Short title: Neofunctionalization of LncRNA GhDAN1

Author for contact:
Xueying Guan

College of Agriculture and Biotechnology
Zhejiang University
Hangzhou, 310058, China

E-mail: xueyingguan@zju.edu.cn

Neofunctionalization of a Polyploidization-Activated Cotton Long intergenic non-coding RNA DANI During Drought Stress Regulation

Xiaoyuan Tao,1 Menglin Li,2 Ting Zhao,1 Shouli Feng,1 Hailin Zhang,1 Luyao Wang,3 Jin Han,1 Mengtao Gao,2 Kening Lu,2 Quanjia Chen,3 Baoliang Zhou,2 and Xueying Guan1*

1 College of Agriculture and Biotechnology, Zhejiang University, Hangzhou, 310058, China

2 State Key Laboratory of Crop Genetics and Germplasm Enhancement, Cotton Hybrid R & D Engineering Center (the Ministry of Education), College of Agriculture, Nanjing Agricultural University, Nanjing, 210095, China

3 College of Agriculture, Xinjiang Agricultural University, Engineering Research Center of Ministry of Cotton Education, 311 Nongda East Road, Urumqi, 830052, P. R. China.

© The Author(s) (2021). Published by Oxford University Press on behalf of American Society of Plant Biologists. This is an Open Access article distributed under the terms of the Creative Commons Attribution-Non-Commercial-NoDerivs licence (http://creativecommons.org/licenses/by-nc-nd/4.0/), which permits non-commercial reproduction and distribution of the work, in any medium, provided the original work is not altered or transformed in any way, and that the work is properly cited. For commercial re-use, please contact journals.permissions@oup.com.
**One sentence summary:** A hybridization- and polyploidization-activated IncRNA neofunctionalized into a regulator of drought responses by forming RNA-DNA triplexes targeting genes mainly in the auxin response pathway.

**Footnotes:**

X. G. and X. T. conceptualized this project and designed the experiments. X. T., M. L., and H. Z. performed the experiments and generated the data. T. Z., X. T. and S. F. carried out the bioinformatic analysis. X. G., X. T., M. L., H. Z., Q. C., B. Z., and L. W. analyzed the data. M. L., L. W., J. H., M. G and K. L. conducted the functional IncRNA screening. B. Z. provided the (G. arboreum × G. raimondii) F1 leaves for the study. X. G., X. T., and M. L. wrote the manuscript. All authors discussed results and commented on the manuscript. All authors read and approved the final manuscript.

This work was financially supported in part by grants from the National Key Research and Development Program (2016YFD0100605), National Natural Science Foundation of China (NSFC, 31971985, 31900395, 31901430), Dabeinong Funds for Discipline Development and Talent Training in Zhejiang University, Fundamental Research Funds for the Central Universities, and the funding of "Leading Innovative and Entrepreneur Team Introduction Program of Zhejiang" (2019R01002).

The author responsible for the distribution of materials integral to the findings presented in this article in accordance with the policy described in the Instructions for Authors (www.plantphysiol.org) is Xueying Guan (xueyingguan@zju.edu.cn).
Abstract

The genomic shock of whole-genome duplication and hybridization introduces great variation into transcriptomes, for both coding and non-coding genes. An altered transcriptome provides a molecular basis for improving adaptation during the evolution of new species. The allotetraploid cotton, together with the putative diploid ancestor species compose a fine model for study the rapid gene neofunctionalization over the genome shock. Here we report on Drought-Associated Non-coding gene 1 (DAN1), a long intergenic non-coding RNA (lincRNA) that arose from the cotton progenitor A-diploid genome after hybridization and whole-genome duplication events during cotton evolution. DAN1 in allotetraploid upland cotton (Gossypium hirsutum, Gh) is a drought-responsive lincRNA predominantly expressed in nucleoplasm. A chromatin isolation by RNA purification (ChIRP) profiling and electrophoretic mobility shift assay (EMSA) analysis demonstrated that GhDAN1 RNA can bind with DNA fragments containing AAAG motifs, similar to DNA-binding one zinc finger (Dof) transcription factor binding sequences. The suppression of GhDAN1 mainly regulates genes with AAAG motifs in auxin-response pathways, which are associated with drought stress regulation. As a result, GhDAN1-silenced plants exhibit improved tolerance to drought stress. This phenotype resembles the drought-tolerant phenotype of the A-diploid cotton ancestor species, which has an undetectable expression of DAN1. The role of DAN1 in cotton evolution and drought tolerance regulation suggests that the genomic shock of interspecific hybridization and whole-genome duplication
(WGD) stimulated neofunctionalization of non-coding genes during the natural evolutionary process.

Keywords: Cotton, IncRNA, neofunctionalization, drought, RNA-DNA triplexes

Introduction

Interspecific hybridization and whole-genome duplication (WGD), also known as polyploidization, can stimulate the formation of new species via dramatic alterations in expression of coding and non-coding genes, neofunctionalization of redundant genes, and other variations, which allow for fast adaptation to a new habitat (Adams and Wendel, 2005). Over a long span of evolution, redundant genes generated by WGD might remain functional due to their biological importance. Alternatively, some duplicated genes lose their function because of expressional suppression (defunctionalization), sequence mutation (pseudogene), or gene loss because of redundancy (Adams et al., 2004; Conrad and Antonarakis, 2007). Some genes, especially coding genes, might gain new functions (i.e., undergo neofunctionalization) by means of new expression patterns or gene structures (Conrad and Antonarakis, 2007; Semon and Wolfe, 2008). When comparing gene expression activity in the hybrid or allotetraploid genomes with that in the parental genomes, the differentially expressed genes can be classified as having additive expression or non-additive expression (Chen, 2010). Genes with non-additive expression are understood to be the major contributors to the heterosis of hybridization and polyploidization (Chen, 2010).
Some epigenomic marks, such as small RNAs (Ha et al., 2009) and chromatin modifications (Ha et al., 2009; Zhang et al., 2016), are also believed to play critical roles as buffers in response to the “genomic shock” caused by hybridization and WGD (Mcclintock, 1984; Chen, 2007). Previously, we reported that lncRNA in the cotton lineage also proactively follows a non-additive expression pattern in interspecies hybridization and polyploidization (Zhao et al., 2018). Unique expression patterns and functions of lncRNA are commonly reported in animal and plant genomes (Li et al., 2014; Necsulea et al., 2014; Deng et al., 2018; Zhao et al., 2018). Whether these non-additive or newly stimulated and activated lncRNAs arising during genomic shock are biologically important for the hybrid and polyploid species remains largely unknown.

The putative diploid ancestors of the current allotetraploid Gossypium (cotton) lineage species are Gossypium herbaceum (A₁) or Gossypium arboreum (A₂) and Gossypium raimondii (D₃) (Cronn et al., 1996; Zhang et al., 2015). The cotton diploid ancestors formed an allotetraploid by hybridization and polyploidization approximately one to two million years ago, and from these ancestors, seven current allotetraploid species were derived (Wendel and Grover, 2015; Wang et al., 2018). Two diploid species (G. herbaceum, A₁ and G. arboreum, A₂) and two allotetraploid species (Gossypium hirsutum, AD₁ and Gossypium barbadense, AD₂) have been independently domesticated for farming. Diploid cultivars from G. arboreum (A₂) are low in product yield and have a high tolerance to drought stress (Zhang, 2010; Chen et al., 2015; Jia et al., 2018). An allotetraploid species (upland cotton, G. hirsutum) is the dominant
cultivated species of cotton worldwide; it has high productivity and accounts for 90% of the cotton grown. LncRNAs have been shown to play crucial regulatory roles in plant–environmental interactions, including seed dormancy (Qin et al., 2017), cold sensing and flowering (Swiezewski et al., 2009), phosphate nutrition homeostasis (Franco-Zorrilla et al., 2007), pathogen interaction (Yang et al., 2019), pathogen resistance (Zhang et al., 2018), and so on. We employed cotton lineage as a model system, using functional tests of abiotic stress regulation to examine the evolutionary path from diploid to allotetraploid species in order to identify a lncRNA activated in hybridization and polyploidization. Here, we report on a long intergenic non-coding RNA (lincRNA) responsible for drought stress regulation, the Drought-Associated Non-coding gene 1 (DAN1).

Results

GhDANI is found to be associated with the drought tolerance trait

We employed a simplified model system with G. arboreum (Ga, A2), G. raimondii (Gr, D5), (Ga x Gr)F1, and G. hirsutum (Gh, AD1) to mimic cotton’s evolutionary lineage (Fig. 1A) and demonstrate that lncRNA follows a non-additive expression pattern from diploid to allotetraploid by interspecies hybridization (Zhao et al., 2018). Diploid cultivars from G. arboreum (A2) show high tolerance to drought stress in comparison to allotetraploid cotton (Chen et al., 2015) (Fig. 1B). To address the potential link between the lncRNA reprogramming and the alteration of the drought tolerance trait, we conducted a low through-put functional screening of the lncRNAs in
Gh with virus-induced gene silencing (VIGS) technology. A lincRNA responsive to drought stress, the *DANI* gene, was selected for further study because of its stable phenotyping result from at least five independent biological replicates (Supplemental Fig. S1). The *GhDANI*-silenced plants showed improved tolerance to drought stress (Fig. 1C and D), with increased glutathione (GSH) (Fig. 1E) and proline contents in drought-treated roots (Fig. 1F), but without obvious changes in its photosynthesis traits (Supplemental Fig. S2).

Allotetraploid *DANI* (*GhDANI*) is a non-coding gene with a predicted length of 8,581 base pairs (bp). At least three different isoforms of transcripts were predicted from the *GhDANI* locus. The longest of these isoforms (6,779 nt) that had abundant transcripts was confirmed in the results of both the lncRNA-seq and RT-qPCR assays, and was further analyzed in our study (Supplemental Fig. S3). Like mRNA, *GhDANI* has a poly(A) tail (Supplemental Fig. S4A). However, according to the coding potential calculation, *GhDANI* does not seem to contain an open reading frame encoding any known peptide or protein. This is also in agreement with the low frequency of fragments found in the ribosome profiling (Supplemental Fig. S4B). Synteny analysis indicated that *GhDANI* is inherited from the A diploid genome progenitor (Fig. 1G).

*GhDANI* was predominantly expressed in leaf and stem tissues (Fig. 1H) and was spontaneously suppressed by drought and salt treatments (Fig. 1I). RNA-seq and RNA polymerase II (Pol II) ChIP-seq analysis showed that *DANI* was very probably transcribed by Pol II. Its transcription was only slightly activated in the F1 hybrid but was highly expressed in the allotetraploid Gh genome (Fig. 1J and K). Interestingly, the
CG and CHG methylation levels on the promoter region and 5′ end region of the \textit{DAN1} gene were lower in Gh compared with those in Ga and (Ga x Gr)\textsubscript{F1} (Fig. 1L and Supplemental Fig. S5), which was in agreement with our previous finding that the activation of IncRNA transcription is predominantly associated with DNA demethylation after hybridization and polyploidization (Zhao et al., 2018).

\textbf{\textit{DAN1} is active in the allotetraploid cotton species}

To further determine the behavior of \textit{DAN1} in the cotton lineage, we examined five accessions of \textit{G. herbaceum} (A\textsubscript{1}), 17 accessions of \textit{G. arboreum} (A\textsubscript{2}), 28 accessions of \textit{G. hirsutum} (AD\textsubscript{1}), and four accessions of \textit{G. barbadense} (AD\textsubscript{2}). We also examined one accession each of the other allotetraploid species, including \textit{Gossypium tomentosum} (AD\textsubscript{3}), \textit{Gossypium mustelinum} (AD\textsubscript{4}), \textit{Gossypium darwinii} (AD\textsubscript{5}), \textit{Gossypium ekmanianum} (AD\textsubscript{6}), and \textit{Gossypium stephensii} (AD\textsubscript{7}) (Fig. 2A and Supplemental Table S1). With this selection, we were able to examine a diverse cotton population that included wild species, primitive landraces, and cultivars of the A diploid species and allotetraploid species. The primary sequences of \textit{DAN1} were conserved in the A diploid species and allotetraploid species (Supplemental Fig. S6). In all of the A diploid species examined, the transcription of \textit{DAN1} in leaf tissue was either below detection or had very low expression (Fig. 2B). However, the expression of \textit{DAN1} in leaf tissue was commonly 10- to 80-fold higher in allotetraploid species than in the A diploid species (Fig. 2B). To determine the variations in epigenetic modification underlying \textit{DAN1} activation, we further examined the histone
modification profiles of H3K4me3 (active marker) and H3K27me3 (repressive marker) using ChIP-seq with the leaves harvested from the cotton lineage collection. Results indicated that the active expression of \textit{DAN1} in allotetraploid species was consistent with the presence of the active histone modification marker H3K4me3 at the 5' end of the gene (Fig. 2C); no obvious differences in the H3K27me3 signal were found (Supplemental Fig. S7).

