs indicates distinct roles in the process of biotic stimulus. Further functional analysis showed that two core lncRNAs *GhlncNAT-ANX2-* and *GhlncNAT-RLP7-* silenced seedlings displayed an enhanced resistance towards *V. dahlia*e and *Botrytis cinerea*, possibly associated with the increased expression of *LOX1* and *LOX2*. This study represents the first characterization of lncRNAs involved in resistance to fungal disease and provides new clues to elucidate cotton disease response mechanism.

INTRODUCTION

The transcriptional landscape in eukaryotes has been extensively studied using RNA sequencing (RNA-seq) and reveals that RNA molecules are transcribed ranging from protein-coding mRNAs to non-coding transcripts (Yamada *et al.*, 2003; Chekanova *et al.*, 2007; Kapranov *et al.*, 2007; Berretta and Morillon, 2009; Ponting *et al.*, 2009; Sanchez-Leon *et al.*, 2012; Zhu *et al.*, 2014). Non-coding RNAs are classified into two types, containing either short sequences (less than 200 nt), or long non-coding RNAs (lncRNAs, longer than

200 nt) (Bertone *et al.*, 2004; Guttman *et al.*, 2009; Cabili *et al.*, 2011; Wang *et al.*, 2014a; Zhou *et al.*, 2014). lncRNAs can in turn be classified into long intergenic non-coding RNAs (lincRNAs), natural antisense transcripts (NATs) and intronic RNAs (incRNAs) based on genome location (Ponting *et al.*, 2009; Chen, 2012; Rinn and Chang, 2012; Dogini *et al.*, 2014; Ma *et al.*, 2014; Weick and Miska, 2014; Wang *et al.*, 2015a). Studies of the biological roles of lncRNAs are challenging because of their diverse expression and regulation patterns across a wide range of cells and tissues (Orom and Shiekhattar, 2011). lncRNAs realized their functions mostly as signals, decoys, guides and scaffolds (Wang and Chang, 2011).

Although a large number of lncRNAs have been identified from sequencing data, only a few lncRNAs are functionally well-characterized in plants. Two lncRNAs from *Arabidopsis*, *COOLAIR* and *COLDAIR*, have been characterized from *FLOWERING LOCUS C (FLC)* that acts as a floral repressor (Swiezewski *et al.*, 2009; Heo and Sung, 2011). In rice, *LONG-DAY-SPECIFIC MALE-FERTILITY-ASSOCIATED RNA* (*LDMAR*), exerting like a structure lncRNA, regulates photoperiod-sensitive male sterility (PSMS) (Ding *et al.*, 2012). In *Medicago truncatula*, the lncRNA *Enod40* involves symbiotic interactions with soil-rhizobia in nodule formation by regulating the relocalization of a nuclear RBP (Campalans *et al.*, 2004). Several lncRNAs responding to *Fusarium oxysporum* infection have been identified in *Arabidopsis*, but with unknown function (Zhu *et al.*, 2014). Recently, lncRNA ELF18-INDUCED LONG-NONCODING RNA1 (ELENA1) identified in *Arabidopsis* enhanced the resistance against *Pseudomonas syringe* via interacting with Mediator subunit 19a to regulate *PR1* (Seo *et al.*, 2017). These findings highlight the essential function and increasing attention of lncRNAs in plant biology and in controlling important agronomic traits.

Cotton (*Gossypium* spp.) has long been widely cultivated for its renewable textile fiber and seeds oil. More than 90% of cultivated cotton was allotetraploid, which originated from the accidently merging of two progenitor donors with A genome and D genome respectively (much like modern *G. arboretum* and *G. raimondii*) 1-2 million years ago (Wendel & Cronn, 2003; Wendel *et al.*, 2012; Zhang *et al.*, 2015b). It takes thousands of years for human to domesticate cotton from wild to modern cultivated cotton, which produces the spinnable, fine white fibers (Wang *et al.*, 2017). However, China now faces the huge economic loss resulting

from the sharply decreased cotton yield and quality which were destroyed by *Verticillium* wilt (VW). VW is caused by soil-borne fungus *Verticillium dahliae*, which worldwide invades more than 400 plant species hosts (Zhang *et al.*, 2016; Li *et al.*, 2017). This disease will lead to chlorosis and wilt of leaves or defoliation, the browning of vascular and even death ultimately (Xu *et al.*, 2011; Li *et al.*, 2014). It has been the major challenge for cotton and deserves enormous researches to control efficiently.

Plants possess a multilayered immune system to counteract pathogens through both constitutive and inducible defenses, such as physical and chemical barriers, pattern recognition receptors (PRRs), resistance genes (R genes) encoding proteins containing a nucleotide-binding site (NBS) with leucine-rich repeats (LRRs) (Jones and Dangl, 2006; Bent and Mackey, 2007; Yang *et al.*, 2013). However, the recognition of apoplastic pathogen effectors is mediated by receptor-like proteins (RLPs), such as Ve1 (de Jonge *et al.*, 2012). In addition, some plant hormones, such as jasmonic acid (JA), salicylic acid (SA) and ethylene (ET), act as immunity signal molecules (Bari and Jones, 2009). SA can activate effective defense responses against hemibiotrophs and biotrophs, which are important for the establishment of systemic acquired resistance (SAR) (Dempsey and Klessig, 2012; Yang *et al.*, 2015). JA functions with ethylene to activate resistance against necrotrophic pathogens (Thaler *et al.*, 2004; Cacas *et al.*, 2016).

Contrasted with protein-coding genes, immunity related lncRNAs are less well documented in plant immunity. However, advanced sequencing data will unveil profiles of lncRNAs and provide new insights and promising lncRNA candidates in this area. There were some small RNAs identified related to cotton defense againt *V. dahliae* (Yin *et al.*, 2012; He *et al.*, 2014), but information related to lncRNAs in cotton was restricted to fiber development (Wang *et al.*, 2015b). The availability of the complete genome sequences of *G. barbadense* and *G. hirsutum* has made it possible to conduct a genome-wide comparative analysis of lncRNAs associated with disease response (Yuan *et al.*, 2015; Zhang *et al.*, 2015b).

Here, we reported the first charaterization of resistance-associated lncRNAs in two distinct cotton species, *G. barbadense* (which is resistant to VW) and *G. hirsutum* (which is susceptible). We showed that the different resistance responses were caused by genomic divergence between the two tetraploid cotton species. We related disease response to lncRNA profile, and identified functional lncRNAs in the cotton immune response following infection by *V. dahliae*.

RESULTS

Identification and characterization of lncRNAs in cotton root

Verticillium dahliae primarily infects cotton from roots, thus we are interested in analyzing lncRNAs profiles in roots. Two cotton species, *G. barbadense* (resistant) and *G. hirsutum* (susceptible), were inoculated for sequencing root samples (Figure 1a and S1). We generated twelve high-depth transcriptomes consisting of more than 1.5 billion clean reads, of which six were produced from *G. barbadense* and the other six were produced from *G. hirsutum* (Figure S1). We used an integrated approach (see Experimental procedures) to identify high-confidence lncRNAs for each cotton species. Four classes of lncRNAs were identified, and the majority of them were long intergenic noncoding RNAs (lincRNAs) and long noncoding natural antisense transcripts (lncNATs). In total, there were 13,452 loci of lincRNAs and 1,297 loci of lncNATs in *G. barbadense*, and 14,547 loci of lincRNAs and 1,406 loci of lncNATs in *G. hirsutum* (Table 1). The numbers of lincRNAs in the At subgenome were larger than those in the Dt subgenome, for *G. barbadense* and *G. hirsutum* (Figure 1b). However, the numbers of lncNATs in the At and Dt subgenome were similar (Figure 1b).

To prove that the full transcriptomes from libraries with only removal of rRNAs contain more complete non-coding message, we also sequenced two stranded libraries in which only mRNAs with poly(A) tails were retained for comparison. As expected, we found more lncRNAs were identified in libraries with the removal of rRNAs than in the regular stranded libraries following the same identification procedure (Table S1). For example, more than 32% of lincRNAs and 159% (namely 1.6 fold) of lncNATs were identified in the full transcriptome of sample Y12m (Table S1).

GC content, which reflects the biased intergenomic nonreciprocal DNA exchanges (Guo *et al.*, 2014), was investigated for lncRNAs. The result showed that lincRNAs and lncNATs exhibited lower GC content than protein-coding genes in both cultivars (Figure 1c). lincRNAs were found to have a lower GC content than lncNATs. There was no difference in GC content between *G. barbadense* and *G. hirsutum* both for lincRNAs and lncNATs.

The average length of protein-coding transcripts (1180 bp) was similar to sequence length of lncNATs (1061 bp in *G. barbadense*, 1150 bp in *G. hirsutum*), but was larger than those of lincRNAs both in *G. barbadense* and *G. hirsutum* (678 bp in *G. barbadense*, 729 bp in *G. hirsutum*). lncNATs and protein-coding transcripts exhibited a similar trend of length distribution. In contrast, lincRNAs showed an earlier peak primarily because of the large population of short sequences (Figure 1d). Analysis of exon number distribution revealed that all types of single-exonic transcripts represented the largest proportion (Figure 1e). The ratio of single-exonic lncRNAs was extremely high, especially for lncNATs in *G. barbadense* (72.6%). However, single-exonic protein-coding transcripts had the lowest ratio (29.9%).

Biased expression of lncRNAs upon infection in co-existing subgenomes

Homoeologous expression bias was found to exist widely in allopolyploids species, presenting underexplored scale in transcriptomic diversity and evolution process (Yoo *et al.*, 2013; Yuan *et al.*, 2015). We found that, in *G. barbadense*, the induced ratio of lincRNAs from the Dt subgenome is 0.094, while the ratio from the At subgenome was 0.082 (Figure 2a). In *G. hirsutum*, induced ratios of lincRNAs from the Dt subgenome and the At subgenome were 0.127 and 0.113, respectively (Figure 2b).

Using the reciprocal best match alignment, there are 1757 homoeologous lincRNA pairs between the At subgenome and Dt subgenome in *G. barbadense*. We obtained 187 lincRNAs pairs for At-biased induced expression and 337 lincRNAs pairs for Dt-biased induced expression (Figure 2c,e). We also found that 485 pairs showed no-biased expression and the other 1400 pairs showed a chimeric expression pattern (Figure 2d,f). Simultaneously in *G. hirsutum*, among 2026 homoeologous lincRNA pairs, we found 223 lincRNAs pairs of At-biased, 352 pairs of Dt-biased induced pattern after inoculation.

