# Systematic analysis of NAC transcription factors in *Gossypium barbadense* uncovers their roles in response to Verticillium wilt

Zhanji Liu[1*], Mingchuan Fu[1], Hao Li[1], Yizhen Chen[1], Liguo Wang[1], Renzhong Liu[1]

[1]Key Laboratory of Cotton Breeding and Cultivation in Huang-Huai-Hai Plain, Ministry of Agriculture, Cotton Research Center of Shandong Academy of Agricultural Sciences, Jinan 250100, China

*Corresponding Author:

Zhanji Liu

202 Gongyebei Road, Jinan, Shandong, 250100, China

E-mail address: scrcliuzhanji@sina.com

## ABSTRACT

25  As one of the largest plant-specific gene families, the NAC transcription factor gene family plays

26  important roles in various plant physiological processes that are related to plant development,

27  hormone signaling, and biotic and abiotic stresses. However, systematic investigation of the NAC

28  gene family in sea-island cotton (*Gossypium babardense* L.) has not been reported, to date. The

29  recent release of the complete genome sequence of sea-island cotton allowed us to perform

30  systematic analyses of *G. babardense* NAC (*GbNAC*) genes. In this study, we performed a

31  genome-wide survey and identified 270 *GbNAC* genes in the sea-island cotton genome. Genome

32  mapping analysis showed that *GbNAC* genes were unevenly distributed on 26 chromosomes.

33  Through phylogenetic analyses of GbNACs along with their Arabidopsis counterparts, these

34  proteins were divided into 10 groups (I–X), and each contained a different number of GbNACs

35  with a similar gene structure and conserved motifs. One hundred and fifty-four duplicated gene

36  pairs were identified, and almost all of them exhibited strong purifying selection during evolution.

37  In addition, various *cis*-acting regulatory elements in *GbNAC* genes were found to be related to

38  major hormones, defense and stress responses. Notably, transcriptome data analyses unveiled the

39  expression profiles of 62 *GbNAC* genes under Verticillium wilt (VW) stress. Furthermore, the

40  expression profiles of 15 *GbNAC* genes tested by quantitative real-time PCR (qPCR) demonstrated

41  that they were sensitive to methyl jasmonate (MeJA) and salicylic acid (SA) treatments and that

42  they could be involved in pathogen-related hormone regulation. Taken together, the genome-wide

43  identification and expression profiling pave new avenues for systematic functional analysis of

44  *GbNAC* candidates, which may be useful for improving cotton defense against VW.

45  **Keywords:** sea-island cotton; NAC transcription factor; genome-wide survey; Verticillium wilt

46

47

## INTRODUCTION

49   The plant-specific *NAC* genes (*NA**M*, no apical meristem; *A**TAF*, Arabidopsis transcription

50   activation factor; and *C**UC*, cup-shaped cotyledon) form one of the largest families of transcription

51   factors (*Nuruzzaman et al., 2013*). Typically, NAC proteins harbor a highly conserved NAC

52   domain at the N-terminal and a variable transcriptional regulatory region (TR) at the C-terminal.

53   The NAC domain can be further divided into five subdomains (A–E) and functions as DNA

54   binding, nuclear localization, and formation of homodimers or heterodimers, while the TR region

55   is responsible for transcription regulation as either an activator or a repressor (*Olsen et al., 2005*).

56   *NAM*, the first *NAC* gene, was discovered in Petunia and functions in determining positions of

57   shoot apical meristems and primordia (*Souer et al., 1996*). Since then, a large number of *NAC*

58   genes have been identified from diverse plant species (*Nuruzzaman et al., 2010*; *Liu et al., 2019*).

59   Accumulated evidences indicate that the functions of *NAC* genes are associated with almost every

60   biological process in plants, such as leaf senescence (*Wang et al., 2015*; *Zhao et al., 2016*),

61   secondary cell wall formation (*Zhang et al., 2018*), and hormone signaling (*Takasaki et al., 2015*).

62   Notably, a number of NAC genes, especially ATAF subfamily members, were proven to serve as

63   critical regulators in plant defense against biotic and abiotic stresses (*Nuruzzaman et al., 2013*;

64   *Karanja et al., 2017*). In Arabidopsis, *ANAC032* was induced by bacterial pathogen *Pst*

65   (*Pseudomonas syringae* pv. *tomato* DC3000) infection, and SA and jasmonic acid (JA) treatments.

66   Furthermore, transgenic Arabidopsis plants overexpressing *ANAC032* showed strongly enhanced

67   resistance to *Pst*, while the *ANAC032* knockout mutant exhibited increased susceptibility to *Pst*

68   (*Allu et al., 2016*). In rice, the expressions of *SNAC1* were enhanced under drought, salt, and cold

69   stresses. Overexpression *SNAC1* in rice resulted in increased tolerance to drought (*Hu et al., 2006*).

70   Similarly, transgenic *OsNAC111* showed improved resistance to blast fungus in rice by regulating

71   the expression of several defense-related genes (*Yokotani et al., 2014*).

72   Cotton is the most important natural fiber crop, and is also used as a food crop due to high levels

73   of vegetable oil and protein in cottonseeds (*Li et al., 2009*). However, cotton production can be

74   dramatically decreased by the occurrence of VW, a devastating vascular disease caused by

75   *Verticillium dahliae* (*Sun et al., 2013*). Typically, infected plants show leaf chlorosis, leaf

76   shedding, vascular discoloration, wilting, and plant death. In general, upland cotton (*Gossypium*

77   *hirsutum*), accounting for about 90% of annual world cotton production, is susceptible to VW,

78   whereas sea-island cotton, accounting for approximately 5% of annual world cotton output, is

79   immune to VW. Thus, extensive efforts have been made to investigate the molecular mechanism

80   of sea-island cotton resistance to VW (*Zhang et al., 2015*). Recently, a *NAC* gene *GbNAC1* was

81   identified from *G. barbadense*. Overexpression of *GbNAC1* in Arabidopsis can significantly

82   enhance resistance to VW, implying that *G. barbadense NAC* genes might play pivotal roles in

83   biotic stress resistance (*Wang et al., 2016*). The availability of diploid and tetraploid cotton

84   genomes has allowed scientists to identify the *NAC* gene family members at the genome-wide

85   scale, such as 145 genes in *G. raimondii* (*Shang et al., 2013*), 141 in *G. arboreum* (*Shang et al.,*

86   *2016*; *Fan et al., 2018*), and 283 in *G. hirsutum* (*Sun et al., 2018*). However, systematic analysis

87   of *NAC* genes in *G. babardense* has not been completed. *G. babardense* (AD$_2$) and *G. hirsutum*

88   (AD$_1$) are allotetraploids, which evolve from transoceanic hybridization between an A-genome

89   species immigrated from Africa and a native American D-genome species, while *G. arboreum*

90   (A$_2$) and *G. raimondii* (D$_5$) are diploids resembling the A-genome and D-genome progenitor,

91   respectively (*Hu et al., 2019*). In this study, we performed a genome-wide survey and identified

92   270 *GbNAC* genes in the latest *G. babardense* genome released by *Wang et al.* (2019). These

93   *GbNAC* genes were classified into 10 groups on the basis of sequence similarity. Sequence

94   comparison among *GbNAC* genes revealed the presence and distribution of duplicated genes.

95   Additionally, to identify *GbNAC* candidate genes associated with VW resistance, we analyzed the

96   expression patterns of *GbNAC* genes using available transcriptome data of *G. babardense* cv. 7124

97   inoculated with the fungal pathogen *V. dahliae*. Subsequently, qPCR-based gene expression

98   profiling demonstrated that selected *GbNAC* candidate genes might be involved in MeJA and SA

99   regulation. This study provides comprehensive information about sea-island cotton *NAC* genes, as

100    well as a foundation for in-depth functional analysis of novel *GbNAC* candidate genes, which may

101    be useful for the improvement of pathogen resistance in cotton.

## MATERIALS AND METHODS

### Identification of *NAC* genes in the sea-island cotton genome

104    We downloaded the genome sequences (v. HAU) of sea-island cotton from the CottonGen

105    database (https://www.cottongen.org/, *Wang et al., 2019*). The Hidden Markov Model (HMM)

106    profile of the NAC domain (PF02365) was retrieved from the Pfam database

107    (http://pfam.xfam.org/, *Finn et al., 2016*). The HMMER program was used to search NAC protein

108    in the sea-island cotton genome (*Finn et al., 2011*) with an *E*-value cutoff of $e^{-5}$. Then, all the

109    putative proteins were confirmed by the Pfam and SMART database (http://smart.embl-

110    heidelberg.de/, *Letunic and Bork, 2018*). The MW (molecular weight) and pI (theoretical

111    isoelectric point) of each NAC protein were predicted by the online software ExPASy

112    (https://www.expasy.org/, *Artimo et al., 2012*). The TMHHM server (v. 2.0,

113    http://www.cbs.dtu.dk/services/TMHMM/) was used to identify membrane-bound NAC proteins

114    (*Krogh et al., 2001*). CELLO (v. 2.5, subCELlular Localization predictor,

115    http://cello.life.nctu.edu.tw/, *Yu et al., 2006*) was used to predict the subcellular localization of

116    GbNACs.

### Multiple alignments, phylogenetic analysis, gene duplication and synteny analysis

118    Multiple sequence alignments were performed with the NAC domain sequences of the GbNAC

119    proteins using MEGA X (https://www.megasoftware.net/, *Kumar et al., 2018*). A phylogenetic

120    tree was constructed by the neighbor-joining method with the following parameters: Poisson

121    correction, pairwise deletion, and 1000 bootstrap replicates. Gene duplications were analyzed with

122    two major criteria, that is, the length of the aligned sequence covers more than 75% of the longer

123    gene and similarity of the aligned regions is greater than 75% (*Vatansever et al.,* 2016). Alignment

124    of the coding sequences of duplicated genes was performed by the Clustal X (v. 2.0) program

125   (*Larkin et al.,* 2007), and the values of nonsynonymous (Ka) and synonymous (Ks) substitution

126   rates were calculated using KaKs_Calculator package (*Zhang et al.,* 2006) *via* model averaging.

127   The approximate date of duplication events (million years ago, Mya) was estimated using the

128   formula T = Ks/2λ × $10^{-6}$, on the basis of molecular clock rate of 2.6 × $10^{-9}$

129   substitutions/synonymous site for cotton (*Liu et al.,* 2015). The relationships of duplicated genes

130   were illustrated with the Circos program (*Krzywinski et al.,* 2009). MCScanX was used to detect

131   the synteny of *NAC* genes between *G. barbadense* and the other plant species (*Wang et al., 2012*).