We also compared the drought tolerance characteristics of diploid and allotetraploid cotton species with varying levels of \textit{DAN1} expression. Results showed a negative correlation ($R^2 = 0.6473$) between the expression of \textit{DAN1} and the drought tolerance in the species we tested (Fig. 2D, E and Supplemental Fig. S8). These data confirmed that \textit{DAN1} was activated after hybridization and polyploidization in the \textit{Gossypium} lineage and is one of the factors that contributes to the valuable characteristic of drought tolerance in A diploid species, which most allotetraploid cultivated cotton species have lost.

\textit{GhDAN1} can form RNA-DNA triplexes to anchor the AAAG DNA target site and regulate gene expression in \textit{trans}

To determine the function of \textit{GhDAN1} in drought tolerance, the gene expression profiles of \textit{GhDAN1}-silenced plants were examined using mRNA-seq and were compared with the VIGS control group. The silenced expression of \textit{GhDAN1} by VIGS did not significantly affect the activity of the 25 flanking genes of \textit{GhDAN1} (which covered approximately 8 Mb of the flanking region) in the chromosome (Supplemental
Fig. S9), which suggests that *GhDAN1* might not mediate the control of drought
tolerance in *cis*. *GhDAN1* transcripts were preferentially present in the nucleus (Fig.
3A) and more specifically, in the nucleoplasm (Fig. 3B). An immunofluorescence in
situ hybridization assay with RNA probes (RNA-FISH) using isolated nuclei showed
that the *GhDAN1* signal was dispersed throughout the nucleus (Fig. 3C). Our data are
consistent with a model in which the *GhDAN1* binding sites were concentrated in
multiple specific spots on different chromosomes. The above analysis demonstrated
that *GhDAN1* is a nucleus-localized lincRNA, which might be required for the ability
of a lincRNA to act in *trans*.

Generally, the lncRNA affects the transcription in *trans* by recruiting a regulatory
ribonucleoprotein (RNP) complex to the target gene. In such cases, lncRNA acts as a
partner of the RNPs (Rinn and Chang, 2012). LncRNA can also act in *trans* by
inhibiting the binding of a transcriptional regulatory factor by acting as a decoy or can
inhibit its activity by direct active-site occlusion to prevent its activity (Rinn and
Chang, 2012; Long et al., 2017). To investigate the possible molecular mechanism of
*GhDAN1* in *trans*, chromatin isolation by RNA purification (ChIRP) assays were
performed to map the regions of chromatin interacting with the *GhDAN1* transcripts.
We set up the ChIRP reactions using anti-sense oligonucleotide probes that separated
into two groups (even and odd), which was followed by DNA-seq analysis. Only those
peaks shared by both the odd and even probe sets were considered to be valid peaks
(see the section on Materials and Methods for a detailed ChIRP-seq workflow). Based
on the filter parameters, 12,142 valid peaks were identified, with 10,941 (90.1%) peaks
located intergenically, and 1,201 (9.9%) peaks located near 945 coding genes (including the 3-kb 5’ flanking region, gene body, and 3-kb 3’ flanking region, designated as the gene-associated \textit{DAN1} ChIRP peaks) (Supplemental Fig. S10A). An AG-rich motif similar to the binding motif of DNA binding with one finger (Dof)-type plant-specific transcription factors was established by MEME software using the sequences of the top 1,214 valid peaks (i.e., the top 10% of the total valid peaks as ranked by fold enrichment) (Fig. 3D). To determine that this established motif resulted from the ChIRP binding but not from random sequences, the occurrence of the motif in the 1,201 gene-associated \textit{DAN1} ChIRP peaks were analyzed by FIMO scanning. Results indicated that this AG-rich motif was enriched in the peaks (±250 bp from the peak summits) compared with the control flanking regions (500 bp upstream and downstream from the peaks) (Fig. 3E). The latest studies have revealed that multiple mammalian lncRNAs (e.g., \textit{MEG3, HOTAIR,} and \textit{PARTICLE}) (Mondal et al., 2015; Oleary et al., 2015; Kalwa et al., 2016) are capable of regulating gene expression by forming RNA-DNA triplexes. We therefore scanned the \textit{GhDAN1} RNA sequence for possible triplex-forming oligonucleotides (TFOs) using Triplexator software (Buske et al., 2012). High-score TFOs were obtained (Supplemental Table S2) with TC-rich sequences. Among these TFOs, \textit{GhDAN1} TFO motif 1 (TFO1), which was a 19-bp TC-rich motif at the 5’ end of \textit{GhDAN1} (from positions +118 to +137), exhibited the triplex target site (TTS) sequence that was most similar to the ChIRP-established AG-rich motif (Fig. 3D). Accordingly, we speculated that the TC-rich \textit{GhDAN1} RNA
might be recruited to the AG-rich genomic DNA targeting sites through RNA-DNA triplex formation following the Hoogsteen configuration (Li et al., 2016).

To validate the ability of \( \text{DAN1} \) lincRNA to form triplex structures, an electrophoretic mobility shift assay (EMSA) was performed. It examined the triplex formation of a 178-nt \( \text{GhDAN1} \) single-stranded RNA fragment (+9 to +186) containing \( \text{GhDAN1} \) TFO (+118 to +137) with the digoxigenin (DIG)-labeled AG-rich double-stranded DNA TTSs. The results indicated that \( \text{GhDAN1} \) RNA can bind to the double-stranded DNA probe and has a sequence identical to that of the predicted \( \text{DAN1} \) TTS motif (probe 1). To evaluate the sequence discrepancy of the RNA-DNA triplex formation, the probe with mismatches of the predicted TTS motif were also examined with EMSA. We observed that an alternative probe with two mismatches (probe 2) preserved the triplex formation (Fig. 3F). Therefore, the formation of the RNA-DNA triplex exhibited a mild tolerance to sequence variation. In the control, RNA-DNA triplex binding was retarded upon the addition of a corresponding unlabeled cold probe serving as a competitor (Fig. 3F). This RNA-DNA triplex structure was sensitive to RNase A but resistant to RNase H (Fig. 3G). Because the RNA strand of an RNA-DNA hybrid can be specifically degraded by RNase H, these results suggested that the EMSA shift band detected \textit{in vitro} was not due to the RNA-DNA hybrid formed by Watson–Crick pairing. Because \( \text{GhDAN1} \) recognizes DNA sites similar to the Dof TFs binding motif, we speculated that the pathway corresponding to Dof TFs in plants may be coordinated by \( \text{GhDAN1} \) regulation.
Genes with the *GhDAN1* TTS motif were sensitive to the silencing of *GhDAN1* and significantly enriched in the auxin response pathway

The mRNA-seq analysis with leaves of VIGS-silenced plants and TRV control plants before drought treatment was performed in order to determine the early response genes under *GhDAN1* regulation. As a result, 493 differentially expressed genes (DEGs; $P < 0.001$, false discovery rate [FDR] < 0.05) were identified in *GhDAN1*-silenced plants in comparison to the control group, including 343 upregulated and 150 downregulated genes (Fig. 4A, Supplemental Fig. S11 and Supplemental Table S3). Gene ontology (GO) enrichment analysis showed that the DEGs were significantly enriched in the category of “response to hormone stimulus” (FDR < 1.0E-07), especially in the “response to auxin stimulus” (FDR < 1.0E-13) for the upregulated DEGs (Fig. 4C). The downregulated DEGs were enriched in the developmental and reproductive processes (Fig. 4C). We further validated the main auxin biosynthesis pathway in plants, in which tryptophan is first converted to indole-3-pyruvate (IPA) by the TAA/TAR family of amino transferases, and IAA is subsequently produced from IPA by the YUC family of flavin monooxygenases (Mano and Nemoto, 2012), and the results indicated that none of the TAA/TAR or YUC genes were differentially expressed in *GhDAN1*-silenced plants in comparison to the control group (Supplemental Fig. S12A). Considering the importance of abscisic acid (ABA) in drought stress responses, the expression profile of genes in ABA biosynthesis and catabolism pathways were summarized, and the results showed that none of these genes
were differentially expressed (Supplemental Fig. S12B). Thus, GhDAN1 was unlikely to affect phytohormone biosynthesis, including for auxin and ABA.

To establish the link between the ChIRP peak-associated genes and the DEGs in GhDAN1-silenced plants, we compared the ChIRP peaks and mRNA-seq data. Of the 945 genes with DAN1 ChIRP peaks, 3 overlapped with the DEGs. They were (1) a 50S ribosomal protein L9 gene encoding one of the primary rRNA-binding proteins; (2) a copper transporter 5 gene that responded to abscisic acid and was involved in the transport of copper for protein maturation and copper homeostasis in Arabidopsis thaliana (Sancenon et al., 2003); and (3) a germin-like protein subfamily 1 member 13, the GLP6 gene that is involved in manganese and iron binding and nutrient reservoir activity, and which may play a role in plant defense (Davidson et al., 2010) (Supplemental Fig. S10B). These three genes were all significantly upregulated in GhDAN1-silenced plants (Supplemental Table S3).

Because of the low resolution of the genes that were associated with ChIRP peaks and were also differentially expressed in GhDAN1-silenced plants, we evaluated, as an alternative, the effects of the GhDAN1 RNA-DNA triplex on the transcriptional activity of the potential target-coding genes. This was done by examining the presence of the GhDAN1 TTS motif in the promoters of the DEGs. The promoters of the total 42,269 expressed coding genes (FPKM>1) in the Gh genome were used as background controls. Results indicated that the TTS motif was significantly enriched in the promoter regions of the DEGs (Fig. 4B; \( P < 1.0E-05 \), Fisher’s exact test), indicating
that the genes containing the *GhDAN1* TTS motif were dominant among the DEGs that were responsive in the early stages following *GhDAN1* regulation. Multiple DEGs related to the auxin response pathway, ethylene response pathway, brassinosteroid response and biosynthesis pathway, dehydration response/oxidation resistance, and cell wall biosynthesis contained the *GhDAN1* TTS motif in their promoters, suggesting that they are the primary targets of *GhDAN1* (Fig. 4D). Orthologs of 13 TTS motif containing-DEGs in other plants were reported to be involved in drought stress tolerance (Supplemental Table S4). Based on all of this evidence, it can be assumed that *GhDAN1* preferentially regulates genes containing the *GhDAN1* TTS motif, and that these genes are significantly enriched in the auxin response pathway.