To further examine this biased distribution of disease-response loci, we collected evidence as complete as possible from prior quantitative trait locus (QTL) mapping results for *Verticillium wilt* resistance, and found the biased distribution in two subgenomes (Total number At: 76; Dt: 97; summarized in Figure S2 and detailed in Table S2) (Wang *et al.*, 2008; Yang *et al.*, 2008; Jiang *et al.*, 2009; Fang *et al.*, 2013a; Fang *et al.*, 2013b; Zhiyuan *et al.*, 2013; Wang *et al.*, 2014b; Zhang *et al.*, 2014a; Zhao *et al.*, 2014; Zhang *et al.*, 2015a; Fang *et al.*, 2017b; Li *et al.*, 2017).

Previous studies had shown that neighbor protein-coding genes might have functional connections with lncRNAs and might have similar expression profiles (Wierzbicki *et al.*, 2008; Wang *et al.*, 2015b; Engreitz *et al.*, 2016; Luo *et al.*, 2016). The possibility was measured by calculating the Pearson correlation coefficients (r_p) for three groups: lincRNAs and their adjacent protein-coding genes (lincRNA-PCgene: 5928 pairs); lncNATs and their paired protein-coding genes on opposite strand (lncNAT-PCgene: 1407 pairs); protein-coding genes and nearest protein-coding genes (PCgene-PCgene pairs: randomly selected 5000 pairs). In contrast with the random PCgene pairs, there were higher positive correlations in identified lncRNA-associated pairs (Figure S3). For instance, we noticed the high ratio of positively correlated lincRNA-PCgene pairs (10% vs 5%; $r_p>0.8$) and lncNAT-PCgene pairs (12% vs 6%; $r_p>0.8$).

Gene Ontology (GO) enrichment of lncRNAs was putatively conducted according to the functional annotations of neighbor protrein-coding genes. The results displayed that At-biased induced lincRNAs were enriched in kinase activator activity, fructose-bisphosphate aldolase activity and structure-specific DNA binding (Table 2). Nevertheless, Dt-biased induced lincRNAs were enriched in signal transducer activity, MAP kinase activity and superoxide dismutase copper chaperone activity. This illustrates distinct disease response mechanisms of two divergent subgenomes, which may have resulted from asymmetric evolution during allopolyploid formation and long-term domestication (Zhang *et al.*, 2015b; Wang *et al.*, 2017).

Comparison of pathogen induced expression profiles of lncRNAs in two cotton species

The global expression patterns of lncRNAs in *G. barbadense* and *G. hirsutum* were respectively found to fall into three classes, as determined by a K-means method (Figures 3a and S4). Type I and Type II clusters represent positively and negatively induced lncRNAs, respectively. Type III represents the complex expression patterns during pathogen infection. For instance, there were 632 lncRNAs which were down-regulated in 6 hours post infection and then slightly up-regulated in later two time points (Figure 3a).

Intriguingly, we found distinct numbers of differentially expressed lncRNAs in two cotton species during the invasion of pathogens (*P*-value <0.05; log2 ratio of 1). There were a total of 1236 and 1907 differentially expressed lincRNAs in *G. barbardense* and *G. hirsutum* respectively (Figure 3b). The up-regulated lincRNAs occupied a large proportion (*G. hirsutum*: 69%; *G. barbardense*: 56%). In addition, there were 63 and 128 differentially expressed lncNATs in *G. barbardense* and *G. hirsutum* separately (Figure 3c). In 12 hours post infection, the number of differentially expressed lncNATs in *G. hirsutum* was even twice of that in *G. barbadense* (Figure 3c). It seemed that more lncRNAs were differentially expressed in susceptible species, which suggests fiercer disease response.

To compare the potential functions of lncRNAs between two different cotton species, the homologous lncRNAs between two cottons were identified by reciprocal BLAST alignment with the best hit. The differentially expressed homologous lncRNAs (3411 pairs) were divided into sixteen groups (I to XVI) (Figure 4). Groups I to III and VII to VIII contained up-regulated lncRNAs in *G. barbadense* and *G. hirsutum* separately; Group IV to VI and IX to XI exhibited a down-regulated pattern; Group XII to XIV showed a high level of expression in *G. barbadense* but a low level of expression in *G. hirsutum*; Group XV displayed a low level of expression in *G. barbadense* but a high level of expression in *G. hirsutum*; and Group XVI had a complex expression pattern, which was distinct from the other groups. GO enrichment analysis was conducted to infer the potential biology processes of lncRNAs for all groups except for group III (*P*-value < 0.01). For instance, group II was enriched in antioxidant activity and ncRNA 3'-end processing (Figure 4).

Characterization of species-conserved and species-specific lncRNAs in *G. barbadense* and *G. hirsutum*

Genetic variation is required for rapid adaptation and evolution in the battle between plants and pathogens (de Jonge *et al.*, 2013), which is expected to be reflected in lineage-specific (LS) genomic regions. To explore whether the LS lncRNAs contribute to pathogen resistance, we compared LS lncRNAs with core lncRNAs, i.e. those common between cotton species. We identified 9443 unique loci of core lncRNAs in *G. barbadense* and 9937 unique loci in *G. hirsutum*. LS lncRNAs were also identified in *G. barbadense* (3943 unique loci) and *G. hirsutum* (5183 unique loci) (Table 3). Intriguingly, we found that a higher ratio of LS lncRNAs was differentially induced compared with core lncRNAs in both cultivars (Table 3). We also found that LS lincRNAs showed higher expression levels than core lincRNAs in both cultivars (Wilcoxon rank sum test, *, P-value < 0.01; **, P-value < 0.001; Figure 5a,b), except at 6 hours post infection (6 hpi). LS lncNATs in *G. hirsutum* also exhibited a significantly stronger pathogen induction, but no such significant difference was seen in *G. barbadense* (Figure 5a,b). These suggest that LS lncRNAs have greater expression changes towards pathogen infection than core lncRNAs.

To further elucidate evolution force of species-conserved and species-specific lncRNAs, the existence frequency of transposable elements (TEs) and polymorphic single-nucleotide polymorphisms (SNP) were computed. TEs have long been recognized as a driving-force for genome variation. To explore whether TEs contribute to the evolution of species-specific lncRNAs, we calculated the occupation of TEs in LS lncRNAs and core lncRNAs. Unexpectedly, there was no obvious difference of TE distribution between LS and core lncRNAs in gene body and genic flanking regions (Figure S5). Therefore, the genetic variation was not mainly caused by TE insertions. In addition, SNP frequency in LS and core regions was calculated according to resequencing data using 58 *G. hirsutum* accessions and 70 *G. barbadense* accessions (Fang *et al.*, 2017a; Wang *et al.*, 2017). It was found that SNP frequency was consistently higher in LS lncRNAs than core lncRNAs in all comparisons (*P*-value <2.2e-16) (Figures 5c and S6). These results indicated that SNP widely contributed to the variation of LS lncRNAs and might evolve more rapidly than core lncRNAs.

GO enrichment analysis (*P*-value <0.01) showed that core lncRNAs were enriched in 'ncRNA metabolic process' and 'RNA methylation' (Figure 5d). LS lncRNAs were preferentially enriched in 'defense response process' and 'response to biotic stimulus' (Figure 5d).

Pairs of lncRNAs and neighbor genes and their expressions after inoculation

We collected differentially expressed gene pairs between protein-coding genes and neighbor lncRNAs after inoculation for further functional identification. These pairs were divided into two groups, lincRNA/protein-coding gene pairs and lncNAT/protein-coding gene pairs. A total of 63 pairs of lincRNA/protein-coding genes and 29 pairs of lncNAT/protein-coding genes were identified (Figure S7a,b). We found that a large number of gene pairs were enriched in plant-pathogen interaction pathways, plant hormone signal transduction and starch and sucrose metabolism (Figure S7c,d), suggesting their functional implications in responding to *V. dahliae* infection. Genes participated in plant-pathogen interaction were selected for further validation.

Antisense expression is enriched when genes respond to environmental factors and stresses (Xu *et al.*, 2011; Qi and Arkin, 2014; Luo *et al.*, 2016). Then the expression of lncNATs and neighbor paired protein-coding genes was investigated. It was found that their expression patterns were complex following infection, including reverse, similar or non-related patterns in the two species (Figure S8). For instance, Gh_A03G1709 and its paired lncNAT (XLOC_005731) showed a similar induced pattern in *G. hirsutum*, with both of them being up-regulated during pathogen invasion (Figure S8).

To validate the expression patterns of ten pairs from previously identified 29 pairs of lncNATs and the associated protein-coding genes (Figure S9), qRT-PCR experiment of protein-coding gene Gh_D06G1866 (named P2) and its overlapping lncNAT XLOC_051276 (named L2) were performed in both cotton cultivars (Figure S9). Expression of more genes were identified by qRT-PCR, including Gh_A01G1977 and its paired lncNAT XLOC_002524 (named P3 and L3), Gh_A03G0544 and XLOC_006187 (named P4, L4), Gh_A08G0154 and XLOC_019529 (named P6, L6), Gh_D08G1915 and XLOC_056034 (named P9, L9), Gh_D03G0546 and XLOC_040782 (named P10, L10), Gh_A03G1307 and XLOC_006730 (named P11, L11), Gh_A04G1172 and XLOC_007816 (named P12, L12), Gh_D05G3796 and XLOC_081611 (named P14, L14) and Gh_A13G0172 with XLOC_033015 (named P15, L15) (Figure S9). The majority (96%) of qRT-PCR results showed a strong correlation (r = 0.8) with the transcriptome sequencing data (Figure S9).

Functional candidate lncRNAs in resistance to *V. dahliae*

To annotate candidate genes that were associated with disease-induced response, we adopted a phylogenetic approach using known homologous genes in *Arabidopsis*. CrRLK1L family RLKs are regulated by the steroid hormones brassinosteroids (BRs) including several important receptor-like kinase genes, such as *ANXUR2* (*ANX2*), *ANXUR1* (*ANX1*), and *FERONIA* (*FER*) (Lindner *et al.*, 2012). They are known to play roles in fertilization by controlling the timing of pollen tube rupture (Miyazaki *et al.*, 2009). Some members regulate the development of cell wall, such as *THESEUS1* (*THE1*) controlling lignin accumulation (Hematy *et al.*, 2007).

