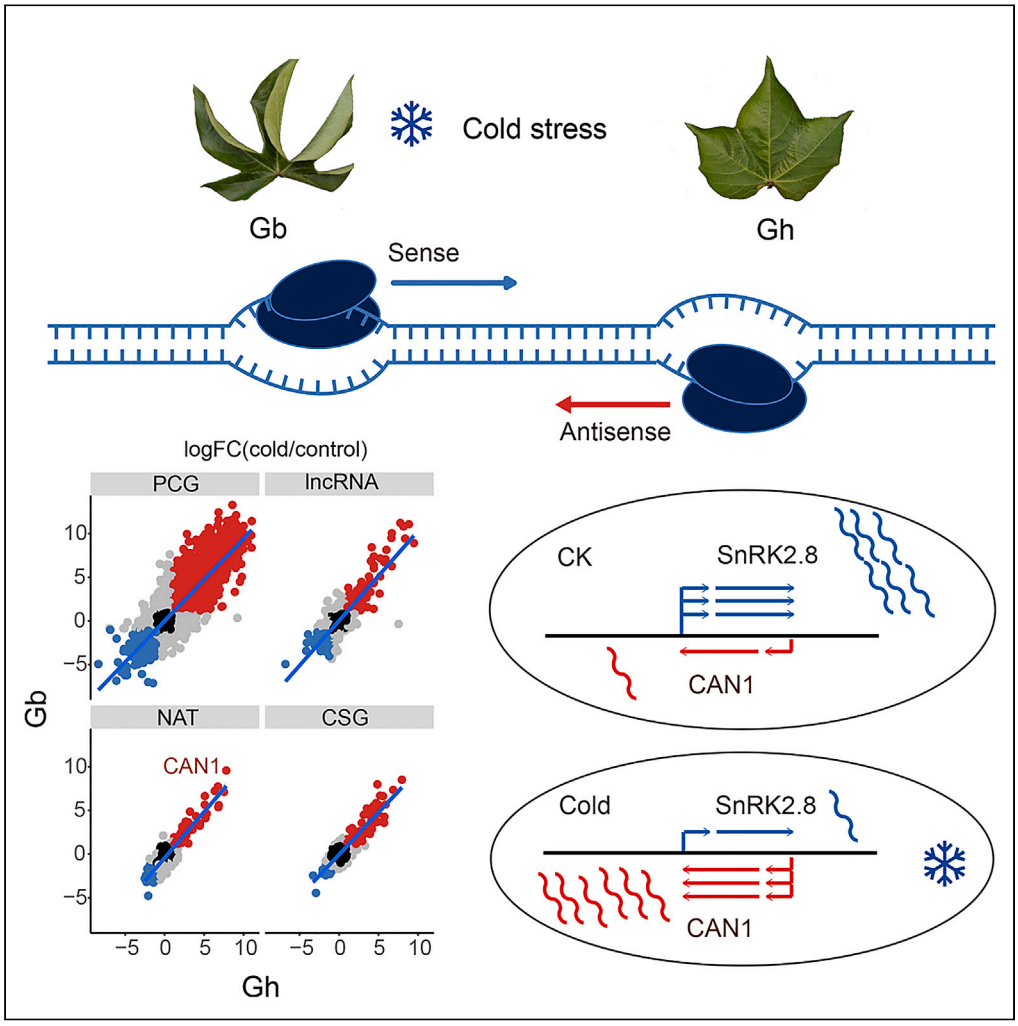
iScience

CellPress
OPEN ACCESS

Article

Global identification of natural antisense transcripts in *Gossypium hirsutum* and *Gossypium barbadense* under chilling stress



Shouli Feng, Xuan Long, Mengtao Gao, Yongyan Zhao, Xueying Guan

xueyingguan@zju.edu.cn

Highlights

Natural antisense transcripts are profiled for two cotton species under chilling

A conservative NAT pattern between *G. hirsutum* and *G. barbadense* was established

NAT expression is positively correlated with CSG expression

CAN1 is potentially involved in regulating chilling tolerance in cotton

Feng et al., iScience *26*, 107362
August 18, 2023
https://doi.org/10.1016/j.isci.2023.107362

iScience

CellPress
OPEN ACCESS

Article

Global identification of natural antisense transcripts in *Gossypium hirsutum* and *Gossypium barbadense* under chilling stress

Shouli Feng,[1,5] Xuan Long,[1] Mengtao Gao,[2] Yongyan Zhao,[1,3] and Xueying Guan[1,3,4,6,*]

SUMMARY

Natural antisense transcripts (NATs) in model plants have been recognized as important regulators of gene expression under abiotic stresses. However, the functional roles of NATs in crops under low temperature are still unclear. Here, we identified 815 and 689 NATs from leaves of *Gossypium hirsutum* and *G. barbadense* under chilling stress. Among those, 224 NATs were identified as interspecific homologs between the two species. The correlation coefficients for expression of NATs and their cognate sense genes (CSG) were 0.43 and 0.37 in *G. hirsutum* and *G. barbadense*, respectively. Furthermore, expression of interspecific NATs and CSGs alike was highly consistent under chilling stress with correlation coefficients of 0.90–0.91. Four cold-associated NATs were selected for functional validation using virus-induced gene silencing (VIGS). Our results suggest that *CAN1* engage in the molecular regulation of chilling stress by regulating *SnRK2.8* expression. This highly conserved NAT have valuable potential for applications in breeding cold-tolerant cotton.

INTRODUCTION

Cotton (*Gossypium* spp.) originated from tropical and subtropical regions, is an important cash crop throughout the world, and provides natural and renewable raw materials for the textile industry. Suitable temperatures for cotton growth and development lie in the range of about 22°C–32°C,[1–3] and temperatures below 10°C will cause cold damage. Two important cultivated tetraploid species, *Gossypium hirsutum* L. (AD_1) and *G. barbadense* L. (AD_2), are currently in widespread use for global cotton farming. After a long period of domestication and improvement, cotton has gradually adapted to diversified climates and is grown in the Yellow River valley, Yangtze River valley, northwest, south, and even northeast China (Figure S1A). However, both *G. hirsutum* and *G. barbadense* are chilling sensitive crops, and *G. barbadense* was more sensitive to low temperature stress.[4] In addition, the Xinjiang region accounts for more than 80% of the total cotton produced in China. Cotton in Xinjiang is usually sown in mid-April and covered with mulch to keep it warm, with the seedlings emerging from the mulch in early May. From 1951 to 2021, the daily minimum temperature in Xinjiang was lower than 10°C on most days in April, and also on some days in May (Figure S1B). The cold damage that occurs in the Xinjiang spring incurs great losses with regard to cotton quality and yield.[5] Due to the relatively short history of their cultivation and domestication in cold regions, there is urgent need to improve the cold tolerance of tropical- and subtropical-originating crops such as cotton, adapting them to the wider modern cultivation area and increasingly severe global climate anomalies. Understanding the molecular mechanism of cold tolerance in cotton seedlings is an important basis for breeding new varieties with cold tolerance, and also an important strategy to rescue the losses resulting from cold damage.

Plants have evolved sophisticated molecular regulatory mechanisms to respond and adapt to low temperature. Under low temperature, plant genes encoding transcription factors such as C-repeat (CRT)-binding factors (CBFs) are rapidly induced; their encoded proteins then directly bind to CRT/DRE *cis*-elements and activate a set of cold-regulated (*COR*) genes, leading to the cold tolerance response.[6,7] Recently, long non-coding RNAs have emerged as new regulation components associated with abiotic and biotic stresses in plant genomes.[8] Long non-coding genes have high interspecific specificity and are able to rapidly generate new functions, which is the basis of rapid environmental adaptation in new species.[9,10] Natural antisense transcripts (NATs) comprise a unique type of lncRNA, being usually transcribed from the antisense strands

[1]Zhejiang Provincial Key Laboratory of Crop Genetic Resources, The Advanced Seed Institute, Plant Precision Breeding Academy, College of Agriculture and Biotechnology, Zhejiang University, Hangzhou 300058, China

[2]State Key Laboratory of Crop Genetics and Germplasm Enhancement, College of Agriculture, Nanjing Agricultural University, Nanjing, Jiangsu 210095, China

[3]Hainan Institute of Zhejiang University, Building 11, Yonyou Industrial Park, Yazhou Bay Science and Technology City, Yazhou District, Sanya, Hainan 572025, China

[4]Hainan Yazhou Bay Seed Lab, Yazhou Bay Science and Technology City, Yazhou District, Sanya, Hainan 572025, China

[5]Xianghu Laboratory, Hangzhou 311231, China

[6]Lead contact

*Correspondence: xueyingguan@zju.edu.cn

https://doi.org/10.1016/j.isci.2023.107362

of protein-coding genes (PCGs). These transcripts enact *cis*-regulation on their cognate sense genes (CSGs) through several mechanisms and so directly regulate corresponding biological functions.[11–13] When the CSG in question is a stress-regulated gene, production of the NAT helps the organism rapidly develop environmental adaptation. Previous studies have shown that NATs are induced by low temperature and play an important role in plant growth, development, and stress responses. For example, cold-induced *FLC* antisense transcripts have an early role in the epigenetic silencing of *FLC*, and their induction predates and is independent of other vernalization markers.[14] MAS generated from the antisense strand of the *MADS AFFECTING FLOWERING4* (*MAF4*) locus serves to activate *MAF4* by recruiting WDR5a, the core component of COMPASS-like complexes, to enhance histone 3 lysine 4 trimethylation (H3K4me3) on MAF4.[15] Moreover, *MAS* and *MAF4* are both activated by low temperature, which has the effect of prohibiting premature flowering.[15] *SVALKA* is a cold-responsive lncRNA in *Arabidopsis svalka*, which exhibits decreased cold tolerance. It is transcribed from the antisense strand between *CBF3* and *CBF1*, primarily from the *CBF1*-proximal promoter, and hence inhibits transcription of *CBF1* through RNA transcriptase II collision.[16] *DgTCP1* and the associated NAT *DglncTCP1* respond to low temperature and increase cold tolerance in chrysanthemum.[17] *DglncTCP1* acts as a scaffold to recruit the histone methyltransferase DgATX to *DgTCP1* to enhance H3K4me3 level and thus activate *DgTCP1* expression.[17] These reported case studies all support that NATs are actively involved with cold stress regulation in the plant kingdom. Therefore, it is reasonable to analyze the molecular function of NATs in cotton seedlings under chilling stress with the goal of ultimately improving cotton cold tolerance and guiding the breeding of cold-tolerant varieties of cotton.

It is generally believed that genes with interspecific conservation have important functions. Compared with PCGs, NATs, and lncRNAs evidence rapid evolution in their sequences and poor interspecific conservation. Between humans and mice, a genome-wide analysis identified 313 conserved NATs, which accounted for only 7.99%–10.30% of the total NATs identified[18]; in contrast, the proportion of homologous coding genes conserved between humans and mice is about 80%.[19] In addition, only eighteen human, ten mouse, and four rat NATs are conserved in chicken,[18] indicating low interspecific conservation of NATs across vertebrates. For comparison, among long non-coding RNAs in rice and maize, the proportion of conserved lincRNAs and NATs was about 5% and 3.5%, respectively.[20]

Whether NATs have a certain degree of interspecific transcriptional conservation is the main basis for evaluating the importance of their function, and such conservation analysis can promote our understanding in the following respects: (1) potential biological functions, (2) mode in gene regulation, and (3) clues to origin and evolution. Therefore, interspecific comparative genomic analyses among species with relatively close evolutionary relationships can assist in identifying functional NATs. *G. hirsutum* and *G. barbadense* are two allotetraploid cotton species that were formed by the hybridization and polyploidization of A and D diploid ancestors, with an estimated divergence time of 0.4–0.6 million years ago (MYA).[4] Comparative genomic analysis has indicated high collinearity and homology of PCGs between *G. hirsutum* and *G. barbadense*,[4,21] which provides a reliable basis for the definition and screening of NATs conserved between these species.