To determine whether the DEGs with the *GhDAN1* TTS motif in *GhDAN1*-silenced plants are subject to drought stress regulation, we examined the transcriptome analysis of drought treatment in upland cotton TM-1. Among the 145 TTS-containing genes up-regulated in *GhDAN1*-silenced plants, 41 (28.3%) and 25 (17.2%) showed significantly increased and suppressed expression in response to dehydration treatment, respectively (Supplemental Fig. S13A, Supplemental Table S5). Among the 72 TTS-containing genes down-regulated in *GhDAN1*-silenced plants, 15 (20.8%) and 8 (11.1%) showed significantly increased and suppressed expression in response to dehydration treatment, respectively (Supplemental Fig. S13A, Supplemental Table S5). For comparison, in the total 42,269 expressed coding genes (FPKM>1) in the Gh genome, 4441(10.5%) and 4443(10.5%) were significantly up-regulated and down-regulated, respectively, in response to dehydration treatment.
Statistical analysis indicated that the TTS motif-containing DEGs that were up-regulated in both *GhDAN1*-silenced plants and dehydration-treated Gh plants were significantly enriched compared with dehydration up-regulated genes at the whole genome background (Supplemental Fig. S13A, *P* < 1.0E-05, Fisher’s Exact Test), indicating that the major effect of *GhDAN1* is targeting these up-regulated genes. Totally 89 of the TTS motif-containing DEGs in *GhDAN1*-silenced plants were dehydration-responsive, which are enriched in the GO term “response to auxin stimulus” (FDR < 0.05; Supplemental Fig. S13B). We found that six out of 13 documented drought tolerance–related genes are both up-regulated or down-regulated in *GhDAN1*-silenced plants and drought-treated, Gh wild-type plants (Supplemental Table S4). We further selected 11 candidate genes from the DEGs with the *GhDAN1* TTS motif to examine their functions under drought stress using the VIGS assay (Supplemental Fig. S14). A representative gene encoding thiamine thiazole synthase 2 (*THI1-2, GH_A05G1689*), which was responsive to both *GhDAN1* silencing and drought treatment, was demonstrated to be involved in drought tolerance regulation in the VIGS and dehydration assays (Fig. 4E and F). These findings confirmed that *GhDAN1* can regulate drought stress regulation by targeting genes with the TTS motif.

**Working model of *DAN1*-mediated drought-response dynamic**

Given all the data we obtained for *DAN1*, a most likely working model for *DAN1* evolution and its role in drought tolerance regulation is proposed. *DAN1* originated from an A diploid ancestor genome after polyploidization due to DNA CG
demethylation and H3k4me3 modification at the 5′ end of the gene” (Figs. 1l and 2c). It obtained new functions associated with stress adaptation over the evolution of the cotton allotetraploid. In favorable conditions, *GhDAN1* is transcribed at a sufficient level and engages with genomic regions with AAAG motifs by forming RNA–DNA triplexes. When plants encounter conditions of drought stress, the *GhDAN1* transcript levels are reduced (Fig. 1i), thus releasing the AAAG motifs, possibly for the Dof TFs. Because members of the Dof TFs are characterized as being activators and suppressors in plants (Noguero et al., 2013), the expression of growth/development-related genes (downregulated DEGs in Fig. 4c and d) could be decreased by possible Dof suppressors, and the expression of drought defense–related genes (upregulated DEGs in Fig. 4c and d) could be promoted by possible Dof activators. In an orchestrated result, plants could successfully adapt to environmental dynamics (Fig. 5). This allows plants to achieve a balance between the energy they expend in growth and the stress tolerance they gain. *GhDAN1* activation is a remarkable case of a gene that, during hybridization and WGD, was selected for and fixed in such a way that it could maintain its growth and fitness when encountering environmental stress.

**Discussion**

**WGD-induced *GhDAN1* plays a regulatory role in drought stress tolerance**

Whole-genome duplication (WGD) in plants that is led by polyploidization and interspecies hybridization usually leads to dramatic changes, both genetically and physically (Chen, 2013). The encounter of two genomes in one nucleus proactively
stimulates the non-coding region of the genome to transcribe RNAs (Zhao et al., 2018). As has been previously reported, these activated RNAs are mainly transcribed by Pol II (Zhao et al., 2018). These newly transcribed regions of the genome are also highly associated with markers of epimodification, such as DNA methylation (Zhao et al., 2018). We proposed that the total effects of the epimodifications to chromatin and the transcriptome provided a reservoir for variations in heredity. The transcriptional activity and function of \textit{DAN1} in the cotton lineage supports this hypothesis. The expression of \textit{DAN1} in diploid species was almost undetectable. In all allotetraploid species, however, \textit{DAN1} expression is active. \textit{DAN1} activation in the synthetic hybrid of the A$_2$ and D$_5$ diploids also provided strong evidence to suggest that the \textit{DAN1} locus is sensitive to genomic shock. Meanwhile, the primary sequence of \textit{DAN1} was conserved in the cotton lineage from the A diploid to the A subgenome in the allotetraploid over one to two million years. Therefore, \textit{DAN1} displays a unique evolutionary pattern wherein RNA expression is activated with no fundamental changes to the sequence. The activation and function of \textit{GhDAN1} in allotetraploid cotton species demonstrates that the non-coding transcripts activated by genomic shock can cause the emergence of new functional and beneficial genes, which can be fixed in the natural population during evolution.

In recent reports, there is controversy about the sequence and functional conservation of lncRNAs. The primary sequences of lncRNAs evolved faster relative to mRNAs, both in tetrapod animals (Necsulea et al., 2014), and the relatively closely related plant species (Zhao et al., 2018; Barrera-Redondo et al., 2019). The functional
characterization of the known lncRNAs in one species is rarely identified in other relatively close species, suggesting that lncRNA has very strong species specificity. A recent report in *Drosophila* found that functional tests (using CRISPR-CAS9 technologies) of 30 relatively conserved lncRNAs resulted in no observed developmental functions for any of them (Goudarzi et al., 2019). Here we argue that according to its primary sequence, *DANI* is a conserved gene in the cotton lineage. In terms of transcriptional activity, however, *DANI* is a novel gene in the allotetraploid, harbored in a non-coding region. The activity of *DANI* we are reporting here demonstrates the flexibility of the genome in employing the non-coding region in response to environmental stimuli.

**The molecular basis of *DANI*-mediated gene expression regulation in trans**

Previous studies have shown that multiple mammalian lncRNAs (e.g., *MEG3*, *HOTAIR*, and *PARTICLE*) (Mondal et al., 2015; Oleary et al., 2015; Kalwa et al., 2016) can regulate gene expression by RNA–DNA triplex formation, indicating that this type of formation may be a general mechanism of target site recognition by lncRNAs. *GhDANI* could bind to an AAAG-rich DNA motif corresponding to the Dof TFs’ core binding sites in chromatin by forming an RNA–DNA triplex. The triplex can engage *cis* elements on the promoters of stress-responsive and developmentally related genes in order to prevent gene transcription under certain conditions. Dof TFs are plant specific, and their involvement in a wide range of processes is suggested by their binding to the AAAG consensus sequence in target promoters; this implies that they have important
roles, such as regulating stress responses, photosynthesis, or flower induction. Dof TFs can act as either positive or negative regulators of gene expression (Noguero et al., 2013). It was documented that a Dof protein DNA binding sequence (ACTTTA), which is targeted by a tobacco Dof protein NtBBF1, acts as the cis-regulatory element required in auxin induction of rolB (Baumann et al., 1999), which linked the relationship between Dof TFs and the auxin response pathway. More transcriptional association between Dof TFs and their potential target genes involved in auxin signaling has been established (Pattison et al., 2015). As the first plant hormone to be identified, auxin and its response pathway not only play important roles in plant growth regulation, but also interconnect and stimulate different metabolic pathways under abiotic stresses (Tognetti et al., 2012; Du et al., 2013; Tiwari et al., 2017; Wang et al., 2019). Recently, Auxin-sensitive Aux/IAA proteins were found to mediate drought tolerance in Arabidopsis by regulating glucosinolate levels, which provide a connection between auxin signaling and drought response (Salehin et al., 2019). GhDAN1 has the ability to “sequester” the core binding sites of Dof TFs, thus retarding positive or negative regulation by TFs and coordinating proper gene expression during drought stress. Accordingly, the genes in auxin response pathway corresponding to Dof TFs in plants were coordinated by GhDAN1 regulation. We infer that by coordinating transcription factors such as Dof for the binding of a specific DNA motif, GhDAN1 forms a negative feedback loop in response to abiotic stress.

Triplex-forming motifs identified in the mammalian IncRNAs MEG3, HOTAIR, and PARTICLE transcripts were seen to be mostly AG-rich sequences, which formed
antiparallel triplexes with their AG-rich DNA target sequences by forming antiparallel reverse Hoogsteen hydrogen bonds with the purines of the DNA duplex (Mondal et al., 2015; Oleary et al., 2015; Kalwa et al., 2016). The lncRNAs Fendrr and Khps1, with TC-rich TFOs in their transcripts, bind to AG-rich DNA target sequences by means of parallel Hoogsteen hydrogen bonds with the purines of the DNA duplex (Grote et al., 2013; Postepska-Igielska et al., 2015). Our example of GhDAN1 forms an RNA–DNA triplex according to the rules of parallel Hoogsteen configuration, and this indicates a common functional strategy of lncRNA in both plant and animal genomes.

Interestingly, a new AG-rich motif in a plant with a sequence of AAAG, rather than simple AG repeats, might represent a divergence in patterns of RNA–DNA triplex formation between plant and animal genomes.

Triplex-forming lncRNAs are also reported to recruit specific proteins (e.g., PRC2) to the correct locus on chromatin in response to stimuli or oncogene expression (Mondal et al., 2015; Oleary et al., 2015; Kalwa et al., 2016). The RNA–DNA triplex formed by DAN1 could also be involved in site-specific protein binding in response to environmental signals or abiotic stress. Further examination of DAN1-binding proteins is the next critical question to address.

The potential of non-coding genes in abiotic stress

Drought causes severe losses in global crop production (Comas et al., 2013). It is a major unsolved issue in modern agriculture. The study of mechanisms of drought tolerance is needed in order to improve crop adaptability under drought stress.
Knowledge of drought tolerance variations among *Gossypium* species may help us to understand how genomic shock leads to the activation and fixation of specific functional genes, and it may also shed light on our progress in the genetic breeding of cultivated cotton species with altered characteristics that would improve stress tolerance. According to the phenotypic analysis, the A diploid cotton species we tested showed high drought tolerance, whereas most cultivated allotetraploids seemed to have lost this valuable characteristic (this change has often been recognized in previous reports) (Zhixia, 2010; Chen et al., 2015; Jia et al., 2018). Plants might take on defensive strategies that sacrifice their biomass accumulation and developmental processes in stress conditions, a tradeoff we have called *fitness cost* (Bostock et al., 2014; Cipollini et al., 2017). According to our data, *GhDANI* may act as a modulator of the auxin pathways, thereby adjusting the balance between biomass accumulation and stress tolerance processes. To optimize the balance between stress tolerance and biomass accumulation after hybridization and polyploidization, the allotetraploid genome could activate non-coding genes such as *DANI* to create a novel regulatory feedback loop in response to drought. Considering the evolutionary profile of *DANI* and its mode of action, we have inferred that *GhDANI* activation is a remarkable case of a gene that, during hybridization and WGD, was selected for and fixed to maintain the balance between its growth and fitness costs when encountering environmental stress.

**Materials and Methods**
Cotton species and plant culture

Seed and leaf sources of the varieties of diploid and allotetraploid cotton used in this study, including wild species, landraces, and cultivars, are listed in Supplemental Table S1. Plant material of the interspecies hybrid F₁ of *G. arboreum* (A₂, 2n = 2x = 26) and *G. raimondii* (D₅, 2n = 2x = 26) was used as previously described (Zhao et al., 2018). All leaf samples were collected in identical field growth conditions, frozen in liquid nitrogen immediately, and stored at −70 °C until they were prepared for RNA isolation and *DAN1* gene expression analysis. For the VIGS assay, cotton seedlings were grown in pots at 22 °C in a greenhouse in a 16/8-hour light/dark cycle with 60% humidity. Seedlings with mature cotyledons but without visible true leaves (7 days after germination) were infiltrated in the VIGS assay. For drought treatment, abiotic treatment, and leaf collection for the RNA-FISH, ChIP, and ChIRP assays, seedlings were grown in the same greenhouse conditions as for the VIGS, and the treatment or sample collection was performed when the seedlings had two or three true leaves (i.e., they were 4-week-old seedlings).