In this study, we explored the function of one core lncRNAs *GhlncNAT-ANX2* (*L2*), involved in plant-pathogen interaction, which was differentially regulated by pathogen. *L2* was firstly differentially up-regulated at 6 hpi and then *P2* was later down-regulated in 12 hpi and 24 hpi in *G. hirsutum* after *V. dahliae* invasion (Figure S9). However, *L2* was slightly down-regulated at 6 hpi and 24 hpi in *G. barbadense* (Figure S9). We found that *L2*-related protein-coding gene and *GhANX2* (*P2*) belonged to the CrRLK1L family of RLKs, with the highest similarity with *ANX2* (Figure S10a). Virus-induced gene silencing (VIGS) of *L2* plants showed enhanced resistance to *V. dahliae*, with reduced wilting and leaf defoliation

(Figures 6a and S11a). Moreover, a fungal recovery assay on inoculated stem tissue showed reduced infection, and a reduced vascular browning phenotype also suggests an effect on infectivity (Figure 6a). The disease index (DI) and infected proportion of *L2*-suppressed seedlings was sharply reduced compared to controls at all stages of *V. dahliae* infection (Figures 6b and S11b).

Cell surface-located receptor like proteins (RLPs) have dual functions in plant development and immunity. Until now, only one locus (Ve1) has been shown to confer full resistance to *V. dahliae*, and is also known as RLP (Fradin *et al.*, 2009). Moreover, some homologs of *Ve1* from cotton may also be involved in disease resistance (Zhang *et al.*, 2011; Zhang *et al.*, 2012). In this study, another core lncRNAs *GhlncNAT-RLP7* (*L3*) involved in plant-pathogen interaction were differentially regulated when infected. *L3* was sharply up-regulated at 6 hpi and 12 hpi in *G. hirsutum*, but only slightly up-regulated at 12 hpi in *G. barbadense* (Figure S9). In this study, *L3* paired protein-coding gene *GhRLP7* (*P3*) was identified as *GhRLP7*, with the highest identity to *AtRLP7* in *Arabidopsis* (Figure S10b). *L3*-silenced plants showed an enhanced resistance compared with the control, with less wilting and etiolated leaves (Figures 6a and S11a). In addition, fewer colonies in a fungal recovery assay and less browning of the vascular bundles were detected (Figure 6a). The DI and infection ratio also suggested enhanced resistance (Figures 6b and S11b).

After validation by efficient silencing of these target lncNATs, the expression changes of their paired neighbor protein-coding genes were also checked (Figure 6c, d). Compared with the control, *P2* had a higher expression level in *L2*-silenced plants (Figure 6d). Similarly, in *L3*-silenced seedlings, *P3* was dramatically up-regulated (Figure 6g, h). These results suggest that the influence of lncNATs on its neighbor protein-coding genes seems to be negative.

Additionally, we detected the expression change of lipoxygenase 1 (*LOX1)* and lipoxygenase 2 (*LOX2)* both in *L2*- and *L3*-silenced seedlings. JA is a positive regulator of cotton immunity that regulates plant resistance to pests and pathogens (Rodriguez-Saona *et al.*, 2001; Gao *et al.*, 2013; Gao *et al.*, 2016), so the up-regulation of JA pathway genes, like *LOX1* and *LOX2*, might contribute to the enhanced resistance in *L2*- and *L3*-silenced plants (Figure 6e, f, i, j).

To further confirm our results, we performed the *in vitro* inoculation of *Botrytis cinerea* on cotton leaves. Consistent with *V. dahliae* inoculation results, both *L2-* and *L3-*silenced plants showed less necrosis, meaning an enhanced resistance towards *B. cinerea in vitro* (Figure 7a). The trypan blue staining results also supported this observation (Figure 7b). These findings were further supported by the statistics of symptom area, which showed a significantly smaller range of necrosis in both *L2-* and *L3-*silenced leaves (ANOVA, **P<0.01) (Figure 7c).

DISCUSSION

The availability of allotetraploid cotton genome sequences provides extensive information for genic regions and their functional annotations. Non-coding regions, which comprise a large proportion of the genomes, have not previously been well characterized. A study on the expression of lncRNAs that have a 3'-end poly(A) during cotton fiber development has been carried out (Zhang *et al.*, 2014b; Wang *et al.*, 2015b), but little is known about lncRNAs lacking a poly(A) tail. LncRNAs identified in the current study provides a comprehensive picture of non-coding genomic regions in cotton.

Accumulating evidence suggests a link between lncRNAs and human disease (Wapinski and Chang, 2011). Nevertheless, there is limited information about plant disease related lncRNAs. In our research, the global expression contour of disease-induced lncRNAs in allotetraploid cotton was firstly established. Intriguingly, the ratio of induced lncRNAs from the Dt subgenome was higher than that in the At subgenome in both cultivars '7124' (*Verticillium*-resistant *G. barbadense*) and 'YZ-1' (susceptible *G. hirsutum*). Besides, more Dt-biased homoeologous lincRNA pairs and Dt-located VW resistance loci based on reference collection were found. This phenomenon suggests that a distinct disease response and function might exist in the two subgenomes, providing a possibility for functional divergence of homoeologous genes following polyploidization. Allotetraploid cotton is formed from two divergent A and D genomes (Wendel and Cronn, 2003; Wendel *et al.*, 2012). Allopolyploidization leads to exclusive either At-biased or Dt-biased expression of homoeologous genes and asymmetric evolution of two subgenomes derived from different selective pressures (Yuan *et al.*, 2015; Zhang *et al.*, 2015b). Expression of A-homoeologous

positive selection genes (PSGs) is enriched during fiber elongation, while D-homoeologous PSG expression is enriched in response to superoxide and other stresses (Zhang *et al.*, 2015b). This predominant expression of the Dt subgenome in disease response provides a basis for understanding the evolution of the immune system in cotton.

Our research studied the distinct characters of species-specific lncRNAs and species-conserved lncRNAs in both cultivars. The comparison results unveiled genetic variation sources, which were mainly contributed by SNP but not TE insertion. The observed pathogen-inducible expression and GO enrichment of LS lncRNAs suggest an important role of LS genomic regions in responding to disease. The GO analysis of core lncRNAs indicates their potential correlation with RNA methylation. RNA m(6)A methylation directs the translational regulation of abiotic stress response (Zhou *et al.*, 2015; Xiang *et al.*, 2017), but its roles in immunity have not been discovered and deserves future investigation.

Validating the functions of lncRNAs represents a major challenge in understanding RNA-mediated gene regulation (Luo *et al.*, 2016). Although thousands of lncRNAs have been identified from transcriptome profilings, the functions of the vast majority of them remain unknown (Luo *et al.*, 2016). A variety of methods for investigating genomic location, chromatin features, tissue-specific expression, subcellular localization and co-expression of lncRNAs have been developed to predict and categorize their functions (Ponjavic *et al.*, 2009; Mondal *et al.*, 2010; Cabili *et al.*, 2011; Luo *et al.*, 2016). lncRNAs may function in modulating the transcription of their nearby genes (Luo *et al.*, 2016). In this study, lncRNAs have been categorized into several functional groups based on annotation of nearby protein-coding genes, which potentially helped predict unrecognized roles in response to pathogen inoculation. Therefore, we test expressional change of nearby protein-coding genes and detected the significant up-regulation when lncRNAs transcripts were silenced by VIGS, although more evidence should be provided. For example, chromatin isolation by RNA purification (ChIRP) and RNA immunoprecipitation (RIP) are necessary for elucidating this potential regulatory mechanism (Chu *et al.*, 2011; Quinn *et al.*, 2014).

This study represents the first to characterize the expression landscape of lncRNAs involved in plant responses to infection by *Verticillium* wilt. The enhanced resistance to *V. dahliae* and *B. cinerea* after silencing lncRNAs provides a possible road to improve the broad spectrum resistance towards multiple fungal pathogens. The identification of lncRNAs expressed in the context of plant defense may in the longer term provide new approaches for the genetic improvement of disease resistance traits in cotton. Future studies will be directed to understand the mechanism by which lncRNAs may regulate gene expression.

EXPERIMENTAL PROCEDURES

Plant material and fungal pathogen inoculation

Cotton seedlings of *G. barbadense* cv. 7124 and *G. hirsutum* cv. YZ1 were grown in vermiculite-filled pots and watered with Hoagland's solution under greenhouse conditions of 25℃ for 2 weeks, under a photoperiod of 14 h light and 10 h dark. *V. dahliae* were cultivated in Potato Dextrose Agar (PDA) medium for 3-4 days from storage at -80℃, and then high activity hyphae were collected and then cultivated in Czapk's medium for 3 days at 25 ℃. 10^6 spores per ml in deionized water were used as the final concentration for inoculation.

When two fully expanded leaves appeared, whole plants were taken from the vermiculite for inoculation using a dipping infection method with the spores of *V. dahliae*, and the inoculated plants were returned to the pots. Roots were harvested at 6 hours, 12 hours and 24 hours post inoculation. Plants treated with distilled water were collected at different time points for use as mock treatments. All samples were stored at -80 °C until further use.

Stranded RNA library construction and sequencing

High-quality RNA was extracted using a guanidine thiocyanate method (Zhu *et al.*, 2005). The Stranded libraries only removing rRNAs were constructed using the Ribo-Zero Kit (Illumina, San Diego, CA, USA) following the manufacturer's instructions. Sequencing was performed on the Illumina Hiseq^{TM} 2000 system in the Beijing Genomic Institute (BGI). The regular stranded libraries in which only mRNAs with poly(A) tails retained were constructed using Illumina TruSeq Stranded RNA Kit (Illumina, San Diego, CA, USA) and performed on the Illumina Hiseq^{TM} 2000 system.

LncRNA identification and classification

All sequence data were firstly processed by filtering the low-quality reads (the ratio of base with $Q < 10$ should be greater than 50% of whole read) and adapters. Reads derived from rRNA were removed by SOAP alignment. We mapped those reads to the cotton genome (*Gossypium hirsutum* L. cv. TM-1) by applying TOPHAT (Trapnell *et al.*, 2009; Zhang *et al.*, 2015b). Each transcriptome was assembled separately by CUFFLINKS while background noise was filtered based on Fragments Per Kilobase of transcript per Million base pairs sequenced (FPKM), length, coverage and status threshold (FPKM>0.5; length>200; coverage>1; status: OK) (Trapnell *et al.*, 2010). The separated gene models from the same cultivar were merged together using the CUFFMERGE procedure. Novel transcripts were detected by CUFFCOMPARE. The coding potential capability was calculated by Coding Potential Calculator (value < 0). Finally, the class code "u" represents the long intergenic noncoding RNAs (lincRNAs), class code "x" represents long noncoding natural antisense transcripts (lncNAT), class code "j" represents the sense transcripts, and class code "i" represents the intronic transcripts. The lincRNA/protein-coding gene pairs were restricted to nearby 5 kb regions and non-overlapping with 1 kb away from protein-coding genes.