In this study, in order to explore the interspecific conservation of NATs associated with cold stress in cotton, we employed strand-specific RNA sequencing (ssRNA-Seq) to systematically identify NATs in *G. hirsutum* and *G. barbadense*. In total, we characterized 815 and 689 NATs from *G. hirsutum* and *G. barbadense*, respectively. Of those, 224 were interspecific homologs. We then selected four interspecific homologous NATs and corresponding CSGs for validation of their functions under chilling stress at seedling stage using virus-mediated gene silencing (VIGS) technology.

RESULTS

Identification of natural antisense transcripts (NATs) in cotton

To identify NATs at the genome-wide level in cotton, we reconstructed *G. hirsutum* and *G. barbadense* transcriptomes using high-depth strand-specific RNA sequencing technology. cDNA libraries were constructed for rRNA-depleted RNAs prepared from *G. hirsutum* and *G. barbadense* leaves under favorable (28°C) or chill stress (4°C) conditions. A total of 305.92 million (M) and 320.08 M clean reads was obtained from *G. hirsutum* and *G. barbadense*, respectively (Table S1). Among those, 279.23 M clean reads from *G. hirsutum* aligned to 82,144 genomic loci, and 284.63 M clean reads from *G. barbadense* aligned to 83,366 genomic loci. Based on transcript abundance, there was high consistency between biological

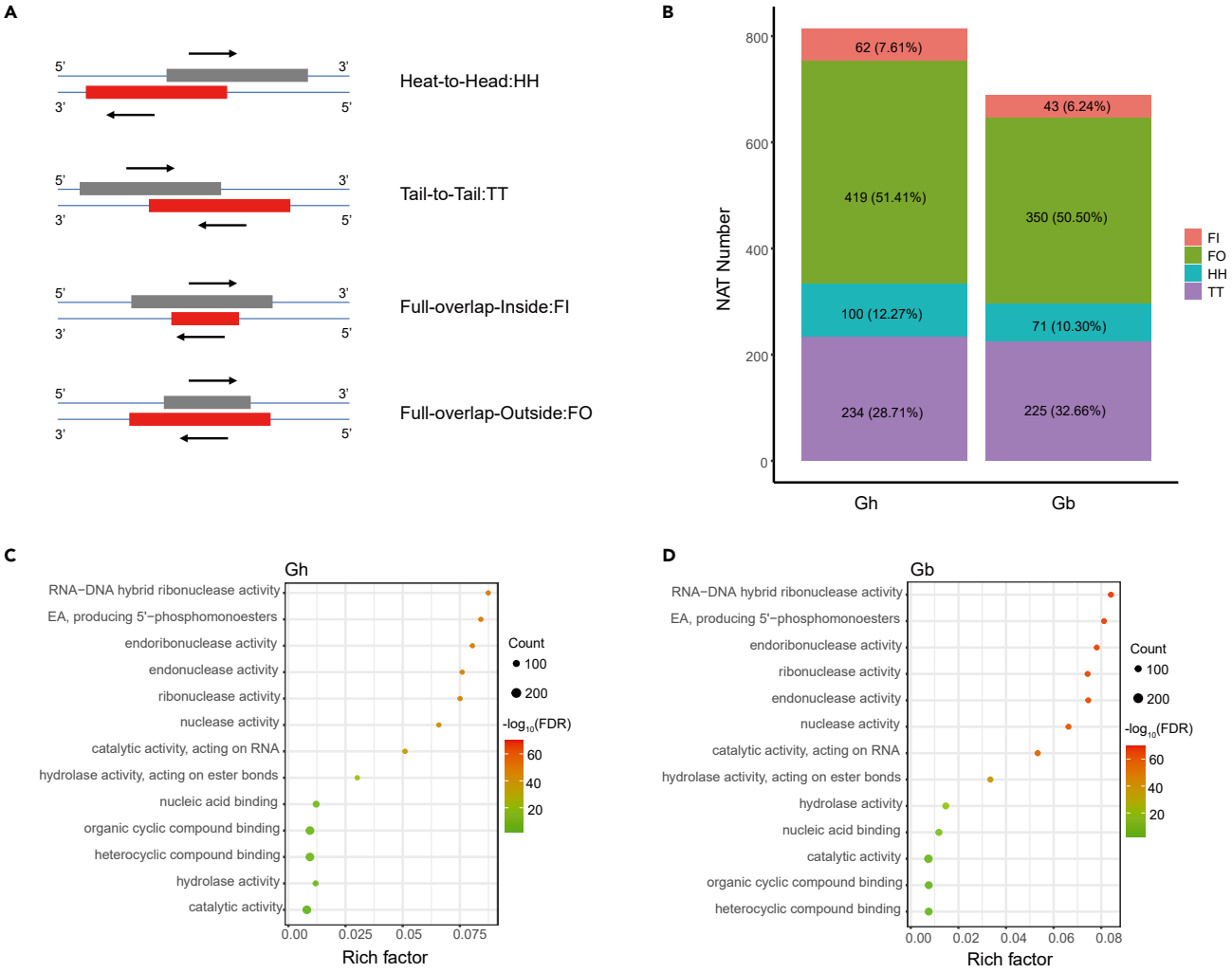


Figure 1. Natural antisense transcripts (NATs) from *G. hirsutum* and *G. barbadense*

(A) Schematic diagram of the four classes of NATs.

(B) The number and proportion of NATs in each class.

(C and D) GO terms enriched among NATs detected from *G. hirsutum* and *G. barbadense*.

replicates (Figure S2). After removing low-abundance transcripts (TPM_{MAX} < 1), 38,054 PCGs, 3,596 lncRNAs, and 815 NATs were identified in *G. hirsutum*; meanwhile, 36,461 PCGs, 2,839 lncRNAs, and 689 NATs were obtained for *G. barbadense*. The NATs were further classified into four categories: Head-to-Head (HH), Tail-to-Tail (TT), Full-overlap-Inside (FI) and Full-overlap-Outside (FO) (Figure 1A). The most numerous type was FO, with 419 (51.41%) NATs in *G. hirsutum* and 350 (50.50%) in *G. barbadense* (Figure 1B). Meanwhile, FI was the least represented at 62 (7.61%) and 43 (6.24%) NATs, respectively. The features of the identified NATs and PCGs were further evaluated in terms of average size (nucleotide, nt) and exon number. TT and FO NATs in *G. hirsutum* were significantly longer than PCGs and lncRNAs, with mean lengths of 2,209 nt and 2,708 nt versus the 1,705 nt of coding genes and 1,049 nt of lncRNAs (p value < 0.001, Mann-Whitney U-Test) (Figure S3). Similarly, TT and FO NATs in *G. barbadense* had mean lengths of 2,242 nt and 2,460 nt versus the 1,648 nt of coding genes and 1,010 nt of lncRNAs (p value < 0.001, Mann-Whitney U-Test) (Figure S2). Meanwhile, HH and FI NATs in *G. hirsutum* averaged 1,835 and 1,773 nt in length, which was not significantly different from PCGs. In *G. barbadense*, the average length of HH NATs was 2,267 nt, much longer than coding genes (p value < 0.01, Mann-Whitney U-Test), but the mean length of FI transcripts was 1,626, not significantly different from coding genes (Figure S3). NATs had fewer exons than coding genes (p value < 0.001, Mann-Whitney U-Test) (Figure S3), and were not evenly distributed across chromosomes. Chromosome

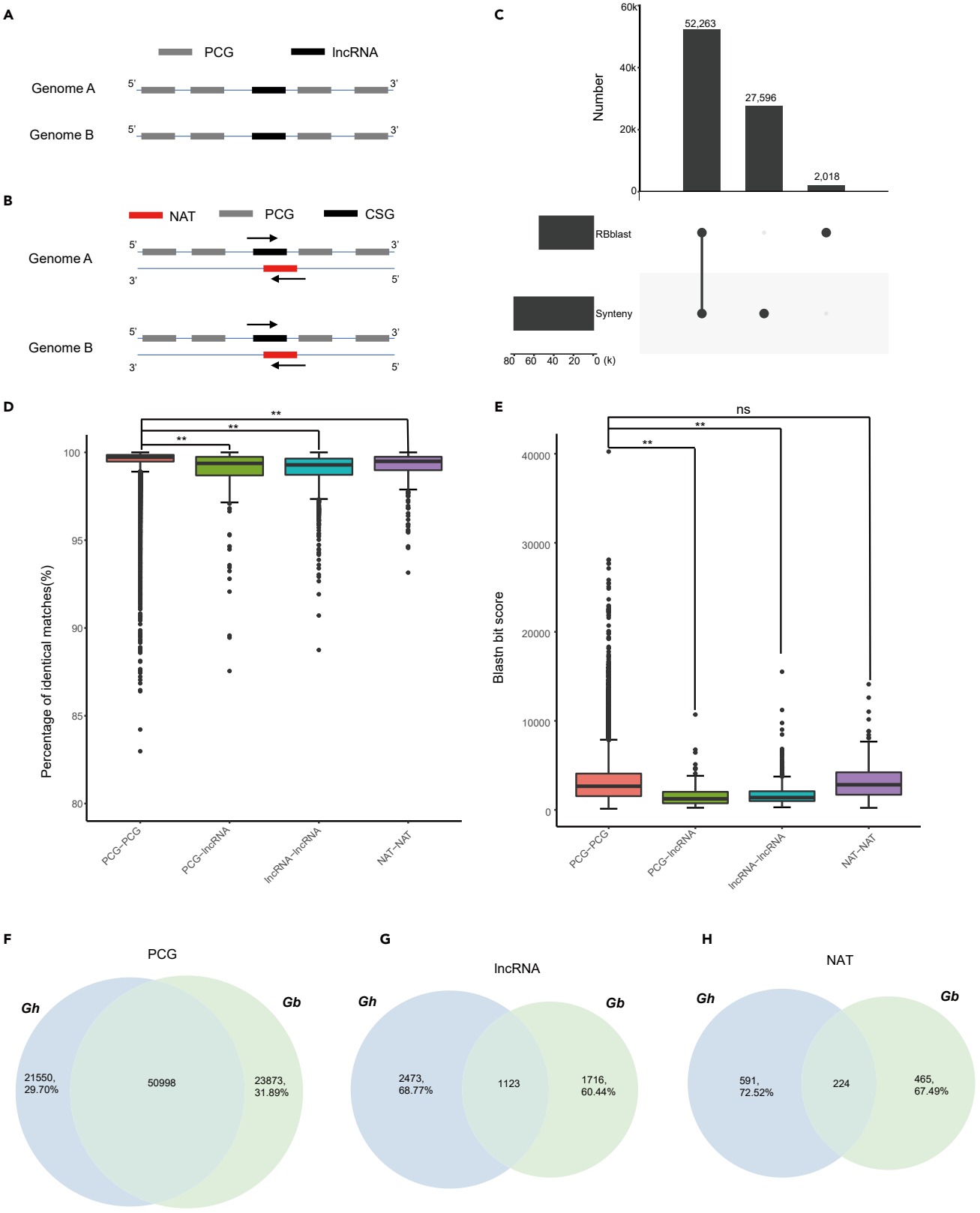

A
PCG
lncRNA
Genome A
Genome B
5'
3'
B
NAT
PCG
CSG
Genome A
Genome B
C
Number
60k
40k
20k
0
52,263
27,596
2,018
RBblast
Synteny
80 60 40 20 0 (k)
D
Percentage of identical matches(%)
100
95
90
85
80
**
PCG-PCG
PCG-lncRNA
lncRNA-lncRNA
NAT-NAT
E
Blastn bit score
40000
30000
20000
10000
0
ns
**
F
PCG
Gh
Gb
21550,
29.70%
50998
23873,
31.89%
G
lncRNA
2473,
68.77%
1123
1716,
60.44%
H
NAT
591,
72.52%
224
465,
67.49%

Figure 2. Homologous NATs between *G. hirsutum* and *G. barbadense*

(A) The schematic of identify homologous genes/lncRNAs.