Virus-induced gene silencing

The preparation of tobacco rattle virus (TRV) vectors and *Agrobacterium tumefaciens* bacteria for our experiments in cotton plants was conducted as previously reported (Zhao et al., 2018). Fragments of candidate genes were amplified and separately cloned into a *pTRV2* vector by the infusion method to construct recombinant *pTRV2* vectors.
Of these, the recombinant pTRV2–GhCLA vector was used as a positive control in order to monitor the effect of the VIGS. The primer pairs used to construct recombinant pTRV2 vectors are listed in Supplemental Table S6. The vectors, including pTRVI, recombinant pTRV2, and pTRV2 empty vectors, were then separately introduced into specimens of A. tumefaciens strain GV3101 by electroporation. The specimens of A. tumefaciens GV3101 containing pTRV1 and recombinant pTRV2 were cultured overnight at 28°C in an antibiotic selection medium with rifampicin (25 mg/L) and kanamycin (50 mg/L). Agrobacterium cells were then centrifuged and resuspended in an infiltration buffer (10 mM MgCl₂, 10 mM MES, and 200 mM acetosyringone) adjusted to an OD₆₀₀ of 2. Specimens of A. tumefaciens GV3101 with pTRV1 vectors were mixed individually with either recombinant pTRV2 vectors containing candidate genes or a positive marker gene or an empty vector pTRV2 at a ratio of 1:1. The mixture was left in darkness at 28°C for 3 hours. Finally, 7-day-old seedlings were infiltrated by inserting the Agrobacterium suspension into the cotyledons with a syringe. Four weeks after infiltration, the second true leaves from each seedling were collected for gene expression analysis, or the intact seedlings were subsequently subjected to drought treatment. For each vector or gene, 15 to 30 Gh TM-1 seedlings were infiltrated for subsequent phenotypical analysis. The VIGS assay for GhDAN1 was performed with at least five biological replicates.

**Drought treatment**

Downloaded from https://academic.oup.com/plphys/advance-article/doi/10.1093/plphys/kiab179/6237502 by guest on 23 April 2021
For each independent biological experiment, 12 to 15 4-week-old cotton seedlings (having at least two fully expanded true leaves) were subjected to a cessation in watering. After 10 to 20 days of water deprivation, the phenotype was observed, and a sample collection or re-watering treatment for 2 to 3 days was performed when needed.

**Abiotic treatment**

Twelve to fifteen 4-week-old cotton seedlings (having at least two fully expanded true leaves) were used in each abiotic treatment group. Different foliar sprays composed of 150 mM NaCl, 1 μM abscisic acid, 1 mM ethephon, or 2 mM H₂O₂ were applied, and the leaves were collected 6 hours later for gene expression analysis. Plants sprayed with distilled water were used as controls. Seedlings were treated in a growth chamber at 42 °C or 4 °C for 6 hours before analysis.

**LncRNA-seq and pipeline for LncRNA identification**

LncRNA sequencing and the pipeline for LncRNA identification in the cotton lineage were described previously (Zhao et al., 2018). The total RNA was isolated from the plant tissues using the Spectrum™ Plant Total RNA Kit (Sigma-Aldrich, St. Louis, MO, USA). Ribosomal RNA was removed using the Epicentre Ribo-Zero™ rRNA Removal Kit (Epicentre, Madison, WI, USA). LncRNA libraries were generated from rRNA-depleted RNA using the NEBNext® Ultra™ Directional RNA Library Prep Kit.
for Illumina® (New England Biolabs, Inc., Beverly, MA, USA). Strand-specific sequencing was performed with the Illumina HiSeq 2000 system (paired-end 125-bp reads; Illumina, Inc., San Diego, CA, USA). The quality control of strand-specific RNA-seq reads, the mapping of reads to the reference genomes, and the assembly of all of the transcripts were as described (Zhao et al., 2018). The coding potential of the transcripts was calculated using the Coding Potential Calculator (CPC) software (Kong et al., 2007). All transcripts with CPC scores above zero were discarded. The remaining transcripts were subjected to HMMER (v. 3.1b2) to exclude the transcripts containing known protein domains (cutoff < 0.001) (Finn et al., 2011). The remaining transcripts were candidate IncRNAs.

**Analysis of whole-genome MethylC-seq reads**

MethylC-seq data for diploid *G. arboreum*, (*G. arboreum* x *G. raimondii*)F₁, and *G. hirsutum* were retrieved from a previous study (Song et al., 2017). Methylation levels in the *DANI* promoter and gene body regions were calculated using in-house Perl and R scripts.

**Physiological trait measurements**

The total GSH content was measured using a glutathione assay kit (Sigma-Aldrich cat. no. CS0260). For the sample preparation, 50 mg of ground tissue was homogenized
with 0.5 mL 5% (w/v) 5-sulfosalicylic acid (SSA) solution to obtain the supernatant. The reaction procedure was performed according to the user manual. After the reaction, a plate reader was set to 412 nm with a kinetic read at 1-minute intervals for 10 minutes. The standard curve was established by using a series of concentrations of GSH (3.125, 6.25, 12.5, 15, and 50 μM). Based on the standard, the total GSH content of the samples was calculated. GSH content values from three independent biological replicates were collected for each treatment.

The proline concentration was measured based on its reaction with ninhydrin according to the following method (Lee et al., 2018). Fresh tissue samples (~0.1 g) were homogenized in 1 mL of extracting solution (3% (w/v) SSA) on ice, followed by incubation at 100°C for 10 minutes. Samples were then centrifuged at 25°C at 4000 rpm for 10 minutes. An aliquot of 0.25 mL of the supernatant was transferred to a new tube and combined with 0.25 mL glacial acetic acid and 0.25 mL 2.5% (w/v) acid ninhydrin. The reaction was inoculated at 100°C for 0.5 hours and terminated in an ice bath. An amount of 0.5 mL of toluene was added to the reaction, and the solution was vortexed for 30 seconds. The absorbance at 520 nm was recorded. The standard curve was established by using a series of concentrations of proline (0, 1, 2, 4, 6, 8, 10, and 15 μg/mL). Based on the standard, the proline content of the samples was calculated. The proline content values were collected from three independent biological replicates for each treatment.
Photosynthetic parameters were determined using intact, fully expanded, functional leaves (the third or fourth topmost leaves). Measurements of the net photosynthetic rate (Pn), stomatal conductance (Gs), transpiration rate (Tr), and intercellular CO₂ concentration (Ci) were gathered with an LI-6400 portable photosynthesis system (LI-COR Biosciences, Lincoln, NE, USA) according to instructions in the user manual. Photosynthetic parameters of 15 to 30 seedlings were collected for each treatment.

**Fractionation of RNA**

Isolation of nuclei, cytoplasm, nucleoplasm, and nucleoli from the cotton leaves was performed as previously described (Xing et al., 2017) with modifications. One gram of cotton leaves was ground into a fine powder using liquid nitrogen. The ground powder was transferred into a 50-mL ice-cold centrifuge tube containing 30 mL ice-cold Tris buffer (10 mM Tris pH 8.0, 10 mM MgCl₂, 0.1 mM PMSF, 5 mM β-mercaptoethanol). After being agitated gently, the mixture was immediately filtered through four layers of Miracloth into a new 50-mL tube and centrifuged at 100x g for 5 minutes at 4 °C to derive a pellet. Two milliliters of a nuclei isolation buffer (10 mM Tris pH 8.0, 10 mM MgCl₂, 0.1 mM PMSF, 5 mM β-mercaptoethanol, 1% (v/v) Triton X-100, 1x proteinase inhibitors, 5 U/μL RNAse inhibitors) was added to resuspend the pellet. The suspension was incubated on ice for 10 minutes and then centrifuged at 100x g for 3 minutes at 4 °C to form a pellet of the nuclei. An aliquot of
0.5 mL supernatant (the cytoplasmic fraction) was transferred to a new 2-mL tube for cytoplasmic RNA extraction. The nuclei pellet was resuspended with 1 mL 340 mM sucrose solution containing 5 mM MgCl₂, and 250 μL of the lysate was transferred to a new 2-mL tube for RNA extraction of the nuclei. The rest of the lysate of the nuclei were broken by sonication until intact nuclei could not be detected in the suspension by a microscope. A solution of 750 μL 880 mM sucrose containing 5 mM MgCl₂ was gently added to the sonicated nuclei and then centrifuged for 20 minutes at 2000 rpm at 4 °C to form a pellet of the nucleoli. The supernatant was the nucleoplasmic fraction. Fractionated RNAs of the nuclei, cytoplasm, nucleoplasm, and nucleoli were extracted using the PureLink Plant RNA Reagent (Invitrogen™ cat. no. 12322012, Thermo Fisher Scientific, USA). Fractionated RNAs from the same amount of cells were used for cDNA synthesis and RT-qPCR. Each fractionated RNA had three biological replicates for analysis.

RNA-FISH Assay

The RNA-FISH assay was performed as previously described (Zhu et al., 2016) using isolated nuclei. For plant material cross-linking, 1 g of fresh cotton leaves was cut into pieces and transferred to a 50-mL RNase/DNase-free tube and cross-linked with 30 mL of fixation buffer (10 mM Tris pH 8.0, 1% w/v formaldehyde, 0.1 mM PMSF, 5 mM β-mercaptoethanol) for 15 minutes by vacuum infiltration in an exicator at room temperature. After cross-linking, 2.5 mL 2 M glycine was added, followed by vacuum
infiltration for 5 minutes to quench the reaction. The tissue was then washed three times with plenty of ddH$_2$O and briefly dried between sheets of clean absorbent paper for 30 seconds to remove most of the water. For nuclei isolation, tissue reserved from the cross-linking step was ground using liquid nitrogen to form a fine-grained dry powder. Nuclei were isolated as described above in the section entitled Fractionation of RNA. A Quasar 570–labeled \textit{GhDAN1}-specific probe set was ordered. A hybridization assay was performed with 2 mL of the nuclei suspension using Stellaris™ RNA-FISH Hybridization Buffer, Wash Buffer A and Wash Buffer B (Biosearch Technologies, Beverly, MA, USA, cat. nos. SMF-HB1-10, SMF-WA1-60, and SMF-WB1-20) strictly following the protocol for hybridization, washing, and DAPI staining according to the Stellaris™ RNA-FISH kit before proceeding to imaging. For probes labeled with Quasar 570, an excitation line of 561 nm was applied and signals were detected at 570-640 nm; for DAPI, an excitation line of 405 nm was applied and signals were detected at 420-480 nm.

**Chromatin isolation by RNA purification**

The chromatin isolation by RNA purification (ChIRP) assay was performed as described (Chu et al., 2012; Ariel et al., 2020) using the Magna ChIRP RNA Interactome Kit (Millipore Sigma, Burlington, MA, USA, cat. no. 17-10495), with minor modifications. Antisense oligo probes for \textit{GhDAN1} were designed using the online probe designer.
The probes were designed against every 100 bp over the 2 kb region on GhDAN1 (as indicated in Supplemental Fig. S3) and numbered based on their position along the RNA from 5’ to 3’. GhDAN1 ChIRP reactions were performed using “even” numbered probe pool and “odd” numbered probe pool. Real RNA-dependent signal will be present in both “even” and “odd” pools as shared DNA peaks, while the random noises will be unique to each pool [52]. Two grams of fresh cotton leaves was sufficient for the ChIRP reaction with the gene-specific probes (odd and even probe sets) and input control. The plant material cross-linking and nuclei isolation was performed as described above in the section entitled RNA-FISH Assay. Next, the pellet of nuclei from each 1 g of tissue was resuspended in 2 mL of complete lysis buffer (from the Millipore ChIRP Kit), followed by immediate sonicating in a Bioruptor™ (Diagenode, Denville, NJ, USA) to shear the DNA (aliquot 350 μL in each tube for sonication) to 100 to 500 bp in length. The sonicated chromatin was centrifuged for 10 minutes at 12,000x g at 4 °C to collect the supernatant. Ultimately, approximately 2 mL of the sonicated chromatin was obtained from each 1 g of fresh leaves. The tubes were snap frozen in liquid nitrogen and stored at –80 °C. For chromatin isolation by RNA purification, each 1-mL sample of sonicated chromatin was transferred to a 15-mL RNase/DNase-free tube to be used in one ChIRP reaction. The Millipore ChIRP Kit user manual guidelines were followed carefully for completion of the ChIRP reaction assay and DNA isolation, followed by next-generation sequencing (NGS). Magna ChIRP™ Negative Control Probe Set (LacZ) in the Magna ChIRP RNA Interactome
Kit was used as negative control. For the ChIRP-seq bioinformatic analysis, the first step after quality control and trimming of the adaptor was to map the clean data of the NGS to the reference cotton genome (TM-1) (Hu et al., 2019) using hierarchical indexing for spliced alignment of transcripts 2 (HISAT2) (the no-spliced-alignment parameter) (Pertea et al., 2016). Second, peaks were called separately using data from the odd and even probe sets by MACS2 (using the p-value 1 x 10^{-5} parameter) (Feng et al., 2011). Third, the peaks shared by the odd and even pools (designated as ChIRP valid peaks) were generated using intersectBed in the BEDTools suite (Quinlan and Hall, 2010). The peaks with a fold enrichment greater than or equal to 2 were selected as the final valid ChIRP peaks. The motif discovery was performed using MEME Suite 5.0.5 (Bailey et al., 2009) with the top 10% most-enriched valid peaks. (MEME Top1000validpeaks.fasta -dna -nmotifs 20 -minsites 50 -minw 6 -maxw 20 -maxsize 6000000); motif scanning was performed using FIMO in MEME Suite 5.0.5 (using default settings); and motif comparison was performed using Tomtom in MEME Suite 5.0.5 (using default settings). Two biological replicates for ChIRP reaction were performed, and similar AAAG rich motif was repetitively established by MEME software.

