

Supplementary Files

The Catalase Gene Family in Cotton: Genome-Wide Characterization and Bioinformatics Analysis

Wei Wang, Yingying Cheng, Dongdong Chen, Dan Liu, Mengjiao Hu, Jie Dong, Xiaopei Zhang, Lirong Song and Fafu Shen *

Contents

Supplementary file 1: Table S1. Summary of some plant *CAT* genes in relation to the biotic and abiotic stress resistance.

Supplementary file 2: Table S2. Primers used in this study.

Supplementary file 3: Table S3. Gene numbers of *CAT* gene family in four other genomes descended from common eudicot genome ancestor with *Gossypium* in Rosids.

Supplementary file 4: Table S4. Orthologous *CAT* gene pairs among *G. arboreum* L., *G. raimondii* Ulbr., *G. hirsutum* L. and *G. barbadense* L..

Supplementary file 5: Table S5. The details of *GhCAT* and *GbCAT* gene family expression difference in different tissues and/or organs, at developmental processes of ovules and fibers, and under different stress treatments.

Supplementary file 6: Table S6. List of identified TFBSs in the putative promoter regions of *GhCAT* and *GbCAT* gene family using PlantTFDB 4.0 (<http://planttfdb.cbi.pku.edu.cn/>).

Supplementary file 7: Table S7. The predicted details of miRNA-mediated targeting regulatory relations with the principle transcripts of *CAT* gene family in upland cotton and sea-island cotton using the psRNATarget web server (<http://plantgrn.noble.org/psRNATarget/home>).

Supplementary file 8: Table S8. The predicted details of miRNA-mediated targeting regulatory relations with all full-length isoforms of *CAT* gene family in upland cotton and sea-island cotton using the psRNATarget web server (<http://plantgrn.noble.org/psRNATarget/home>).

(All the tables in additional file see a separate excel file.)

Supplementary file 9: Figure S1. (A) Blast hits of *GbCAT7* (GOBAR_AA20422) in the upland cotton genome database [cDNA, *G. hirsutum* (AD1), NAU] and the location of the designed primer pairs for cloning the actual ORF of *GbCAT7*. (B) PCR result shows that the actual TSS of *GbCAT7* is the ATG contained by the F3 forward primer (green dotted underline in A). (C) Verification of the amplified line in Figure S1 (B) by PCR-based sequencing.

Supplementary file 10: Figure S2. The supporting alignment result of the coding sequence of *GbCAT7*

(GOBAR_AA20422) and Gh_A07G1555. Query, the coding sequence of *GbCAT7*; Subject, the coding sequence of Gh_A07G1555.

Supplementary file 11: Figure S3. Phylogenetic tree of *Gossypium* and four other genomes descended from common eudicot genome ancestor in Rosids and the number of *CAT* genes identified from the genomes of these plants.

Supplementary file 12: Figure S4. Transcriptional profiling of *GhCATs* under different abiotic stress treatments.

Supplementary file 13: Figure S5. Gene Ontology (GO) annotation of the TFs binding to the upland cotton and sea-island cotton *CAT* genes' putative promoter regions based on their cellular component, molecular function, and biological process.

Supplementary file 14: Figure S6. All potentially alternative spliced isoform structure of *GhCATs* and *GbCATs* full-length transcripts and identification of full-length isoforms targeted by cotton miRNAs.