(B) The schematic of identify homologous NATs.

(C) Upset plot displaying synteny and RBblast gene overlap.

(D) Percentage of match bases within homologous sequences.

(E) Blast bit scores of homologous sequences.

(F–H) Venn plot indicating sharing of PCGs, lncRNAs, and NATs between *G. hirsutum* and *G. barbadense*. Wilcoxon test: **, p value < 0.01.

A04 harbored the fewest NATs in both *G. hirsutum* and *G. barbadense*, with 14 and 7, respectively, and A05 the most, at 62 and 45, respectively (Figure S4).

To investigate the putative function of NATs, we performed Gene Ontology (GO) enrichment analysis on the CSGs of all respective NATs in *G. hirsutum* and *G. barbadense*. Interestingly, similar GO terms were enriched in both species, such as "RNA-DNA hybrid ribonuclease activity," "endoribonuclease activity, producing 5′-phosphomonoesters," and "endoribonuclease activity" (Figures 1C and 1D). We next performed GO enrichment analysis on each of the four NAT classes. Surprisingly, we found only the FO class to exhibit enrichment in *G. hirsutum,* namely for the term "RNA-DNA hybrid ribonuclease activity," while all four classes exhibited similar enrichments in *G. barbadense* (Figure S5). RNA-DNA hybrids can form from RNA transcripts. In eukaryotes, nascent RNA transcripts are quickly exported out of nuclei, otherwise RNA-DNA hybrid complexes will accumulate and then impair genome stability.[22] The similarity of GO enrichment results in both cotton species indicates that NATs is not transcribed noise, they may play important roles in removing RNA-DNA hybrid complexes to protect transcript and genome stability.

Conserved NATs identified between *G. hirsutum* and *G. barbadense*

NATs are transcribed from the opposite strands of PCGs or lncRNAs. To identify homologous NATs between *G. hirsutum* and *G. barbadense*, we first identified interspecific homologous PCGs and lncRNAs based on synteny and reciprocal best blast (Figures 2A, 2B, and S6A). This yielded 79,859 synteny anchors and 54,281 reciprocal best hits, of which 52,263 overlapped pairs were identified as homologous PCGs or lncRNAs (Figure 2C). These included 50,998 (97.58%), 1,123 (2.15%), and 142 (0.27%) homologous sequences categorized as PCG-PCG, lncRNA-lncRNA, and PCG-lncRNA, respectively (Figure S6B). The median percentage of identical matches was 99.72, 99.29, and 99.37 for PCG-PCG, lncRNA-lncRNA, and PCG-lncRNA pairs, respectively (Figure 2D). In addition, the median blastn bit score was 2,654, 783, and 1,244 for PCG-PCG, lncRNA-lncRNA, and PCG-lncRNA pairs, respectively (Figure 2E). These results indicated high sequence similarity between homologous PCGs or lncRNAs, and constituted a reliable basis for subsequent identification of interspecific homologous NATs.

NATs corresponding to homologous genes were defined as homologous NATs. In total, 224 homologous NATs were identified between *G. hirsutum* and *G. barbadense* (Figures 2H and S6C). The sequence similarity of these NATs was compared as aforementioned, with median percentage of identical matches and blastn bit score 99.49 and 2,823, respectively (Figures 2D and 2E; Table S2). The sequence similarity of homologous NATs between *G. hirsutum* and *G. barbadense* was comparable to that of PCGs, but higher than lncRNAs (Figure 2E). We further compared the proportions of homologous PCGs, lncRNAs, and NATs, and found that about 70% of PCGs are interspecific homologs (Figure 2F), while proportion of homologous lncRNAs or NATs was significantly less at about 31.23%–39.56% (Figure 2G) and 27.48%–32.51% (Figure 2H), respectively. This indicates that there is less interspecific conservation of both lncRNAs and NATs. With regard to categorization, the homologous NATs comprised 122, 68, 17, and 17 FO, TT, HH, and FI NATs, respectively. In addition, five NATs were transcribed from the antisense strands of lncRNAs, and 219 from the antisense of PCGs (Figure S6C). GO enrichment analysis of homologous NAT CSGs unsurprisingly revealed enriched terms to include "RNA-DNA hybrid ribonuclease activity" and "endoribonuclease activity, producing 5′-phosphomonoesters" (Figure S7A). The specific NATs from *G. hirsutum* and *G. barbadense* were also enriched in the same GO terms, indicating that conserved NATs do not collectively perform a unique function (Figures S7B and S7C).

Conserved NATs identified between subgenomes in *G. arboreum*, *G. hirsutum*, and *G. barbadense*

Examining NATs by subgenome revealed 49 NATs as conserved between A_T (A subgenome in tetraploid) and D_T (D subgenome in tetraploid) in *G. hirsutum* (Figure S8A) and 39 in *G. barbadense* (Figure S8B). These subgenome-conserved NATs comprised about 8.19%–11.54% of all identified NATs (Figures S8A

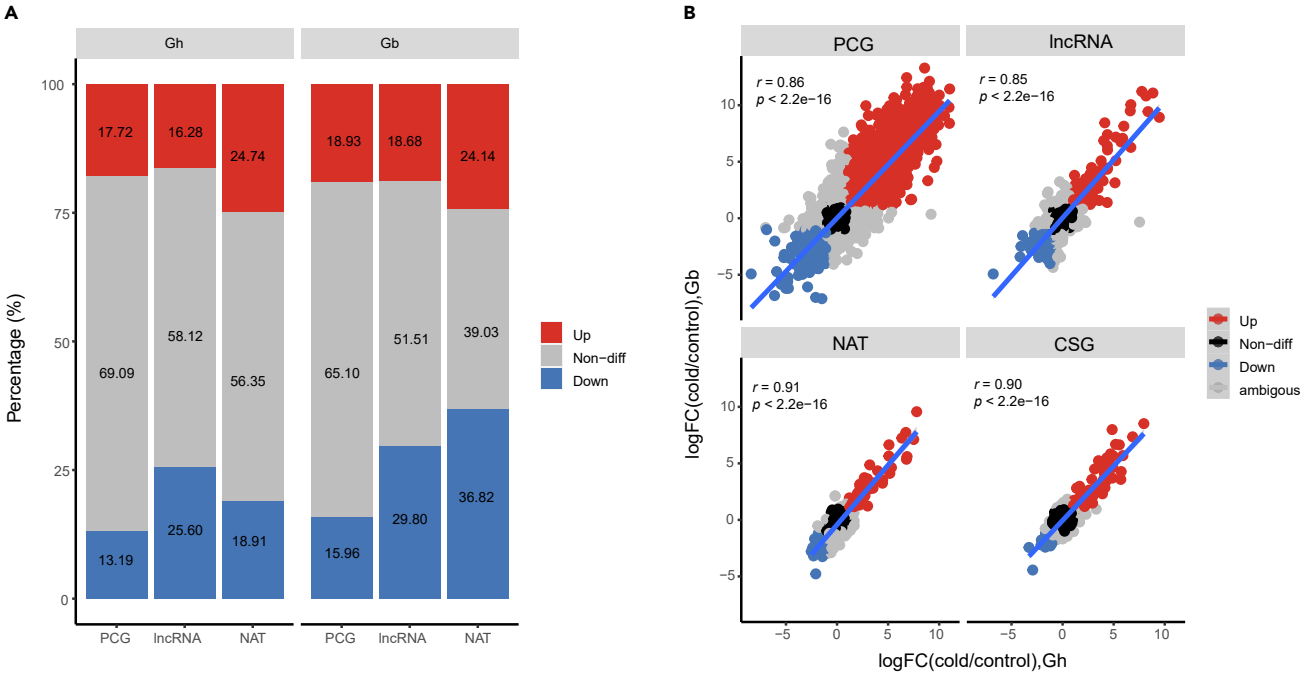


Figure 3. Conservation of the expression trends of homologous NATs and CSGs between *G. hirsutum* and *G. barbadense*

(A) Percentages of PCGs, lncRNAs, and NATs that are differentially expressed in *G. hirsutum* and *G. barbadense* during cold stress.

(B) Correlation and regression analysis of logFC(cold/control) for PCGs, lncRNAs, NATs, and CSGs between *G. hirsutum* and *G. barbadense*. The blue line is the regression curve; the shading indicates 95% confidence intervals. The *R* value is the correlation coefficient, both the *r* and p value were obtained from Pearson correlation test.

and S8B), which was significantly less than the proportion conserved between *G. hirsutum* and *G. barbadense* (~27.48%–32.51%). To further assess how many *G. hirsutum* and *G. barbadense* conserved NATs were inherited from ancestral species, we determined those between *G. arboreum* and the A_T subgenome of each tetraploid species. This revealed 43 conserved NATs between *G. arboreum* and *G. hirsutum* A_T and 37 between *G. arboreum* and *G. barbadense* A_T (Figures S8C–S8E), comprising about 9.75%–15.87% of all identified NATs (Figure S8).

Conserved NATs of *G. hirsutum* and *G. barbadense* have similar expression patterns under chilling stress

The differential expression of PCGs, lncRNAs, and NATs was evaluated under chilling stress. In *G. hirsutum*, a total of 7,948(17.72%)/669(16.28%)/259(24.74%) up-regulated and 5,917(13.19%)/1052 (25.6%)/198(18.91%) down-regulated PCGs, lncRNAs, and NATs were identified, respectively (Figure 3A; Table S3). The corresponding totals in *G. barbadense* were 8,574(18.93%)/667(18.68%)/240(24.14%) up-regulated and 7,229(15.96%)/1064(29.80%)/366(36.82%) down-regulated PCGs, lncRNAs, and NATs, respectively (Figure 3A). The proportion of up-regulated PCGs was similar to that of lncRNAs in both *G. hirsutum* and *G. barbadense*, while the proportion of up-regulated NATs was higher. Similarly, the proportion of down-regulated lncRNAs and NATs was higher than that of PCGs (Table 1). To validate this observation, we performed 1000 bootstrap re-samples of random PCGs with the same sample size as NATs, and found that the proportion of up-regulated and down-regulated NATs was significantly higher than randomly selected PCGs sets (Table 1). Moreover, we investigated the response intensity of CSGs and NAT under cold stress, and found that the response intensity of up-regulated and down-regulated NATs were higher than that of CSGs (Figure S9). These results suggested that NAT expression is more responsive than PCG expression under chilling stress. Similarly, it has been reported that NAT is more sensitive to drought response in maize.[23] Notably the proportion of up-regulated NATs was similar in *G. hirsutum* and *G. barbadense*, while *G. barbadense* evidenced a much greater proportion of down-regulated NATs than did *G. hirsutum* (Figure 3A). This was mainly driven by the FO category, which featured many more down-regulated NATs in *G. barbadense* (Figure S10).

Table 1. The number and ratio of cold responsive transcripts in cotton

	TM-1		H7124	
Category	Up-regulated	Down-regulated	Up-regulated	Down-regulated
Sig.PCGs	7948(17.72)	5917(13.19)	8574(18.93)	7229(15.96)
Sig.lncRNA	669(16.28)	1052(25.6)	667(18.68)	1064(29.8)
Sig.CSGs	137(16.81)	58(7.12)	88(12.77)	76(11.03)
Sig.NATs	259(24.74)	198(18.91)	240(24.14)	366(36.82)
p value (χ^2 test)	6.05×10^{-9}	9.5×10^{-8}	4.17×10^{-5}	1.03×10^{-68}
Sig.bootstrap	144.49(17.73)	107.56(13.20)	130.78(18.98)	109.99(15.96)
p value (t-test)	0	0	0	0

The ratio of cold responsive transcripts is in parentheses (%). Sig.: Significantly responded to cold stress. *p* value (χ^2 test): p value of χ^2 test for assessing differences between significantly up- or down-regulated PCGs and NATs s identified in cotton. p value (t-test): p value of t-test for assessing differences between significantly up- or down-regulated PCGs of bootstrap samples compared with NATs identified in cotton.