132   **Gene structure, chromosomal mapping, and conserved motif analysis**

133   The gene structures were determined using the CDS and DNA sequences of *GbNAC* genes and

134   visualized by the Gene Structure Display Server (http://gsds.cbi.pku.edu.cn/, *Hu et al., 2015*). The

135   positions of these *GbNAC* genes were determined by using the nucleotide sequence as a query to

136   search against the *G. barbadense* genome. In addition, chromosomal localization map was

137   constructed by using the MapChart (v. 2.32) program (*Voorrips 2002*). In order to identify the

138   conserved motifs among all the *GbNAC* genes, their protein sequences were subjected to the online

139   software MEME (http://meme-suite.org/tools/meme, *Bailey et al., 2015*) using default parameters

140   with exception for number of motif. The number of motifs was set to 20.

141   *Cis*-acting regulatory element and miRNA target analysis

142   For *cis*-acting regulatory element analysis, we retrieved 1,500 bp DNA sequences up-stream from

143   the transcription start site from the newly released *G. barbadense* genome sequences (*Wang et al.,*

144   *2019*) and then screened them in the PlantCare

145   (http://bioinformatics.psb.ugent.be/webtools/plantcare/html/, *Lescot et al., 2002*) and PLACE

146   (https://www.dna.affrc.go.jp/PLACE/, *Higo et al., 1999*) databases. For miRNA target analysis,

147   we downloaded all the mature miRNA sequences of Gossypium from the miRBase database (v.

148   22.1) (http://www.mirbase.org/). The online sever psRNATarget

149   (http://plantgrn.noble.org/psRNATarget/, *Dai et al., 2018*) was used to predict the miRNA target.

150   **Gene expression analysis under Verticillium wilt stress**

151   The transcriptome data of sea-island cotton were obtained from the Sequence Read Archive under

152   accession          No.          SRP03537          at          the          NCBI          website

153   ([http://www.ncbi.nlm.nih.gov/sra/?term=SRP035371](http://www.ncbi.nlm.nih.gov/sra/?term=SRP035371), *Chen et al., 2015*), where the plants were

154   inoculated with *V. dahliae*. Briefly, two-week-old seedlings of *G. barbadense* resistant cultivar

155   7124 were inoculated with the high virulence V991 defoliating strain of *V. dahliae* ($5 \times 10^6$

156   spores/mL) by the root-dip method for 2, 6, 12, 24, 48, and 72 h. Then, the samples, including the

157   mock-inoculated (control) and six inoculated, were collected for RNA sequencing. We retrieved

158   the expression data of *GbNAC* genes from the root under *V. dahliae* infection (*Chen et al., 2015*).

159   The hierarchical clustering and the heatmap-based expression profiles of *GbNAC* genes were

160   performed using ClustVis (*Metsalu and Vilo 2015*). The Venn map of differentially expressed

161   *GbNAC* genes was constructed using the UpSetR package (*Lex et al., 2014*).

162   **RNA isolation and qPCR analysis**

163   *G. barbadense* cv. 7124 seeds were cultivated in commercial soil at 28°C with a photoperiod of

164   16 h light/8 h dark. Two-week-old seedlings were gently uprooted, rinsed and cultivated in

165   Hoagland solution for two days. Then these seedlings were treated with Hoagland solution

166   containing 0.1 mM MeJA and 1 mM SA, respectively. Roots were sampled from three biological

167   replicates after treatment for 1, 2, 6, and 12 h, then immediately immersed in liquid nitrogen and

168   stored at -80°C for qPCR.

169   The total RNA from the root samples treated with MeJA and SA and the control (roots from

170   Hoagland solution without hormone) was extracted using Trizol (Invitrogen). The first strand

171   cDNA was generated from 1 µg of total RNA using a PrimerScript™ 1st Strand cDNA Synthesis

172   Kit (Takara, Dalian, China). qPCR was performed with three replicates using an ABI QuantStudio

173   5 Real-Time PCR System (Thermo Fisher Scientific, Waltham, MA, USA) and SYBR® Premix Ex

174   Taq™ (Takara, Dalian, China). The amplification procedure was as follows: one cycle at 95°C for

175   3 min; then 40 cycles at 95°C for 15 s, 60°C for 15 s. The cotton *actin* (AF059484) was selected

176   as the internal reference gene (*Zhang et al., 2013*). Gene expression levels were calculated

177    according to the $2^{-\Delta\Delta Ct}$ method described by *Livak & Schmittgen* (*2001*). The primers used for

178    qPCR are listed in Table S1.

## RESULTS

**Identification and phylogenetic analysis of NAC family members in *Gossypium barbadense***

181    A total of 270 NAC proteins were identified from *G. barbadense* and named as GbNAC001 to

182    GbNAC270. All of the 270 GbNACs contained the NAC domain (PF02365) based on Pfam and

183    SMART tests. The lengths of GbNAC proteins ranged from 154 (GbNAC059) to 959

184    (GbNAC175) amino acids with MW from 17.65 to 107.75 kDa, and pI from 4.67 to 9.79.

185    Subcellular localization of GbNACs was predicted using the online software CELLO

186    (http://cello.life.nctu.edu.tw/). Among the 270 GbNAC proteins, three were predicted to be

187    mitochondrial proteins (GbNAC031, 032, and 255); five were located in the chloroplast

188    (GbNAC014, 036, 116, 174, and 249); 10 were extracellular; 23 were cytoplasmic, and the rest

189    were localized in the nucleus. These results are similar to those of cucumber (*Liu et al., 2018*).

190    Detailed information including gene locus, chromosome location, exon number, sequence length,

191    MW, pI, Arabidopsis orthologous locus and subcellular location of all identified GbNAC proteins

192    is provided in Supplementary Table S2.

193    The NAC domain sequences of GbNAC proteins were used to construct a neighbor-joining

194    phylogenetic tree. As a result, the 270 GbNACs were classified into 10 groups which were named

195    as I to X (Figure 1). Group VI contained the most NAC members with 47 GbNACs, followed by

196    group I with 39 GbNACs. Group III had the least NAC members with only ten GbNACs.

197    Additionally, 35 GbNACs with high similarity to Arabidopsis ATAF members were clustered in

198    group IV. Arabidopsis ATAF members play pivotal roles in the responses to biotic and abiotic

199    stresses. These results suggest that NAC members in group IV may have similar functions with

200    Arabidopsis ATAF members.

**Chromosomal locations, duplications, and synteny analysis of the *GbNAC* genes**

202    To determine the chromosomal distribution of the *GbNAC* genes, we searched the sea-island cotton

203    genome database using blastn and the DNA sequence of each *GbNAC* gene. The results suggested

204    that 263 *GbNAC* genes (97.4%) were mapped to 26 chromosomes. Specifically, 132 *GbNAC* genes

205    were distributed in the A-subgenome, and 131 were located in the D-subgenome (Figure 2). In

206    addition, five genes (*GbNAC011*, *GbNAC012*, *GbNAC049*, *GbNAC050*, and *GbNAC069*) were

207    anchored in four A-subgenome scaffolds, and two genes (*GbNAC148* and *GbNAC262*) were found

208    in two D-subgenome scaffolds. The number of *GbNAC* genes distributed on each chromosome

209    was uneven. Chromosome D11 contained the highest number of *NAC* genes, with 15 *GbNAC*

210    genes. In contrast, chromosome A10 contained the least number of *NAC* genes, with only five

211    genes. Interestingly, many *GbNAC* genes were clustered within a short distance, such as the top of

212    A11 and D11 and the bottom of A08 and D08.

213    Gene duplication including tandem duplication, segmental duplication, and whole-genome

214    duplication (WGD) is a major driving force in the evolution of plants. The origin of multigene

215    family is due to gene duplication that arose from region-specific duplication or WGD (*Du et al.,*

216    *2012*). To reveal the expansion mechanism of the *GbNAC* gene family, gene duplication analysis

217    was performed using blastn and the coding sequences (cds) of all *GbNAC* genes. In all, we

218    identified 148 pairs (212 *GbNAC* genes) of segmental duplications, three pairs of tandem

219    duplications (GbNAC011/GbNAC012, GbNAC024/GbNAC025, and GbNAC082/GbNAC083),

220    and one triplicate repeat of tandem duplications (GbNAC212/GbNAC213/GbNAC214) (Figure

221    3). One hundred and thirty-two duplication gene pairs occur between the A-subgenome and the D-

222    subgenome, while only 12 and 10 duplication gene pairs occur within the A-subgenome and the

223    D-subgenome, respectively. These genes represent approximately 81.9% (221 of 270) of the

224    *GbNAC* genes, indicating their origin may be from sea-island cotton genome duplication events.

225    Using the Ka and Ks of each duplicated *GbNAC* gene pair, we found that the Ks values of all gene

226    pairs were between 0.008 and 0.912. Specifically, the Ks values of 87 (58.78%) gene pairs were

227    less than 0.05. Additionally, the Ka/Ks value of each gene pair was calculated and the Ka/Ks values

228    of 140 gene pairs (94.59%) were less than 1, which indicated these genes had evolved under strong

229    purifying selection. Furthermore, eight gene pairs (Ka/Ks >1) may evolve under strong positive

230    selection after duplication. Moreover, we also calculated the approximate date of duplication

231    events. The duplication events of *GbNAC* genes occurred from 1.58 Mya (Ks = 0.008) to 175.29

232    Mya (Ks = 0.912), with a mean of 59.31 Mya (Ks = 0.154). Detailed information including

233    duplication gene pairs, chromosome location, duplication type, Ka, Ks, Ka/Ks and approximate

234    duplication date (Mya) of all identified duplicated *GbNAC* genes is provided in Supplementary

235    Table S3.

236    To detect the synteny of *NAC* genes, we performed a collinearity analysis between *G. barbadense*

237    and the other four plant species (*Arabidopsis thaliana*, *G. arboretum*, *G. raimondii* and *G.*

238    *hirsutum*) using MCScanX. Previously, 117, 141, and 145 *NAC* genes were revealed in *A. thaliana*

239    (*AtNAC, Nuruzzaman et al., 2010*), *G. arboretum* (*GaNAC, Shang et al., 2016*), and *G. raimondii*

240    (*GrNAC, Shang et al., 2013*), respectively. For *NAC* genes in *G. hirsutum* (*GhNAC*), we identified

241    272 *GhNAC* genes in the upland cotton genome that was released recently (*Wang et al., 2019*).

242    Totally, 945 *NAC* genes were used to evaluate synteny relationship in this study. As a result, we

243    found 421 paired collinearity relationships between 247 *GbNAC* and 234 *GhNAC* genes, 241 pairs

244    between 241 *GbNAC* and 133 *GrNAC* genes, 181 pairs between 120 *GbNAC* and 54 *AtNAC* genes,

245    and 142 pairs between 141 *GbNAC* and 81 *GaNAC* genes (Figure 4 and Table S4). Notably, 69

246    *GbNAC* genes are collinear with NAC genes from the other four species (Table S4).