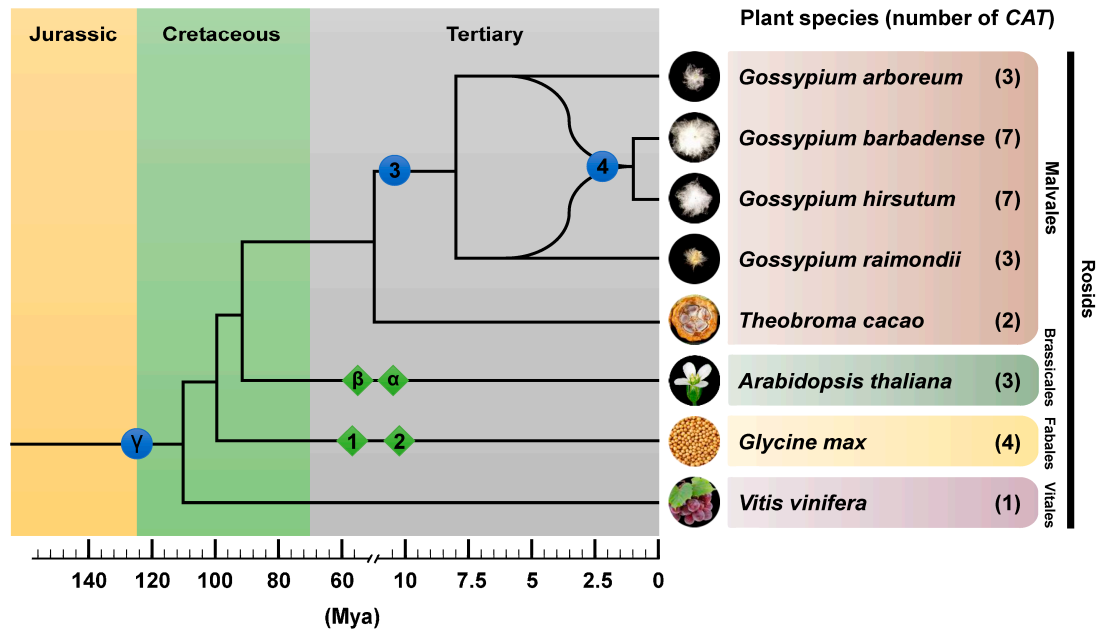
Supplementary file 15: Figure S7. The relative expression levels of miRNAs and their putative targeting *CAT* genes under *V. dahliae* infection treatment.

Supplementary file 16: Figure S8. (A, B) Amplification of *GhCAT3* (Gh_A05G1539) and *GhCAT7* (Gh_A07G1556) by PCR with the designed primer pairs for cloning their transcripts. PCR results were aligned with gene structure annotation and were verified by Sanger sequencing.

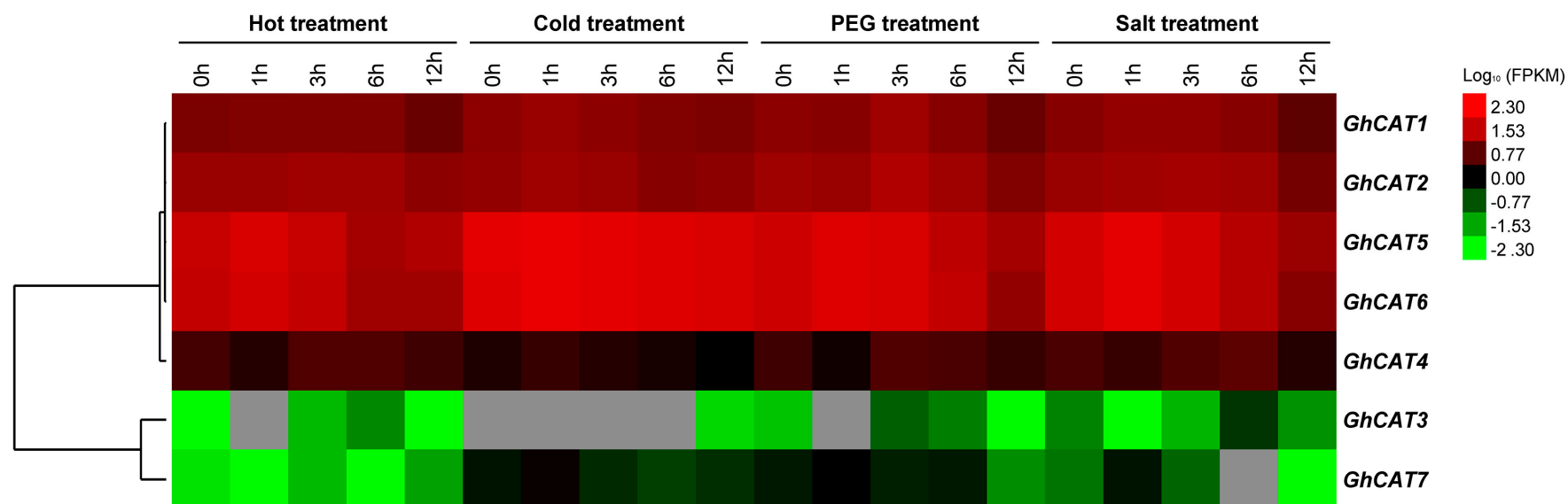
▼ Gh_A07G1555.1 LOC:A07|57085879..57086511|+;