To compare whether the response patterns of homologous NATs under chilling stress were similar between *G. hirsutum* and *G. barbadense*, we counted up the co-differentially expressed PCGs, lncRNAs, NATs, and CSGs and performed Pearson correlation and linear regression analysis of $\log_2$(Fold Change) (logFC) between *G. hirsutum* and *G. barbadense*. Of the 33,090 expressed homologous PCGs, 4,226, 19,194, and 2,457 were up-regulated, non-different, and down-regulated in both species, while only 19 and 32 homologous PCGs displayed opposite differential expression patterns (Figure 3B). Among the 1,073 expressed homologous lncRNAs, 111, 491, and 196 were up-regulated, non-different, and down-regulated in both species, while only four homologous lncRNAs displayed opposing differential expression patterns (Figure 3B). Of the 224 expressed homologous NATs, 37, 69, and 41 were up-regulated, non-different, and down-regulated in both species, and no homologous NATs were found to display opposite differential expression patterns (Figure 3B). The logFC correlation coefficients of PCGs, lncRNAs, NATs, and CSGs between *G. hirsutum* and *G. barbadense* were 0.86, 0.85, 0.91, and 0.90, respectively (p value $< 2.2 \times 10^{-16}$) (Figure 3B). These results indicated that homologous PCGs, lncRNAs, and NATs of *G. hirsutum* and *G. barbadense* have similar expression patterns under chilling stress. The similarity in expression patterns between *G. hirsutum* and *G. barbadense* could either have originated from diploid ancestor species, or it may have arisen following polyploidization and subsequently been retained during species divergence. That is, these conserved NATs in both *G. hirsutum* and *G. barbadense* have maintained similar expression patterns over millions of years of natural selection, indicating their potential significance in imparting cotton resistance to cold stress.

The expression of NATs is positively correlated with their cognate sense genes

To explore the gene regulation function of NATs, we first examined the expression correlation between NATs and their CSGs. The correlation coefficient between logFC of NATs and logFC of CSGs was 0.43 (p value $< 2.2 \times 10^{-16}$) in *G. hirsutum* and 0.37 (p value $< 2.1 \times 10^{-13}$) in *G. barbadense*, which indicated that NAT and CSG expression tend to a positive correlation under chilling stress (Figures 4A and 4B). CSGs were then divided into three classes based on the expression pattern of NATs: Up, Down, and Non-diff. Relative to the Non-diff genes, the logFC of Up class CSGs was significantly higher, while the logFC of Down class CSGs was significantly lower in both *G. hirsutum* and *G. barbadense* (Figure 4C). However, individual NAT classes varied in their adherence to this overall pattern. In *G. hirsutum*, the respective correlation coefficients of HH, TT, FI, and FO NATs were 0.41, −0.02, 0.45, and 0.61, and the corresponding p values were 0.0007, 0.80, 0.0015, and $< 2.2 \times 10^{-16}$ (Figure S11). Similarly, the correlation coefficients of HH, TT, FI, and FO NATs in *G. barbadense* were 0.38, 0.14, 0.32, and 0.58, and the p values were 0.009, 0.09, 0.06, and 2.73×10^{-12}, respectively (Figure S12). These results suggested that FO NATs and their paired genes have consistent expression trends in *G. hirsutum* and *G. barbadense*, while TT NATs have relatively diversified expression.

NATs are involved in chill-tolerance regulation

To validate the function of the characterized NATs in chill-tolerance regulation, we selected conserved NATs based on expression levels and differentially expressed NATs and CSGs. This yielded four

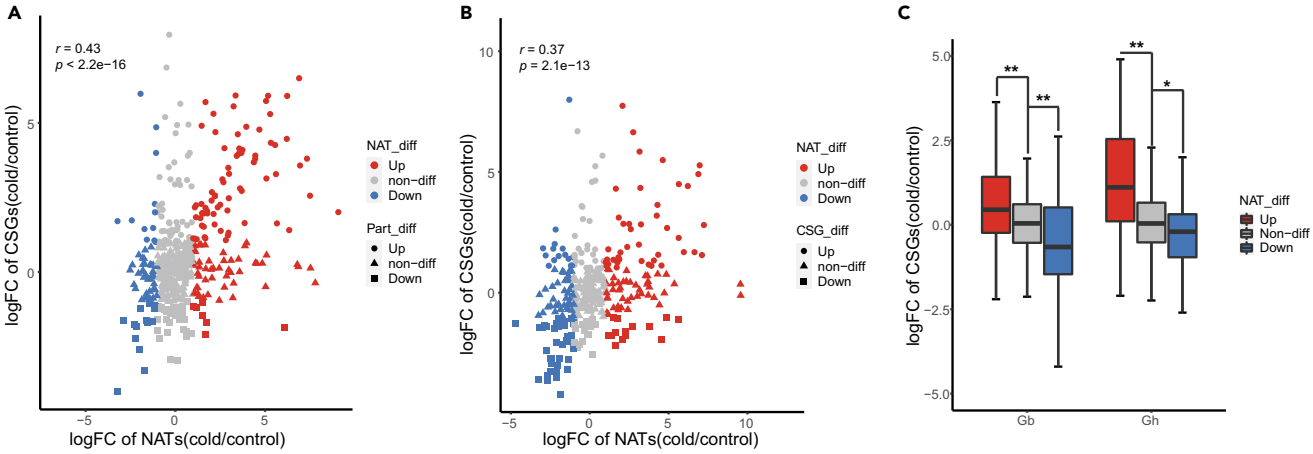


Figure 4. The positive expression trend of NATs and their cognate sense genes

(A and B) Correlation analysis of logFC of NATs(cold/control) and logFC of CSGs(cold/control) in *G. hirsutum* and *G. barbadense*, respectively. The R value is the correlation coefficient, both the *r* and p value were obtained from Pearson correlation test.

(C) Boxplot showing the distribution of logFC of CSGs based on differentially expressed NATs. Wilcoxon test: **, p value < 0.01, *, p value < 0.05.

differentially expressed conserved NATs, termed cold-associated NAT 1 through 4 (*CAN1-4*) (Figure S13). These four NATs are abundantly expressed, and compared with the control, NATs and CSGs were differentially expressed in both *G. hirsutum* and *G. barbadense* under chilling stress. The observation of similar expression patterns in both cotton species suggests that these transcripts play conserved roles in coping with chilling injury. *CAN1*, which was up-regulated during chill stress, is transcribed from the antisense strand of a PCG homologous to *SnRK2.8*, which encodes a serine/threonine-protein kinase (Figures 5 and S13). *CAN2*, which was down-regulated during chilling stress, corresponds to a CSG that was up-regulated under chilling stress and is homologous to *BBX27*, a B-box zinc finger family protein (Figures S13 and S14). *CAN3* was down-regulated during chill stress, as was its CSG *TOM3*, which encodes a G-protein coupled receptor (GPCR) (Figures S13 and S14). *CAN4*, which was up-regulated during chill stress, corresponds to the gene *MAG2*, which encodes a protein required for protein transport between the Golgi and the endoplasmic reticulum (Figure S13 and S14).

To investigate the roles of these NATs in cotton cold tolerance, the four NATs and their CSG were silenced using VIGS technology. Compared with empty vector-infected plants (*TRV2::0*), *TRV2::CAN1*- and *TRV2::SnRK2.8*-infected plants demonstrated reduced transcription of *GhCAN1* and *GhSnRK2.8*, respectively (Figures 6A and 6B). Interestingly, the level of *GhCAN1* transcripts was also decreased in *TRV2::SnRK2.8*-infected plants (Figure 6A). However, the expression of *GhSnRK2.8* in *TRV2::CAN1*-infected plants was slightly increased (Figure 6B). The chilling injury index and injured leaf area of *TRV2::CAN1*-infected plants were significantly less than with *TRV2::0*, indicating silencing of *CAN1* to increase tolerance to cold injury (Figures 6C–6F). Conversely, the chilling injury index and injured leaf area of *TRV2::SnRK2.8*-infected plants were significantly higher than with *TRV2::0*, indicating silencing of *SnRK2.8* to increase sensitivity to cold injury (Figures 6C–6F). With silencing of *CAN2*, *CAN3*, and *CAN4*, all plants were more sensitive to chilling stress (Figures S15–S17). In addition, silencing of *TRV2::BBX27* slightly increased cold tolerance (Figure S15), that of *TRV2::TOM3* had no significant effect (Figure S16), and loss of *TRV2::MAG2* resulted in cold sensitivity (Figure S17).

DISCUSSION

In 43 years ago, the first natural antisense transcript (NAT) was reported in plasmids,[24] and NATs were subsequently found in both mammals[25] and plants.[26] Recently, these transcripts (NATs) have been increasingly identified to function in various biological processes in plants.[14–17,27–29] However, due to their regulating gene expression in varied ways[30] and being characterized by poor interspecific conservation,[18,20] the general laws and interspecies universality of functional NATs has not been well elucidated. Generally, genes/lncRNAs/NATs that have been retained after a long period of natural selection in the process of species divergence are more likely to have important potential functions. Other species-specific NATs could either be transcriptional noise, or they could be the molecular basis that shapes species specificity. Therefore,

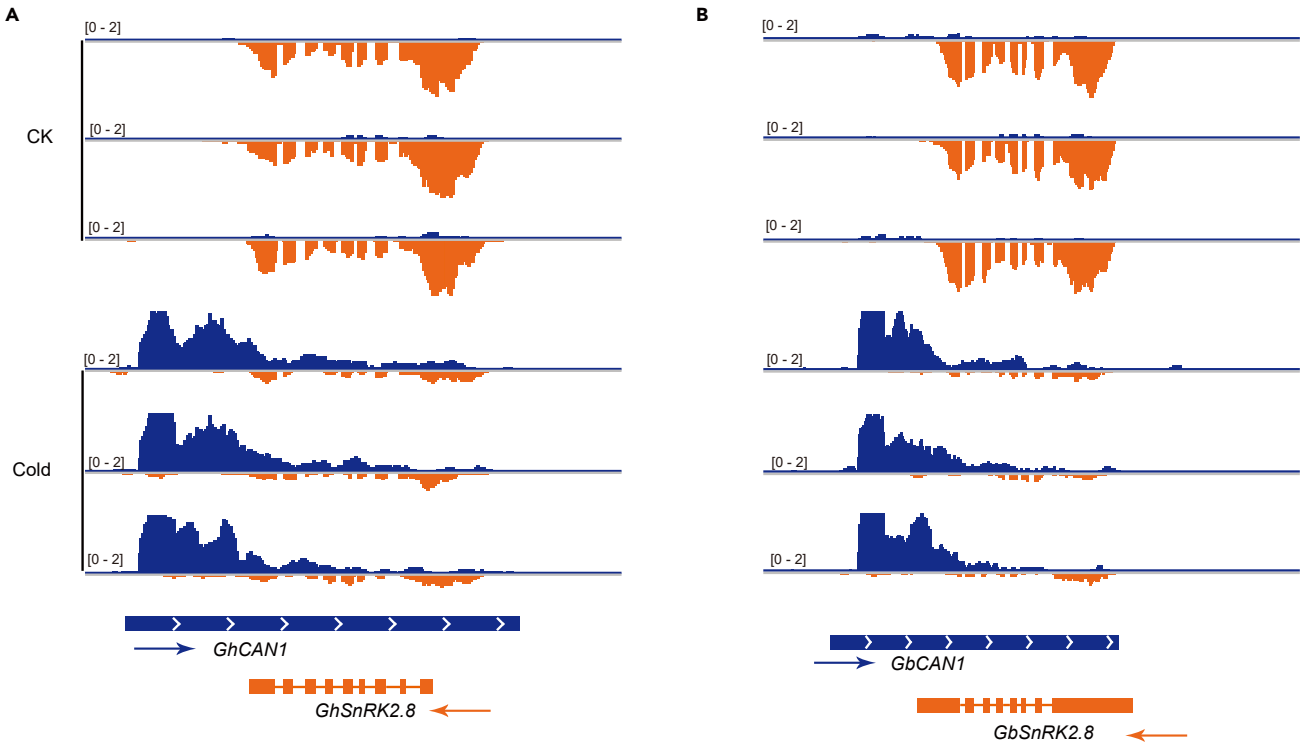


Figure 5. Expression pattern of *CAN1* and its cognate sense genes in *G. hirsutum* and *G. barbadense*

(A and B) IGV pile display showing the expression of *CAN1* and its cognate sense gene in *G. hirsutum* and *G. barbadense*, respectively. The blue pile represents NAT expression and the orange pile coding gene expression. Arrows indicate the direction of transcription. "0–2" represents the peak distribution range of CPM (counts per million mappings) within each 50-base bin.

interspecific conservation analysis is an important tool for analyzing the regulation modes of NATs and exploring functional NATs. However, there are yet few reports focusing on interspecific-conserved NATs in plants.