247    ***GbNAC* gene structures and conserved motifs**

248    To better understand the relationship between gene function and evolution among the *GbNAC*

249    genes, the exon/intron organization and conserved motifs were analyzed (Figure 5). The number

250    of exons ranged from 1 to 10. Most *GbNAC* genes (182/270, 67.41%) had three exons (Figure 5),

251    although *GhNAC059*, *GhNAC105*, and *GhNAC238* contained only one exon, and *GhNAC129* and

252    *GhNAC175* contained 10 exons, the highest number of all genes. We also found that *GbNAC* genes

253    in the same group had a similar gene structure. For example, all 12 members in group IX had three

254    exons. Among the 39 members in group I, all had three exons except *GbNAC060* and *GbNAC145*

255    (Figure 5).

256    Twenty conserved motifs were identified among the 270 GbNAC proteins (Figure S1 and Table

257    S5). As a result, all GbNAC proteins contain a conserved NAC domain at the N-terminal, which

258    includes five subdomains (A−E, Figure 6). Notably, the members with high similarity in the same

259    group shared a common motif composition. For example, GbNAC213 and GbNAC214 were found

260    to contain the same three motifs (Figure 5). This finding indicates that these genes may have

261    similar functions. Most GbNAC proteins (257/270, 95.19%) contain 2−5 conserved motifs.

262    However, seven GbNAC proteins (GbNAC033, GbNAC068, GbNAC092, GbNAC109,

263    GbNAC200, GbNAC226, and GbNAC242) contain only one motif and six GbNAC proteins

264    (GbNAC017, GbNAC020, GbNAC074, GbNAC160, GbNAC206, and GbNAC253) contain six

265    motifs. Additionally, we found that subdomains A (246/270), and/or C (200/270), and/or D

266    (250/270) are present in most GbNAC proteins. Furthermore, we found that 231 GbNAC proteins

267    contained the conserved histidine residue in subdomain D, which influences homodimerization

268    and DNA binding of NAC proteins (*Kang et al., 2018*).

269    Membrane-bound NAC proteins feature a distinctive transmembrane motif (TM) at either the C

270    terminal region or the N terminal region and play vital roles in plant defense against abiotic stresses

271    (*Kim et al., 2010*; *Sun et al., 2018*). In the sea-island cotton genome, 22 membrane-bound GbNAC

272    proteins were identified (Table S2). Notably, 20 membrane-bound GbNAC proteins contain one

273    TM at the C terminal and two GbNAC members (GbNAC110 and GbNAC243) contain one TM

274    at the N terminal. These membrane-bound GbNAC proteins are distributed in three groups: seven

275    in group V, 11 in group VII, and four in group VIII.

276    ***Cis*-acting regulatory element and miRNA target analysis**

277    To better understand the transcriptional regulation mechanisms of *GbNAC* genes, we characterized

278    the *cis*-acting regulatory elements within a 1500 bp upstream region from the transcription start

279    site using PlantCARE and PLACE database (Table S6). A large number of *cis*-acting regulatory

280    elements were identified in promoter sequences of 270 *GbNAC* genes (Figure S3 and Table S7).

281    Common regulatory elements such as TATA-box and CAAT-box were present in all *GbNAC*

282    genes. Meanwhile, we identified 11 *cis*-acting regulatory elements related to hormone responses.

283    Among these, AuxRR-core and TGA-element, auxin-responsive element; ABRE, MYB, and

284    MYC, *cis*-element involved in abscisic acid (ABA) responsiveness; GARE, P-box, and TATC-

285    box, involved in gibberellin responsiveness; ERE, ethylene responsive element; CGTCA-motif,

286    MeJA responsive elements; and TCA-element, involved in SA responsiveness. Notably, all of the

287    270 *GbNAC* genes contained at least one hormone-responsive element (Figure S3). MYB, MYC,

288    and ERE were available in at least 80% of the *GbNAC* genes. Additionally, promoter sequences

289    of some *GbNAC* genes also contained several elements involved in biotic and abiotic stress

290    responses, including pathogen defense (AT-rich and TC-rich repeat), drought (MBS), cold (DRE

291    and LTR), anaerobic stress (ARE), and wounding (WUN-motif). Thus, *GbNAC* genes could be

292    regulated by diverse hormone and environmental changes.

293    Recent reports have defined a subset of genes from the NAC-domain gene family as potential

294    targets of miRNAs. To determine the involvement of miRNAs in regulating the expression of

295    *GbNAC* genes, putative miRNA targets were determined in the 270 *GbNAC* genes using the online

296    sever psRNATarget with the default parameters except Expection ($\leq$ 3.0). Totally, 43 *GbNAC*

297    genes were predicted as the targets of 21 known miRNAs (Table S8). Seven *GbNAC* genes were

298    each predicted to be the targets of two miRNAs and *GbNAC007* was the target of three miRNAs

299    (gra-miR8634, gra-miR8786a and gra-miR8786b). Specifically, miR164 targets 14 *GbNAC* genes,

300    which are all from group II (Table S8). In Arabidopsis, the miR164 family (ath-miR164a/b/c)

301    guides the cleavage of the transcripts of five *NAC* genes (*NAC1/At1G56010*, *CUC1/At3g15170*,

302    *CUC2/At5g53950*, *ANAC080/At5g07680*, and *ANAC100/At5g61430*) that function in the

303    regulation of plant growth and development such as lateral root emergence, formation of

304    vegetative and floral organs, and age-dependent cell death (*Fang et al., 2014*; *Hernandez & Sanan-*

305    *Mishra 2017*). The 14 miR164-targeted *GbNAC* genes and the five Arabidopsis *NAC* genes were

306    clustered into group II, indicating that these genes have high sequence identity and may have a

307    similar function.

308    **Expression profile of *GbNAC* genes in response to Verticillium wilt and hormones**

309    We used publicly available transcriptome data to assess expression of *GbNAC* genes in roots under

310    VW stress. As a result, 239 (88.52%) *GbNAC* genes were identified to be expressed in infected

311    root samples (Figure 7). Three patterns (Pattern I–III) of expression were revealed. Genes from

312    Pattern I have low expression levels in the control and then were up-regulated gradually during

313    VW infection. Genes from Pattern II, in contrast to Pattern I, were highly expressed in the control

314    and then were subsequently down-regulated during VW infection. Genes from Pattern III showed

315    high expression levels in the control, then were down-regulated at the early stages (2, 6 and 12 h)

316    of inoculation, and finally up-regulated at the late stages (24, 48 and 72 h) of inoculation. There

317    are 130 duplicated gene pairs among the 239 *GbNAC* genes. Most duplicated gene pairs (80.77%)

318    demonstrated similar expression patterns, suggesting that duplicated genes are functionally

319    redundant. However, some duplicated genes have a divergent expression. For example,

320    *GbNAC082* and *GbNAC083* are tandem duplication genes. *GbNAC082* had considerably high

321    expression in the control and was up-regulated by VW stress at 2 h, while *GbNAC083* had low

322    expression in the control and was down-regulated remarkably by VW stress at 2 h.

323    The expression of 192 *GbNAC* genes was significantly altered ($|\log_2^{Fold}| \geq 1$) in roots inoculated

324    with *V. dahliae* as compared to the control in at least one time point (Figure S2). 103, 130, 123,

325    135, 130, and 128 *GbNAC* genes were significantly induced at 2, 6, 12, 24, 48, and 72 h after

326    inoculation, respectively, which indicates the number of differentially expressed *GbNAC* genes

327    was similar among different time points. Additionally, the expression of 62 *GbNAC* genes

328    including 15 up-regulated and 47 down-regulated genes were significantly altered at all of the six

329    time points (Figure S2 and Table S9). Eighteen duplicated gene pairs were revealed among the 62

330    *GbNAC* genes. Notably, *GbNAC* genes in each duplicated gene pair had similar expression

331    patterns. The expression of previously reported sea-island cotton *NAC* gene *GbNAC1* (*GbNAC220*

332    in this study) was also altered. *GbNAC1*, a positive regulator involved in cotton resistance to *V.

333    dahliae*, was down-regulated in infected roots compared to the control, which is consistent with

334    qPCR results reported by *Wang et al.* (2016). These results imply that *GbNAC* genes could play

335 crucial roles in the defense of cotton against VW.

336 Phytohormones, such as MeJA and SA, regulate plant defenses against diverse pathogens. In order

337 to identify hormone-responsive *GbNAC* genes, *G. barbadense* cv. 7124 seedlings were treated

338 with MeJA and SA, and the changes in transcript abundance of 15 genes selected from the 62

339 differentially expressed *GbNAC* genes were analyzed by qPCR. As shown in Figure 8, all the

340 *GbNAC* genes tested were sensitive to the hormones MeJA and SA, but the levels of sensitivity

341 were substantially different at the four time points. In general, most tested *GbNAC* genes were up-

342 regulated under MeJA treatment, but down-regulated under SA treatment. Specifically, the

343 expression of two *GbNAC* genes (*GbNAC014* and *GbNAC164*) from group IV were up-regulated

344 at all the four time points, and had at least 186-fold and 225-fold increase at 6 h compared with

345 control, respectively (Figure 8).

346 **DISCUSSION**

347 **Characterization of *GbNAC* genes**

348 In this study, we performed a genome-wide analysis of the sea-island cotton *GbNAC* gene family

349 to investigate their potential functions in response to VW. As a result, 270 *GbNAC* genes were

350 revealed in the *G. barbadense* genome, which is similar to that found in *G. hirsutum* (283 *NAC*

351 members, *Sun et al., 2018*), but is about twice as much as that found in *G. raimondii* (145, *Shang*

352 *et al., 2013*) and *G. arboreum* (141, *Shang et al., 2016*; *Fan et al.,* 2018). The difference in the

353 size of *NAC* genes is because the genomes of tetraploid species possess more *NAC* genes than the

354 diploid species, presumably because an allopolyploidization event occurred in *G. babardense* and

355 *G. hirsutum* approximately 1.7–1.9 Mya (*Hu et al., 2019*). *GbNAC* genes were unevenly

356 distributed on 26 chromosomes, and a number of *GbNAC* genes were clustered on the top or

357 bottom of specific chromosomes (Figure 2). These results are similar to those of the two diploid

358 cotton species, *G. raimondii*, and *G. arboreum* (*Shang et al., 2013*; *2016*).