Identification of species-common (core) and species-specific (LS) lncRNAs

As mentioned above, all separated transcriptome gtf files of *G. barbadense* were merged into one gtf file using CUFFMERGE with parameter -g. Simultaneously, all individual transcriptome gtf files of *G. hirsutum* were merged following the same procedure. These merged transcriptomes from *G. barbadense* and *G. hirsutum* made it possible to compare the loci of lncRNAs from different cultivars using CUFFCOMPARE. The class code "u" represents the specific lncRNAs between *G. barbadense* and *G. hirsutum*. Beyond this, sequences with similarity were discarded to ensure the reliability of identified specific lncRNAs according to reciprocal BLASTN results with E threshold (*E* value < 1e-10). The class code "=" represents the core lncRNAs between *G. barbadense* and *G. hirsutum* that share fully equal loci. A reciprocal BLASTN (*E* value < 1e-10) was also run to improve the confidence of identified core lncRNAs, and only those with high sequence similarity were retained for further analysis.

Expression analysis

We applied CUFFMERGE to merge multiple assemblies to get merged transcripts separately for two cotton cultivars. The expression of all identified lncRNAs was processed by CUFFDIFF and genes expressed differentially were obtained by the following criteria: adjusted *P* value less than 0.001 and at least two-fold FPKM change (Trapnell *et al.*, 2010). The expression of lncRNAs were normalized and then clustered into several groups by K-means in Gene Expression Similarity Investigation Suite software (Genesis; http://genome.tugraz.at/genesisclient/genesisclient_description.shtml).

GO enrichment analysis

All GO terms of listed genes were annotated using Blast2GO (https://www.blast2go.com) by comparing to the reference genome background ($P < 0.01$).

Phylogenetic analysis

Protein sequences were aligned by Clustalx (http://www.clustal.org). Phylogenetic trees were constructed using a unweighted paired-group method with arithmetic means (UPGMA) followed by bootstrap test in MEGA4 (http://www.megasoftware.net/mega4/) and visualized in FigTree (http://tree.bio.ed.ac.uk/software/figtree/) .

Virus-induced gene silencing (VIGS) vector construction and genetic transformation

The lncNATs and their paired protein-coding genes always have overlapping regions with each other, so it's essential to investigate the genomic organization and the overlap between them. We designed specific primers to amplify fragments (avoiding overlapping and conserved regions) to construct VIGS vectors as indicated in the scheme design to ensure the silencing specificity for each gene (Figure S12). Non-overlapping regions were identified by genic genomic locations; Non-conserved regions were found according to the NCBI Conserved Domain Search web service (http://www.ncbi.nlm.nih.gov/Structure/cdd/wrpsb.cgi). Primer pairs that were used to construct vectors are provided in Table S3. The preparation of TRV vectors and *Agrobacterium tumefaciens* in experiments were conducted as previously reported (Fradin *et*

al., 2009; Gao *et al.*, 2013). The fragment of candidate genes armed with infusion connections were separately inserted into the TRV:00 vectors. Positively ligated plasmids were transformed into *A. tumefaciens* GV3101. TRV1 vectors were respectively mixed with TRV vectors that comprised candidate genes or empty vector TRV:00 using equal amounts, then agro-infiltration by syringe was used to infiltrate 10-day-old seedling cotyledons of *G. barbadense* cv. 7124 (Gao *et al.*, 2013). TRV:CLA1 (*CLOROPLASTOS ALTERADOS 1*) was utilized as the positive control and the empty vector TRV:00 as the negative control.

VIGS plant inoculation and fungal recovery assay

Two weeks after infiltration, the bleaching phenotype of positive controls appeared. Then, we started to prepare inoculation. The seedlings were pulled out of the pot carefully, then the plant roots were dipped into the distilled water with 10^5 spores per litre for 1 minute. After that, each four inoculated seedlings were re-planted in one larger pot. We performed the inoculation with at least 16 plants for each treatment using *V. dahliae* isolate V991 with at least three biological replicates. Disease index (DI) for plant populations was calculated as previously described (Gao *et al.*, 2013). The higher score the population had, the lower the resistance. Similarly, the rate of diseased plant was used to estimate the susceptibility of the whole population.

After inoculation by *V. dahliae* for two weeks, the fresh stems on cotyledon nodes were collected from the same position on each plant and sterilized by 84 disinfectant for 5 minutes. After washing 3 times by sterilized water, disinfected edges were removed and the stems were cut into small pieces. Stem samples were inoculated on PDA medium and cultured at 25℃ for 3~5 days. The fungi in stem were inspected by light microscopy (Leica MZFLIII, Germany).

Similarly, we conducted the inoculation of *Botrytis cinerea* when bleaching phenotype of positive control appeared. *B. cinerea* for leaves inoculation was cultivated at 25℃ for 3~5 days. Only outermost part was utilized to guarantee the high pathogenicity of fungi. Then leaves with *B. cinerea* were cultivated in 25℃ for 3 days, then the disease symptom area in each leaves were calculated in photographs. Staining of leaves by trypan blue were boiled in

lactophenol-trypan blue for 15 minutes for the first step and then destained by chloral hydrate overnight as previously described (Gao *et al.*, 2013).

Real-time (RT) PCR analysis

Total plant RNA was extracted from cotton root using a guanidine thiocyanate method. The first stranded cDNA was synthesized from 2μg RNA using the M-MLV reverse transcript system (Promega, USA). We designed gene specific primers (design strategy was similar as mentioned above) to conduct the qRT-PCR verifications as indicated in the scheme design to ensure measure specificity for each gene (Figure S12). Quantitative real-time (RT) PCR was run at 95℃ for 3 minutes followed by 28-35 cycles at 95℃ for 20s, 55-60℃ for 20s and 72℃ for 20s. Quantitative RT-PCR was conducted on an ABI 7500 Real Time PCR system (Applied Biosystems, USA) with the iTag™ Universal SYBR® Green Supermix (Bio-Rad, USA). Gene expression levels were normalized to *UB7* expression (Tan *et al.*, 2013).

Data access

The stranded RNA-seq data generated from *G. barbadense* and *G. hirsutum* were submitted to NCBI Sequence Read Archive database with the BioProject ID PRJNA360482.

CONFLICT OF INTEREST

The authors have declared that no competing interests exist.

SUPPORTING INFORMATION LEGENDS

Additional Supporting Information may be found online in the supporting information tab for this article:

Figure S1. Summarized data for sequenced samples.

Figure S2. Summary of reported genetic mapping results about *Verticillium* wilt resistance loci.

Figure S3. Distribution of pearson correlation coefficient for putative paired and random pairs.

Figure S4. The global expression profiles of lncRNAs in *G. hirsutum*.

Figure S5. Distribution of transposon elements overlapping with or located within lincRNAs and lncNATs.

Figure S6. SNP distribution of lineage specific (LS) lncRNAs and core lncRNAs.

Figure S7. Functional implication of differentially induced pairs of lincRNAs and lncNATs.

Figure S8. Examples of plant pathogen interaction pathways that candidate genes are involved in.

Figure S9. Expression validation and correlation between qRT-PCR and transcriptomic analysis.

Figure S10. Phylogenetic trees of candidate lncNAT-paired protein coding genes.

Figure S11. Phenotypes and proportion statistics of infected plants.

Figure S12. The genomic location and scheme design of primers for verifying lncRNAs and protein-coding genes.

Table S1. Distribution of identified lncRNA number in two types of stranded libraries.

Table S2. *Verticilium* wilt resistance loci in cotton modified from references.

Table S3. List of PCR primers used in this study.

REFERENCES

Bari, R. and Jones, J.D. (2009) Role of plant hormones in plant defence responses. *Plant Mol. Biol.* **69**, 473-488.

Bent, A.F. and Mackey, D. (2007) Elicitors, Effectors, and R Genes: The new paradigm and a lifetime supply of questions. *Annu. Rev. Phytopathol.* **45**, 399-436.

Berretta, J. and Morillon, A. (2009) Pervasive transcription constitutes a new level of eukaryotic genome regulation. *EMBO Rep*. **10**, 973-982.

Bertone, P., Stolc, V., Royce, T.E. *et al.* (2004) Global identification of human transcribed sequences with genome tiling arrays. *Science*, **306**, 2242-2246.

Cabili, M.N., Trapnell, C., Goff, L., Koziol, M., Tazon-Vega, B., Regev, A. and Rinn, J.L. (2011) Integrative annotation of human large intergenic noncoding RNAs reveals global properties and specific subclasses. *Genes Dev*. **25**, 1915-1927.

Cacas, J.L., Pre, M., Pizot, M., Cissoko, M., Diedhiou, I., Jalloul, A., Doumas, P., Nicole, M. and Champion, A. (2016) GhERF-IIb3 regulates the accumulation of jasmonate and leads to enhanced cotton resistance to blight disease. *Mol. Plant. Pathol*. **18**, 825-836.

Campalans, A., Kondorosi, A. and Crespi, M. (2004) *Enod40*, a short open reading frame-containing mRNA, induces cytoplasmic localization of a nuclear RNA binding protein in *Medicago truncatula*. *Plant Cell*, **16**, 1047-1059.

Chekanova, J.A., Gregory, B.D., Reverdatto, S.V. *et al.* (2007) Genome-wide high-resolution mapping of exosome substrates reveals hidden features in the *Arabidopsis* transcriptome. *Cell*, **131**, 1340-1353.

Chen, X. (2012) Small RNAs in development - insights from plants. *Curr. Opin. Genet. Dev.* **22**, 361-367.

Chu, C., Qu, K., Zhong, F.L., Artandi, S.E. and Chang, H.Y. (2011) Genomic maps of long noncoding RNA occupancy reveal principles of RNA-chromatin interactions. *Mol. Cell*. **44**, 667-678.

de Jonge, R., Bolton, M.D., Kombrink, A., van den Berg, G.C., Yadeta, K.A. and Thomma, B.P. (2013) Extensive chromosomal reshuffling drives evolution of virulence in an asexual pathogen. *Genome Res*. **23**, 1271-1282.

de Jonge, R., van Esse, H.P., Maruthachalam, K. *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.

Dempsey, D.A. and Klessig, D.F. (2012) SOS - too many signals for systemic acquired resistance? *Trends Plant Sci.* **17**, 538-545.

Ding, J., Lu, Q., Ouyang, Y., Mao, H., Zhang, P., Yao, J., Xu, C., Li, X., Xiao, J. and Zhang, Q. (2012) A long noncoding RNA regulates photoperiod-sensitive male sterility, an essential component of hybrid rice. *Proc. Natl. Acad. Sci. USA*, **109**, 2654-2659.

Dogini, D.B., Pascoal, V.D., Avansini, S.H., Vieira, A.S., Pereira, T.C. and Lopes-Cendes, I. (2014) The new world of RNAs. *Genet. Mol. Biol.* **37**, 285-293.