Interspecific-conserved NATs in cotton

In this study, 815 and 689 NATs were identified from *G. hirsutum* and *G. barbadense*, respectively (Figure 2). In previous studies, 2,486 and 4,718 NATs were identified in *G. arboreum*[31] and *G. barbadense*,[32] respectively, which was significantly more than the number reported here. This could be attributable to the following reasons. Firstly, it may be due to the tissue-specific expression patterns of non-coding genes. Here, we identified NATs from control and chilling-stressed leaves specifically, while previous studies examined different tissues such as leaves and ovules and fibers from multiple developmental stages of *G. arboreum* and *G. barbadense*.[31,32] The number of NATs in this study is less than report in *Arabidopsis* (Ivanov et al., 2021) and similar to report in maize,[20] *Brassica rapa*,[33] and *Salvia miltiorrhiza*.[28] Secondly, persistently low expressed genes/NATs are unlikely to be assessed as significant DE because low counts do not provide enough statistical evidence to make a reliable determination. In this work, NATs with low expression levels (TPM < 1) were removed from subsequent analysis, similar to previous studies.[15,20]

In addition, we identified 224 NATs conserved between *G. hirsutum* and *G. barbadense* (Figure 2). The expression changes of these NATs under cold stress were highly consistent between species, with logFC correlation coefficients as high as 0.91 (p value < 2.2×10^{-16}) (Figure 3B). This is consistent with and even higher than the trends of conserved PCGs and lncRNAs, for which the correlation coefficients were 0.86 and 0.85, respectively (Figure 3B). This similar interspecies expression pattern of homologous NATs suggests that the response pattern of NATs under cold stress was preserved during the divergence of *G. hirsutum* and *G. barbadense*. Allotetraploid cotton originated from genomic hybridization and polyploidy between an ancestral A genome diploid species (*G. arboreum* or *G. herbaceum*) and a D genome species (*G. raimondii*) at ~1.7–1.9 MYA, with the divergence of *G. hirsutum* and *G. barbadense* occurring

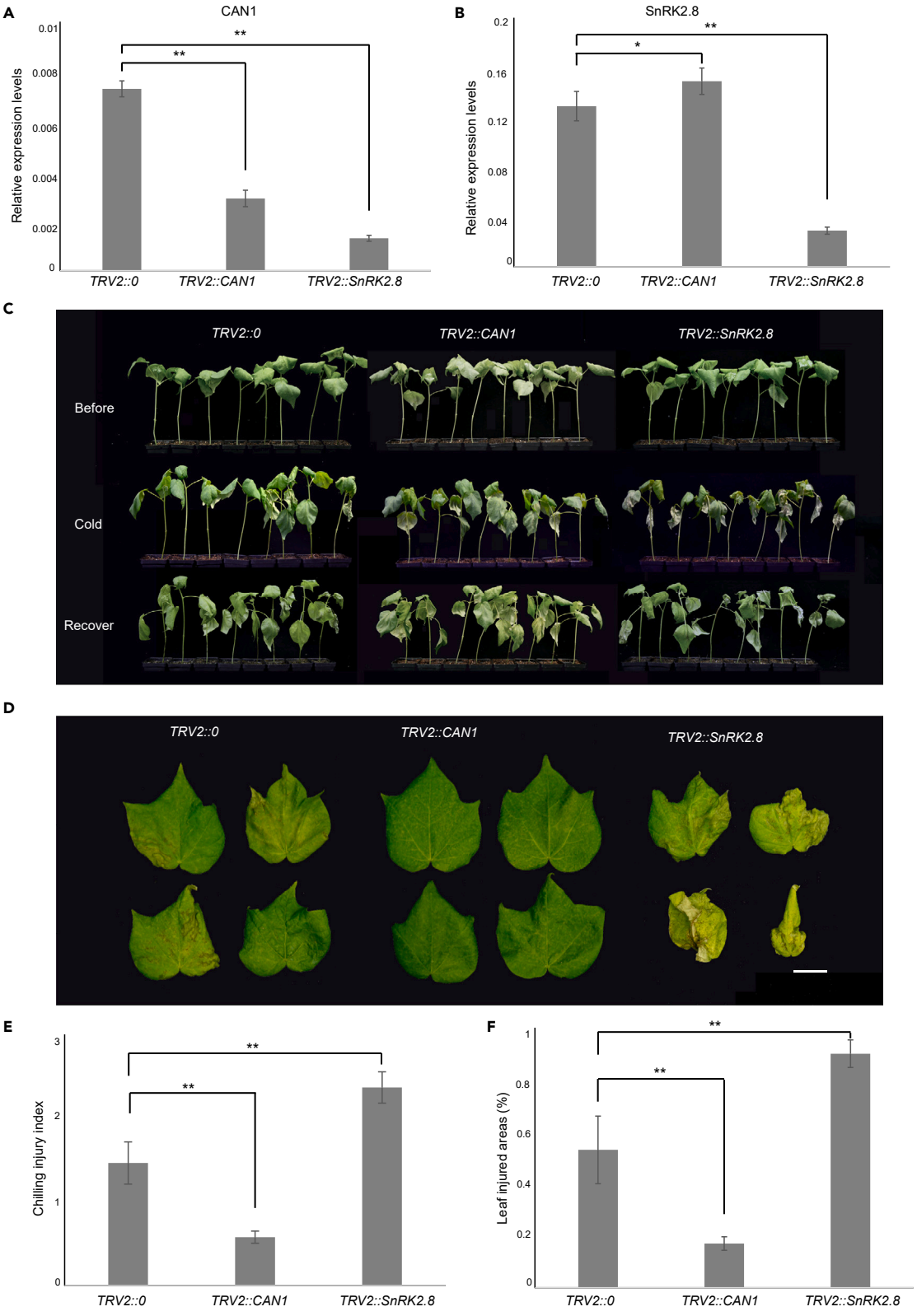

A
CAN1
Relative expression levels
0.01
0.008
0.006
0.004
0.002
0
**
**
TRV2::0
TRV2::CAN1
TRV2::SnRK2.8
B
SnRK2.8
Relative expression levels
0.2
0.16
0.12
0.08
0.04
0
*
**
TRV2::0
TRV2::CAN1
TRV2::SnRK2.8
C
TRV2::0
TRV2::CAN1
TRV2::SnRK2.8
Before
Cold
Recover
D
TRV2::0
TRV2::CAN1
TRV2::SnRK2.8
E
Chilling injury index
3
2
1
0
**
**
TRV2::0
TRV2::CAN1
TRV2::SnRK2.8
F
Leaf injured areas (%)
1
0.8
0.6
0.4
0.2
0
**
**
TRV2::0
TRV2::CAN1
TRV2::SnRK2.8

Figure 6. Silencing *GhCAN1* enhances chill tolerance in *G. hirsutum*

(A and B) qRT-PCR of *GhCAN1* and *GhSnRK2.8*, respectively. *t*-test, *, p value < 0.05, **, p value < 0.01. The error bar is standard error.

(C) Cold injury phenotype of plants after silencing *GhCAN1* and *GhSnRK2.8*.

(D) Leaf chilling injury phenotype after silencing *GhCAN1* and *GhSnRK2.8*.

(E and F) Chilling injury index and leaf injured area after silencing *GhCAN1* and *GhSnRK2.8*. *t*-test, **, p value < 0.01. The error bar is standard error.

subsequently at ~0.4–0.6 MYA.[4] The relatively high proportion of conserved NATs in *G. hirsutum* and *G. barbadense* (27.48%–32.51%) may be due to their relatively short period of divergence.

Diploid cotton, which includes the A and D genome species, diversified at ~4.7–7.1 MYA.[21] The conserved NATs between the A_T and D_T subgenomes indicates that they originated before the divergence of the A and D genome diploid species. In total, 49 and 39 subgenome-conserved NATs were found in *G. hirsutum* and *G. barbadense,* respectively (Figures S8A and S8B), constituting about 8.19%–11.54% of all identified NATs (Figures S8A and S8B), which was significantly less than the proportion shared between *G. hirsutum* and *G. barbadense* (~27.48%–32.51%). The significant difference in the number and proportion of conserved NATs between subgenomes and between species may be due to the decrease of interspecific conserved NATs with longer divergence time, or it may be due to a large number of NATs originating from hybridization and polyploidy and then being retained. The interspecies hybridization and polyploidization always introduce huge genome-wide variations, namely genome shock, normally resulting demethylations of retroelements,[34] changes in histone modification,[35] relaxation of imprinting genes,[36] silencing and activation of homologous genes[37,38] and so on. Latest researches reveal that, genome shock is often accompanied by the emergence of newly activated lncRNAs and neofunctionalization.[9,39] To further explore the origin and evolution of conservative NATs, we analyzed the number shared between *G. arboreum* and the A_T subgenome. This yielded 43 conserved NATs between *G. arboreum* and *G. hirsutum* A_T, and 37 between *G. arboreum* and *G. barbadense* A_T (Figures S8C–S8E); these comprised about 9.75%–15.87% of all identified NATs (Figure S8). The A genome (*G. arboreum*) and the A_T subgenome were separated at ~0.8–1.0 MYA after genomic hybridization and polyploidy, which is significantly later than divergence of the A and D genomes (~6.2–7.1 MYA).[4] However, the number and proportion of NATs conserved between *G. arboreum* and the A_T subgroups were similar to the corresponding values between A_T and D_T subgroups (37–43, 9.75%–15.87% vs. 39–49, 8.19%–11.54%). These results seem to suggest that a large number of NATs were generated after hybridization and polyploidy and retained during the divergence of *G. hirsutum* and *G. barbasense*.

