RESEARCH PAPER

Genome-wide identification of GhAAI genes reveals that GhAAI66 triggers a phase transition to induce early flowering

Ghulam Qanmber1,*, Lili Lu1,*, Zhao Liu1,*, Daoqian Yu1, Kehai Zhou1, Peng Huo1, Fuguang Li1,2,†, and Zuoren Yang1,2,†
1 State Key Laboratory of Cotton Biology, Institute of Cotton Research, Chinese Academy of Agricultural Sciences, Anyang, 455000, Henan, China
2 Zhengzhou Research Base, State Key Laboratory of Cotton Biology, Zhengzhou University, Zhengzhou, 4550001, Henan, China
* These authors contributed equally to this work.
† Correspondence: ayfig@163.com or yangzuoren4012@163.com

Received 22 November 2018; Editorial decision 10 May 2019; Accepted 11 May 2019

Editor: Zoe Wilson, University of Nottingham, UK

Abstract
Plants undergo a phase transition from vegetative to reproductive development that triggers floral induction. Genes containing an AAI (α-amylase inhibitor) domain form a large gene family, but there have been no comprehensive analyses of this gene family in any plant species. Here, we identified 336 AAI genes from nine plant species including 122 AAI genes in cotton (Gossypium hirsutum). The AAI gene family has evolutionarily conserved amino acid residues throughout the plant kingdom. Phylogenetic analysis classified AAI genes into five major clades with significant polyploidization and showing effects of genome duplication. Our study identified 42 paralogous and 216 orthologous gene pairs resulting from segmental and whole-genome duplication, respectively, demonstrating significant contributions of gene duplication to expansion of the cotton AAI gene family. Further, GhAAI66 was preferentially expressed in flower tissue and as responses to phytohormone treatments. Ectopic expression of GhAAI66 in Arabidopsis and silencing in cotton revealed that GhAAI66 triggers a phase transition to induce early flowering. Further, GO (Gene Ontology) and KEGG (Kyoto Encyclopedia of Genes and Genomes) analysis of RNA sequencing data and qRT–PCR (quantitative reverse transcription–PCR) analysis indicated that GhAAI66 integrates multiple flower signaling pathways including gibberellin, jasmonic acid, and floral integrators to trigger an early flowering cascade in Arabidopsis. Therefore, characterization of the AAI family provides invaluable insights for improving cotton breeding.

Keywords: Early flowering, ectopic expression, floral integrators, gene duplication, Gossypium hirsutum, phylogenetic analysis.

Introduction
Annual plants undergo a major phase transition triggering vegetative to reproductive development in their life cycle. The process of phase transition is rarely reversible; however, it ensures optimal timing of the transition for pollination as well as seed development (Boss et al., 2004). Of these, proper timing of flowering is most important for ensuring reproductive success (Lee and Lee, 2010). Genetic and physiological approaches to investigate flowering mechanisms have shown that various environmental as well as endogenous factors determine the phase transition and mediate the flowering signaling cascade (Boss et al., 2004). During the past decade, multiple studies have been carried out to provide a molecular understanding of control of flowering time by various comprehensive approaches (Mouradov et al., 2002; Ratcliffe and
Riechmann, 2002; Simpson and Dean, 2002; Henderson et al., 2003; Yanovsky and Kay, 2003). The integration of multiple pathways collectively regulates a set of common targets that promotes the floral transition. These include light quality, photoperiod, ambient temperature, hormone signaling, and biosynthesis (Simpson and Dean, 2002; Boss et al., 2004) that induce or repress the expression of genes involved in the flowering signaling cascade. These floral pathway integrators include FLOWERING LOCUS T (FT), LEAFY (LFY), and SUPPRESSOR OF OVEREXPRESS OF CONSTANS (SOC1) (Nilsson et al., 1998; Kardailsky et al., 1999; Kobayashi et al., 1999; Lee et al., 2000; Samach et al., 2000). Further, gibberellin (GA) induces flowering signals in many plant species, such as GA-promoted flowering in Arabidopsis. Additionally, exogenous treatment with GA in Arabidopsis accelerates flowering under short-day conditions (Blažquez et al., 1998; Gocal et al., 2001). Previously it has been reported that GA and jasmonic acid (JA) antagonistically regulate flowering time in Arabidopsis as JA represses FT expression and mediates signaling cascades to delay flowering (Zhai et al., 2015).