Engreitz, J.M., Haines, J.E., Perez, E.M., Munson, G., Chen, J., Kane, M., McDonel, P.E., Guttman, M. and Lander, E.S. (2016) Local regulation of gene expression by lncRNA promoters, transcription and splicing. *Nature*, **539**, 452-455.

Fang, H., Zhou, H., Sanogo, S., Flynn, R., Percy, R.G., Hughs, S.E., Ulloa, M., Jones, D.C. and Zhang, J. (2013a) Quantitative trait locus mapping for *Verticillium* wilt resistance in a backcross inbred line population of cotton (*Gossypium hirsutum × Gossypium barbadense*) based on RGA-AFLP analysis. *Euphytica*, **194**, 79-91.

Fang, H., Zhou, H., Sanogo, S., Lipka, A.E., Fang, D.D., Percy, R.G., Hughs, S.E., Jones, D.C., Gore, M.A. and Zhang, J. (2013b) Quantitative trait locus analysis of *Verticillium* wilt resistance in an introgressed recombinant inbred population of Upland cotton. *Molecular Breeding*, **33**, 709-720.

Fang, L., Gong, H., Hu, Y. *et al.* (2017a) Genomic insights into divergence and dual domestication of cultivated allotetraploid cottons. *Genome Biol.* **18**, 33.

Fang, L., Wang, Q., Hu, Y. *et al.* (2017b) Genomic analyses in cotton identify signatures of selection and loci associated with fiber quality and yield traits. *Nat. Genet.* **49**, 1089-1098.

Fradin, E.F., Zhang, Z., Juarez Ayala, J.C., Castroverde, C.D., Nazar, R.N., Robb, J., Liu, C.M. and Thomma, B.P. (2009) Genetic dissection of *Verticillium* wilt resistance mediated by tomato Ve1. *Plant Physiol.* **150**, 320-332.

Gao, W., Long, L., Xu, L., Lindsey, K., Zhang, X. and Zhu, L. (2016) Suppression of the homeobox gene *HDTF1* enhances resistance to *Verticillium dahliae* and *Botrytis cinerea* in cotton. *J. Integr. Plant. Biol.* **58**, 503-513.

Gao, W., Long, L., Zhu, L.F., Xu, L., Gao, W.H., Sun, L.Q., Liu, L.L. and Zhang, X.L. (2013) Proteomic and virus-induced gene silencing (VIGS) Analyses reveal that gossypol, brassinosteroids, and jasmonic acid contribute to the resistance of cotton to *Verticillium dahliae*. *Mol. Cell. Proteomics,* **12**, 3690-3703.

Guo, H., Wang, X., Gundlach, H., Mayer, K.F., Peterson, D.G., Scheffler, B.E., Chee, P.W. and Paterson, A.H. (2014) Extensive and biased intergenomic nonreciprocal DNA exchanges shaped a nascent polyploid genome, *Gossypium* (cotton). *Genetics*, **197**, 1153-1163.

Guttman, M., Amit, I., Garber, M. *et al.* (2009) Chromatin signature reveals over a thousand highly conserved large non-coding RNAs in mammals. *Nature*, **458**, 223-227.

He, X., Sun, Q., Jiang, H. *et al.* (2014) Identification of novel microRNAs in the *Verticillium* wilt-resistant upland cotton variety KV-1 by high-throughput sequencing. *SpringerPlus*, **3**, 564.

Hematy, K., Sado, P.E., Van Tuinen, A., Rochange, S., Desnos, T., Balzergue, S., Pelletier, S., Renou, J.P. and Hofte, H. (2007) A receptor-like kinase mediates the response of *Arabidopsis* cells to the inhibition of cellulose synthesis. *Curr. Biol.* **17**, 922-931.

Heo, J.B. and Sung, S. (2011) Vernalization-mediated epigenetic silencing by a long intronic noncoding RNA. *Science*, **331**, 76-79.

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

Jones, J.D. and Dangl, J.L. (2006) The plant immune system. *Nature*, **444**, 323-329.

Kapranov, P., Willingham, A.T. and Gingeras, T.R. (2007) Genome-wide transcription and the implications for genomic organization. *Nat. Rev. Genet.* **8**, 413-423.

Li, C., He, X., Luo, X., Xu, L., Liu, L., Min, L., Jin, L., Zhu, L. and Zhang, X. (2014) Cotton WRKY1 mediates the plant defense-to-development transition during infection of cotton by *Verticillium dahliae* by activating *JASMONATE ZIM-DOMAIN1* expression. *Plant Physiol.* **166**, 2179-2194.

Li, T., Ma, X., Li, N., Zhou, L., Liu, Z., Han, H., Gui, Y., Bao, Y., Chen, J. and Dai, X. (2017) Genome-wide association study discovered candidate genes of *Verticillium* wilt resistance in upland cotton (*Gossypium hirsutum* L.). *Plant Biotechnol. J.*

Lindner, H., Muller, L.M., Boisson-Dernier, A. and Grossniklaus, U. (2012) CrRLK1L receptor-like kinases: not just another brick in the wall. *Curr. Opin. Plant Biol.* **15**, 659-669.

Luo, S., Lu, J.Y., Liu, L. *et al.* (2016) Divergent lncRNAs regulate gene expression and lineage differentiation in pluripotent cells. *Cell Stem Cell*, **18**, 637-652.

Ma, L., Bajic, V.B. and Zhang, Z. (2014) On the classification of long non-coding RNAs. *RNA Biol.* **10**, 924-933.

Miyazaki, S., Murata, T., Sakurai-Ozato, N., Kubo, M., Demura, T., Fukuda, H. and Hasebe, M. (2009) *ANXUR1* and *2*, sister genes to *FERONIA/SIRENE*, are male factors for coordinated fertilization. *Curr. Biol.* **19**, 1327-1331.

Mondal, T., Rasmussen, M., Pandey, G.K., Isaksson, A. and Kanduri, C. (2010) Characterization of the RNA content of chromatin. *Genome Res.* **20**, 899-907.

Orom, U.A. and Shiekhattar, R. (2011) Long non-coding RNAs and enhancers. *Curr. Opin. Genet. Dev.* **21**, 194-198.

Ponjavic, J., Oliver, P.L., Lunter, G. and Ponting, C.P. (2009) Genomic and transcriptional co-localization of protein-coding and long non-coding RNA pairs in the developing brain. *PLoS Genet.* **5**, e1000617.

Ponting, C.P., Oliver, P.L. and Reik, W. (2009) Evolution and functions of long noncoding RNAs. *Cell*, **136**, 629-641.

Qi, L.S. and Arkin, A.P. (2014) A versatile framework for microbial engineering using synthetic non-coding RNAs. *Nat. Rev. Microbiol.* **12**, 341-354.

Quinn, J.J., Ilik, I.A., Qu, K., Georgiev, P., Chu, C., Akhtar, A. and Chang, H.Y. (2014) Revealing long noncoding RNA architecture and functions using domain-specific chromatin isolation by RNA purification. *Nat. Biotechnol.* **32**, 933-940.

Rinn, J.L. and Chang, H.Y. (2012) Genome regulation by long noncoding RNAs. *Annu. Rev. Biochem.* **81**, 145-166.

Rodriguez-Saona, C., Crafts-Brandner, S.J., Pare, P.W. and Henneberry, T.J. (2001) Exogenous methyl jasmonate induces volatile emissions in cotton plants. *J. Chem. Ecol.* **27**, 679-695.

Sanchez-Leon, N., Arteaga-Vazquez, M., Alvarez-Mejia, C. *et al.* (2012) Transcriptional analysis of the *Arabidopsis* ovule by massively parallel signature sequencing. *J. Exp. Bot.* **63**, 3829-3842.

Seo, J.S., Sun, H.X., Park, B.S., Huang, C.H., Yeh, S.D., Jung, C. and Chua, N.H. (2017) ELF18-INDUCED LONG-NONCODING RNA associates with mediator to enhance expression of innate immune response genes in *Arabidopsis*. *Plant Cell*, **29**, 1024-1038.

Swiezewski, S., Liu, F., Magusin, A. and Dean, C. (2009) Cold-induced silencing by long antisense transcripts of an *Arabidopsis* Polycomb target. *Nature*, **462**, 799-802.

Tan, J., Tu, L., Deng, F., Hu, H., Nie, Y., Zhang, X. (2013) A genetic and metabolic analysis revealed that cotton fiber cell development was retarded by flavonoid naringenin. *Plant Physiol*. **162**, 86-95.

Trapnell, C., Pachter, L., Salzberg, S.L. (2009) TopHat: discovering splice junctions with RNA-Seq. *Bioinformatics*, **25**, 1105-1111.

Trapnell, C., Williams, B.A., Pertea, G., Mortazavi, A., Kwan, G., van Baren, M.J., Salzberg, S.L., Wold, B.J., Pachter, L. (2010) Transcript assembly and quantification by RNA-Seq reveals unannotated transcripts and isoform switching during cell differentiation. *Nat. Biotechnol.* **28**, 511-515.

Thaler, J.S., Owen, B. and Higgins, V.J. (2004) The role of the jasmonate response in plant susceptibility to diverse pathogens with a range of lifestyles. *Plant Physiol.* **135**, 530-538.

Wang, K.C. and Chang, H.Y. (2011) Molecular mechanisms of long noncoding RNAs. *Mol. Cell*, **43**, 904-914.

Wang, H., Chung, P.J., Liu, J., Jang, I.C., Kean, M.J., Xu, J. and Chua, N.H. (2014a) Genome-wide identification of long noncoding natural antisense transcripts and their responses to light in *Arabidopsis*. *Genome Res.* **24**, 444-453.

Wang, H.M., Lin, Z.X., Zhang, X.L., Chen, W., Guo, X.P., Nie, Y.C. and Li, Y.H. (2008) Mapping and quantitative trait loci analysis of *Verticillium* wilt resistance genes in cotton. *J. Integr. Plant Biol.* **50**, 174-182.

Wang, H., Niu, Q.W., Wu, H.W., Liu, J., Ye, J., Yu, N. and Chua, N.H. (2015a) Analysis of non-coding transcriptome in rice and maize uncovers roles of conserved lncRNAs associated with agriculture traits. *Plant J.* **84**, 404-416.

Wang, M., Tu, L., Lin, M. *et al.* (2017) Asymmetric subgenome selection and cis-regulatory divergence during cotton domestication. *Nat. Genet*. **49**, 579-587.