359 According to the multiple sequence alignment, all the GbNAC proteins contain the highly

360    conserved NAC domain with 150−160 amino acids, which can be further divided into five distinct

361    subdomains (A−E) at the N-terminal. However, a number of GbNAC proteins have an atypical

362    NAC domain pattern. For example, three GbNAC proteins (GbNAC068, GbNAC200, and

363    GbNAC226) have only conserved subdomain A, while four GbNAC proteins (GbNAC033,

364    GbNAC092, GbNAC109, and GbNAC242) have only subdomain D. In addition, the previously

365    reported GbNAC1 lacks the conserved subdomains B, C, and E (*Wang et al., 2016*). NAC proteins

366    lacking one to four subdomains were also observed in Arabidopsis, rice, and radish (*Ooka et al.,*

367    *2003*; *Nuruzzaman et al. 2010*; *Karanja et al., 2017*). Thus, the atypical NAC domain pattern

368    appears to be common among NAC proteins from diverse plant species. Previous studies revealed

369    that subdomains A, C, and D have higher levels of conservation than subdomains B and E, and

370    play important roles in the function of NAC genes (*Ooka et al., 2003*). Subdomain A has the ability

371    to form a helical structure and is involved in the formation of homodimers or heterodimers with

372    other NAC domain proteins (*Jensen et al., 2010*). Subdomain D contains the nuclear localization

373    signal and subdomain C is associated with DNA binding activity (*Hernandez & Sanan-Mishra*

374    *2017*). In this study, most GbNAC proteins (196/270) contain subdomains A, C, and D, whereas

375    only 48 GbNAC proteins have subdomain B (Figure 5). Furthermore, we found 231 GbNAC

376    proteins contained a conserved histidine residue in subdomain D. A recent study revealed that the

377    conserved histidine residue is present in 80% of Arabidopsis NAC members and functioned as a

378    switch to regulate both pH-dependent homodimerization and DNA binding (*Kang et al.*, 2019).

379    Thus, the subdomain D may function not only in nuclear localization but also in the formation of

380    functional dimers and DNA-binding.

381    *NAC* gene duplication was investigated in sea-island cotton genome. One hundred and fifty-four

382    duplicated *NAC* gene pairs including 148 segmental duplication pairs and 6 tandem duplication

383    pairs were revealed in sea-island cotton (Figure 3). Therefore, it can be concluded that segmental

384    duplication dominates the expansion of the *NAC* gene family in the *G. barbadense* genome.

385    Furthermore, most duplicated gene pairs had undergone strong purifying selection during

386    evolution, indicating that purifying selection played pivotal roles in the confinement of the *GbNAC*

387   gene functions, which was further confirmed by the similar expression pattern of most duplication

388   gene pairs. Previous studies have showed that the A and D ancestor genomes diverged

389   approximately 6.2–7.1 Mya (*Hu et al., 2019*). In this study, about half of the duplication events

390   occurred after the divergence of the two diploid progenitors (Table S3). Synteny analysis indicated

391   that the collinearity of *NAC* genes for the four *Gossypium* species is highly conserved. However,

392   the level of collinearity is different. *G. barbadense* and *G. hirsutum* have the highest level of

393   collinearity, while *G. barbadense* and *G. arboretum* have the lowest level of collinearity (Table

394   S4). The difference is probably attributed to the genomic characteristics and evolution history of

395   cotton. *G. barbadense* and *G. hirsutum* evolved from a common tetraploid ancestor and diverged

396   approximately 0.4–0.6 Mya (*Hu et al., 2019*), which leads to high collinearity for the two tetraploid

397   species. Nevertheless, the A subgenome of *G. barbadense* has large chromosome inversions in

398   comparison with *G. arboretum* due to chromosomal rearrangement after allopolyploidization

399   (*Wang et al., 2019*), which results in relatively low collinearity between *G. barbadense* and *G.*

400   *arboretum*.

401   **Differential expression of *GbNAC* genes in response to Verticillium wilt**

402   Cotton VW is a destructive soil-borne fungal disease and dramatically reduces the yield and quality

403   of cotton. VW was first reported in Virginia, USA, in 1914 and now occurs worldwide. Over the

404   past century, substantial efforts have been made to develop ways to control VW, and a number of

405   genes, such as *Gh_A10G2076*, *GhATAF1*, *GbERF1*, *GbWRKY1*, and *GbRLK*, have been revealed

406   to be associated with VW resistance (*Li et al., 2017*; *He et al., 2016*; *Guo et al., 2016*; *Li et al.,*

407   *2014*; *Jun et al., 2015*). However, whether the NAC genes could play possible role in response to

408   VW in *G. barbadense* is still in question. In this study, 62 *GbNAC* genes were identified to be

409   significantly up- or down-regulated in roots inoculated with *V. dahliae* at all six time-points

410   (Figure S2 and Table S9). These genes are all from the 10 phylogenetic groups, except group V,

411   and contain at least one *cis*-regulatory element involved in stress responses (Table S7). We also

412   found that *GbNAC* genes in each duplicated gene pair had similar expression patterns in response

413   to VW infection. The functional redundancy of duplicated genes may be related to their similar

414 gene structure, motifs, and *cis*-regulatory elements. In Arabidopsis, five *NAC* genes (*ANAC016*,

415 *ANAC036*, *ANAC037*, *ANAC061*, *ANAC081*, and *ANAC091*) were significantly up-regulated after

416 chitin treatment (*Libault et al., 2007*). Chitin is an elicitor of plant defense responses and its

417 elicitation plays important role in plant defense to fungal pathogens. Among the 62 *GbNAC* genes,

418 10 genes clustered with *ANAC081* are from group IV; 13 genes along with *ANAC036* and

419 *ANAC061* belong to group VI; 4 genes and *ANAC091* are from group VII, and 3 genes and

420 *ANAC037* are from group I (Table S9).

421 Previously, an upland cotton ATAF subfamily *NAC* gene, *GhATAF1* (DT549350), was reported

422 to be up-regulated by *V. dahliae* inoculation. Cotton plants overexpressing *GhATAF1* increased

423 susceptibility to pathogen *V. dahliae* (*He et al., 2016*). In our study, *GbNAC164*, an ortholog to

424 *GhATAF1*, was down-regulated by *V. dahliae* infection. The different expression pattern may be

425 caused by cotton genotypes and/or *V. dahliae* strains. *GbNAC164* was investigated in *G.*

426 *barbadense* cv. 7124 treated with a highly virulent *V. dahliae* strain V991, while *GhATAF1* was

427 analyzed in *G. hirsutum* cv. YZ1 treated with a moderately aggressive *V. dahliae* strain ICD3-2.

428 Sea-island cotton *GbNAC1* (*GbNAC220* in this study) belongs to the Tobacco elicitor-responsive

429 gene encoding NAC-domain protein (TERN) subgroup and was down-regulated by VW. Cotton

430 plants silencing *GbNAC1* reduce resistance to VW, whereas transgenic Arabidopsis lines

431 overexpressing *GbNAC1* enhance resistance to VW compared to wild type (*Wang et al., 2016*).

432 *GbNAC220* was also down-regulated by VW in our study. In addition, MeJA and SA are important

433 pathogen-related hormonal regulators. Among the 62 *GbNAC* genes, 31 and 22 genes contain

434 MeJA- and SA-responsive elements, respectively, and 12 genes contain both MeJA- and SA-

435 responsive elements (Table S7), indicating the expression of these genes may be regulated by

436 MeJA and/or SA. These results were further verified by qPCR analysis. The qPCR results

437 indicated *GbNAC003*, *GbNAC137*, *GbNAC140*, *GbNAC215*, and *GbNAC248* were sensitive to

438 MeJA and SA treatments (Figure 8), which was in agreement with the results of *cis*-regulatory

439 element analysis (Table S7). *GbNAC164* was up-regulated by MeJA and SA treatment. Moreover,

440 *GbNAC164* keeps high expression level more durable by SA than by MeJA (Figure 8). This result

441 was consistent with *GhATAF1* (*He et al., 2016*). Thus, these findings suggest that up-regulating of

442 the 62 *GbNAC* genes may result in increased or decreased VW resistance in cotton and these genes

443 can be candidate genes for in-depth study on VW resistance.

444 Interestingly, *GhATAF1* was also up-regulated by ABA, cold, and salt treatments. Overexpression

445 of *GhATAF1* confers transgenic cotton improved salt tolerance (*He et al., 2016*). Similarly,

446 Transgenic *GbNAC1* results in enhanced drought tolerance in Arabidopsis (*Wang et al., 2016*).

447 The *NAC* genes *ANAC019*, *ANAC055* and *ANAC072* from Arabidopsis were up-regulated by

448 drought, salt, and ABA treatments. Transgenic either *ANAC019*, *ANAC055* or *ANAC072* conferred

449 Arabidopsis plants enhanced drought tolerance. The finding suggests that the 62 *GbNAC* genes

450 may play role in response to abiotic stresses.

451 **CONCLUSIONS**

452 In this study, the plant-specific *NAC* gene family in sea-island cotton was characterized with

453 particular focus on their responses to VW infection. A total of 270 *GbNAC* genes were identified

454 and characterized in sea-island cotton. The gene structure, chromosomal distribution, gene

455 duplication, conserved motif, *cis*-elements, and expression profiles of the *GbNAC* genes were

456 analyzed. Furthermore, expression profile analyses revealed that 62 *GbNAC* genes may play

457 crucial role in response to VW infection. However, further functional data are required to evaluate

458 each *GbNAC* gene. Overall, our results will provide new insights for plant engineering programs

459 so that economically important traits for cotton can be developed, including improved resistance

460 to VW.

461 REFERENCES

462 **Allu AD, Brotman Y, Xue GP, Balazadeh S. 2016.** Transcription factor *ANAC032* modulates

463 JA/SA signaling in response to *Pseudomonas syringae* infection. *EMBO Reports* **17 (11):** 1578-

464 1589.

465 **Artimo P, Jonnalagedda M, Arnold K, Baratin D, Csardi G, de Castro E, Duvaud S, Flegel**

466 **V, Fortier A, Gasteiger E, Grosdidier A, Hernandez C, Ioannidis V, Kuznetsov D, Liechti R,**

467 **Moretti S, Mostaguir K, Redaschi N, Rossier G, Xenarios I, Stockinger H. 2012.** ExPASy:

468 SIB bioinformatics resource portal. *Nucleic Acids Research* **40 (W1):** W597-W603.

469 **Bailey TL, Johnson J, Grant CE, Noble WS. 2015.** The MEME Suite. *Nucleic Acids Research*

470 **43 (W1):** W39-W49.

471 **Chen JY, Huang JQ, Li NY, Ma XF, Wang JL, Liu C, Liu YF, Liang Y, Bao YM, Dai XF.**

472 **2015.** Genome-wide analysis of the gene families of resistance gene analogues in cotton and their

473 response to verticillium wilt. *BMC Plant Biology* **15: 148.**

474 **Dai X, Zhuang Z, Zhao PX. 2018.** psRNATarget: a plant small RNA target analysis server (2017

475 release). *Nucleic Acids Research* **46 (W1):** W49-W54.

476 **Du H, Yang SS, Liang Z, Feng BR, Liu L, Huang YB, Tang YX. 2012.** Genome-wide analysis

477 of the MYB transcription factor superfamily in soybean. *BMC Plant Biology* **12:** 106.

478 **Fan K, Bibi N, Gan S, Li F, Yuan S, Ni M, Wang M, Shen H, Wang X. 2015.** A novel NAP

479 member GhNAP is involved in leaf senescence in *Gossypium hirsutum*. *Journal of Experimental*

480 *Botany* **66 (15):** 4669-4685.