Cotton is an important fiber crop and an ideal source of oilseed and feed. Although cotton is primarily cultivated as an annual crop, it is naturally a short-day photoperiodic perennial. Due to its perennial growth habit, crop management strategies are complicated. In this regard, more determinate cotton plant architecture is desired. Some genes, such as cotton GhTFL1-like genes, delay flowering due to ectopic expression in transgenic Arabidopsis (McCarty et al., 2016; Prewitt et al., 2018). Previously, expression of Arabidopsis FT in cotton led to altered plant architecture in terms of reducing indeterminate growth as well as perennial traits of cotton plant. Moreover, Arabidopsis FT expression uncoupled flowering from photoperiod in photoperiodic cotton and high florigen-synchronized flowering coupled with compressed growth habits in domesticated day-neutral cotton lines (McCarty and Ayre, 2012; McCarty et al., 2013). Cotton GhLFY, a homolog of florigca/leafy (LFY), is expressed in the shoot apex and plays an important role in flower initiation (Li et al., 2013). Breeding cotton for plant architecture, growth habit, and tolerance to environmental and hormonal stresses is a high priority for plant breeders. The α-amylase inhibitor (AAI) domain is a plant lipid transfer protein (LTP), hydrophobic seed protein, and a trypsin α-amylase protein gene family with 68 members in Arabidopsis. In Arabidopsis, LTP3 served as a target of MYB96 and mediated freezing and drought tolerance (Guo et al., 2013). AtLTP2 played a structural role in maintaining the integrity of adhesion between the hydrophobic cuticle and the hydrophilic underlying cell wall below in Arabidopsis (Jacq et al., 2017). Another study reported that a gain-of-function mutation of AtLTP5 disturbed pollen tube tip growth and subsequent fertilization in Arabidopsis (Chae et al., 2009). Cotton GhLTPG1 was reported to regulate cotton fiber elongation by mediating transport of phosphatidylinositol monophosphates (Deng et al., 2016). However, the functions of plant AAI domain-containing genes remain largely unknown in cotton.

Here, we identified 336 AAI domain-containing genes in nine different plant species including monocots, dicots, moss, and ferns. We also exclusively identified 122 AAI genes in Gossypium hirsutum. We then conducted a phylogenetic analysis and identified conserved amino acid residues, protein motif distribution pattern, encoded proteins, gene structure, chromosomal location, gene duplication, synteny analysis, and Ka/Ks values. Moreover, spatial expression patterns of GhAAI genes and responses under various phytohormone treatments were monitored. We also performed ectopic expression of GhAAI66 in Arabidopsis and silencing by virus-induced gene silencing (VIGS), conducted Kyoto Encyclopedia of Genes and Genomes (KEGG) and Gene Ontology (GO) analysis of RNA sequencing (RNA-seq) data of OE-GhAAI66 lines, and confirmed our results using quantitative reverse transcription–PCR (qRT–PCR) analysis to produce a proposed working model of GhAAI66 in Arabidopsis. This work will provide a basic foundation to improve cotton breeding for hormonal responses and cotton plant growth habit.

Materials and methods

Gene identification and analysis of conserved residues

First we identified AAI domain-containing genes in Arabidopsis down-loaded from TAIR 10 (http://www.arabidopsis.org). The retrieved Arabidopsis AAI genes were then used as a query to identify AAI genes in G. hirsutum (NAU, version 1.1), G. arboreum (ICR, version 1.0), G. raimondii (JGI, version 2.0), Theobroma cacao (version 10), Oryza sativa (version 10), Zea mays (version 10), Physcomitrella patens (version 3.3), Selaginella moellendorffii (version 1.0), and the alga Chlamydomonas reinhardtii (version 1.0). We downloaded the G. arboreum database from ftp://bioinfo.aist-nara.ac.jp/ftp/Gh AA166/ and G. raimondii and G. hirsutum from COTTONGEN (https://www.cottongen.org/). The databases for other plant species were downloaded from Phytozone v11 (https://phytozone.jgi.doe.gov/pz/portal.html). Next, we confirmed all putative AAI gene sequences using Interproscan 63.0 (http://www.ebi.ac.uk/InterProScan/) (Jones et al., 2014) and SMART (http://smart.embl-heidelberg.de/) (Letunic et al., 2015). Basic properties of GhAAI genes were estimated using ExPASy ProtParam (http://us.expasy.org/tools/protparam.html) and we used softberry (www.softberry.com) to predict the subcellular localization. Further, conserved amino acid residues analysis was conducted using the online tool WEBLOG (Crooks et al., 2004) after multiple alignment of conserved domain regions by ClustalX with default parameters (Thompson et al., 1997).