Wang, M., Yuan, D., Tu, L., Gao, W., He, Y., Hu, H., Wang, P., Liu, N., Lindsey, K. and Zhang, X. (2015b) Long noncoding RNAs and their proposed functions in fibre development of cotton (*Gossypium* spp.). *New Phytol.* **207**, 1181-1197.

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

Wapinski, O. and Chang, H.Y. (2011) Long noncoding RNAs and human disease. *Trends Cell Biol.* **21**, 354-361.

Weick, E.M. and Miska, E.A. (2014) piRNAs: from biogenesis to function. *Development*, **141**, 3458-3471.

Wendel, J.F. and Cronn, R.C. (2003) Polyploidy and the evolutionary history of cotton. *Advances in Agronomy*, **78**, 139-186.

Wendel, J.F., Flagel, L.E. and Adams, K.L. (2012) Jeans, Genes, and Genomes: Cotton as a Model for Studying Polyploidy. In: *Polyploidy and Genome Evolution* (Pamela, S.S. and Douglas, E.S., eds), pp.181-207. Berlin Heidelberg: Springer Verlag.

Wierzbicki, A.T., Haag, J.R. and Pikaard, C.S. (2008) Noncoding transcription by RNA polymerase Pol IVb/Pol V mediates transcriptional silencing of overlapping and adjacent genes. *Cell*, **135**, 635-648.

Xiang, Y., Laurent, B., Hsu, C.H. *et al.* (2017) RNA m6A methylation regulates the ultraviolet-induced DNA damage response. *Nature*, **543**, 573-576.

Xu, L., Zhu, L., Tu, L., Liu, L., Yuan, D., Jin, L., Long, L. and Zhang, X. (2011) Lignin metabolism has a central role in the resistance of cotton to the wilt fungus *Verticillium dahliae* as revealed by RNA-Seq-dependent transcriptional analysis and histochemistry. *J. Exp. Bot.* **62**, 5607-5621.

Yamada, K., Lim, J., Dale, J.M. *et al.* (2003) Empirical analysis of transcriptional activity in the *Arabidopsis* genome. *Science*, **302**, 842-846.

Yang, C., Guo, W., Li, G., Gao, F., Lin, S. and Zhang, T. (2008) QTLs mapping for *Verticillium* wilt resistance at seedling and maturity stages in *Gossypium barbadense* L. *Plant Science*, **174**, 290-298.

Yang, L., Li, B., Zheng, X.Y., Li, J., Yang, M., Dong, X., He, G., An, C. and Deng, X.W. (2015) Salicylic acid biosynthesis is enhanced and contributes to increased biotrophic pathogen resistance in *Arabidopsis* hybrids. *Nat. Commun.* **6**, 7309.

Yang, S., Li, J., Zhang, X., Zhang, Q., Huang, J., Chen, J.Q., Hartl, D.L. and Tian, D. (2013) Rapidly evolving R genes in diverse grass species confer resistance to rice blast disease. *Proc. Natl. Acad. Sci. USA*, **110**, 18572-18577.

Yin, Z., Li, Y., Han, X. and Shen, F. (2012) Genome-wide profiling of miRNAs and other small non-coding RNAs in the *Verticillium dahliae*-inoculated cotton roots. *PloS One*, **7**, e35765.

Yoo, M.J., Szadkowski, E. and Wendel, J.F. (2013) Homoeolog expression bias and expression level dominance in allopolyploid cotton. *Heredity*, **110**, 171-180.

Yuan, D., Tang, Z., Wang, M. *et al.* (2015) The genome sequence of Sea-Island cotton (*Gossypium barbadense*) provides insights into the allopolyploidization and development of superior spinnable fibres. *Sci. Rep.* **5**, 17662.

Zhang, B., Yang, Y., Chen, T. *et al.* (2012) Island cotton *Gbve1* gene encoding a receptor-like protein confers resistance to both defoliating and non-defoliating isolates of *Verticillium dahliae*. *PloS One*, **7**, e51091.

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

Zhang, T., Hu, Y., Jiang, W. *et al.* (2015b) Sequencing of allotetraploid cotton (*Gossypium hirsutum* L. acc. TM-1) provides a resource for fiber improvement. *Nat. Biotechnol.* **33**, 531-537.

Zhang, T., Jin, Y., Zhao, J.H., Gao, F., Zhou, B.J., Fang, Y.Y. and Guo, H.S. (2016) Host-induced gene silencing of the target gene in fungal cells confers effective resistance to the cotton wilt disease pathogen *Verticillium dahliae*. *Mol. Plant*, **9**, 939-942.

Zhang, X., Yuan, Y., Wei, Z., Guo, X., Guo, Y., Zhang, S., Zhao, J., Zhang, G., Song, X. and Sun, X. (2014a) Molecular mapping and validation of a major QTL conferring resistance to a defoliating isolate of *Verticillium* wilt in cotton (*Gossypium hirsutum* L.). *PloS One*, **9**, e96226.

Zhang, Y., Wang, X., Yang, S., Chi, J., Zhang, G. and Ma, Z. (2011) Cloning and characterization of a *Verticillium* wilt resistance gene from *Gossypium barbadense* and functional analysis in *Arabidopsis thaliana*. *Plant Cell Rep.* **30**, 2085-2096.

Zhang, Y.C., Liao, J.Y., Li, Z.Y., Yu, Y., Zhang, J.P., Li, Q.F., Qu, L.H., Shu, W.S. and Chen, Y.Q. (2014b) Genome-wide screening and functional analysis identify a large number of long noncoding RNAs involved in the sexual reproduction of rice. *Genome Biol.* **15**, 512.

Zhao, Y., Wang, H., Chen, W. and Li, Y. (2014) Genetic structure, linkage disequilibrium and association mapping of *Verticillium* wilt resistance in elite cotton (*Gossypium hirsutum* L.) germplasm population. *PloS One*, **9**, e86308.

Zhiyuan, N., Zhao, R., Chen, H., Ai, N., Zhang, X., Zhao, J., Mei, H., Wang, P., Guo, W. and Zhang, T. (2013) Molecular tagging of a major quantitative trait locus for broad-spectrum resistance to *Verticillium* wilt in upland cotton cultivar prema. *Crop Science*, **53**, 2304.

Zhou, J., Wan, J., Gao, X., Zhang, X., Jaffrey, S.R. and Qian, S.B. (2015) Dynamic m(6)A mRNA methylation directs translational control of heat shock response. *Nature*, **526**, 591-594.

Zhou, Z.Y., Li, A.M., Adeola, A.C., Liu, Y.H., Irwin, D.M., Xie, H.B. and Zhang, Y.P. (2014) Genome-wide identification of long intergenic noncoding RNA genes and their potential association with domestication in pigs. *Genome Biol. Evol.* **6**, 1387-1392.

Zhu, L., Tu, L., Zeng, F., Liu, D., Zhang, X. (2005) An improved simple protocol for isolation of high quality RNA from *Gossypium* spp. suitable for cDNA library construction. *Acta Agronomica Sinica*, **31**, 1657–1659.

Zhu, Q.H., Stephen, S., Taylor, J., Helliwell, C.A. and Wang, M.B. (2014) Long noncoding RNAs responsive to *Fusarium oxysporum* infection in *Arabidopsis thaliana*. *New Phytol.* **201**, 574-584.

Table 1. Number of major types of lncRNAs.

Cotton Species	lincRNA	lncNAT	sense	intronic
Gossypium barbadense	13452	1297	260	200
Gossypium hirsutum	14547	1406	262	198

lincRNA, long intergenic noncoding RNAs.

lncNAT, long noncoding natural antisense transcripts.

Table 2. The Gene Ontology of At- and Dt-biased lncRNAs.

Type	GO-ID	Term	*P*-value
At-bias	GO:0019209	kinase activator activity	2.33E-03
	GO:0003690	double-stranded DNA binding	2.66E-03
	GO:0030983	mismatched DNA binding	2.66E-03
	GO:0004332	fructose-bisphosphate aldolase activity	1.41E-02
	GO:0043566	structure-specific DNA binding	1.75E-02
Dt-bias	GO:0004871	signal transducer activity	6.84E-03
	GO:0008173	RNA methyltransferase activity	7.45E-03
	GO:0004707	MAP kinase activity	1.90E-02
	GO:0005057	receptor signaling protein activity	1.90E-02
	GO:0016532	superoxide dismutase copper chaperone activity	2.25E-02

Table 3. The identification of core and lineage-specific (LS) lncRNAs.

Classification	H Core	H LS	Y Core	Y LS
Total number	9443	3943	9937	5183
Induced number	565	514	975	725
Induced Ratio	6%	12%	9%	12%

H, *Gossypium barbadense.*

Y, *Gossypium hirsutum.*

Core, Conserved sequence between two cotton species.

LS, Lineage specific sequence between two cotton species.

FIGURE LEGENDS

Figure 1. Identification and characterization of long noncoding RNAs (lncRNAs) in *Gossypium barbadense* and *G. hirsutum*. (a) Characterization of resistance to *Verticillium dahliae* in *G. barbadense* and *G. hirsutum*. (b) Distribution of long intergenic noncoding RNAs (lincRNAs) and long noncoding natural antisense transcripts (lncNATs) in the At subgenome, Dt subgenome and ungrouped scaffolds separately for *G. barbadense* (Gb) and *G. hirsutum* (Gh). (c) The GC content of different genes in cotton. (d) Density plot showing transcript length distribution of lincRNAs, lncNATs and protein-coding genes. (e) Exon number distribution of lincRNAs, lncNATs and protein-coding genes.