NAT expression is positively correlated with CSG expression

NATs exert biological functions through a variety of molecular mechanisms.[29] In this study, we found the expression trends of NATs and CSGs to be consistent, with logFC of NATs and logFC of CSGs having a correlation coefficient of 0.43 (p value $<2.2 \times 10^{-16}$) and 0.37 (p value $= 2.1 \times 10^{-13}$) in *G. hirsutum* and *G. barbadense*, respectively (Figures 4A and 4B). For up-regulated NATs, logFC of CSGs was significantly higher than in non-differentially expressed NATs (Figure 4C). This is consistent with reports from previous studies in *Arabidopsis*[15] and *Salvia miltiorrhiza*.[28] In addition, positive correlations between sense/antisense are consistent with model presented in human and yeast, while antisense transcription increases sense transcript stability in chromatin-dependent manner (Brown et al., 2018). In *Arabidopsis*, MAS positively regulated the transcription level of MAF4 by regulating the H3K4me3 modification level at the MAF4 site, indicating that changing the histone modification level was also one of the factors positively correlated with sense/antisense expression.[15] Among NAT types, FO NATs exhibited the greatest consistency with CSG expression, while the expression trend in TT NATs was the least consistent (Figures S11 and S12). Overall, these results indicate that NAT-CSG pairs have a strong tendency to positively correlated expression patterns during chill stress in both *G. hirsutum* and *G. barbadense*. Elucidating the regulatory mechanism that governs NATs under cold stress will provide a new direction for improving chill tolerance of cotton. In total, this study identified 259 (24.74%) and 240 (24.14%) NATs that were up-regulated and 198 (18.91%) and 366 (36.82%) that were down-regulated under chilling stress in *G. hirsutum* and *G. barbadense*, respectively (Figure 3A).

NATs are potential loci with utility for genetic improvement

CBF has been widely studied as a core functional gene in cold stress.[6,7] In the regulatory network governing plant response to cold stress, OST1(SnRK2.6) is located upstream of CBF, and is responsible for cold injury signal transduction and *CBF* activation.[40] In this study, we found that a cold-induced NAT (*CAN1*) is

transcribed from the antisense strand of *SnRK2.8*, another member of the SnRK family (Figure 5). Expression of *GhSnRK2.8* was slightly increased in *TRV2::CAN1*-infected plants, suggesting a potential regulatory role of *CAN1* on *SnRK2.8* (Figure 6B). In addition, silencing *CAN1* significantly enhanced the cold tolerance of plants, while silencing *SnRK2.8* made them more sensitive (Figures 6E and 6F). Sucrose nonfermenting1 (SNF1)-related protein kinases (SnRKs) are essential for regulating plant responses to stress conditions.[41] Heterologous expression of *AcSnRK2.11* and *TaSnRK2.4* increases freezing tolerance in tobacco[42] and *Arabidopsis*,[43] respectively. Interestingly, where we observed *SnRK2.8* to be cold-repressed in cotton, a previous study in grape reported it to be a cold-inducible gene, and to have higher expression in cold-tolerant varieties.[44] Another study in tea plants found *SnRK2.8* to be dramatically up-regulated during 7-day long-term chill acclimation and 6-h short-term cold stimulus, and higher expression of *SnRK2.8* to correlate with cold tolerance.[45] In wheat, *TaSnRK2.8* was induced in response to cold stress, and overexpression of TaSnRK2.8 resulted in enhanced tolerance to freezing.[46] However, the current work found *CAN1* to be a cold-induced gene and *SnRK2.8* a cold-repressed gene in cotton, which may be due to the influence of *CAN1* on the expression of *SnRK2.8*. While SnRK2.8 has been demonstrated to be involved in cold stress regulation in other crops, this study is the first to report the existence of a cold-induced NAT of *SnRK2.8*. Although the specific mechanism of this regulatory relationship is still unclear, this study provides a new perspective for further research on NAT regulation of *SnRK2.8* in response to cold injury stress, and also provides a new molecular basis for the cultivation of cold-tolerant cotton varieties.

NATs play important roles in various stress responses. *Delay of germination 1* (DOG1) is the major quantitative trait locus for seed dormancy in *Arabidopsis thaliana* that is reported to be expressed exclusively in seeds and suppressed in seedlings by an antisense transcript (*asDOG1*).[47] The suppression is released by down-regulated *asDOG1*, which is induced by the plant hormone abscisic acid (ABA) and drought.[48] Loss of *asDOG1* leads to constitutive high-level *DOG1* expression, resulting in increased drought tolerance.[48] Conversely, DOG1 inactivation causes enhanced drought sensitivity.[48] Another example of functional NAT in crop is ZmNAC48 in maize, which encodes an NAC transcript factor playing crucial roles in response to abiotic stress.[29] *cis-NATZmNAC48* is located in the second intron and third exon of *ZmNAC48*, and negatively regulates *ZmNAC48* expression by generating small interfering RNAs.[29] Similarly, drought response NATs was identified in *Oryza nivara* and *O. sativa*.[49] A pair of natural antisense transcripts could be generated by overlapping coding genes. *P5CDH* (Delta(1)-pyrroline-5-carboxylate dehydrogenase) was a stress-related gene, and overlapping with *SRO5*.[50] *SRO5* is induced by salt and leads to degradation of *P5CDH* through formation small interfering RNAs, resulting in proline accumulation and increased *Arabidopsis* salt-tolerance.[50] The collective of data suggest that stress-induced NATs may influence plant resistance by regulating the expression of sense genes, and cold-induced CAN1 may have a similar regulatory mechanism in inhibiting *SnRK2.8* in cotton. Furthermore, NATs were reported response to heat stress,[51,52] cadmium exposure,[53] and phosphate fluctuations.[54] These findings suggest that NATs have tremendous potential for crop genetic improvement under various stress.

Limitation of the study

Natural antisense transcripts related to cotton cold stress were identified by transcriptome strand-specific library sequencing. However, RNA-seq lacks the ability to define 5′- and 3′-ends precisely to accurately define transcript boundaries. Combined with high-quality 5′and 3′ tag sequencing, identification of new transcripts including NAT will be more accurate.

STAR★METHODS

Detailed methods are provided in the online version of this paper and include the following:

- KEY RESOURCES TABLE
- RESOURCE AVAILABILITY
  - Lead contact
  - Materials availability
  - Data and code availability
- EXPERIMENTAL MODEL AND SUBJECT DETAILS
  - Plant materials and growth conditions
- METHOD DETAILS
  - Library construction and sequencing
  - NAT identification

- Identification of NATs conserved between *G. hirsutum* and *G. barbadense*
- Virus-induced gene silencing (VIGS)
- Quantitative RT-PCR

- QUANTIFICATION AND STATISTICAL ANALYSIS

SUPPLEMENTAL INFORMATION

Supplemental information can be found online at https://doi.org/10.1016/j.isci.2023.107362.

ACKNOWLEDGMENTS

This work was financially supported in part by grants from the National Key R&D Program of China (2022YFF1001400), National Natural Science Foundation of China (NSFC, 31971985, 3200379), Central Government Guided Local Science and Technology Development Funds (2023ZY1016), the Fundamental Research Funds for the Central Universities (2022FZZX07-04), JCIC-MCP, Hainan Provincial Natural Science Foundation of China (323CXTD385), Hainan Yazhou Bay Seed Lab, JBGS (B21HJ0403), Supported by High-performance Computing Platform of YZBSTCACC.

AUTHOR CONTRIBUTIONS

Conceptualization, X.G. and S.F.; Investigation, S.F., M.G., and X.L.; Visualization, S.F. and Y.Z.; Writing – Original Draft, X.G. and S.F.; Writing – Review & Editing, X.G. and S.F.; Funding Acquisition, X.G. and S.F.

DECLARATION OF INTERESTS

The authors declare no competing financial interests, and the authors have a patent related to this work.

INCLUSION AND DIVERSITY

We support inclusive, diverse, and equitable conduct of research.

Received: February 9, 2023
Revised: May 17, 2023
Accepted: July 10, 2023
Published: July 13, 2023

REFERENCES

1. Burke, J.J., Mahan, J.R., and Hatfield, J.L. (1988). Crop-Specific Thermal Kinetic Windows in Relation To Wheat and Cotton Biomass Production. Agron. J. *80*, 553–556. https://doi.org/10.2134/agronj1988.00021962008000040001x.

2. Reddy, V.R., Reddy, K.R., and Baker, D.N. (1991). Temperature Effect on Growth and Development of Cotton during the Fruiting Period. Agron. J. *83*, 211–217. https://doi.org/10.2134/agronj1991.00021962008300010050x.

3. Reddy, K.R., Reddy, V.R., and Hodges, H.F. (1992). Temperature Effects on Early Season Cotton Growth and Development. Agron. J. *84*, 229–237. https://doi.org/10.2134/agronj1992.00021962008400020021x.

4. Hu, Y., Chen, J., Fang, L., Zhang, Z., Ma, W., Niu, Y., Ju, L., Deng, J., Zhao, T., Lian, J., et al. (2019). *Gossypium barbadense* and *Gossypium hirsutum* genomes provide insights into the origin and evolution of allotetraploid cotton. Nat. Genet. *51*, 739–748. https://doi.org/10.1038/s41588-019-0371-5.

5. Li, X., He, Q., Yuan, Y., and Tang, F. (2003). Cold disasters, the most serious meteorological disasters to the cotton production in Xinjiang, China. Proceeding of the SPIE *4890*, 406–411. https://doi.org/10.1117/12.466189.

6. Stockinger, E.J., Gilmour, S.J., and Thomashow, M.F. (1997). *Arabidopsis thaliana CBF1* encodes an AP2 domain-containing transcriptional activator that binds to the C-repeat/DRE, a *cis*-acting DNA regulatory element that stimulates transcription in response to low temperature and water deficit. Proc. Natl. Acad. Sci. USA *94*, 1035–1040. https://doi.org/10.1073/pnas.94.3.1035.

7. Liu, Q., Kasuga, M., Sakuma, Y., Abe, H., Miura, S., Yamaguchi-Shinozaki, K., and Shinozaki, K. (1998). Two transcription factors, DREB1 and DREB2, with an EREBP/AP2 DNA binding domain separate two cellular signal transduction pathways in drought- and low-temperature-responsive gene expression, respectively, in *Arabidopsis*. Plant Cell *10*, 1391–1406. https://doi.org/10.1105/tpc.10.8.1391.

8. Nejat, N., and Mantri, N. (2018). Emerging roles of long non-coding RNAs in plant response to biotic and abiotic stresses. Crit. Rev. Biotechnol. *38*, 93–105. https://doi.org/10.1080/07388551.2017.1312270.

9. Zhao, T., Tao, X., Feng, S., Wang, L., Hong, H., Ma, W., Shang, G., Guo, S., He, Y., Zhou, B., and Guan, X. (2018). LncRNAs in polyploid cotton interspecific hybrids are derived from transposon neofunctionalization. Genome Biol. *19*, 195. https://doi.org/10.1186/s13059-018-1574-2.

10. Sarropoulos, I., Marin, R., Cardoso-Moreira, M., and Kaessmann, H. (2019). Developmental dynamics of lncRNAs across mammalian organs and species. Nature *571*, 510–514. https://doi.org/10.1038/s41586-019-1341-x.

11. Li, Y., Li, X., Yang, J., and He, Y. (2020). Natural antisense transcripts of MIR398 genes suppress microR398 processing and attenuate plant thermotolerance. Nat. Commun. *11*, 5351. https://doi.org/10.1038/s41467-020-19186-x.

12. Wan, Q., Guan, X., Yang, N., Wu, H., Pan, M., Liu, B., Fang, L., Yang, S., Hu, Y., Ye, W., et al. (2016). Small interfering RNAs from bidirectional transcripts of *GhMML3_A12* regulate cotton fiber development. New

Phytol. *210*, 1298–1310. https://doi.org/10.1111/nph.13860.

13. Faghihi, M.A., and Wahlestedt, C. (2009). Regulatory roles of natural antisense transcripts. Nat. Rev. Mol. Cell Biol. *10*, 637–643. https://doi.org/10.1038/nrm2738.

14. Swiezewski, S., Liu, F., Magusin, A., and Dean, C. (2009). Cold-induced silencing by long antisense transcripts of an *Arabidopsis* Polycomb target. Nature *462*, 799–802. https://doi.org/10.1038/nature08618.