Phylogenetic analysis and determination of protein motif distribution and gene structure

For phylogenetic analysis, AAI genes were aligned and a phylogenetic tree was constructed using MEGA 7.0 (Kumar et al., 2016) with Neighbor-Joining (NJ) and minimum evolution (ME) methods. Further, 1000 bootstrap replicates were used to determine support values for the inferred phylogenetic trees. The phylogenetic tree was then visualized using TreeView1.6 (http://tree.bio.ed.ac.uk/software/treeweview/).

For determining the distribution pattern of protein motifs and gene structure analysis, a BED file of putative GhAAI sequences was used in Gene Structure Display Server 2.0 (http://gds.cbi.pku.edu.cn/index.php) (Hu et al., 2015) to obtain gene structure. The online MEME program (http://meme-suite.org/tools/meme) (Bailey et al., 2006) was used for protein motifs as described previously (Li et al., 2019).

Chromosomal location, gene duplication, and synteny analysis

Chromosomal location information for all GhAAI genes was obtained from a gff3-file of cotton genome annotation data (ftp://ftp.bioinfo.wsu.edu/species/Gossypium_hirsutum/NAU-NBI_G) and genes were mapped on the chromosomes using MapInspect software (https://
Ultra™ RNA Library Prep Kit for Illumina® (New England Biolabs, Ipswich, MA) was used to generate transcriptomic libraries of total extracted RNA were generated using a NEB-Next Oligo( dA) Tailing Module and the NEBNext Multiplex Library Prep Kit for Illumina. For RNA-seq analysis, whole plant seedlings of Arabidopsis Col-0 (Col-0) plants were collected 20 d after germination. Sequencing libraries were constructed using the clf (Cauliflower Mosaic Virus) CaMV 35S promoter and the construct was introduced into Agrobacterium tumefaciens strain GV3101. Transformation into Arabidopsis plants was performed using the floral dip method (Zhang et al., 2007). For ectopic expression of GhAA166, we used Arabidopsis thaliana Columbia-0 (Col-0) plants to generate transgenic OE-GhAA166 lines. We amplified cDNA of full-length GhAA166 genes by PCR with gene-specific primers. The amplified PCR product was then cloned into pCAMBIA-2301 to construct the vector driven by the constitutive Cauliflower mosaic virus (CaMV) 35S promoter and the construct was introduced into Agrobacterium tumefaciens strain GV3101. Transformation into Arabidopsis plants was employed using the floral dip method (Yang et al., 2015). Arabidopsis seeds were germinated on 1/2 Murashige and Skoog (MS) medium and grown under long-day conditions (16 h light/8 h dark) at 23 °C. Transgenic plants were selected on solid 1/2 MS medium plates supplemented with 50 µg mL^-1^ kanamycin. For VIGS, a highly specific region of 300 bp from the coding sequence (CDS) was cloned into the CLCRV binary vector and transformed into GV3101. Inoculation in CRI24 cotton plants was performed as described previously (Gao et al., 2011).

**Results**

### Identification of AAI genes and analysis of conserved amino acid residues in AAI domains

In this study, we identified 122 AAI genes in *G. hirsutum*, 34 in *G. arboresum*, and 33 in *G. raimondii*. Additionally, we identified AAI genes in different dicotyledons (68 genes in Arabidopsis and 22 in *T. cacao*), monocotyledons (18 in *O. sativa* and 21 in *Z. mays*), moss (11 in *P. patens*), and fern (7 in *S. moellendorffii*). However, no AAI gene was identified in algae (*C. reinhardtii*).