481 **Fan K, Li F, Chen J, Li Z, Lin W, Cai S, Liu J, Lin W. 2018.** Asymmetric evolution and

482 expansion of the NAC transcription factor in polyploidized cotton. *Frontiers in Plant Science* **9:**

483 47.

484 **Fang Y, Xie K, Xiong L. 2014.** Conserved miR164-targeted NAC genes negatively regulate

485 drought resistance in rice. *Journal of Experimental Botany* **65 (8):** 2119-2135.

486 **Finn RD, Coggill P, Eberhardt RY, Eddy SR, Mistry J, Mitchell AL, Potter SC, Punta M,**

487 **Qureshi M, Sangrador-Vegas A, Salazar GA, Tate J, Bateman A. 2016.** The Pfam protein

488 families database: towards a more sustainable future. *Nucleic Acids Research* **44 (Database issue):**

489 D279-D285.

490 **Guo W, Jin L, Miao Y, He X, Hu Q, Guo K, Zhu L, Zhang X. 2016.** An ethylene response-
491 related factor, GbERF1-like, from *Gossypium barbadense* improves resistance to *Verticillium*
492 *dahliae* via activating lignin synthesis. *Plant Molecular Biology* **91 (3):** 305-318.

493 **Hernandez Y, Sanan-Mishra N. 2017.** miRNA mediated regulation of NAC transcription factors
494 in plant development and environment stress response. *Plant Gene* **11 (Part B):** 190-198.

495 **He X, Zhu L, Xu L, Guo W, Zhang X. 2016.** *GhATAF1*, a NAC transcription factor, confers
496 abiotic and biotic stress responses by regulating phytohormonal signaling networks. *Plant Cell*
497 *Reports* **35 (10):** 2167-2179.

498 **Higo K, Ugawa Y, Iwamoto M, Korenaga T. 1999.** Plant cis-acting regulatory DNA elements
499 (PLACE) database: 1999. *Nucleic Acids Research* **27 (1):** 297-300.

500 **Hu B, Jin J,Guo AY, Zhang H, Luo J, Gao G. 2015.** GSDS 2.0: an upgraded gene feature
501 visualization server. *Bioinformatics* **31 (1):** 1296-1297.

502 **Hu H, Dai M, Yao J, Xiao B, Li X, Zhang Q, Xiong L. 2006.** Overexpressing a *NAM, ATAF,*
503 and *CUC* (*NAC*) transcription factor enhances drought resistance and salt tolerance in rice.
504 *Proceedings of the National Academy of Sciences of the United States of America* **103 (35):** 12987-
505 12992.

506 **Hu Y, Chen J, Fang L, Zhang Z, Ma W, Niu Y, Ju L, Deng J, Zhao T, Lian J, Baruch K,**
507 **Fang D, Liu X, Ruan YL, Rahman MU, Han J, Wang K, Wang Q, Wu H, Mei G, Zang Y,**
508 **Han Z, Xu C, Shen W, Yang D, Si Z, Dai F, Zou L, Huang F, Bai Y, Brodt A, Ben-Hamo H,**
509 **Zhu X, Zhou B, Guan X, Zhu S, Chen X, Zhang T. 2019.** *Gossypium barbadense* and
510 *Gossypium hirsutum* genomes provide insights into the origin and evolution of allotetraploid
511 cotton. *Nature Genetics* **51 (4):** 739-748.

512 **Jensen MK, Kjaersgaard T, Nielsen MM, Galberg P, Petersen K, O'Shea C, Skriver K. 2010.**
513 The *Arabidopsis thaliana* NAC transcription factor family: structure-function relationships and
514 determinants of ANAC019 stress signaling. *Biochemical Journal* **426 (2):** 183-196.

515   **Jun Z, Zhang Z, Gao Y, Zhou L, Fang L, Chen X, Ning Z, Chen T, Guo W, Zhang T. 2015.**

516   Overexpression of *GbRLK*, a putative receptor-like kinase gene, improved cotton tolerance to

517   Verticillium wilt. *Scientific reports* **5:** 15048.

518   **Kang M, Kim S, Kim HJ, Shrestha P, Yun JH, Phee BK, Lee W, Nam HG, Chang I. 2018.**

519   The C-domain of the NAC transcription factor ANAC019 is necessary for pH-tuned DNA binding

520   through a histidine switch in the N-domain. *Cell Reports* **22 (5):** 1141-1150.

521   **Karanja BK, Xu L, Wang Y, Muleke EM, Jabir BM, Xie Y, Zhu X, Cheng W, Liu L. 2017.**

522   Genome-wide characterization and expression profiling of NAC transcription factor genes under

523   abiotic stresses in radish (*Raphanus sativus* L.) *PeerJ* **5:** e4172.

524   **Kim SG, Lee S, Seo PJ, Kim SK, Kim JK, Park CM. 2010.** Genome-scale screening and

525   molecular characterization of membrane-bound transcription factors in Arabidopsis and rice.

526   *Genomics* **95 (1):** 56-65.

527   **Krogh A, Larsson B, von Heijne G, Sonnhammer EL. 2001.** Predicting transmembrane protein

528   topology with a hidden markov model: application to complete genomes. *Journal of Molecular*

529   *Biology* **305 (3):** 567-580.

530   **Krzywinski M, Schein JE, Birol I, Connors J, Gascoyne R, Horsman D, Jones SJ, Marra**

531   **MA. 2009.** Circos: an information aesthetic for comparative genomics. *Genome Research* **19 (9):**

532   1639-1645.

533   **Kumar S, Stecher G, Li M, Knyaz C, Tamura K. 2018.** MEGA X: molecular evolutionary

534   genetics analysis across computing platforms. *Molecular Biology and Evolution* **35 (6):** 1547-

535   1549.

536   **Larkin MA, Blackshields G, Brown NP, Chenna R, McGettigan PA, McWilliam H, Valentin**

537   **F, Wallace IM, Wilm A, Lopez R, Thompson JD, Gibson TJ, Higgins DG. 2007.** Clustal W

538   and Clustal X version 2.0. *Bioinformatics* **23 (21):** 2947-2948.

539   **Lescot M, Dehais P, Thijs G, Marchal K, Moreau Y, Van de Peer Y, Rouze P, Rombauts S.**

540    **2002.** PlantCARE, a database of plant *cis*-acting regulatory elements and a portal to tools for in

541    silico analysis of promoter sequences. *Nucleic Acids Research* **30 (1):** 325-327.

542    **Letunic I, Bork P. 2018.** 20 years of the SMART protein domain annotation resource. *Nucleic*

543    *Acids Research* **46 (D1):** D493-D496.

544    **Lex A, Gehlenborg N, Strobelt H, Vuillemont R, Pfister H. 2014.** UpSet: visualization of

545    intersecting sets. *IEEE Transactions on Visualization and Computer Graphics* **20 (12):** 1983-1992.

546    **Li C, He X, Luo X, Xu L, Liu L, Min L, Jin L, Zhu L, Zhang X. 2014.** Cotton WRKY1 mediates

547    the plant defense-to-development transition during infection of cotton by *Verticillium dahliae* by

548    activating JASMONATE ZIM-DOMAIN1 expression. *Plant Physiology* **166 (4):** 2179-2194.

549    **Li S, Wang N, Ji D, Xue Z, Yu Y, Jiang Y, Liu J, Liu Z, Xiang F. 2016.** Evolutionary and

550    functional analysis of membrane-bound NAC transcription factor genes in soybean. *Plant*

551    *Physiology* **172 (3):** 1804-1820.

552    **Li T, Ma X, Li N, Zhou L, Liu Z, Han H, Gui Y, Bao Y, Chen J, Dai X. 2017.** Genome-wide

553    association study discovered candidate genes of Verticillium wilt resistance in upland cotton

554    (*Gossypium hirsutum* L.). *Plant Biotechnology Journal* **15 (12):** 1520-1532.

555    **Li W, Zhou Z, Meng Y, Xu N, Fok M. 2009.** Modeling boll maturation period, seed growth,

556    protein, and oil content of cotton (*Gossypium hirsutum* L.) in China. *Field Crops Research* **112**

557    **(2-3):** 131-140.

558    **Libault M, Wan J, Czechowski T, Udvardi M, Stacey G. 2007.** Identification of 118

559    Arabidopsis transcription factor and 30 ubiquitin-ligase genes responding to chitin, a plant-defense

560    elicitor. *Molecular Plant-Microbe Interactions* **20 (8):** 900-911.

561    **Livak KJ, Schmittgen TD. 2001.** Analysis of relative gene expression data using real-time

562    quantitative PCR and the $2^{-\Delta\Delta CT}$ method. *Methods* **25 (4):** 402-408.

563    **Liu M, Ma Z, Sun W, Huang L, Wu Q, Tang Z, Bu T, Li C, Chen H. 2019.** Genome-wide

564  analysis of the NAC transcription factor family in tartary buckwheat (*Fagopyrum tataricum*). *BMC*

565  *Genomics* **20:** 113.

566  **Liu X, Wang T, Bartholomew E, Black K, Dong M, Zhang Y, Yang S, Cai Y, Xue S, Weng**

567  **Y, Ren H. 2018.** Comprehensive analysis of NAC transcription factors and their expression during

568  fruit spine development in cucumber (*Cucumis sativus* L.) *Horticulture Research* **5:** 31.

569  **Liu X, Zhao B, Zheng HJ, Hu Y, Lu G, Yang CQ, Chen JD, Chen JJ, Chen DY, Zhang L,**

570  **Zhou Y, Wang LJ, Guo WZ, Bai YL, Ruan JX, Shangguan XX, Mao YB, Shan CM, Jiang**

571  **JP, Zhu YQ, Jin L, Kang H, Chen ST, He XL, Wang R, Wang YZ, Chen J, Wang LJ, Yu ST,**

572  **Wang BY, Wei J, Song SC, Lu XY, Gao ZC, Gu WY, Deng X, Ma D, Wang S, Liang WH,**

573  **Fang L, Cai CP, Zhu XF, Zhou BL, Jeffrey Chen Z, Xu SH, Zhang YG, Wang SY, Zhang**

574  **TZ, Zhao GP, Chen XY. 2015.** Gossypium barbadense genome sequence provides insight into

575  the evolution of extra-long staple fiber and specialized metabolites. *Scientific reports* **5:** 14139.

576  **Metsalu T, Vilo J. 2015.** ClustVis: a web tool for visualizing clustering of multivariate data using

577  principal component analysis and heatmap. *Nucleic Acids Research* **43 (W1):** W566-W570.

578  **Nuruzzaman M, Manimekalai R, Sharoni AM, Satoh K, Kondoh H, Ooka H, Kikuchi S.**

579  **2010.** Genome-wide analysis of NAC transcription factor family in rice. *Gene* **465 (1-2):** 30-44.