**Triplex prediction**

Prediction of the triplex-forming oligonucleotides (TFOs) in GhDAN1 RNA was performed using Triplexator software (Buske et al., 2012).
EMSA

EMSA for IncRNA and DNA interaction was performed as previously described, with modifications (Postepska-Igielska et al., 2015). A GhDAN1 fragment from +9 to +186 bp was first cloned into the pMD$_{19-T}$ vector to construct pMD$_{19-T}$-GhDAN1 for sequence confirmation. PCR products containing a T7 promoter in the correct orientation were then prepared and purified by phenol–chloroform extraction and ethanol precipitation. RNA synthesis was performed using the HiScribe™ T7 High Yield RNA Synthesis Kit (NEB cat. no. E040S), with the above PCR products used as templates, according to the user manual. As a result, approximately 8 μg/μL (total 30 μL) GhDAN1 RNA was generated from each reaction. The 3’-DIG-labeled single-stranded oligonucleotide was synthesized and stored in ddH$_2$O at a concentration of 100 pmol/μL. DIG-labeled DNA double-stranded probes (4 pmol) and cold probes (40 pmol) were annealed in a 50 μL reaction containing 50 mM NaCl following the prescribed program (95 °C for 10 minutes, 75 °C for 15 minutes, 55 °C for 15 minutes, 45 °C for 15 minutes, and 35 °C for 15 minutes). In the EMSA reaction, 4 pmol (1 μL) of DIG-labeled double-stranded oligonucleotide, comprised of nucleotides containing the GhDAN1 TTS motif that were predicted within the gene promoter (see Supplemental Table S6 for detailed sequences) was incubated for 2 hours at 37°C with 16 μg (~320 pmol) synthetic GhDAN1 fragment in 40 mM Tris-acetate (pH7.5), 20 mM KCl, 10 mM Mg (CH$_3$COO)$_2$, and 10% (v/v) glycerol. Cold
probes (40 pmol) were added for a competitive reaction. RNaseA or RNaseH was added after the reaction when needed to confirm the triplex formation. Triplex formation was confirmed by EMSA using 12% (w/v) polyacrylamide gels containing 40 mM Tris-acetate (pH 7.5) and 10 mM MgCl₂. After electrophoresis in 1x TBE, the DNA/RNA was transferred from the gel to an Amersham™ Hybond™ –N+ membrane in 1x TBE at 400 mA for 1 hour. The probe signal was detected using the DIG Nucleic Acid Detection Kit (Roche cat. no. 11175041910).

**mRNA-seq**

The total RNA was extracted using the PureLink Plant RNA Reagent (Invitrogen™ cat. no. 12322012). Libraries of mRNA were prepared and sequenced using the HiSeq 150-bp paired-end Illumina RNA-seq protocol, with three biological replicates for each treatment. Reads were quality controlled and trimmed using fastp (Chen et al., 2018). The clean reads were mapped onto the cotton genome using HISAT2 (Pertea et al., 2016). Gene expression was calculated using StringTie with parameter (-A) (Pertea et al., 2016). For differential expression analysis, each gene count was calculated using HTseq-count (Anders et al., 2015). DEGs were tested using the edgeR package (Robinson et al., 2010).

**Reverse Transcription Quantitative PCR**
The total RNA was extracted using the PureLink Plant RNA Reagent (Invitrogen™ cat. no. 12322012). cDNAs were synthesized from the total RNA using HiScript II Q RT SuperMix for qPCR (+gDNA wiper) (Vazyme cat. no. R233). Reverse transcription quantitative PCR (RT-qPCR) was performed on the ABI 7900HT Fast Real-Time PCR System using histone h3 (AF024716) as a reference gene. Primers used in this study are presented in Supplemental Table S6.

**Chromatin Immunoprecipitation**

For the ChIP assay, a procedure similar to that of the ChIRP assay was followed to complete the cross-linking of the plant material, nuclei isolation, cell lysis, and chromatin sonication (Haring et al., 2007). The immunoprecipitation of the cross-linked protein/DNA was performed using the Magna ChIP™ A/G Chromatin Immunoprecipitation Kit (Millipore cat. no. 17-10085) according to the user manual instructions. Anti-histone H3 (trimethyl K4) antibodies (ab8580) and anti-histone H3 (trimethyl K27) antibodies (ab6002) were used in the assay. NGS reads were mapped to the cotton genome using Bowtie, allowing up to two mismatches (Langmead et al., 2009). The reference genome for *G. herbaceum* subs. Africanum (A₁₋₂), *G. arboreum* (A₂, accession Shixiya 1), and (*G. arboreum* x *G. raimondii*)F₁ was *G. arboreum* (A₂, accession Shixiya 1) (Du et al., 2018); the reference genome for *G. barbadense* (AD₂, accession H7124) and *G. darwinii* (AD₃) was *G. barbadense* (AD₂, accession H7124) (Hu et al., 2019); and the reference genome for *G. hirsutum* (AD₁, accessions
yucatanense2094, punctatum25 and TM-1), *G. tomentosum* (AD3), *G. mostelinum* (AD4), *G. ekmanianum* (AD6), and *G. stephensii* (AD7) was *G. hirsutum* (AD1, accession TM-1) (Hu et al., 2019). For the uniquely mapped reads, peak calling of the ChIP-seq data was performed using MACS2 (Zhang et al., 2008).

**Phylogenetic analysis**

Phylogenetic tree depicting the evolutionary relationship of *DAN1* in cotton lineage was performed using MEGA software (Tamura et al., 2007).

**Statistical Analyses**

All statistical analysis was done with the software Graphpad Prism (GraphPad Software, La Jolla, CA, USA). The Student’s *t*-test, Mann–Whitney *U*-test, or Fisher’s exact test was used whenever appropriate. The *p*-values of less than .05 were considered statistically significant.

**Accession numbers**

Novel data generated in this study, including raw data of mRNA-seq and ChIRP-seq, have been deposited in the National Center for Biotechnology Information under the accession number PRJNA637977. Pol II ChIP-seq data are under the accession number PRJNA373801, as previously reported (Zhao et al., 2018).
SUPPLEMENTAL DATA:

Supplemental Figure S1. Phenotypes of \textit{GhDAN1}, \textit{CLA1}, and \textit{GoPGF}-silenced plants.

Supplemental Figure S2. Analysis of physiological traits associated with photosynthesis.

Supplemental Figure S3. Alternative splicing of \textit{GhDAN1}.

Supplemental Figure S4. The lincRNA \textit{GhDAN1} has a poly(A) tail and does not encode any peptide/protein.

Supplemental Figure S5. DNA CHG and CHH methylation status of the \textit{DAN1} gene body and promoter region.

Supplemental Figure S6. Sequence alignment of \textit{DAN1} genomic DNA in 13 cotton lineages.

Supplemental Figure S7. Stack plots of ChIP-seq showing the H3K27me3 profile in \textit{DAN1} loci in 12 cotton lineages.

Supplemental Figure S8. Phenotype of A diploid and AD allotetraploid plants before (0 D) and after drought treatment lasting 10 days (10 D).

Supplemental Figure S9. \textit{GhDAN1} is not a regulator in \textit{cis}.

Supplemental Figure S10. Analysis of ChIRP-seq peaks and peak-associated genes.
Supplemental Figure S11. mRNA-seq results of selected DEGs and RT-qPCR confirmation.

Supplemental Figure S12. *GhDAN1* does not affect genes in the auxin biosynthesis or the ABA biosynthesis/catabolism pathways.

Supplemental Figure S13. Profile of the TTS motif-containing DEGs from *GhDAN1*-silenced plants under dehydration treatment.

Supplemental Figure S14. Phenotypes of plants that have undergone VIGS before and after drought treatment.

Supplemental Table S1. Information of cotton accessions used in this study.

Supplemental Table S2. TFOs in *GhDAN1* predicted by Triplexator software.

Supplemental Table S3. List of differentially expressed genes (DEGs) in *GhDAN1*-silenced plants.

Supplemental Table S4. List of orthologous genes of TTS motif containing-DEGs in other plants that were involved in drought stress tolerance.

Supplemental Table S5. Shared DEGs with the TTS motif between *GhDAN1*-silenced plants and dehydration-treated Gh plants.

Supplemental Table S6. Primers used in this study.

**Acknowledgement**
We would like to thank Professor Kunbo Wang from the National Wild Cotton Nursery, Sanya, and Professor Xiongming Du from the National Medium Term Gene Bank of Cotton in China, the Institute of Cotton Research of the Chinese Academy of Agricultural Sciences, for providing the seed and leaf tissues of diploid cotton and wild allotetraploid cotton germplasm. We would also like to thank Professor Liang Wu for providing insightful comments on the manuscript.

FIGURE LEGENDS

Figure 1. GhDAN1 is associated with the drought-tolerance trait in the cotton ancestor. (A) A simplified model system that mimics interspecific hybridization and WGD in the cotton lineage. (B) Drought-tolerance phenotypes of Ga diploid and Gh allotetraploid cotton (wild and cultivar accessions). (C and D) Drought-tolerance phenotype of GhDAN1-silenced plants and GhDAN1 expression. Error bars represent the SD of 15 seedlings in each group of treatment. ** p<0.01, Student's t-test. Total GSH (E) and proline contents (F) before and after drought treatment. Error bars represent the SD of three biological replicates. * p<0.05, Mann-Whitney U-test. (G) Syntenic analysis of DAN1 in diploid Ga and allotetraploid Gh genomes. (H and I) GhDAN1 expression in different tissues and the profile in response to multiple abiotic treatments. Error bars represent the SD of three biological replicates. * p<0.05, Mann-Whitney U-test. (J to L) Stack plots for transcript profile (RNA-seq) (J), Pol II ChIP-seq profile (K) and the
DNA CG-methylation status (L) of \textit{DAN1} loci in Ga (A$_2$), (Ga x Gr)F$_1$, and Gh (AD$_1$) genomes.