1. Score	E value	Identities	Gaps	Strand
742.47 (822)	$0.00 \times 10^{+00}$	411/411 (100.00)	0/411 (0.00)	+/+
Query 139	ATGAGCATTGATGACGCTAATGAGACCGGGATAGATGATCCGAAGGGCCAGGGTTCTTCC	198		
Subject 1	ATGAGCATTGATGACGCTAATGAGACCGGGATAGATGATCCGAAGGGCCAGGGTTCTTCC	60		
Query 199	TCTAGAAAAAGAGGGAGCACTATTGATGTTGCTATTGAGATTGGGTCTGGCAGTGGCAGT	258		
Subject 61	TCTAGAAAAAGAGGGAGCACTATTGATGTTGCTATTGAGATTGGGTCTGGCAGTGGCAGT	120		
Query 259	GGATCAAATGCAAGGAAGAAAAGGGAAATGGAAATTCAGAGGTGAACCTACATGTTAC	318		
Subject 121	GGATCAAATGCAAGGAAGAAAAGGGAAATGGAAATTCAGAGGTGAACCTACATGTTAC	180		
Query 319	GTATGTAGTAAAAATTTACGTCCTGGAAGGCGGTGTTTCGGGCATCTGAAATCCCATCAA	378		
Subject 181	GTATGTAGTAAAAATTTACGTCCTGGAAGGCGGTGTTTCGGGCATCTGAAATCCCATCAA	240		
Query 379	AGGGAAACCCAGGGGCATTGCCTCCGCCAACATTTACGCCAACTGAGGGCTCTCCTGAA	438		
Subject 241	AGGGAAACCCAGGGGCATTGCCTCCGCCAACATTTACGCCAACTGAGGGCTCTCCTGAA	300		
Query 439	AATAATAATGATGATGAAACTAACCCGAGGGAGCAACTAGCTCCCACTCTGTTGAACTTG	498		
Subject 301	AATAATAATGATGATGAAACTAACCCGAGGGAGCAACTAGCTCCCACTCTGTTGAACTTG	360		
Query 499	GCTCTTGAAACCATGCAAAAGATGAGTGAAGACTCGAACATGAGTGTTGTT	549		
Subject 361	GCTCTTGAAACCATGCAAAAGATGAGTGAAGACTCGAACATGAGTGTTGTT	411		

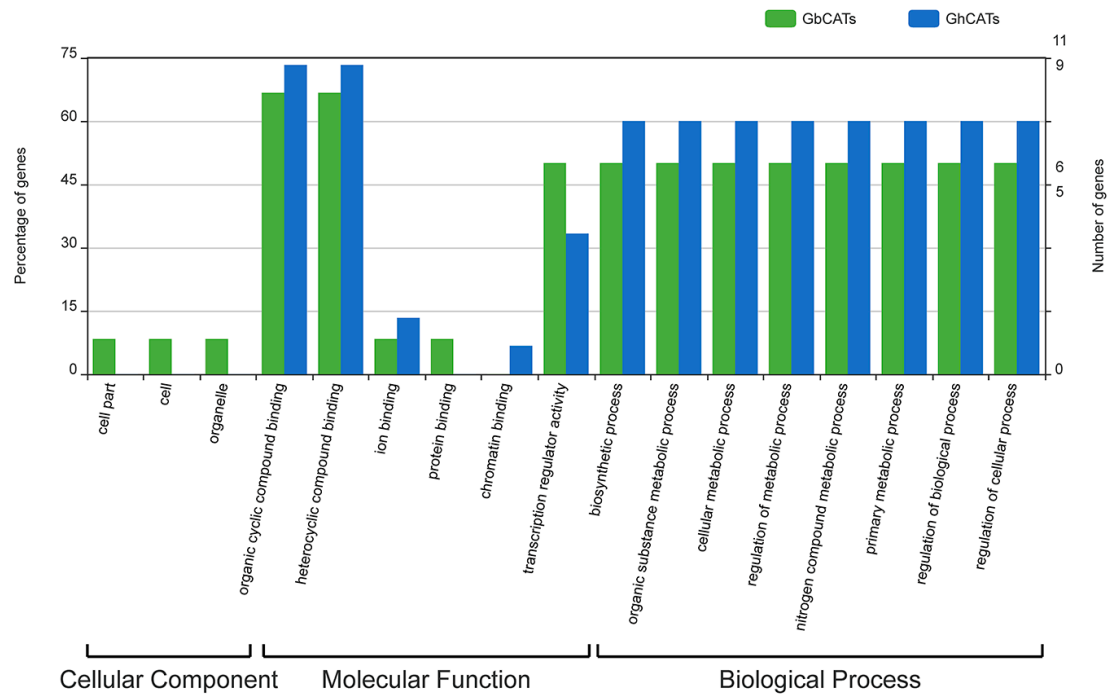
Supplementary file 10: Figure S2. The supporting alignment result of the coding sequence of *GbCAT7* (GOBAR_AA20422) and Gh_A07G1555. Query, the coding sequence of *GbCAT7*; Subject, the coding sequence of Gh_A07G1555.