Figure 2. Comparison of pathogen response for lncRNAs in subgenomes. (a) Ratio of differentially induced lncRNAs in *G. barbadense*. X-axis represents the total number and Y-axis represents the differentially induced number of lncRNAs. (b) Ratio of differentially induced lncRNAs in *G. hirsutum.* (c) The category of At-bias induced lncRNAs from *G. barbadense*. Grey dashed lines mean the cutoff of bias induced expression ($|\log2(At/Dt)| \geqslant 1$). (d) The category of No-bias induced lncRNAs. (e) The category of Dt-bias induced lncRNAs. (f) The category of Chimeric induced lncRNAs.
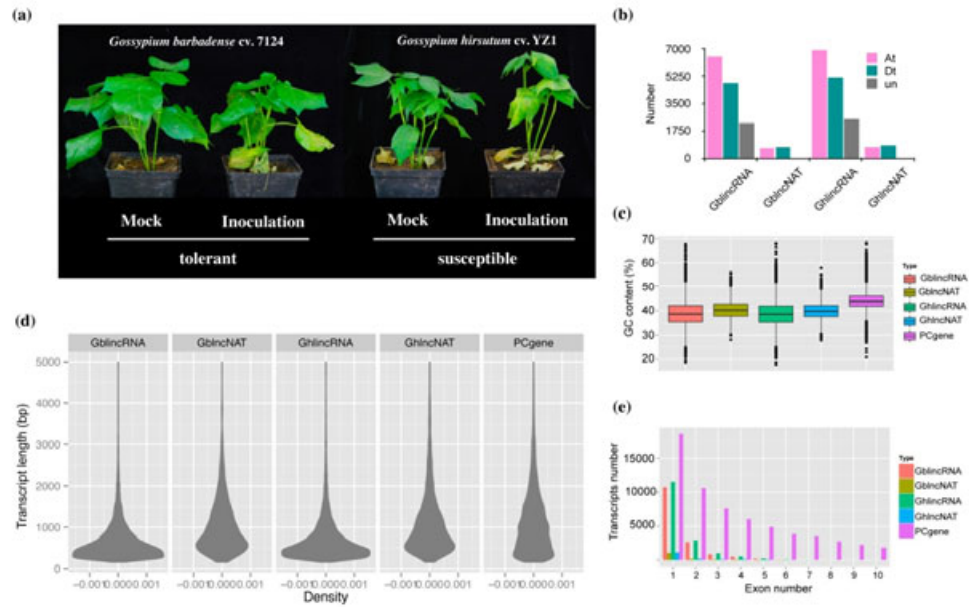
Figure 3. The global expression profiles of lncRNAs and distribution of differentially expressed lncRNAs. (a) Clusters of expressed lncRNAs in *G. barbadense* developed by K-means. "6", "12" and "24" mean hours post infection. "m" and "v" mean mock and seedling roots inoculated with *Verticillium dahliae* V991. (b) The distribution of differentially induced lincRNAs in two different cotton for each time point. (c) The distribution of differentially induced lncNATs.

Figure 4. The comparison of induced pattern for lncRNAs in two different cotton cultivars. All expressed homologous lncRNA pairs between *G. barbadense* and *G. hirsutum* were clustered into 16 groups (I to XVI). Gene ontology (GO) terms are indicated with significant *P* values ($P < 0.01$) for each cluster.

Figure 5. Characterization of core and specific lncRNAs. (a) The charts show changes in the induced expression levels (log2-transformed FPKM) of different classes of core/lineage specific (LS) lncRNAs in *G. barbadense* at three induced stages, 6 hours post infection (hpi), 12 hpi, 24 hpi. (b) Expression change of core/LS lncRNAs in *G. hirsutum*. (c) SNP distribution of lineage specific (LS) lncRNAs and core lncRNAs. Scatter plot showing the correlation between SNP frequency and length of lncRNAs in *G. barbadense*. Significant levels of distribution divergence are indicated as *P* values. (d) Gene ontology enrichment analysis of neighbor protein-coding genes of core lncRNAs and LS lncRNAs ($P < 0.01$).
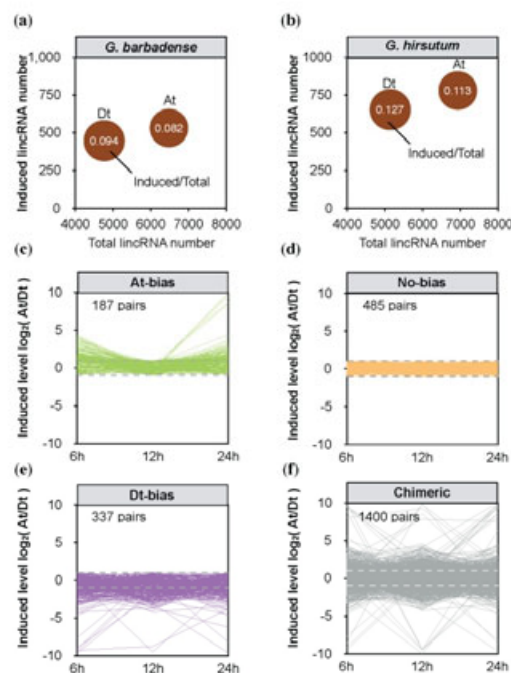
Figure 6. Functional identification of lncRNAs towards *Verticillium dahliae* in *G. barbadense* using a virus induced gene silencing (VIGS) method. (a) Phenotypes of seedlings with lncNAT silencing post inoculation, showing the wilting phenotype, etiolated leaves, fungal recovery assay and stem inspection. *Cloroplastos alterados 1* (*CLA*) used as the positive control. (b) Disease index of infected plants. (c,g) The qRT-PCR verification of *L2* (c) and *L3* (g) silenced by VIGS. (d,h) Expression change level of *P2* (d) after silencing *L2* and P3 (h) after silencing *L3*. (e,i) Transcriptional change of *lipoxygenase 1* (*LOX1*) after silencing *L2* (e) and *L3* (i). (f,j) Transcriptional change of *lipoxygenase 2* (*LOX2*) after silencing *L2* (f) and *L3* (j). Error bars show SDs (n=3).

Figure 7. Functional identification of two lncRNAs towards *Botrytis cinerea* in *G. barbadense* for virus induced gene silencing (VIGS) plants. (a) Disease symptoms of 3 days post inoculation leaves. (b) Trypan blue staining of hyphae cover area. (c) The statistics of disease symptom area. Error bars mean the standard deviation of three biological replicates. Asterisk represents statistically significant differences conducted by ANOVA test (**, *P* <0.01).
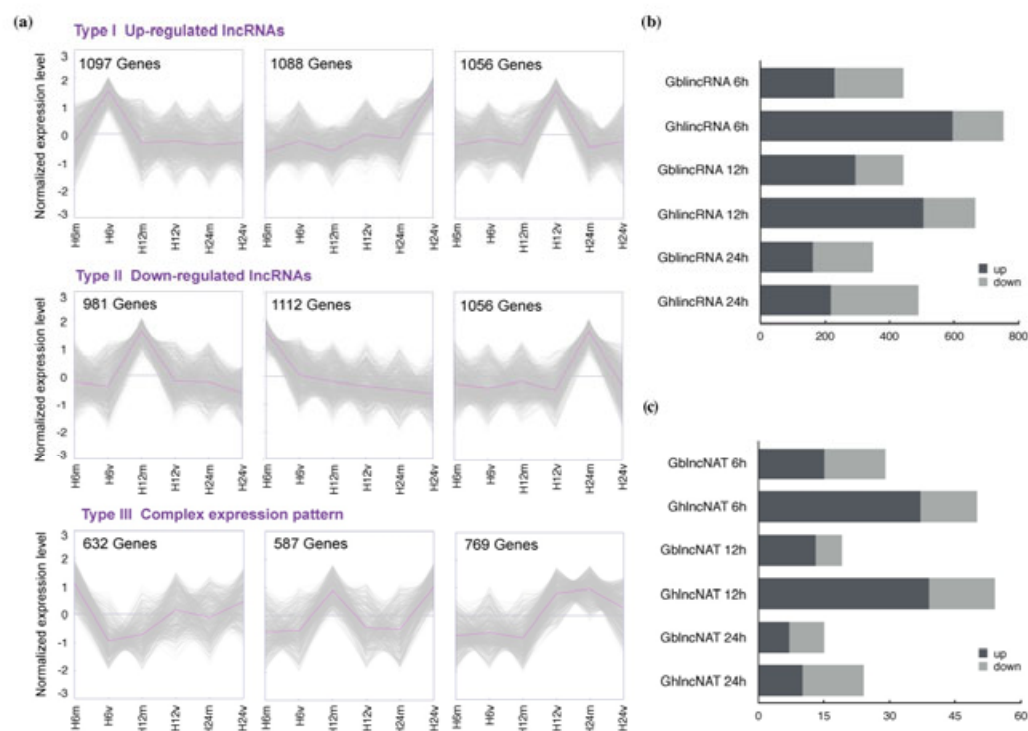
Figure 1 Identification and characterization of long noncoding RNAs (lncRNAs) in *Gossypium barbadense* and *G. hirsutum*. (a) Characterization of resistance to *Verticillium dahliae* in *G. barbadense* and *G. hirsutum*. (b) Distribution of long intergenic noncoding RNAs (lincRNAs) and long noncoding natural antisense transcripts (lncNATs) in the At subgenome, Dt subgenome and ungrouped scaffolds separately for *G. barbadense* (Gb) and *G. hirsutum* (Gh). (c) The GC content of different genes in cotton. (d) Density plot showing transcript length distribution of lincRNAs, lncNATs and protein-coding genes. (e) Exon number distribution of lincRNAs, lncNATs and protein-coding genes.
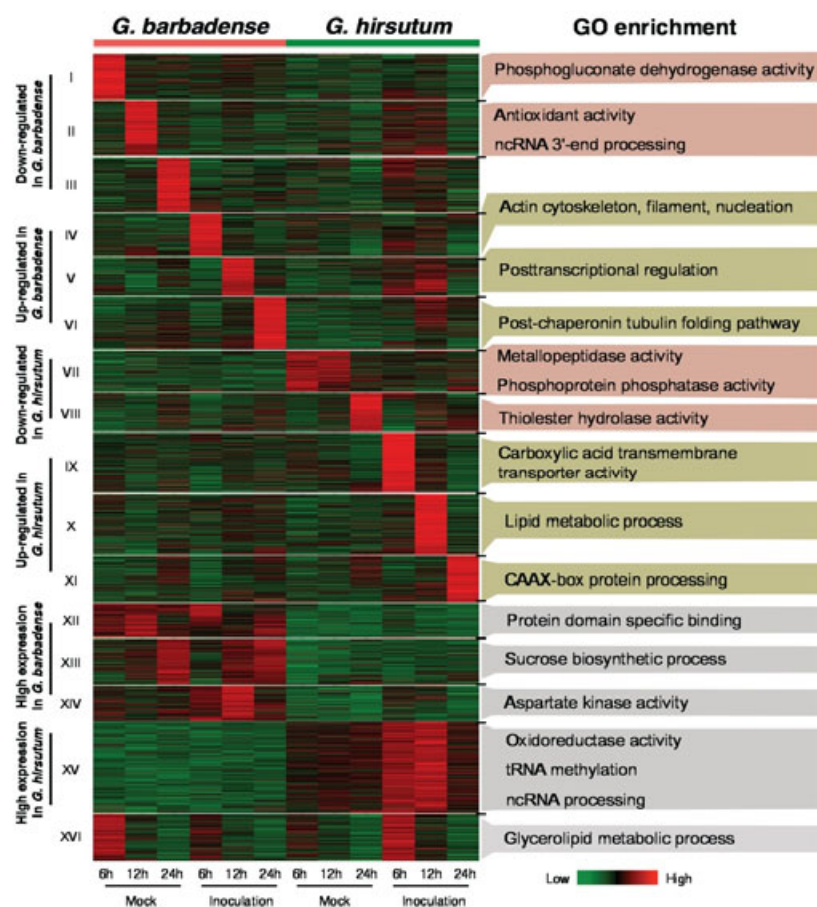
Figure 2 Comparison of pathogen response for lncRNAs in subgenomes. (a) Ratio of differentially induced lncRNAs in *G. barbadense*. X-axis represents the total number and Y-axis represents the differentially induced number of lncRNAs. (b) Ratio of differentially induced lncRNAs in *G. hirsutum*. (c) The category of At-bias induced lncRNAs from *G. barbadense*. Grey dashed lines mean the cutoff of bias induced expression ($|\log2(At/Dt)| \geq 1$). (d) The category of No-bias induced lncRNAs. (e) The category of Dt-bias induced lncRNAs. (f) The category of Chimeric induced lncRNAs.
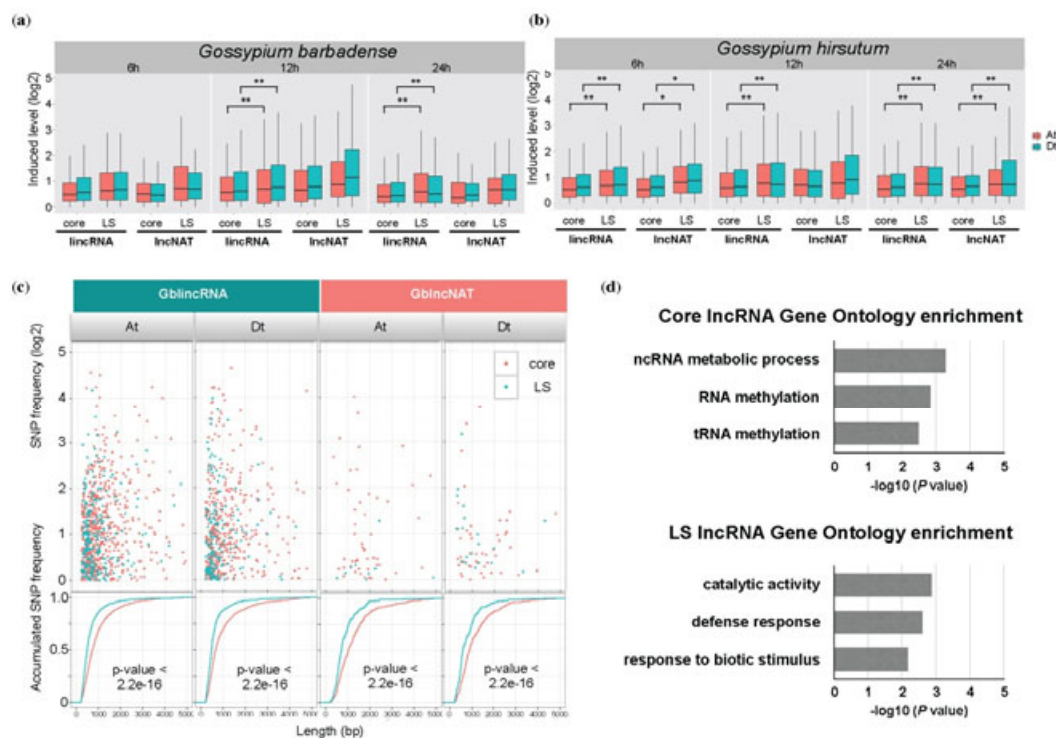
Figure 3 The global expression profiles of lncRNAs and distribution of differentially expressed lncRNAs.