15. Zhao, X., Li, J., Lian, B., Gu, H., Li, Y., and Qi, Y. (2018). Global identification of *Arabidopsis* lncRNAs reveals the regulation of MAF4 by a natural antisense RNA. Nat. Commun. *9*, 5056. https://doi.org/10.1038/s41467-018-07500-7.

16. Kindgren, P., Ard, R., Ivanov, M., and Marquardt, S. (2018). Transcriptional read-through of the long non-coding RNA SVALKA governs plant cold acclimation. Nat. Commun. *9*, 4561. https://doi.org/10.1038/s41467-018-07010-6.

17. Li, X., Yang, Q., Liao, X., Tian, Y., Zhang, F., Zhang, L., and Liu, Q. (2022). A natural antisense RNA improves chrysanthemum cold tolerance by regulating the transcription factor *DgTCP1*. Plant Physiol. *190*, 605–620. https://doi.org/10.1093/plphys/kiac267.

18. Zhang, Y., Liu, X.S., Liu, Q.R., and Wei, L. (2006). Genome-wide in silico identification and analysis of *cis* natural antisense transcripts (*cis*-NATs) in ten species. Nucleic Acids Res. *34*, 3465–3475. https://doi.org/10.1093/nar/gkl473.

19. Mouse Genome Sequencing Consortium, Waterston, R.H., Lindblad-Toh, K., Birney, E., Rogers, J., Abril, J.F., Agarwal, P., Agarwala, R., Ainscough, R., Alexandersson, M., et al. (2002). Initial sequencing and comparative analysis of the mouse genome. Nature *420*, 520–562. https://doi.org/10.1038/nature01262.

20. Wang, H., Niu, Q.W., Wu, H.W., Liu, J., Ye, J., Yu, N., and Chua, N.H. (2015). Analysis of non-coding transcriptome in rice and maize uncovers roles of conserved lncRNAs associated with agriculture traits. Plant J. *84*, 404–416. https://doi.org/10.1111/tpj.13018.

21. Chen, Z.J., Sreedasyam, A., Ando, A., Song, Q., De Santiago, L.M., Hulse-Kemp, A.M., Ding, M., Ye, W., Kirkbride, R.C., Jenkins, J., et al. (2020). Genomic diversifications of five *Gossypium* allopolyploid species and their impact on cotton improvement. Nat. Genet. *52*, 525–533. https://doi.org/10.1038/s41588-020-0614-5.

22. Brambati, A., Zardoni, L., Nardini, E., Pellicioli, A., and Liberi, G. (2020). The dark side of RNA:DNA hybrids. Mutat. Res. Rev. Mutat. Res. *784*, 108300. https://doi.org/10.1016/j.mrrev.2020.108300.

23. Xu, J., Wang, Q., Freeling, M., Zhang, X., Xu, Y., Mao, Y., Tang, X., Wu, F., Lan, H., Cao, M., et al. (2017). Natural antisense transcripts are significantly involved in regulation of drought stress in maize. Nucleic Acids Res. *45*, 5126–5141. https://doi.org/10.1093/nar/gkx085.

24. Itoh, T., and Tomizawa, J. (1980). Formation of an RNA primer for initiation of replication of ColE1 DNA by ribonuclease H. Proc. Natl. Acad. Sci. USA *77*, 2450–2454. https://doi.org/10.1073/pnas.77.5.2450.

25. Williams, T., and Fried, M. (1986). A mouse locus at which transcription from both DNA strands produces mRNAs complementary at their 3′ ends. Nature *322*, 275–279. https://doi.org/10.1038/322275a0.

26. Schmitz, G., and Theres, K. (1992). Structural and functional analysis of the Bz2 locus of *Zea mays*: characterization of overlapping transcripts. Mol. Gen. Genet. *233*, 269–277. https://doi.org/10.1007/BF00587588.

27. Wang, X.J., Gaasterland, T., and Chua, N.H. (2005). Genome-wide prediction and identification of *cis*-natural antisense transcripts in *Arabidopsis thaliana*. Genome Biol. *6*, R30. https://doi.org/10.1186/gb-2005-6-4-r30.

28. Jiang, M., Chen, H., Liu, J., Du, Q., Lu, S., and Liu, C. (2021). Genome-wide identification and functional characterization of natural antisense transcripts in *Salvia miltiorrhiza*. Sci. Rep. *11*, 4769. https://doi.org/10.1038/s41598-021-83520-6.

29. Mao, Y., Xu, J., Wang, Q., Li, G., Tang, X., Liu, T., Feng, X., Wu, F., Li, M., Xie, W., and Lu, Y. (2021). A natural antisense transcript acts as a negative regulator for the maize drought stress response gene *ZmNAC48*. J. Exp. Bot. *72*, 2790–2806. https://doi.org/10.1093/jxb/erab023.

30. Reis, R.S., and Poirier, Y. (2021). Making sense of the natural antisense transcript puzzle. Trends Plant Sci. *26*, 1104–1115. https://doi.org/10.1016/j.tplants.2021.07.004.

31. Zou, C., Wang, Q., Lu, C., Yang, W., Zhang, Y., Cheng, H., Feng, X., Prosper, M.A., and Song, G. (2016). Transcriptome analysis reveals long noncoding RNAs involved in fiber development in cotton (*Gossypium arboreum*). Sci. China Life Sci. *59*, 164–171. https://doi.org/10.1007/s11427-016-5000-2.

32. Wang, M., Yuan, D., Tu, L., Gao, W., He, Y., Hu, H., Wang, P., Liu, N., Lindsey, K., and Zhang, X. (2015). Long noncoding RNAs and their proposed functions in fibre development of cotton (*Gossypium* spp.). New Phytol. *207*, 1181–1197. https://doi.org/10.1111/nph.13429.

33. Mehraj, H., Shea, D.J., Takahashi, S., Miyaji, N., Akter, A., Seki, M., Dennis, E.S., Fujimoto, R., and Osabe, K. (2021). Genome-wide analysis of long noncoding RNAs, 24-nt siRNAs, DNA methylation and H3K27me3 marks in *Brassica rapa*. PLoS One *16*, e0242530. https://doi.org/10.1371/journal.pone.0242530.

34. Madlung, A., Tyagi, A.P., Watson, B., Jiang, H., Kagochi, T., Doerge, R.W., Martienssen, R., and Comai, L. (2005). Genomic changes in synthetic *Arabidopsis* polyploids. Plant J. *41*, 221–230. https://doi.org/10.1111/j.1365-313X.2004.02297.x.

35. Song, Q., and Chen, Z.J. (2015). Epigenetic and developmental regulation in plant polyploids. Curr. Opin. Plant Biol. *24*, 101–109. https://doi.org/10.1016/j.pbi.2015.02.007.

36. Josefsson, C., Dilkes, B., and Comai, L. (2006). Parent-dependent loss of gene silencing during interspecies hybridization. Curr. Biol. *16*, 1322–1328. https://doi.org/10.1016/j.cub.2006.05.045.

37. Kashkush, K., Feldman, M., and Levy, A.A. (2002). Gene loss, silencing and activation in a newly synthesized wheat allotetraploid. Genetics *160*, 1651–1659. https://doi.org/10.1093/genetics/160.4.1651.

38. Wang, J., Tian, L., Lee, H.S., Wei, N.E., Jiang, H., Watson, B., Madlung, A., Osborn, T.C., Doerge, R.W., Comai, L., and Chen, Z.J. (2006). Genomewide nonadditive gene regulation in *Arabidopsis* allotetraploids. Genetics *172*, 507–517. https://doi.org/10.1534/genetics.105.047894.

39. Tao, X., Li, M., Zhao, T., Feng, S., Zhang, H., Wang, L., Han, J., Gao, M., Lu, K., Chen, Q., et al. (2021). Neofunctionalization of a polyploidization-activated cotton long intergenic non-coding RNA DAN1 during drought stress regulation. Plant Physiol. *186*, 2152–2168. https://doi.org/10.1093/plphys/kiab179.

40. Ding, Y., Li, H., Zhang, X., Xie, Q., Gong, Z., and Yang, S. (2015). OST1 kinase modulates freezing tolerance by enhancing ICE1 stability in *Arabidopsis*. Dev. Cell *32*, 278–289. https://doi.org/10.1016/j.devcel.2014.12.023.

41. Chen, X., Ding, Y., Yang, Y., Song, C., Wang, B., Yang, S., Guo, Y., and Gong, Z. (2021). Protein kinases in plant responses to drought, salt, and cold stress. J. Integr. Plant Biol. *63*, 53–78. https://doi.org/10.1111/jipb.13061.

42. Xiang, D.J., Man, L.L., Cao, S., Liu, P., Li, Z.G., and Wang, X.D. (2020). Heterologous expression of an *Agropyron cristatum* SnRK2 protein kinase gene (*AcSnRK2.11*) increases freezing tolerance in transgenic yeast and tobacco. 3 Biotech *10*, 209. https://doi.org/10.1007/s13205-020-02203-7.

43. Mao, X., Zhang, H., Tian, S., Chang, X., and Jing, R. (2010). TaSnRK2.4, an SNF1-type serine/threonine protein kinase of wheat (*Triticum aestivum* L.), confers enhanced multistress tolerance in *Arabidopsis*. J. Exp. Bot. *61*, 683–696. https://doi.org/10.1093/jxb/erp331.

44. Liang, G., Ma, Z., Lu, S., Ma, W., Feng, L., Mao, J., and Chen, B. (2022). Temperature-phase transcriptomics reveals that hormones and sugars in the phloem of grape participate in tolerance during cold acclimation. Plant Cell Rep. *41*, 1357–1373. https://doi.org/10.1007/s00299-022-02862-1.

45. Li, Y., Wang, X., Ban, Q., Zhu, X., Jiang, C., Wei, C., and Bennetzen, J.L. (2019). Comparative transcriptomic analysis reveals gene expression associated with cold adaptation in the tea plant *Camellia sinensis*. BMC Genom. *20*, 624. https://doi.org/10.1186/s12864-019-5988-3.

46. Zhang, H., Mao, X., Wang, C., and Jing, R. (2010). Overexpression of a common wheat

gene *TaSnRK2.8* enhances tolerance to drought, salt and low temperature in *Arabidopsis*. PLoS One *5*, e16041. https://doi.org/10.1371/journal.pone.0016041.

47. Fedak, H., Palusinska, M., Krzyczmonik, K., Brzezniak, L., Yatusevich, R., Pietras, Z., Kaczanowski, S., and Swiezewski, S. (2016). Control of seed dormancy in *Arabidopsis* by a *cis*-acting noncoding antisense transcript. Proc. Natl. Acad. Sci. USA *113*, E7846–E7855. https://doi.org/10.1073/pnas.1608827113.

48. Yatusevich, R., Fedak, H., Ciesielski, A., Krzyczmonik, K., Kulik, A., Dobrowolska, G., and Swiezewski, S. (2017). Antisense transcription represses *Arabidopsis* seed dormancy QTL DOG1 to regulate drought tolerance. EMBO Rep. *18*, 2186–2196. https://doi.org/10.15252/embr.201744862.

49. Xu, Y.C., Zhang, J., Zhang, D.Y., Nan, Y.H., Ge, S., and Guo, Y.L. (2021). Identification of long noncoding natural antisense transcripts (lncNATs) correlated with drought stress response in wild rice (*Oryza nivara*). BMC Genom. *22*, 424. https://doi.org/10.1186/s12864-021-07754-4.

50. Borsani, O., Zhu, J., Verslues, P.E., Sunkar, R., and Zhu, J.K. (2005). Endogenous siRNAs derived from a pair of natural *cis*-antisense transcripts regulate salt tolerance in *Arabidopsis*. Cell *123*, 1279–1291. https://doi.org/10.1016/j.cell.2005.11.035.