Supplementary file 11: Figure S3. Phylogenetic tree of *Gossypium* and four other genomes descended from common eudicot genome ancestor in Rosids and the number of *CAT* genes identified from the genomes of these plants. The phylogenetic tree of the 8 plants used in the study was gathered from the PLAZA database (version 4.0). **WGD**, whole-genome duplication; **Mya**, million years ago; **γ**, one whole genome triplication (WGT) event probably shared by all core eudicots ~130.8 Mya; **α** and **β**, two recent WGDs within the crucifer lineage; **1** and **2**, two *Glycine* WGD events; **3**, one penta- or hexaploid duplication event in the *Gossypium* lineage after splitting from cacao lineage ~16.6 Mya; **4**, neoallopolyploids reuniting divergent *Gossypium* genomes 1-2 Mya.

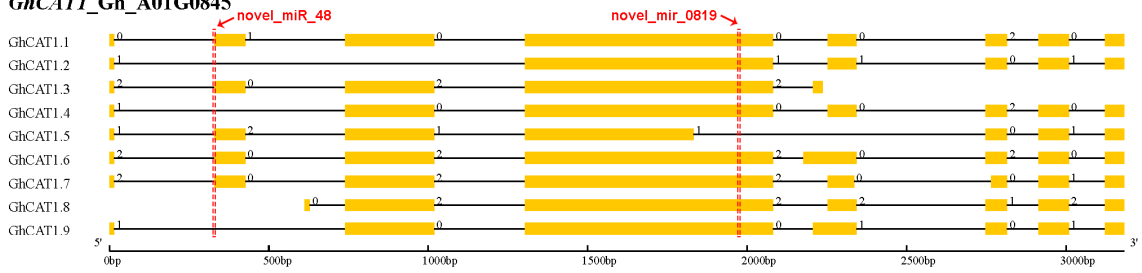


Supplementary file 12: Figure S4. Transcriptional profiling of *GhCATs* under different abiotic stress treatments. The \log_{10} of FPKMs values calculated by RNA-seq data were shown as a heat map. The heat-map showed the hierarchical clustering of the relative expression of 7 *GhCATs* under hot, cold, PEG and salt treatments. The colors of the bar shown to the right of the heat-map varied from red to green representing the relative expression levels from high to low. FPKMs data was obtained from ccNET (<http://structuralbiology.cau.edu.cn/gossypium/>) and CottonFGD (<https://cottonfgd.org/>).