**Figure 2.** \textit{DAN1} is active in the allotetraploid cotton species. (A) Schematic chart showing phylogenetic relationships of diploid ancestors and seven allotetraploid cotton species. (B) \textit{DAN1} transcription activity in two diploid A species and seven allotetraploid species (accessions are listed in order in Supplemental Table S1). Error bars represent the SE of three technical replicates. (C) Enrichment of H3K4me3 ChIP-seq signals (boxed region) near the \textit{DAN1} transcription starting site in the allotetraploid cotton lineage. Fold changes were calculated using normalized peak counts compared with their corresponding mock control. (D) Drought-tolerance phenotype of diploid and allotetraploid cotton cultivars. The number of unwilted leaves after water restriction was calculated (see the Supplemental Fig. S8 for details); Error bars represent the SD of six seedlings from each group. \( p \) value was indicated by comparing Ga (A$_2$) diploid with Gh (AD$_1$) and Gb (AD$_2$) allotetraploid cultivars, respectively. Student's \( t \)-test. (E) Analysis of the correlation between \textit{DAN1} expression and the unwilted leaf count from panel D.

**Figure 3.** \textit{GhDAN1} forms RNA-DNA triplexes to anchor the AAAG DNA target site and regulate gene expression in trans. (A and B) RT-qPCR analysis of \textit{GhDAN1} expression in the nucleus versus cytoplasm fractions, and in the nucleolus versus
nucleoplasm. Histone3 and U3 snoRNA were used as controls. Error bars represent the SE of three technical replicates. (C) RNA-FISH assay showing the GhDANI distribution in the nucleus. Bars= 5 µm. (D) Alignment of the Triplexator software-predicted GhDANI triplex target sites (TTS) with ChIRP-established AG-rich motif and Dot TFs binding motifs. (E) Number and density of TTSs over the gene-associated GhDANI peak summits and neighboring regions. The TTS motif was predicted by FIMO in the peaks (±250 bp from the peak summits) and the flanking control regions (500 bp upstream and downstream from the peaks). The density of the TTS motif was generated by ggplot2 in R using the distances between the predicted TTS motif and the peak summit. (F and G) EMSA assay showed triplex formation of GhDANI RNA with DIG-labeled TTS DNA probes in vitro. The nucleotide in red color in Probe 2 indicated two mismatches compared with predicted TTS motif. The arrows indicated the RNA-DNA triplexes.

**Figure 4.** Genes with GhDANI the triplex target site (TTS) motif were sensitive to silencing of GhDANI and significantly enriched in the auxin response pathway. (A) The column chart shows the number of DEGs in GhDANI-silenced plants and the distribution of genes in each upregulated and downregulated group that were predicted with or without the GhDANI TTS motif in the 2-kb promoters using FIMO scanning (score > 10). (B) Occurrence of GhDANI TTS motif in DEGs and the total coding genes. (C) GO analysis of upregulated and downregulated genes in GhDANI-silenced...
plants showing significant enrichment in the auxin response pathway. (D) Heatmap showing the expression profiles of the representative DEGs with the *GhDAN1* TTS motif. (E) Cotton seedlings were sensitive to drought stress after silencing of the *GhDAN1* downstream gene *GhTHI1-2*. (F) Stack plots of mRNA-seq on *GhTHI1-2* transcription activity during dehydration treatment. The predicted *GhDAN1* TTS motif in the promoter was indicated by the black arrow.

**Figure 5.** Working model of *DAN1*-mediated drought-response dynamics.

**LITERATURE CITED**

Adams KL, Percifield R, Wendel JF (2004) Organ-Specific Silencing of Duplicated Genes in a Newly Synthesized Cotton Allotetraploid. Genetics 168: 2217-2226

Adams KL, Wendel JF (2005) Polyploidy and genome evolution in plants. Current Opinion in Plant Biology 8: 135-141

Anders S, Pyl PT, Huber W (2015) HTSeq--a Python framework to work with high-throughput sequencing data. Bioinformatics 31: 166-169

Ariel F, Lucero L, Christ A, Mammarella MF, Jegu T, Veluchamy A, Mariappan K, Latrasse D, Blein T, Liu C, Benhamed M, Crespi M (2020) R-Loop Mediated trans Action of the APOLO Long Noncoding RNA. Molecular Cell 77: 1055-1065 e1054

Chu C, Quinn J, Chang HY (2012) Chromatin isolation by RNA purification (ChIRP). Journal of Visualized Experiments 61, 3912.

https://doi.org/10.3791/3912
Bailey TL, Boden M, Buske FA, Frith M, Grant CE, Clementi L, Ren J, Li WW, Noble WS (2009) MEME SUITE: tools for motif discovery and searching. Nucleic Acids Research 37: W202-208

Barrera-Redondo J, Ibarra-Laclette E, Vázquez-Lobo A, Gutiérrez-Guerrero YT, Sánchez de la Vega G, Piñero D, Montes-Hernández S, Lira-Saade R, Eguiarte LE (2019) The Genome of Cucurbita argyrosperma (Silver-Seed Gourd) Reveals Faster Rates of Protein-Coding Gene and Long Noncoding RNA Turnover and Neofunctionalization within Cucurbita. Molecular Plant 12: 506-520

Baumann K, De Paolis A, Costantino P, Gualberti G (1999) The DNA Binding Site of the Dof Protein NtBBF1 Is Essential for Tissue-Specific and Auxin-Regulated Expression of the rolB Oncogene in Plants. The Plant Cell 11: 323-333

Bostock RM, Pye MF, Roubtsova TV (2014) Predisposition in Plant Disease: Exploiting the Nexus in Abiotic and Biotic Stress Perception and Response. Annual Review of Phytopathology 52: 517-549

Buske FA, Bauer DC, Mattick JS, Bailey TL (2012) Triplexator: Detecting nucleic acid triple helices in genomic and transcriptomic data. Genome Research 22: 1372-1381

Chen S, Zhou Y, Chen Y, Gu J (2018) fastp: an ultra-fast all-in-one FASTQ preprocessor. Bioinformatics 34: i884-i890

Chen Y, Wang Y, Zhao T, Yang J, Feng S, Nazeer W, Zhang T, Zhou B (2015) A New Synthetic Amphiploid (AADDAA) between Gossypium hirsutum and G. arboreum Lays the Foundation for Transferring Resistances to Verticillium and Drought. PLoS One 10(6): e0128981. doi:10.1371/journal.pone.0128981

Chen ZJ (2007) Genetic and Epigenetic Mechanisms for Gene Expression and Phenotypic Variation in Plant Polyploids. Annual Review of Plant Biology 58: 377-406
Chen ZJ (2010) Molecular mechanisms of polyploidy and hybrid vigor. Trends in Plant Science 15: 57-71

Chen ZJ (2013) Genomic and epigenetic insights into the molecular bases of heterosis. Nature Reviews Genetics 14: 471-482

Cipollini D, Walters D, Voelckel C (2017) Costs of Resistance in Plants: From Theory to Evidence. In Annual Plant Reviews online, J.A. Roberts (Ed.). doi:10.1002/9781119312994.apr0512

Comas LH, Becker SR, Cruz VM, Byrne PF, Dierig DA (2013) Root traits contributing to plant productivity under drought. Frontiers in Plant Science 4: 442

Conrad B, Antonarakis SE (2007) Gene duplication: a drive for phenotypic diversity and cause of human disease. Annual Review of Genomics and Human Genetics 8: 17-35

Cronn RC, Zhao X, Paterson AH, Wendel JF (1996) Polymorphism and concerted evolution in a tandemly repeated gene family: 5S ribosomal DNA in diploid and allopolyploid cottons. Journal of Molecular Evolution 42: 685-705

Davidson RM, Manosalva P, Snelling J, Bruce M, Leung H, Leach JE (2010) Rice Germin-Like Proteins: Allelic Diversity and Relationships to Early Stress Responses. Rice 3: 43-55

Deng P, Liu S, Nie X, Weining S, Wu L (2018) Conservation analysis of long non-coding RNAs in plants. Science China-life Sciences 61: 190-198

Du H, Liu H, Xiong L (2013) Endogenous auxin and jasmonic acid levels are differentially modulated by abiotic stresses in rice. Frontiers in Plant Science 4: 397-397

Du X, Huang G, He S, Yang Z, Sun G, Ma X, Li N, Zhang X, Sun J, Liu M (2018) Resequencing of 243 diploid cotton accessions based on an updated A genome identifies the genetic basis of key agronomic traits. Nature Genetics 50: 796-802
Feng J, Liu T, Zhang Y (2011) Using MACS to Identify Peaks from ChIP-Seq Data. Current Protocols in Human Genetics Chapter 2: Unit 2.14. doi: 10.1002/0471250953.bi0214s34.

Finn RD, Clements J, Eddy SR (2011) HMMER web server: interactive sequence similarity searching. Nucleic Acids Research 39: W29-W37

Franco-Zorrilla JM, Valli A, Todesco M, Mateos I, Puga MI, Rubio-Somoza I, Leyva A, Weigel D, Garcia JA, Paz-Ares J (2007) Target mimicry provides a new mechanism for regulation of microRNA activity. Nature Genetics 39: 1033-1037

Goudarzi M, Berg K, Pieper LM, Schier AF (2019) Individual long non-coding RNAs have no overt functions in zebrafish embryogenesis, viability and fertility. Elife 8: e40815 doi: 10.7554/eLife.40815

Grote P, Wittler L, Hendrix DA, Koch F, Wahrisch S, Beisaw A, Macura K, Blass G, Kellis M, Werber M (2013) The Tissue-Specific lncRNA Fendrr Is an Essential Regulator of Heart and Body Wall Development in the Mouse. Developmental Cell 24: 206-214

Ha M, Lu J, Tian L, Ramachandran V, Kasschau KD, Chapman EJ, Carrington JC, Chen X, Wang XJ, Chen ZJ (2009) Small RNAs serve as a genetic buffer against genomic shock in Arabidopsis interspecific hybrids and allopolyploids. Proceedings of the National Academy of Sciences of the United States of America 106: 17835-17840

Haring M, Offermann S, Danker T, Horst I, Peterhansel C, Stam M (2007) Chromatin immunoprecipitation: optimization, quantitative analysis and data normalization. Plant Methods 3: 11

Hu Y, Chen J, Fang L, Zhang Z, Ma W, Niu Y, Ju L, Deng J, Zhao T, Lian J, Baruch K, Fang D, Liu X, Ruan YL, Rahman MU, Han J, Wang K, Wang Q, Wu H, Mei G, Zang Y, Han Z, Xu C, Shen W, Yang D, Si Z, Dai F, Zou L, Huang F, Bai Y, Zhang Y, Brodt A, Ben-Hamo H, Zhu X, Zhou B, Guan X,
Zhu S, Chen X, Zhang T (2019) Gossypium barbadense and Gossypium hirsutum genomes provide insights into the origin and evolution of allotetraploid cotton. Nature Genetics 51: 739-748

Jia Y, Pan Z, He S, Gong W, Geng X, Pang B, Wang L, Du X (2018) Genetic diversity and population structure of Gossypium arboreum L. collected in China. Journal of Cotton Research 1: 11

Kalwa M, Hanzelmann S, Otto S, Kuo C, Franzen J, Joussen S, Fernandezrebollo E, Rath B, Koch CM, Hofmann A (2016) The lncRNA HOTAIR impacts on mesenchymal stem cells via triple helix formation. Nucleic Acids Research 44: 10631-10643

Kong L, Zhang Y, Ye Z-Q, Liu X-Q, Zhao S-Q, Wei L, Gao G (2007) CPC: assess the protein-coding potential of transcripts using sequence features and support vector machine. Nucleic acids research 35: W345-W349

Langmead B, Trapnell C, Pop M, Salzberg SL (2009) Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. Genome Biol 10: R25

Lee MR, Kim CS, Park T, Choi Y, Lee K (2018) Optimization of the ninhydrin reaction and development of a multiwell plate-based high-throughput proline detection assay. Analytical Biochemistry 556: 57-62

Li L, Eichten SR, Shimizu R, Petsch K, Yeh CT, Wu W, Chettoor AM, Givan SA, Cole RA, Fowler JE, Evans MM, Scanlon MJ, Yu J, Schnable PS, Timmermans MC, Springer NM, Muehlbauer GJ (2014) Genome-wide discovery and characterization of maize long non-coding RNAs. Genome Biology 15: R40

Li Y, Syed J, Sugiyama H (2016) RNA-DNA Triplex Formation by Long Noncoding RNAs. Cell Chemical Biology 23: 1325-1333
Long Y, Wang X, Youmans DT, Cech TR (2017) How do lncRNAs regulate transcription? Science Advances 3: eaao2110