580  **Nuruzzaman M, Sharoni AM, Kikuchi S. 2013.** Roles of NAC transcription factors in the

581  regulation of biotic and abiotic stress responses in plants. *Frontiers in Microbiology* **4:** 248.

582  **Olsen AN, Ernst HA, Leggio LL, Skriver K. 2005.** NAC transcription factors: structurally

583  distinct, functionally diverse. *Trends in Plant Science* **10 (2):** 79-87.

584  **Shang H, Li W, Zou C, Yuan Y. 2013.** Analysis of the NAC transcription factor gene family in

585  *Gossypium raimondii* Ulbr.: chromosomal location, structure, phylogeny, and expression patterns.

586  *Journal of Integrative Plant Biology* **55 (7):** 663-676.

587  **Shang H, Wang Z, Zou C, Zhang Z, Li W, Li J, Shi Y, Gong W, Chen T, Liu A, Gong J, Ge**

588  **Q, Yuan Y. 2016.** Comprehensive analysis of NAC transcription factors in diploid *Gossypium*:

589    sequence conservation and expression analysis uncover their roles during fiber development.

590    *Science China Life Sciences* **59 (2):** 142-153.

591    **Souer E, van Houwelingen A, Kloos D, Mol J, Koes R. 1996.** The no apical meristem gene of

592    Petunia is required for pattern formation in embryos and flowers and is expressed at meristem and

593    primordia boundaries. *Cell* **85 (2):** 159-170.

594    **Sun H, Hu M, Li J, Chen L, Li M, Zhang S, Zhang X, Yang X. 2018.** Comprehensive analysis

595    of NAC transcription factors uncovers their roles during fiber development and stress response in

596    cotton. *BMC Plant Biology* **18:** 150.

597    **Sun Q, Jiang H, Zhu X, Wang W, He X, Shi Y, Yuan Y, Du X, Cai Y. 2013.** Analysis of sea-

598    island cotton and upland cotton in response to *Verticillium dahliae* infection by RNA sequencing.

599    *BMC Genomics* **14:** 852.

600    **Takasaki H, Maruyama K, Takahashi F, Fujita M, Yoshida T, Nakashima K, Myouga F,**

601    **Toyooka K, Yamaguchi-Shinozaki K, Shinozaki K. 2015.** SNAC-As, stress-responsive NAC

602    transcription factors, mediate ABA-inducible leaf senescence. *Plant Journal* **84 (6):** 1114-1123.

603    **Vatansever R, Koc I, Ozyigit II, Sen U, Uras ME, Anjum NA, Pereira E, Filiz E. 2016.**

604    Genome-wide identification and expression analysis of sulfate transporter (*SULTR*) genes in potato

605    (*Solanum tuberosum* L.). *Planta* **244 (6):** 1167-1183.

606    **Voorrips R. 2002.** MapChart: software for the graphical presentation of linkage maps and QTLs.

607    *Journal of Heredity* **93 (1):**77-78.

608    **Wang Y, Tang H, DeBarry JD, Tan X, Li J, Wang X, Lee TH, Jin H, Marler B, Guo H,**

609    **Kissinger JC, Paterson AH. 2012.** MCScanX: a toolkit for detection and evolutionary analysis

610    of gene synteny and collinearity. *Nucleic Acids Research* **40 (7):** e49.

611    **Wang M, Tu L, Yuan D, Zhu D, Shen C, Li J, Liu F, Pei L, Wang P, Zhao G, Ye Z, Huang**

612    **H, Yan F, Ma Y, Zhang L, Liu M, You J, Yang Y, Liu Z, Huang F, Li B, Qiu P, Zhang Q,**

613    **Zhu L, Jin S, Yang X, Min L, Li G, Chen LL, Zheng H, Lindsey K, Lin Z, Udall JA, Zhang**

614    **X. 2019.** Reference genome sequences of two cultivated allotetraploid cottons, *Gossypium*

615    *hirsutum* and *Gossypium barbadense*. *Nature Genetics* **51 (2):** 224-229.

616    **Wang W, Yuan Y, Yang C, Geng S, Sun Q, Long L, Cai C, Chu Z, Liu X, Wang G, Du X,**

617    **Miao C, Zhang X, Cai Y. 2016.** Characterization, expression, and functional analysis of a novel

618    NAC gene associated with resistance to Verticillium wilt and abiotic stress in cotton. *G3: Genes,*

619    *Genomes, Genetics* **6 (12):** 3951-3961.

620    **Yokotani N, Tsuchida-Mayama T, Ichikawa H, Mitsuda N, Ohme-Takagi M, Kaku H,**

621    **Minami E, Nishizawa Y. 2014.** OsNAC111, a blast disease-responsive transcription factor in rice,

622    positively regulates the expression of defense-related genes. *Molecular Plant-Microbe*

623    *Interactions* **27 (10):** 1027-1034.

624    **Yu CS, Chen YC, Lu CH, Hwang JK. 2006.** Prediction of protein subcellular localization.

625    *Proteins: Structure, Function and Bioinformatics* **64 (3):** 643-651.

626    **Zhao F, Ma J, Li L, Fan S, Guo Y, Song M, Wei H, Pang C, Yu S. 2016.** GhNAC12, a neutral

627    candidate gene, leads to early aging in cotton (*Gossypium hirsutum* L). *Gene* **576 (1 Pt 2):** 268-

628    274.

629    **Zhang J, Huang GQ, Zou D, Yan JQ, Li Y, Hu S, Li XB. 2018.** The cotton (*Gossypium*

630    *hirsutum*) NAC transcription factor (FSN1) as a positive regulator participates in controlling

631    secondary cell wall biosynthesis and modification of fibers. *New Phytologist* **217 (2):** 625-640.

632    **Zhang J, Yu J, Pei W, Li X, Said J, Song M, Sanogo S. 2015.** Genetic analysis of Verticillium

633    wilt resistance in a backcross inbred line population and a meta-analysis of quantitative trait loci

634    for disease resistance in cotton. *BMC Genomics* **16:** 577.

635    **Zhang Y, Wang XF, Ding ZG, Ma Q, Zhang GR, Zhang SL, Li ZK, Wu Q, Zhang GY, Ma**

636    **ZY. 2013.** Transcriptome profiling of Gossypium barbadense inoculated with Verticillium dahliae

637    provides a resource for cotton improvement. *BMC Genomics* **14:** 637.

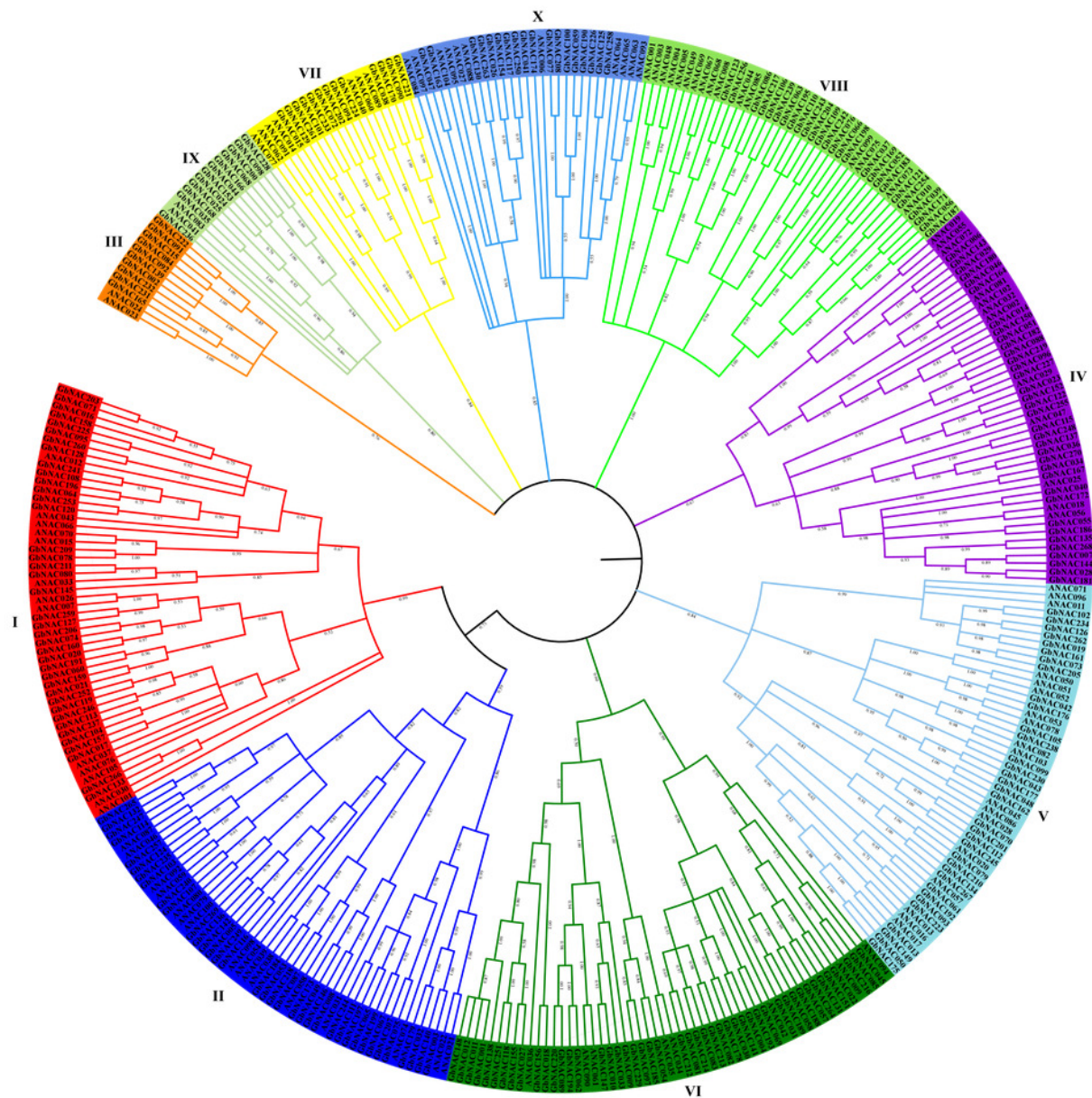
638    **Zhang Z, Li J, Zhao XQ, Wang J, Wong GK, Yu J. 2006.** KaKs_Calculator: calculating Ka and