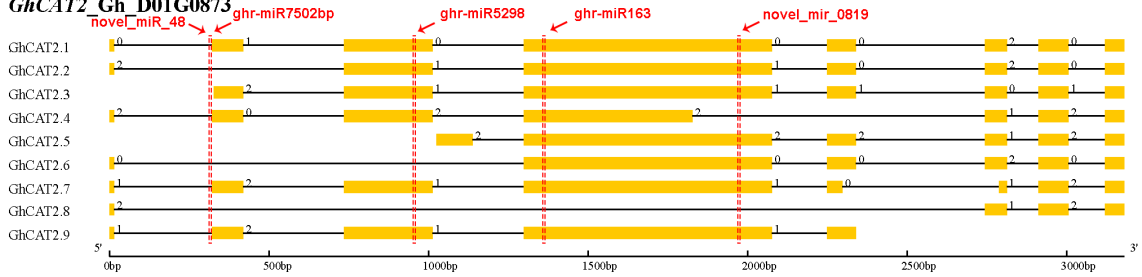


Supplementary file 13: Figure S5. Gene Ontology (GO) annotation of the TFs binding to the upland cotton and sea-island cotton *CAT* genes' putative promoter regions based on their cellular component, molecular function, and biological process. The functional enrichment analysis was carried out by the Gene Ontology Consortium (<http://geneontology.org/>) and were visualized and plotted by WEGO (version 2.0) (<http://wego.genomics.org.cn/>).