Mano Y, Nemoto K (2012) The pathway of auxin biosynthesis in plants. Journal of Experimental Botany 63: 2853-2872

Mcclintock B (1984) The significance of responses of the genome to challenge. Science 226: 792-801

Mondal T, Subhash S, Vaid R, Enroth S, Uday S, Reinius B, Mitra S, Mohammed A, James AR, Hoberg E, Moustakas A, Gyllensten U, Jones SJ, Gustafsson CM, Sims AH, Westerlund F, Gorab E, Kanduri C (2015) MEG3 long noncoding RNA regulates the TGF-beta pathway genes through formation of RNA-DNA triplex structures. Nature Communications 6: 7743

Necsulea A, Soumillon M, Warnefors M, Liechti A, Daish T, Zeller U, Baker JC, Grutzner F, Kaessmann H (2014) The evolution of lncRNA repertoires and expression patterns in tetrapods. Nature 505: 635-640

Noguero M, Atif RM, Ochatt S, Thompson RD (2013) The role of the DNA-binding One Zinc Finger (DOF) transcription factor family in plants. Plant Science 209: 32-45

Oleary VB, Ovsepian SV, Carrascosa LG, Buske FA, Radulovic V, Niyazi M, Moertl S, Trau M, Atkinson MJ, Anastasov N (2015) PARTICLE, a Triplex-Forming Long ncRNA, Regulates Locus-Specific Methylation in Response to Low-Dose Irradiation. Cell Reports 11: 474-485

Pattison RJ, Csukasi F, Zheng Y, Fei Z, van der Knaap E, Catalá C (2015) Comprehensive Tissue-Specific Transcriptome Analysis Reveals Distinct Regulatory Programs during Early Tomato Fruit Development. Plant Physiology 168: 1684-1701
Pertea M, Kim D, Pertea GM, Leek JT, Salzberg SL (2016) Transcript-level expression analysis of RNA-seq experiments with HISAT, StringTie and Ballgown. Nature Protocols 11: 1650-1667

Postepska-Igielska A, Giwojna A, Gasri-Plotnitsky L, Schmitt N, Dold A, Ginsberg D, Grummt I (2015) LncRNA Khps1 Regulates Expression of the Proto-oncogene SPHK1 via Triplex-Mediated Changes in Chromatin Structure. Molecular Cell 60: 626-636

Qin T, Zhao H, Cui P, Albesher N, Xiong L (2017) A Nucleus-Localized Long Non-Coding RNA Enhances Drought and Salt Stress Tolerance. Plant Physiology 175: 1321-1336

Quinlan AR, Hall IM (2010) BEDTools: a flexible suite of utilities for comparing genomic features. Bioinformatics 26: 841-842

Rinn JL, Chang HY (2012) Genome regulation by long noncoding RNAs. Annual Review of Biochemistry 81: 145-166

Robinson MD, McCarthy DJ, Smyth GK (2010) edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. Bioinformatics 26: 139-140

Salehin M, Li B, Tang M, Katz E, Song L, Ecker JR, Kliebenstein DJ, Estelle M (2019) Auxin-sensitive Aux/IAA proteins mediate drought tolerance in Arabidopsis by regulating glucosinolate levels. Nature Communications 10: 4021

Sancenon V, Puig S, Mira H, Thiele DJ, Penarrubia L (2003) Identification of a copper transporter family in Arabidopsis thaliana. Plant Molecular Biology 51: 577-587

Semon M, Wolfe KH (2008) Preferential subfunctionalization of slow-evolving genes after allopolyploidization in Xenopus laevis. Proceedings of the National Academy of Sciences of the United States of America 105: 8333-8338
Song Q, Zhang T, Stelly DM, Chen ZJ (2017) Epigenomic and functional analyses reveal roles of epialleles in the loss of photoperiod sensitivity during domestication of allotetraploid cottons. Genome Biology 18: 99

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

Tamura K, Dudley JT, Nei M, Kumar S (2007) MEGA4: Molecular evolutionary genetics analysis (MEGA) software version 4.0. Molecular Biology and Evolution 24(8):1596–1599

Tiwari S, Lata C, Chauhan PS, Prasad V, Prasad M (2017) A Functional Genomic Perspective on Drought Signalling and its Crosstalk with Phytohormone-mediated Signalling Pathways in Plants. Current genomics 18: 469-482

Tognetti VB, Muhlenbock P, Van Breusegem F (2012) Stress homeostasis - the redox and auxin perspective. Plant Cell and Environment 35: 321-333

Wang K, Wendel JF, Hua J (2018) Designations for individual genomes and chromosomes in Gossypium. Journal of Cotton Research 1: 3 https://doi.org/10.1186/s42397-018-0002-1

Wang P, Shen L, Guo J, Jing W, Qu Y, Li W, Bi R, Xuan W, Zhang Q, Zhang W (2019) Phosphatidic Acid Directly Regulates PINOID-Dependent Phosphorylation and Activation of the PIN-FORMED2 Auxin Efflux Transporter in Response to Salt Stress. The Plant Cell 31: 250-271

Wendel JF, Grover CE (2015) Taxonomy and Evolution of the Cotton Genus, *Gossypium*. In: Cotton, 2, American Society of Agronomy, Inc., Crop Science Society of America, Inc., Soil Science Society of America, Inc., 1-20. https://doi.org/10.2134/agronmonogr57.2013.0020

Xing Y, Yao R, Zhang Y, Guo C, Jiang S, Xu G, Dong R, Yang L, Chen L (2017) SLERT Regulates DDX21 Rings Associated with Pol I Transcription. Cell 169: 664-678
Yang Y, Liu T, Shen D, Wang J, Ling X, Hu Z, Chen T, Hu J, Huang J, Yu W, 
Dou D, Wang MB, Zhang B (2019) Tomato yellow leaf curl virus intergenic 
siRNAs target a host long noncoding RNA to modulate disease symptoms. PLoS 
Pathogens 15: e1007534

Zhang L, Wang M, Li N, Wang H, Qiu P, Pei L, Xu Z, Wang T, Gao E, Liu J 
(2018) Long noncoding RNAs involve in resistance to Verticillium dahliae, a 
fungal disease in cotton. Plant Biotechnology Journal 16: 1172-1185

Zhang Q, Wang D, Lang Z, He L, Yang L, Zeng L, Li Y, Zhao C, Huang H, Zhang 
H, Zhang H, Zhu JK (2016) Methylation interactions in Arabidopsis hybrids 
require RNA-directed DNA methylation and are influenced by genetic variation. 
Proceedings of the National Academy of Sciences of the United States of America 
113: E4248-4256

Zhang T, Hu Y, Jiang W, Fang L, Guan X, Chen J, Zhang J, Saski CA, Scheffler 
BE, Stelly DM, Hulse-Kemp AM, Wan Q, Liu B, Liu C, Wang S, Pan M, 
Wang Y, Wang D, Ye W, Chang L, Zhang W, Song Q, Kirkbride RC, Chen 
X, Dennis E, Llewellyn DJ, Peterson DG, Thaxton P, Jones DC, Wang Q, Xu 
X, Zhang H, Wu H, Zhou L, Mei G, Chen S, Tian Y, Xiang D, Li X, Ding J, 
Zuo Q, Tao L, Liu Y, Li J, Lin Y, Hui Y, Cao Z, Cai C, Zhu X, Jiang Z, Zhou 
B, Guo W, Li R, Chen ZJ (2015) Sequencing of allotetraploid cotton 
(Gossypium hirsutum L. acc. TM-1) provides a resource for fiber improvement. 
Nature Biotechnology 33: 531-537

Zhang Y, Liu T, Meyer CA, Eeckhoute J, Johnson DS, Bernstein BE, Nusbaum C, 
Myers RM, Brown M, Li W, Liu XS (2008) Model-based analysis of ChIP-Seq 
(MACS). Genome Biology 9: R137

Zhao T, Tao X, Feng S, Wang L, Hong H, Ma W, Shang G, Guo S, He Y, Zhou B, 
Guan X (2018) LncRNAs in polyploid cotton interspecific hybrids are derived 
from transposon neo-functionalization. Genome Biology 19: 195
Zhao X, Li J, Lian B, Gu H, Li Y, Qi Y (2018) Global identification of Arabidopsis IncRNAs reveals the regulation of MAF4 by a natural antisense RNA. Nature Communications 9: 5056

Zhang L, Li Fu, Liu C, Zhang C, Wu Z (2010) Isolation and Analysis of Drought-related Gene from Cotton (Gossypium arboreum L.) SSH Library. Cotton Science 22: 110-114

Zhu P, Wang Y, Qin N, Wang F, Wang J, Deng XW, Zhu D (2016) Arabidopsis small nucleolar RNA monitors the efficient pre-rRNA processing during ribosome biogenesis. Proceedings of the National Academy of Sciences of the United States of America 113: 11967-11972
Figure 1. GhDAN1 is associated with the drought-tolerance trait in the cotton ancestor. (A) A simplified model system that mimics interspecific hybridization and WGD in the cotton lineage. (B) Drought-tolerance phenotypes of Ga diploid and Gh allotetraploid cotton (wild and cultivar accessions). (C and D) Drought-tolerance phenotype of GhDAN1-silenced plants and GhDAN1 expression. Error bars represent the SD of 15 seedlings in each group of treatment. ** p < 0.01, Student's t-test. Total GSH (E) and proline contents (F) before and after drought treatment. Error bars represent the SD of three biological replicates. * p < 0.05, Mann-Whitney U-test. (G) Syntenic analysis of DAN1 in diploid Ga and allotetraploid Gh genomes. (H and I) GhDAN1 expression in different tissues and the profile in response to multiple abiotic treatments. Error bars represent the SD of three biological replicates. * p < 0.05, Mann-Whitney U-test. (J to L) Stack plots for transcript profile (RNA-seq) (J), Pol II ChIP-seq profile (K) and the DNA CG-methylation status (L) of DAN1 loci in Ga (A2), (Ga x Gr)F1, and Gh (AD1) genomes.
Figure 2. **DAN1 is active in the allotetraploid cotton species.** (A) Schematic chart showing phylogenetic relationships of diploid ancestors and seven allotetraploid cotton species. (B) **DAN1 transcription activity** in two diploid A species and seven allotetraploid species (accessions are listed in order in Supplemental Table S1). Error bars represent the SE of three technical replicates. (C) Enrichment of H3K4me3 ChIP-seq signals (boxed region) near the **DAN1 transcription starting site** in the allotetraploid cotton lineage. Fold changes were calculated using normalized peak counts compared with their corresponding mock control. (D) Drought-tolerance phenotype of diploid and allotetraploid cotton cultivars. The number of unwilted leaves after water restriction was calculated (see the Supplemental Fig. S8 for details); Error bars represent the SD of 6 seedlings from each group. *p* value was indicated by comparing Ga (A1) diploid with Gh (AD1) and Gb (AD2) allotetraploid cultivars, respectively. Student’s *t*-test. (E) Analysis of the correlation between **DAN1 expression** and the unwilted leaf count from panel D.
Figure 3. \(GhDAN1\) forms RNA-DNA triplexes to anchor the AAAG DNA target site and regulate gene expression in trans. (A and B) RT-qPCR analysis of \(GhDAN1\) expression in the nucleus versus cytoplasm fractions, and in the nucleolus versus nucleoplasm. \(Histone3\) and \(U3\) snoRNA were used as controls. Error bars represent the SE of three technical replicates. (C) RNA-FISH assay showing the \(GhDAN1\) distribution in the nucleus. Bars= 5 \(\mu\)m. (D) Alignment of the Triplexator software-predicted \(GhDAN1\) triplex target sites (TTS) with ChlRP-established AG-rich motif and Dot TFs binding motifs. (E) Number and density of TTSs over the gene-associated \(GhDAN1\) peak summits and neighboring regions. The TTS motif was predicted by FIMO in the peaks (±250 bp from the peak summits) and the flanking control regions (500 bp upstream and downstream from the peaks). The density of the TTS motif was generated by ggplot2 in R using the distances between the predicted TTS motif and the peak summit. (F and G) EMSA assay showed triplex formation of \(GhDAN1\) RNA with DIG-labeled TTS DNA probes in vitro. The nucleotide in red color in Probe 2 indicated two mismatches compared with predicted TTS motif. The arrows indicated the RNA-DNA triplexes.
Figure 4. Genes with 