51. Yu, X., Yang, J., Li, X., Liu, X., Sun, C., Wu, F., and He, Y. (2013). Global analysis of *cis*-natural antisense transcripts and their heat-responsive nat-siRNAs in *Brassica rapa*. BMC Plant Biol. *13*, 208. https://doi.org/10.1186/1471-2229-13-208.

52. Jin, J., Ohama, N., He, X., Wu, H.W., and Chua, N.H. (2022). Tissue-specific transcriptomic analysis uncovers potential roles of natural antisense transcripts in *Arabidopsis* heat stress response. Front. Plant Sci. *13*, 997967. https://doi.org/10.3389/fpls.2022.997967.

53. Oono, Y., Yazawa, T., Kanamori, H., Sasaki, H., Mori, S., and Matsumoto, T. (2017). Genome-wide analysis of rice *cis*-natural antisense transcription under cadmium exposure using strand-specific RNA-Seq. BMC Genom. *18*, 761. https://doi.org/10.1186/s12864-017-4108-5.

54. Cruz de Carvalho, M.H., and Bowler, C. (2020). Global identification of a marine diatom long noncoding natural antisense transcripts (NATs) and their response to phosphate fluctuations. Sci. Rep. *10*, 14110. https://doi.org/10.1038/s41598-020-71002-0.

55. Martin, M. (2011). Cutadapt removes adapter sequences from high-throughput sequencing reads. EMBnet. J. *17*, 10–12. https://doi.org/10.14806/ej.17.1.200.

56. Kim, D., Paggi, J.M., Park, C., Bennett, C., and Salzberg, S.L. (2019). Graph-based genome alignment and genotyping with HISAT2 and HISAT-genotype. Nat. Biotechnol. *37*, 907–915. https://doi.org/10.1038/s41587-019-0201-4.

57. Pertea, M., Pertea, G.M., Antonescu, C.M., Chang, T.C., Mendell, J.T., and Salzberg, S.L. (2015). StringTie enables improved reconstruction of a transcriptome from RNA-seq reads. Nat. Biotechnol. *33*, 290–295. https://doi.org/10.1038/nbt.3122.

58. Pertea, G., and Pertea, M. (2020). GFF Utilities: GffRead and GffCompare. F1000Res *9*. https://doi.org/10.12688/f1000research.23297.2.

59. Haas, B.J., Papanicolaou, A., Yassour, M., Grabherr, M., Blood, P.D., Bowden, J., Couger, M.B., Eccles, D., Li, B., Lieber, M., et al. (2013). De novo transcript sequence reconstruction from RNA-seq using the Trinity platform for reference generation and analysis. Nat. Protoc. *8*, 1494–1512. https://doi.org/10.1038/nprot.2013.084.

60. Mistry, J., Chuguransky, S., Williams, L., Qureshi, M., Salazar, G.A., Sonnhammer, E.L.L., Tosatto, S.C.E., Paladin, L., Raj, S., Richardson, L.J., et al. (2021). Pfam: The protein families database in 2021. Nucleic Acids Res. *49*, D412–D419. https://doi.org/10.1093/nar/gkaa913.

61. Kang, Y.J., Yang, D.C., Kong, L., Hou, M., Meng, Y.Q., Wei, L., and Gao, G. (2017). CPC2: a fast and accurate coding potential calculator based on sequence intrinsic features. Nucleic Acids Res. *45*, W12–W16. https://doi.org/10.1093/nar/gkx428.

62. Wang, Y., Tang, H., Debarry, J.D., Tan, X., Li, J., Wang, X., Lee, T.H., Jin, H., Marler, B., Guo, H., et al. (2012). MCScanX: a toolkit for detection and evolutionary analysis of gene synteny and collinearity. Nucleic Acids Res. *40*, e49. https://doi.org/10.1093/nar/gkr1293.

63. Livak, K.J., and Schmittgen, T.D. (2001). Analysis of relative gene expression data using real-time quantitative PCR and the 2(-Delta Delta C(T)) Method. Methods *25*, 402–408. https://doi.org/10.1006/meth.2001.1262.

STAR★METHODS

KEY RESOURCES TABLE

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Deposited data		
Raw and analyzed data	This paper	Bioproject PRJNA901542
Software and algorithms		
Cutadapt 1.18	Martin[55]	https://cutadapt.readthedocs.io/en/stable/
Hisat2 v2.2.0	Kim et al.[56]	http://daehwankimlab.github.io/hisat2/
StringTie v2.1.2	Pertea et al.[57]	https://ccb.jhu.edu/software/stringtie/
Gffcompare v0.11.2	Pertea and Pertea[58]	http://ccb.jhu.edu/software/stringtie/gffcompare.shtml
TransDecoder 5.5.0	Haas et al.[59]	https://github.com/TransDecoder/TransDecoder/releases
CPC 0.1	Kang et al.[61]	https://github.com/biocoder/CPC2
McScan v1.0.8	Wang et al.[62]	https://github.com/tanghaibao/jcvi/wiki/MCscan-%28Python-version%29
Code for analysis NAT	This paper	https://github.com/ShouliFeng2020/CAN_cotton

RESOURCE AVAILABILITY

Lead contact

Further information and requests for resources and reagents should be directed to and will be fulfilled by lead contact Xueying Guan (xueyingguan@zju.edu.cn).

Materials availability

This study did not generate new unique reagents.

Data and code availability

All RNA sequencing reads have been deposited in the NCBI Short Read Archive (https://www.ncbi.nlm.nih.gov/sra) under Bioproject: PRJNA901542. Sample IDs and metadata can be found in Table S1. The procedure and code used for identifying conservative NATs have been stored in the following GitHub repository: https://github.com/ShouliFeng2020/CAN_cotton.

Any additional information required to reanalyze the data reported in this paper is available from the lead contact upon request.

EXPERIMENTAL MODEL AND SUBJECT DETAILS

Plant materials and growth conditions

G. hirsutum L. acc. Texas Marker-1 (TM-1, AD1) and *G. barbadense* L. acc. Hai7124 were used in this research. All plants were grown under controlled chamber conditions (day/night, 28/26°C, 16/8 h and 70-80% relative humidity). First, 14-day-old seedlings were transferred to a chilling chamber (4°C). Forty-eight hours later, leaves were collected from control or cold treatment plants and immediately placed into liquid nitrogen for flash-freezing. After removal from the liquid nitrogen, all samples were kept in a freezer at -80°C until use.

METHOD DETAILS

Library construction and sequencing

Total RNA was isolated from tissues using an RNA extraction kit (Mofan Bio, RK16-50T). RNA purity was checked using a NanoPhotometer® spectrophotometer (IMPLEN, CA, USA), and RNA integrity was

assessed using the RNA Nano 6000 Assay Kit of the Bioanalyzer 2100 system (Agilent Technologies, CA, USA). After RNA isolation, 20 ng RNA per sample was depleted of ribosomal RNA using the Epicentre Ribo-zero™rRNA Removal Kit (Epicentre, USA). Subsequently, sequencing libraries were generated from the rRNA-depleted RNA with the NEBNext® Ultra™ Directional RNA Library Prep Kit for Illumina® (NEB, USA) following the manufacturer's recommendations. Finally, strand-specific sequencing was performed on the Illumina NovaSeq 6000 platform and 150 bp paired-end reads were generated.

NAT identification

The raw data were firstly processed to remove reads containing adapter sequence, poly-N reads, and low-quality bases using cutadapt.[55] Next, strand-specific clean reads were aligned to *G. hirsutum* and *G. barbadense* references using Hisat2 v2.2.0[56] with parameters '–dta –rna-strandness RF'; other parameters were set as default. StringTie (v2.1.2) was used to assemble transcripts for each gene locus with parameters '–rf -f 0.1 -j 10 -c 10'.[57] After merging transcripts, they were compared with reference genome annotations using Gffcompare.[58] Those transcripts given a class code of "u", "i", or "y" and length greater than 200 bp were selected for open reading frame (ORF) prediction using TransDecoder.[59] Transcripts with a predicted open reading frame of less than 100 amino acids and that had no coding ability predicted by Coding Potential Calculator (CPC) or did not contain a known protein domain were labeled as candidate long non-coding RNAs.[60,61] The GTF files of candidate lncRNAs and reference coding genes were then merged and compared to the StringTie assembled transcripts using Gffcompare. Transcripts with class code "x" were labeled as natural antisense transcripts (NATs) of lncRNA or coding genes.

Identification of NATs conserved between *G. hirsutum* and *G. barbadense*

First, McScan (JCVI Toolkit) was used to establish genome-wide collinearity between the cotton species based on PCG and lncRNA sequences to obtain intergenic collinearity anchors.[62] The allotetraploid cotton has two subgenomes, we only utilized collinearity within the same subgenome for subsequent analysis since our focus was on identifying conserved regions between *G. hirsutum* and *G. barbadense*. Second, reciprocal best blast was performed between PCGs/lncRNA. The overlapped pairs between synteny anchors and reciprocal best hits were defined as homologous lncRNA or gene pairs. If a NAT was contained within both members of a pair, that NAT was defined as a candidate interspecific conserved NAT. Finally, the sequences of the candidate NATs were compared and those with sequence consistency greater than 80% were considered interspecific conserved NATs.

Virus-induced gene silencing (VIGS)

Vector construction for virus-induced gene silencing (VIGS) was carried out using homologous recombination to clone the target fragments into the *TRV2* vector. The target fragments were obtained by PCR with gene-specific primers (Table S3) from *G. hirsutum* leaf cDNA. The generated recombinant vectors were then transformed into *Agrobacterium tumefaciens* GV3101. Briefly, freshly-cultured *TRV1* and *TRV2* resuspensions were mixed at a 1:1 volume ratio. 1∼2 weeks cotton seedlings after germination were then infected using the cotyledon syringe infiltration method. The empty vector *TRV2::0* was used as the negative control, and *TRV2::CLA1* and *TRV2::GhPGF* as positive controls. The seedlings were grown in controlled chambers under a 14-hour light/10-hour dark and a temperature of 22°C. Three to four weeks after the injection treatment, plants were transferred to a chilling chamber of 4°C for 48 h. Subsequently, they were then moved back to a controlled environment at 22°C for 48 h to recover before the assessment of cold damage and physiological phenotype. Samples before chilling treatment were collected for RT-qPCR analysis of target NATs or genes.

Quantitative RT-PCR

Total RNA was extracted with an RNA extraction kit (Mofan Bio, RK16-50T). Reverse transcription was performed with the HiScriptR II Q RT SuperMix for qPCR (R223-01, Vazyme, Nanjing, China). Primers for qPCR (Table S4) were designed with Primer3 (https://bioinfo.ut.ee/primer3-0.4.0/) and the reaction was carried out using ChamQ™ Universal SYBR® qPCR Master Mix (Q711-02/03, Vazyme, Nanjing, China) on a StepOnePlus system (Applied Biosystems, Thermo Fisher Scientific) using histone H3 (AF024716) as a reference gene. The data was analyzed using the ΔΔCt method.[63]

QUANTIFICATION AND STATISTICAL ANALYSIS

Statistical analyses were conducted using Microsoft Excel and R. Differentially expressed genes/NATs were identified using the EdgeR package with significance thresholds set at FDR < 0.05 and absolute logFC > 1. Correlation analysis was carried out using cor.test in R, while regression analysis was performed using geom_smooth (method = ''lm'') in the ggplot2 package. Additionally, the Wilcoxon Rank-Sum Test was applied in R using wilcox. test (* p-value < 0.05, ** p-value < 0.01), and the Student's t-test was conducted in Microsoft Excel with significance thresholds set at *p-value < 0.05 and **p-value < 0.01.