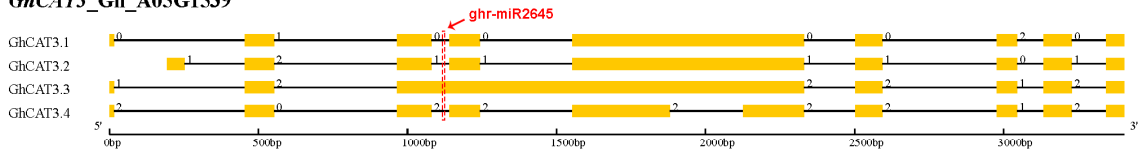
GhCAT1_Gh_A01G0845



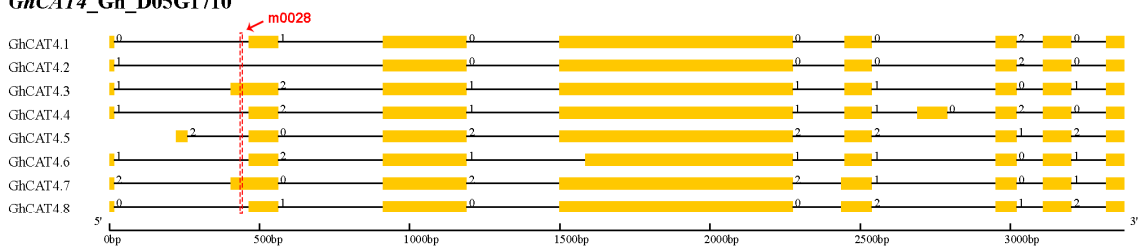
GhCAT2_Gh_D01G0873



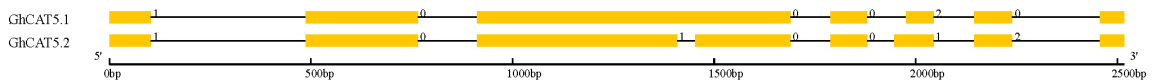
GhCAT3_Gh_A05G1539



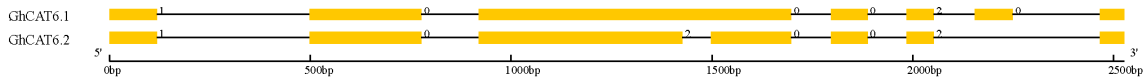
GhCAT4_Gh_D05G1710



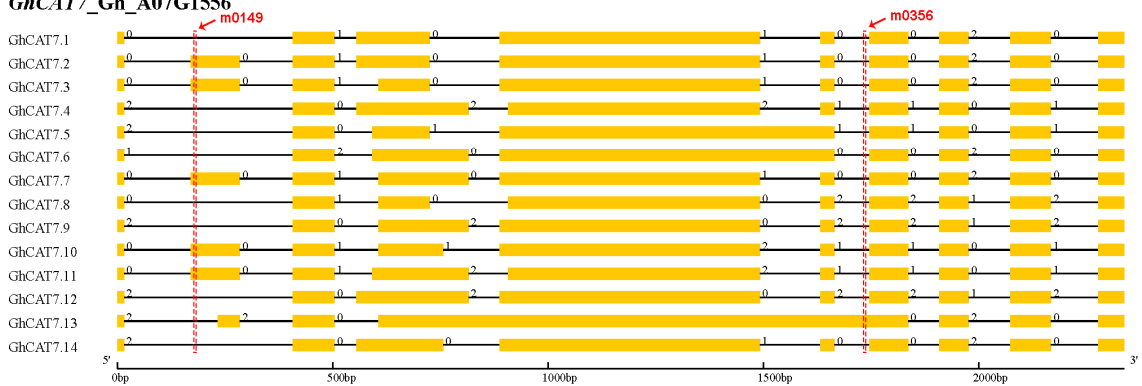
GhCAT5_Gh_A02G1698



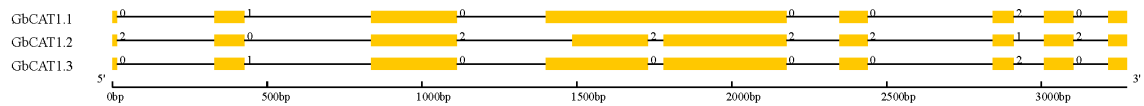
GhCAT6_Gh_D03G0021



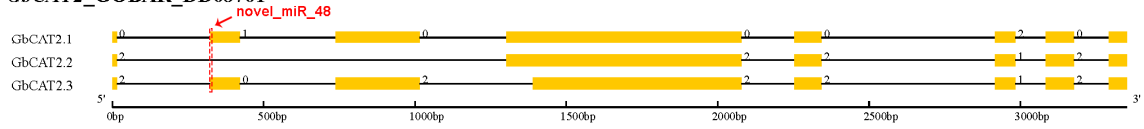
GhCAT7_Gh_A07G1556



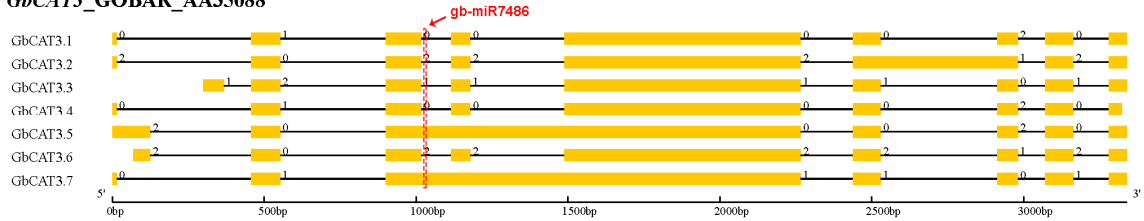
*GbCAT1*_GOBAR_AA22711



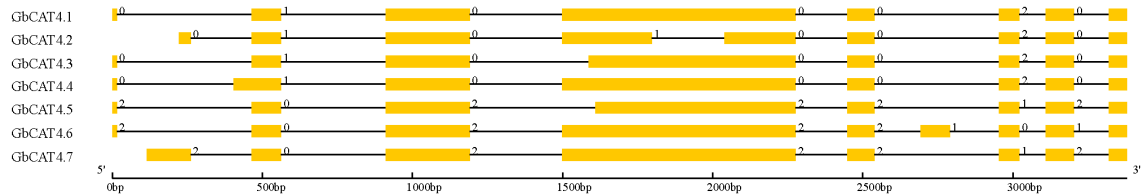
*GbCAT2*_GOBAR_DD08701



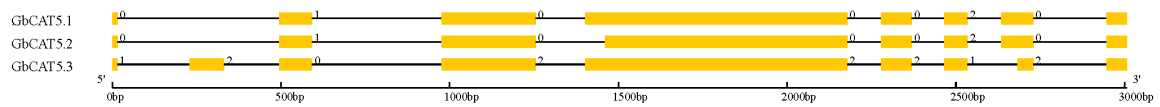
*GbCAT3*_GOBAR_AA35088



*GbCAT4*_GOBAR_DD30012



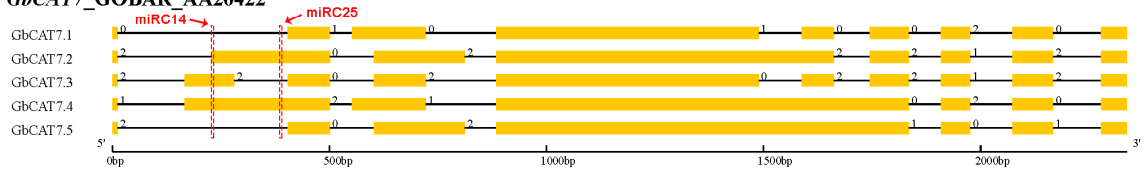
*GbCAT5*_GOBAR_AA31934



*GbCAT6*_GOBAR_AA38517