(a) Clusters of expressed lncRNAs in *G. barbadense* developed by K-means. "6", "12" and "24" mean hours post infection. "m" and "v" mean mock and seedling roots inoculated with *Verticillium dahliae* V991. (b) The distribution of differentially induced lincRNAs in two different cotton for each timepoint. (c) The distribution of differentially induced lncNATs.
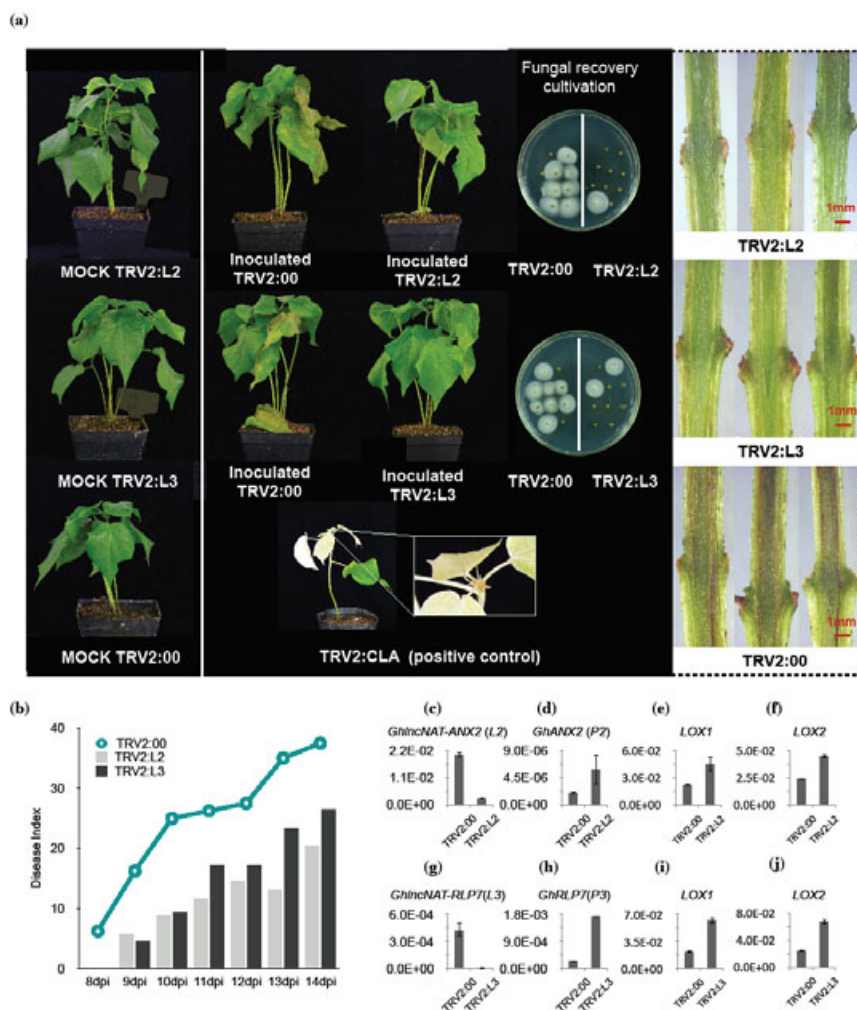
Figure 4 The comparison of induced pattern for lncRNAs in two different cotton cultivars. All expressed homologous lncRNA pairs between *G. barbadense* and *G. hirsutum* were clustered into 16 groups (I to XVI). Gene ontology (GO) terms are indicated with significant *P* values ($P < 0.01$) for each cluster.
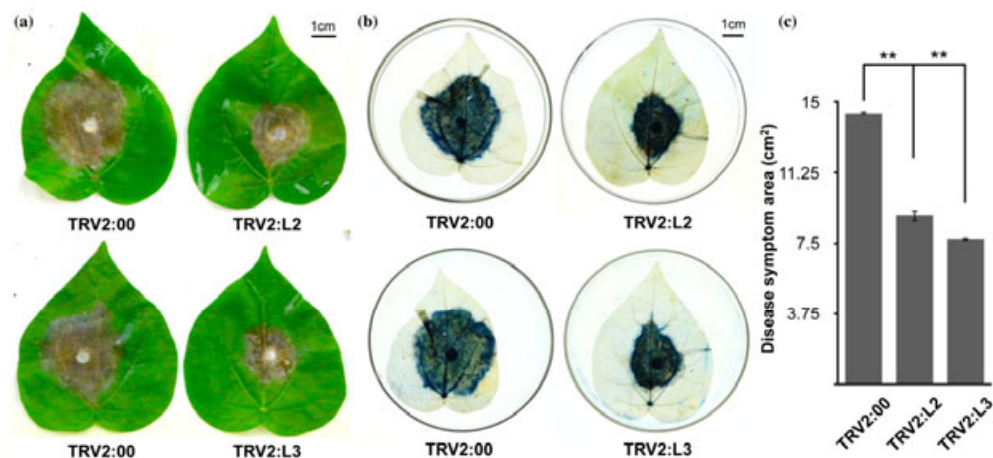
Figure 5 Characterization of core and lineage specific (LS) lncRNAs.

(a) The charts show changes in the induced expression levels (log$_2$-transformed FPKM) of different classes of core/lineage specific (LS) lncRNAs in *G. barbadense* at three induced stages, 6 hours post infection (hpi), 12 hpi, 24 hpi. (b) Expression change of core/LS lncRNAs in *G. hirsutum*. (c) SNP distribution of lineage specific (LS) lncRNAs and core lncRNAs. Scatter plot showing the correlation between SNP frequency and length of lncRNAs in *G. barbadense*. Significant levels of distribution divergence are indicated as *P* values. (d) Gene ontology (GO) enrichment analysis of neighbor protein-coding genes of core lncRNAs and LS lncRNAs (*P* < 0.01).

Figure 6 Functional identification of two lncRNAs towards *Verticillium dahliae* in *G. barbadense* using a virus induced gene silencing (VIGS) method.

(a) Phenotypes of seedlings with lncNAT silencing post inoculation, showing the wilting phenotype, etiolated leaves, fungal recovery assay and stem inspection. *Cloroplastos alterados 1* (*CLA*) used as the positive control. (b) Disease index of infected plants. (c,g) The qRT-PCR verification of *L2* (c) and *L3* (g) silenced by VIGS. (d,h) Expression change level of *P2* (d) after silencing *L2* and *P3* (h) after silencing *L3*. (e,i) Transcriptional change of *lipoxygenase 1* (*LOX1*) after silencing *L2* (e) and *L3* (i). (f,j) Transcriptional change of *lipoxygenase 2* (*LOX2*) after silencing *L2* (f) and *L3* (j). Error bars show SDs (n=3).

Figure 7 Functional identification of lncRNAs towards *Botrytis cinerea* in *G. barbadense* for virus induced gene silencing (VIGS) plants. (a) Disease symptoms of 3 days post inoculation leaves. (b) Trypan blue staining of hyphae cover area. (c) The statistics of disease symptom area. Error bars mean the standard deviation of three biological replicates. Asterisk represents statistically significant differences conducted by ANOVA test (**, *P* <0.01).