639    Ks through model selection and model averaging. *Genomics Proteomics Bioinformatics* **4 (4):**

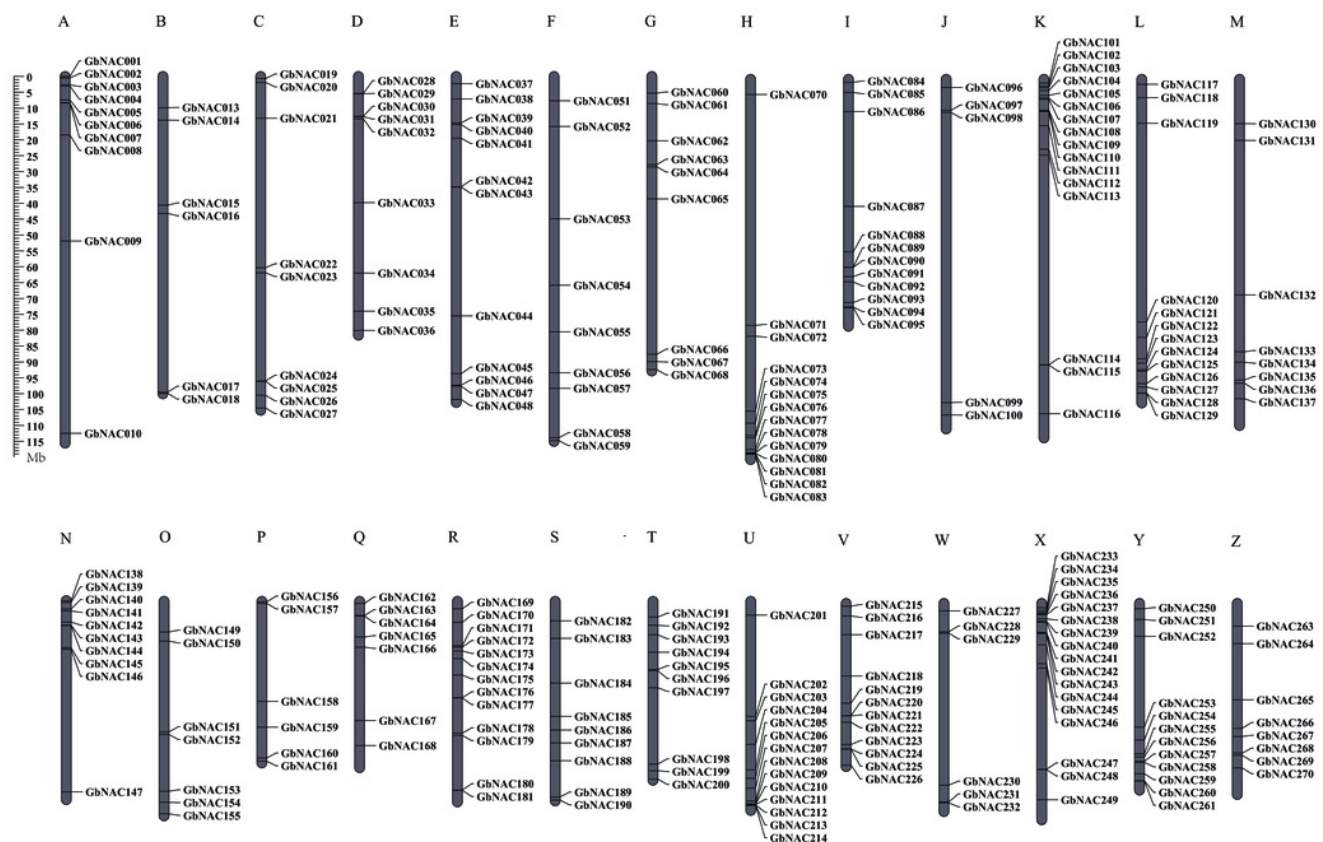
640    259-263.

# Figure 1

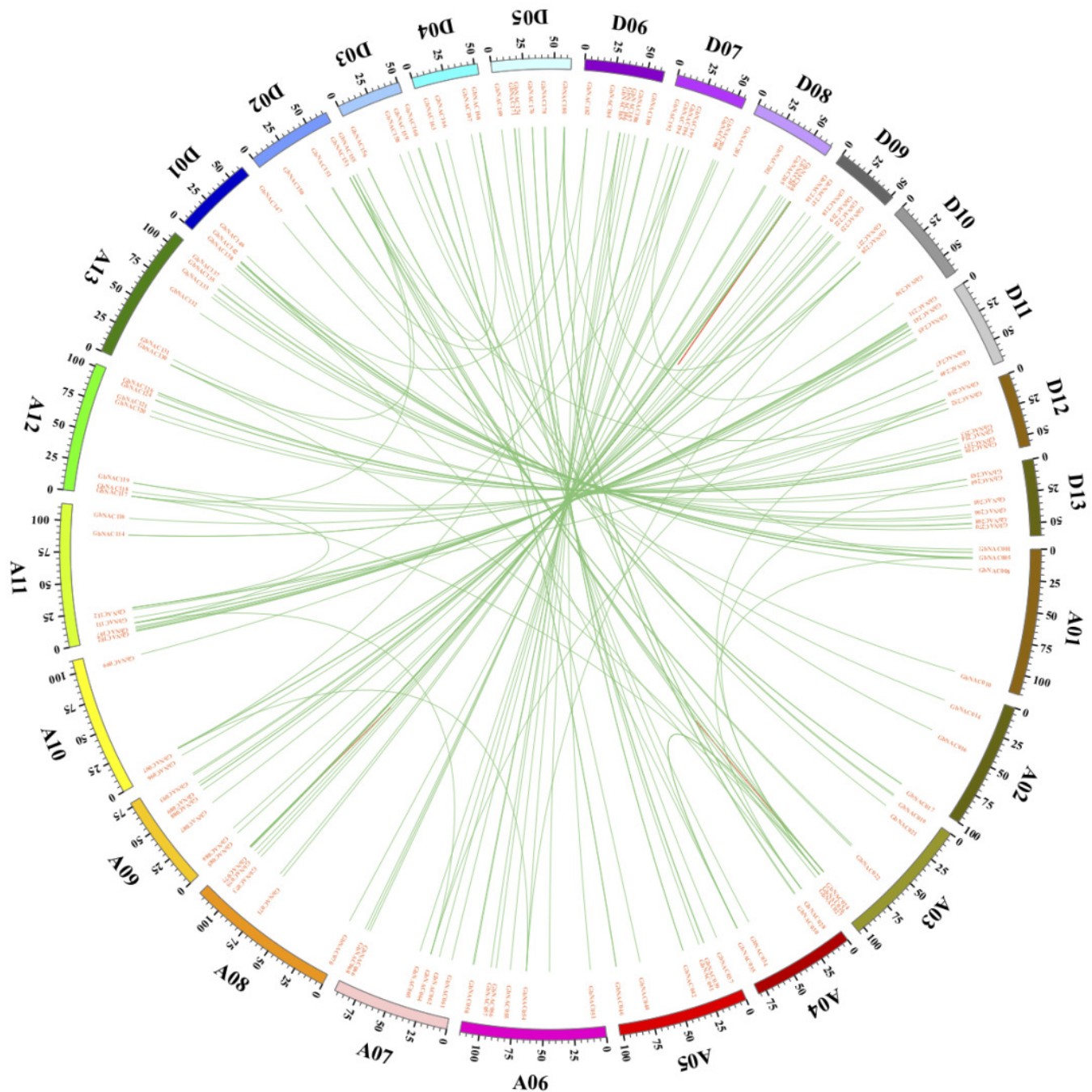
Phylogenetic tree of the 270 GbNAC proteins.

Multiple sequence alignment of NAC domain sequences of *G. barbadense* and *Arabidopsis* was performed using ClustalW. MEGA X was used to construct the neighbor-joining (NJ) tree with 1000 bootstrap replicates. Various colors indicate different groups of GbNACs.

# Figure 2

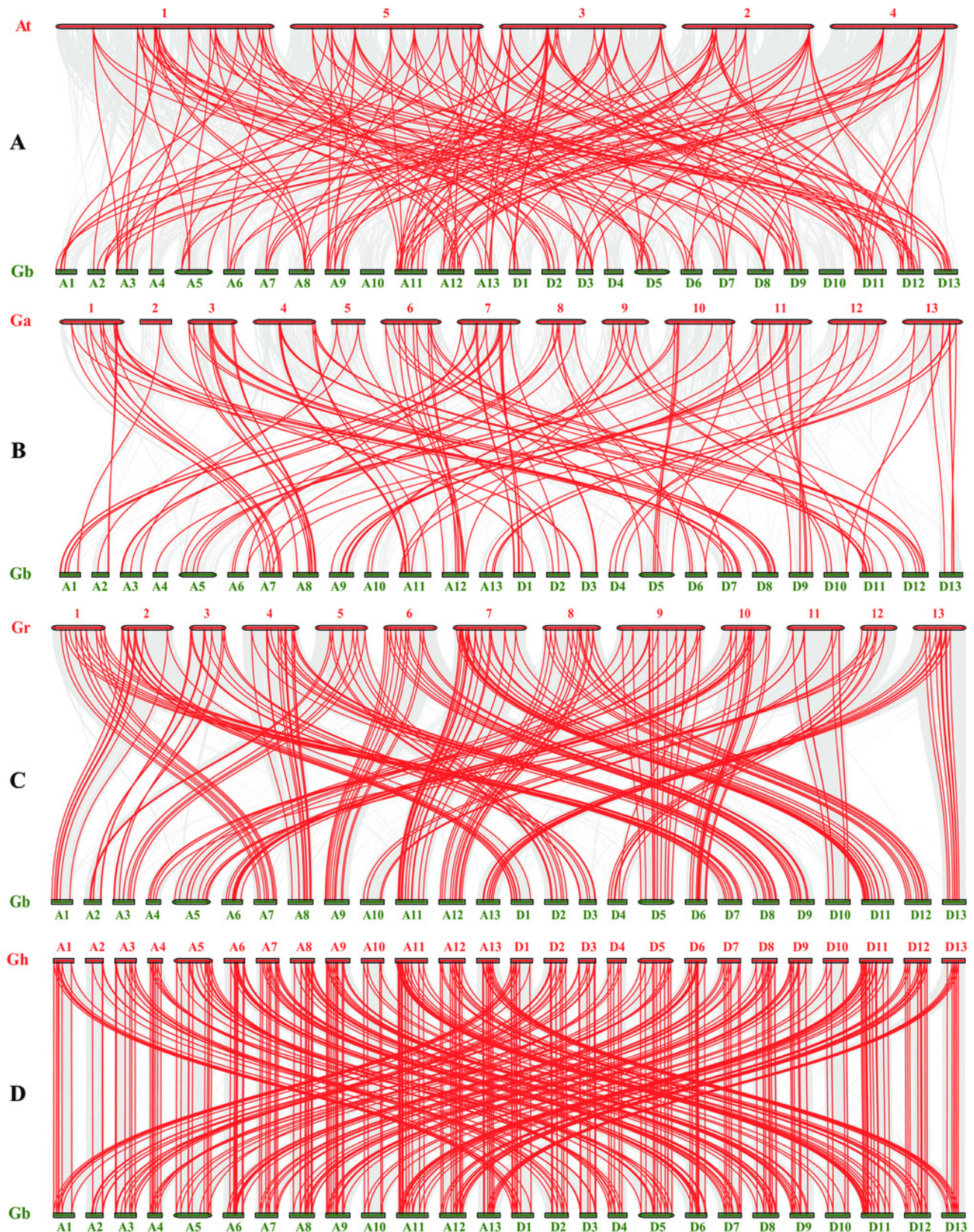
Chromosomal locations of the *GbNAC* genes.