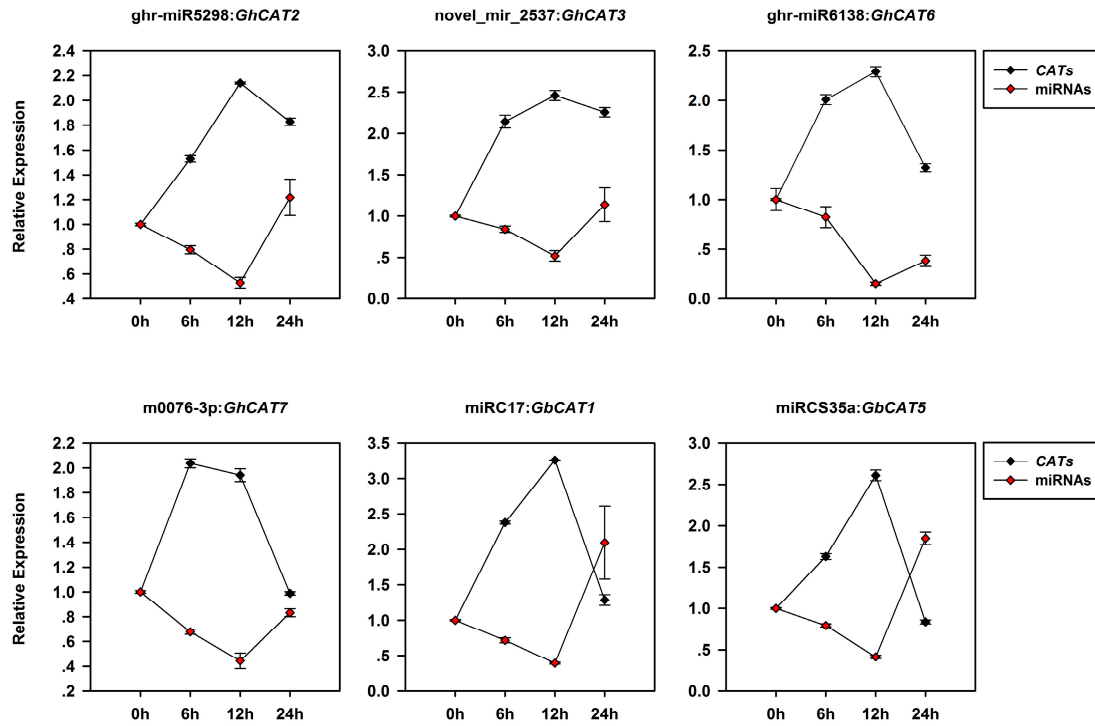


*GbCAT7*_GOBAR_AA20422

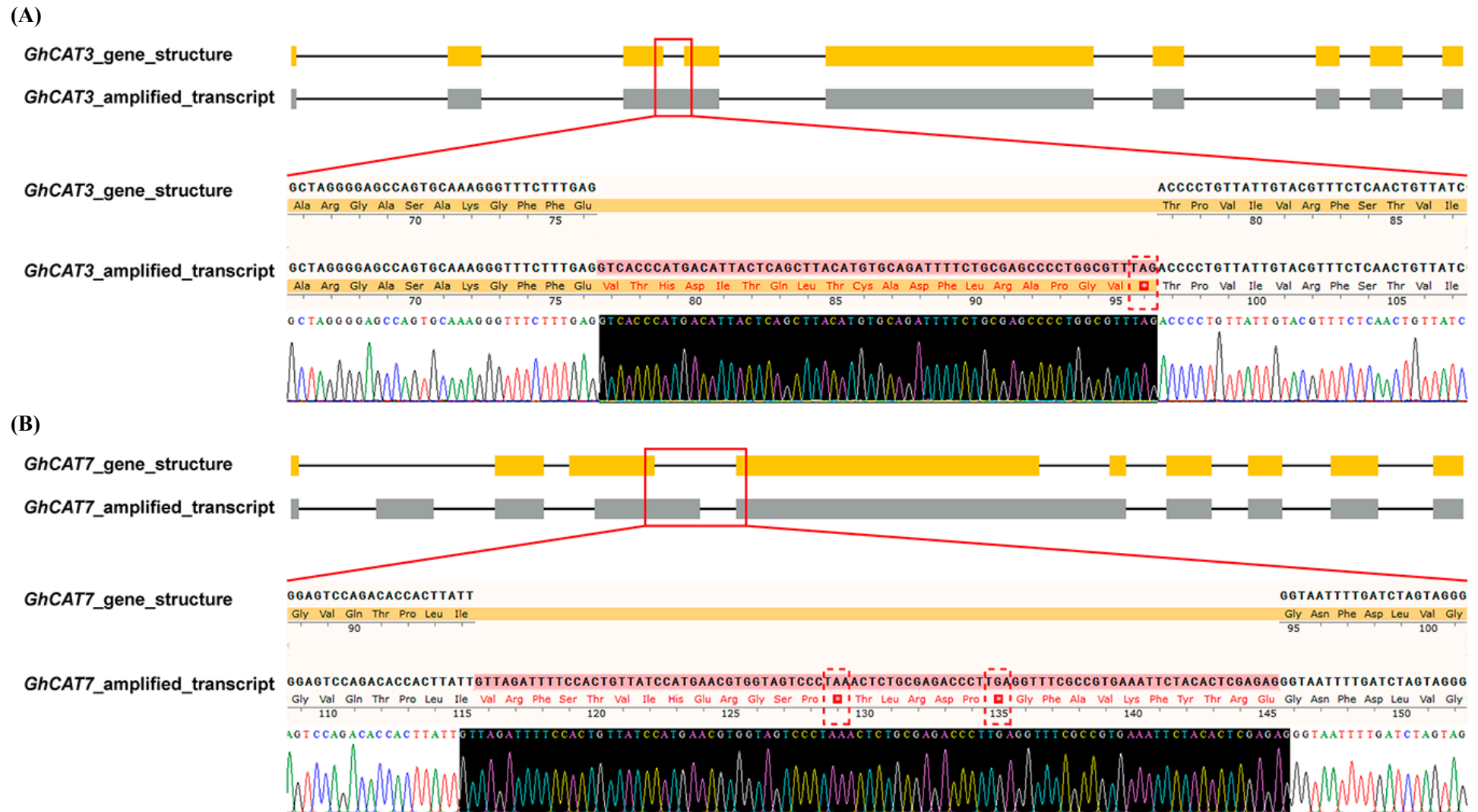


Legend:
 Exon — Intron 0 1 2: intron phase

Supplementary file 14: Figure S6. All potentially alternative spliced isoform structure of *GhCATs* and *GbCATs* full-length transcripts and identification of full-length isoforms targeted by cotton miRNAs. The first gene model of each *CAT* gene was the principle transcript, followed these were potential alternative spliced isoforms predicted with mRNA-seq data. The red dotted rectangles were the predicted miRNA targeting sites.



Supplementary file 15: Figure S7. The relative expression levels of miRNAs and their putative targeting *CAT* genes under *V. dahliae* infection treatment. qPCR quantifications were normalized to the expression of *UBQ7* of cotton. Error bars represent standard deviations of mean value from three biological replicates. ANOVA (analysis of variance) was calculated using DPS and $p < 0.05$ was considered statistically significant.



Supplementary file 16: Figure S8. (A, B) Amplification of *GhCAT3* (Gh_A05G1539) and *GhCAT7* (Gh_A07G1556) by PCR with the designed primer pairs for cloning their transcripts. The largest-rate transcripts of *GhCAT3* and *GhCAT7* were shown (21/24 and 22/48 respectively in *GhCAT3* and *GhCAT7*). PCR results were aligned with gene structure annotation and were verified by Sanger sequencing. Red dotted rectangle represented the premature termination codons (PTC).