**GhDAN1** the triplex target site (TTS) motif were sensitive to silencing of **GhDAN1** and significantly enriched in the auxin response pathway. (A) The column chart shows the number of DEGs in **GhDAN1**-silenced plants and the distribution of genes in each upregulated and downregulated group that were predicted with or without the **GhDAN1** TTS motif in the 2-kb promoters using FIMO scanning (score > 10). (B) Occurrence of **GhDAN1** TTS motif in DEGs and the total coding genes. (C) GO analysis of upregulated and downregulated genes in **GhDAN1**-silenced plants showing significant enrichment in the auxin response pathway. (D) Heatmap showing the expression profile of the representative DEGs with the **GhDAN1** TTS motif. (E) Cotton seedlings were sensitive to drought stress after silencing of the **GhDAN1** downstream gene **THI1-2**. (F) Stack view of mRNA-seq on **THI1-2** transcription activity during dehydration treatment. The predicted **GhDAN1** TTS motif in the promoter was indicated.
Figure 5. Working model of *DAN1*-mediated drought-response dynamics.
Li L, Eichten SR, Shinizu R, Petsch K, Yeh CT, Wu W, Chettoor AM, Givan SA, Cole RA, Fowler JE, Evans MM, Scanlon MJ, Yu J, Schnable PS, Timmermans MC, Springer NM, Muehlbauer GJ (2014) Genome-wide discovery and characterization of maize long non-coding RNAs. Genome Biology 15: R40

Google Scholar: Author Only Title Only Author and Title

Li Y, Syed J, Sugiyama H (2016) RNA-DNA Triplex Formation by Long Noncoding RNAs. Cell Chemical Biology 23: 1325-1333

Google Scholar: Author Only Title Only Author and Title

Long Y, Wang X, Youmans DT, Cech TR (2017) How do IncRNAs regulate transcription? Science Advances 3: eaao2110

Mano Y, Nipomo K (2012) The pathway of auxin biosynthesis in plants. Journal of Experimental Botany 63: 2853-2872

Google Scholar: Author Only Title Only Author and Title

Mcclintock B (1984) The significance of responses of the genome to challenge. Science 226: 792-801

Google Scholar: Author Only Title Only Author and Title

Mondal T, Subhash S, Vaid R, Enroth S, Uday S, Reinius B, Mitra S, Mohammed A, James AR, Hoberg E, Moustakas A, Gyllensten U, Jones SJ, Gustafsson CM, Sims AH, Westerlund F, Gorab E, Kanduri C (2015) MEG3 long noncoding RNA regulates the TGF-beta pathway genes through formation of RNA-DNA triplex structures. Nature Communications 6: 7743

Google Scholar: Author Only Title Only Author and Title

Mecsulea A, Soumillon M, Warnefors M, Liechti A, Daish T, Zeller U, Baker JC, Grutzner F, Kaessmann H (2014) The evolution of lncRNA repertoires and expression patterns in tetrapods. Nature 505: 635-640

Google Scholar: Author Only Title Only Author and Title

Neculce A, Soumillon M, Warnefors M, Liechti A, Daish T, Zeller U, Baker JC, Grutzner F, Kaessmann H (2014) The evolution of lncRNA repertoires and expression patterns in tetrapods. Nature 505: 635-640

Google Scholar: Author Only Title Only Author and Title

Noguero M, Atif RM, Ochatt S, Thompson RD (2013) The role of the DNA-binding One Zinc Finger (DOF) transcription factor family in plants. Plant Science 209: 32-45

Google Scholar: Author Only Title Only Author and Title

Oleary VB, Ovsepian SV, Carrascosa LG, Buske FA, Radulovic V, Niyazi M, Moerlt S, Trau M, Atkinson MJ, Anastasov N (2015) PARTICLE, a Triplex-Forming Long ncRNA, Regulates Locus-Specific Methylation in Response to Low-Dose Irradiation. Cell Reports 11: 474-485

Google Scholar: Author Only Title Only Author and Title

Pattison RJ, Cusakai F, Zheng Y, Fei Z, van der Knaap E, Catala C (2015) Comprehensive Tissue-Specific Transcriptome Analysis Reveals Distinct Regulatory Programs during Early Tomato Fruit Development. Plant Physiology 168: 1684-1701

Google Scholar: Author Only Title Only Author and Title

Pertea M, Kim D, Pertea GM, Leek JT, Salzberg SL (2016) Transcript-level expression analysis of RNA-seq experiments with HISAT, StringTie and Ballgown. Nature Protocols 11: 1650-1667

Google Scholar: Author Only Title Only Author and Title

Postepska-Igielska A, Giwojna A, Gaspi-Plotnitsky L, Schmitt N, Dold A, Ginsberg D, Grummt I (2015) LncRNA Khps1 Regulates Expression of the Proto-oncogene SPHK1 via Triplex-Mediated Changes in Chromatin Structure. Molecular Cell 60: 626-636

Google Scholar: Author Only Title Only Author and Title

Qin T, Zhao H, Cui P, Alfesher N, Xiong L (2017) A Nucleus-Localized Long Non-Coding RNA Enhances Drought and Salt Stress Tolerance. Plant Physiology 175: 1321-1336

Google Scholar: Author Only Title Only Author and Title

Quinlan AR, Hall IM (2010) BEDtools: a flexible suite of utilities for comparing genomic features. Bioinformatics 26: 841-842

Google Scholar: Author Only Title Only Author and Title

Rinn JL, Chang HY (2012) Genome regulation by long noncoding RNAs. Annual Review of Biochemistry 81: 145-166

Google Scholar: Author Only Title Only Author and Title

Robinson MD, McCarthy DJ, Smyth GK (2010) edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. Bioinformatics 26: 139-140

Google Scholar: Author Only Title Only Author and Title

Salehin M, Li B, Tang M, Katz E, Song L, Ecker JR, Kliebenstein DJ, Estelle M (2019) Auxin-sensitive Aux/IAA proteins mediate drought tolerance in Arabidopsis by regulating glucosinolate levels. Nature Communications 10: 4021

Google Scholar: Author Only Title Only Author and Title

Sancenon V, Puig S, Mira H, Thiele DJ, Penarrubia L (2003) Identification of a copper transporter family in Arabidopsis thaliana. Plant Molecular Biology 51: 577-587

Google Scholar: Author Only Title Only Author and Title

Semon M, Wolfe KH (2008) Preferential subfunctionalization of slow-evolving genes after allopolyploidization in Xenopus laevis. Proceedings of the National Academy of Sciences of the United States of America 105: 8333-8338

Google Scholar: Author Only Title Only Author and Title

Song Q, Zhang T, Stelly DM, Chen ZJ (2017) Epigenomic and functional analyses reveal roles of epialleles in the loss of photoperiod sensitivity during domestication of allotetraploid cottons. Genome Biology 18: 99
Swiezewski S, Liu F, Magusin A, Dean C (2009) Cold-induced silencing by long antisense transcripts of an Arabidopsis Polycomb target. Nature 462: 799-802

Tamura K, Dudley JT, Nei M, Kumar S (2007) MEGA4: Molecular evolutionary genetics analysis (MEGA) software version 4.0. Molecular Biology and Evolution 24(8):1596–1599

Tiwari S, Lata C, Chauhan PS, Prasad V, Prasad M (2017) A Functional Genomic Perspective on Drought Signalling and its Crosstalk with Phytohormone-mediated Signalling Pathways in Plants. Current genomics 18: 469-482

Tognetti VB, Muhlenbock P, Van Breusegem F (2012) Stress homeostasis - the redox and auxin perspective. Plant Cell and Environment 35: 321-333

Wang K, Wendel JF, Hua J (2018) Designations for individual genomes and chromosomes in Gossypium. Journal of Cotton Research 1: 3 https://doi.org/10.1186/s42397-018-0002-1

Wang P, Shen L, Guo J, Jing W, Qu Y, Li W, Bi R, Xuan W, Zhang Q, Zhang W (2019) Phosphatidic Acid Directly Regulates PINOID-Dependent Phosphorylation and Activation of the PIN-FORMED2 Auxin Efflux Transporter in Response to Salt Stress. The Plant Cell 31: 250-271

Wendel JF, Grover CE (2015) Taxonomy and Evolution of the Cotton Genus, Gossypium. In: Cotton, 2, American Society of Agronomy, Inc., Crop Science Society of America, Inc., Soil Science Society of America, Inc., 1-20. https://doi.org/10.2134/agronmonogr57.2013.0020

Xing Y, Yao R, Zhang Y, Guo C, Jiang S, Xu G, Dong R, Yang L, Chen L (2017) SLERT Regulates DDX21 Rings Associated with Pol I Transcription. Cell 169: 664-678

Yang Y, Liu T, Shen D, Wang J, Ling X, Hu Z, Chen T, Hu J, Huang J, Yu W, Dou D, Wang MB, Zhang B (2019) Tomato yellow leaf curl virus intergenic siRNAs target a host long noncoding RNA to modulate disease symptoms. PLoS Pathogens 15: e1007534

Zhang L, Wang M, Li N, Wang H, Qiu P, Pei L, Xu Z, Wang T, Gao E, Liu J (2018) Long noncoding RNAs involve in resistance to Verticillium dahliae, a fungal disease in cotton. Plant Biotechnology Journal 16: 1172-1185

Zhang Q, Wang D, Lang Z, He L, Yang L, Zeng L, Li Y, Zhao C, Huang H, Zhang H, Zhang H, Zhu JK (2016) Methylation interactions in Arabidopsis hybrids require RNA-directed DNA methylation and are influenced by genetic variation. Proceedings of the National Academy of Sciences of the United States of America 113: E4248-4256

Zhang T, Hu Y, Jiang W, Fang L, Guan X, Chen J, Zhang J, Saski CA, Scheffler BE, Stelly DM, Hulse-Kemp AM, Wan Q, Liu B, Liu C, Wang S, Pan M, Wang Y, Wang D, Ye W, Chang L, Zhang W, Song Q, Kirkbride RC, Chen X, Dennis E, Llewellyn DJ, Peterson DG, Thaxton P, Jones DC, Wang Q, Xu X, Zhang H, Wu H, Zhou L, Mei G, Chen S, Tian Y, Xiang D, Li X, Ding J, Zuo Q, Tao L, Liu Y, Li J, Lin Y, Hui Y, Cao Z, Cai C, Zhu X, Jiang Z, Zhou B, Guo W, Li R, Chen ZJ (2015) Sequencing of allotetraploid cotton (Gossypium hirsutum L. acc. TM-1) provides a resource for fiber improvement. Nature Biotechnology 33: 531-537

Zhang Y, Liu T, Meyer CA, Eckhoute J, Johnson DS, Bernstein BE, Nusbaum C, Myers RM, Brown M, Li W, Liu XS (2008) Model-based analysis of ChIP-Seq (MACS). Genome Biology 9: R137

Zhao T, Tao X, Feng S, Wang L, Hong H, Ma W, Shang G, Guo S, He Y, Zhou B, Guan X (2018) LncRNAs in polyplid cotton interspecific hybrids are derived from transposon neo-functionalization. Genome Biology 19: 195

Zhao X, Li J, Lian B, Gu H, Li Y, Qi Y (2018) Global identification of Arabidopsis IncRNAs reveals the regulation of MAF4 by a natural antisense RNA. Nature Communications 9: 5056

Zhu P, Wang Y, Qin N, Wang F, Wang J, Deng XW, Zhu D (2016) Arabidopsis small nucleolar RNA monitors the efficient pre-rRNA processing during ribosome biogenesis. Proceedings of the National Academy of Sciences of the United States of America 113: 11967-