Grey bars denote the *G. babardense* chromosome. Scale bar on the left indicates the chromosome lengths (Mb). The letter A-M indicates the A-subgenome chromosome A01-A13, and N-O indicates the D-subgenome chromosome D01-D13, respectively. The vertical red lines on the left of each chromosome indicate tandem duplicated genes.

# Figure 3

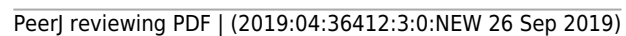
Circos diagram of the *GbNAC* duplication pairs in *G. barbadense*.

143 *GbNAC* duplication pairs are linked with green lines. Scale bar marked on the chromosome indicating chromosome lengths (Mb).

# Figure 4

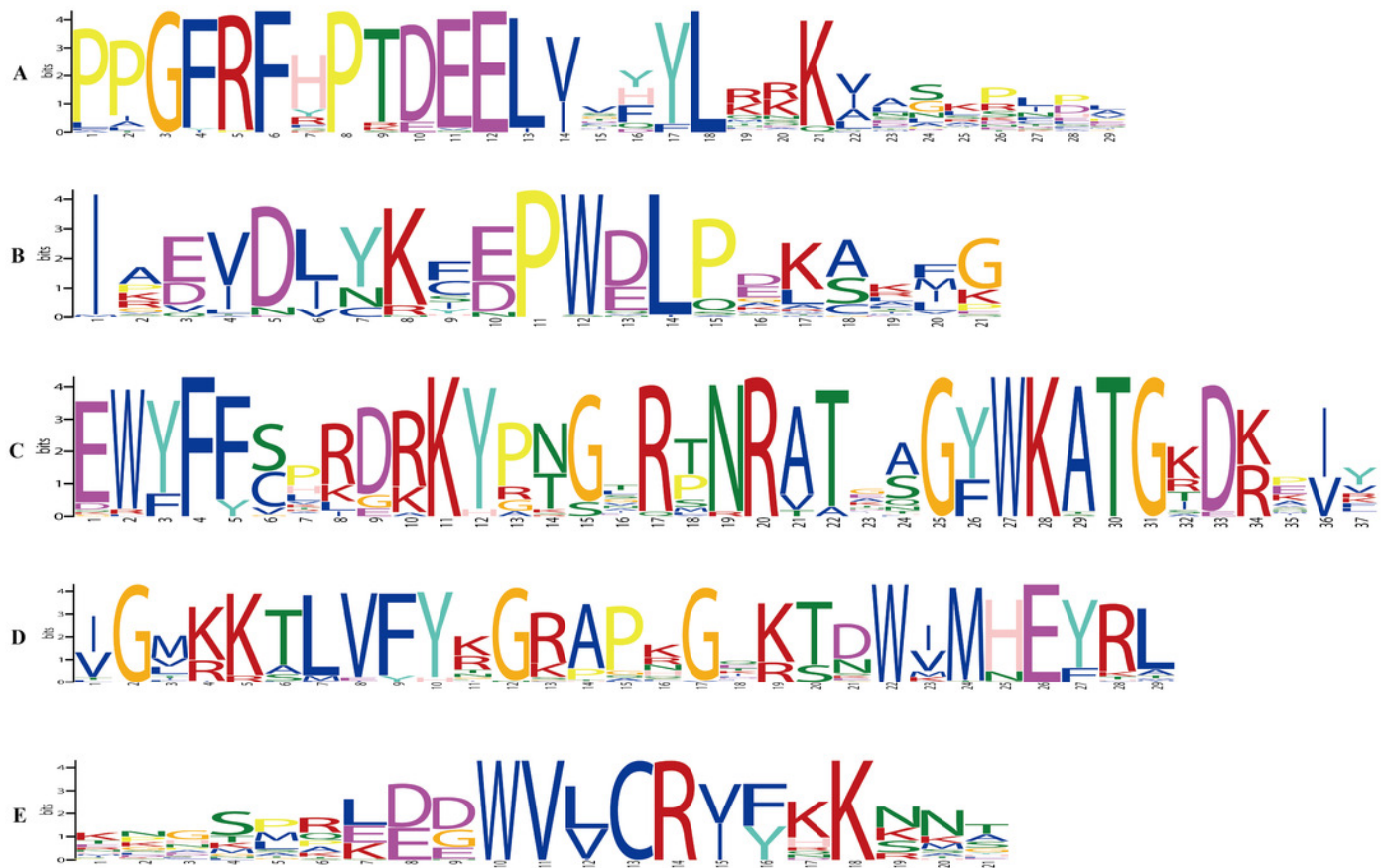
Synteny analysis of NAC genes between Gossypium barbadense and other plant species.

At, Ga, Gr, Gb, and Gh indicate *Arabidopsis thaliana*, *G. arboreum*, *G. raimondii*, *G. barbadense*, and *G. hirsutum*, respectively. A: Collinearity between At and Gb; B: Collinearity between Ga and Gb; C: Collinearity between Gr and Gb; D: Collinearity between Gh and Gb. Gray lines in the background represent the collinear blocks within the genomes of *G. barbadense* and other plant species, while the red lines show the collinear *NAC* gene pairs.

# Figure 5

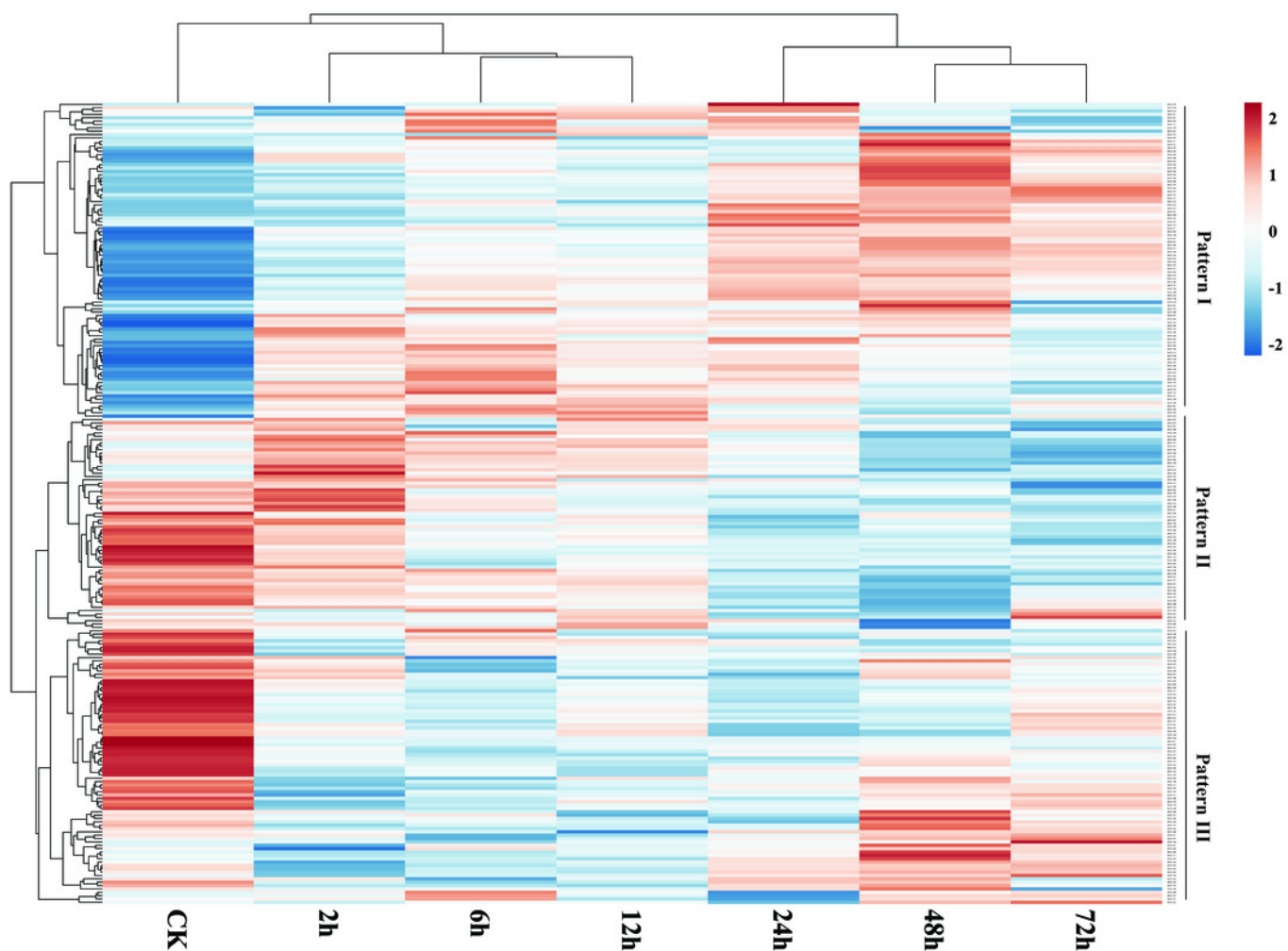
Putative conserved motifs and gene structures of the *GbNAC* genes.

A: phylogenetic tree. Multiple sequence alignment of NAC domain sequences of *G. barbadense* was performed using ClustalW. The neighbor-joining (NJ) tree was constructed using MEGA X with 1000 bootstrap replicates. B: conserved motif. MEME analysis revealed the conserved motifs of the GbNAC proteins. The colored boxes on the right denote 20 motifs. C: gene structure. The yellow boxes, black lines, and green boxes represent exon, intron, and UTR (untranslated region), respectively.

# Figure 6

Conserved subdomains in the NAC domain at the N-terminal of GbNAC proteins

A-E indicate five conserved subdomain A-E, respectively.

# Figure 7

Expression profiles of *GbNAC* genes in roots under Verticillium wilt infection.

The heat map was generated by ClustVis software. The expression data were gene-wise normalized and hierarchically clustered with average linkage. The bar on the right of the heat map indicates relative expression values. Values 2, 0, -2 represent high, intermediated, and low expression, respectively.

# Figure 8

Expression profiles of *GbNAC* genes in response to hormone MeJA and SA.

qPCR was used to analyze the expression profiles of 15 *GbNAC* genes under hormone MeJA and SA treatment. The letter A-O indicates the 15 *GbNAC* genes tested, respectively. The relative expression levels are normalized to *GbActin*. The data represent the mean of three biological replicates. Significant differences were indicated as * (Tukey's HSD, $P < 0.05$) and ** ($P < 0.01$) above the bars between treatment and control.