Next, analysis of conserved amino acid residues within AAI domains of Arabidopsis, *O. sativa*, *G. hirsutum*, *P. patens*, and *S. moellendorffii* depicted the degree of conservation of each residue in the AAI domains of all studied species (Supplementary Fig. S1A–E). The distribution of AAI domain amino acid residues was highly conserved in all species. For instance, amino acid residues such as C, P, and L were highly conserved in the AAI domain across all species and cysteine (C) amino acid residue was equally distributed in the N- and C-terminal ends (four in each) of AAI genes; however, C exhibited enrichment in the middle as compared with the ends. Our results indicated that AAI domain sequences were highly conserved among dicots, monocots, moss, and fern.

### Phylogenetic analysis of AAI genes

To investigate the evolutionary relationship among AAI genes of the nine aforementioned species, an unrooted phylogenetic tree was inferred using the NJ method. AAI genes were classified into five major groups (AAI-a to AAI-e) (Fig. 1). This phylogenetic analysis was then validated by constructing another tree using ME (see Supplementary Fig. S2); phylogenetic trees displayed consistent results including topologies of groups and numbers as well as positions of genes in...
corresponding groups. Group AAI-a was the largest with 93 members; AAI-b contained 65 members; AAI-c had 45 members; AAI-d had 48 members; and AAI-e contained 85 AAI genes. Most importantly, groups AAI-a and AAI-b were only present in dicotyledons and monocotyledons, with a lack of AAI genes from moss and fern. Similarly, groups AAI-c and AAI-d were present in dicotyledons, moss, and ferns, with a lack of AAI genes from monocotyledons. AAI-e was the only group containing AAI genes from moss, fern, dicotyledon, and monocotyledon plant species, illustrating that its evolution occurred before the separation of monocots, dicots, moss, and ferns. The other four AAI groups most probably emerged after the separation of moss, fern, monocot, and dicot plant species.

According to the phylogenetic tree, most orthologous genes between allotetraploids and diploids were clustered close to each other in the same group, showing expansion of the AAI gene family. Further investigation indicated that there were ~71 orthologous gene pairs between allotetraploid and A-genome diploid cotton, while there were 74 orthologous pairs between allotetraploid and D-genome diploid cotton. Additionally, 113 orthologous pairs were identified within At and Dt subgenomes of allotetraploid cotton. Based on this analysis, we deduced that there were more than twice as many

![Phylogenetic analysis by the NJ method and number of AAI genes in nine plant species.](This figure is available in color at JXB online.)
GhAAI genes than GaAAI and GrAAI genes, and that this might be the result of wide-ranging duplication events during polyploidization in G. hirsutum. Moreover, many closely clustered orthologous gene pairs were observed in all studied plant species including Arabidopsis, G. arboreum, G. raimondii, T. cacao, O. sativa, Z. mays, P. patens, and S. moellendorfii. These findings show that gene duplication was the main contributor to expansion of the AAI gene family in the plant kingdom.

It is an established fact that allotetraploid cotton (G. hirsutum) was derived as a result of polyploidization and hybridization between A (G. arboreum) and D (G. raimondii) diploid cotton genomes. To test this hypothesis, we constructed a phylogenetic NJ tree of AAI genes from three cotton species, namely G. hirsutum, G. arboreum, and G. raimondii (Supplementary Fig. S3). The phylogenetic tree divided cotton AAI genes into four main groups, with each group containing the AAI genes from all three cotton species closely clustered, forming orthologous and paralogous gene pairs among and within genomes, respectively. These results strengthen our hypothesis and validate our findings that G. hirsutum evolved from hybridization between G. arboreum and G. raimondii, and subsequent polyploidization.

Protein motifs and gene structure analysis

To better understand the evolutionary relationship among GhAAI gene family members, we generated a NJ tree of all 122 GhAAI genes along with protein motif distribution (Supplementary Fig. S4A) and gene structure (Supplementary Fig. S4B). The GhAAI genes with similar motif distribution patterns were clustered close to each other and made up one clade. Ten different motifs were distributed in all GhAAI proteins, ranging from three to six motifs in each protein. Further, all GhAAI genes had a conserved motif distribution pattern; for instance, motif 4 and motif 5 were found in all proteins. Gene structure analysis showed that GhAAI genes contained three different kinds of proteins, namely LTP2, hydrophobic seed protein, and trypsin α-amylase protein. The genes encoding similar kinds of proteins and intron–exon arrangements were closely clustered, occupying the position in one clade. Further, some genes encoding LTP2 proteins had one to multiple introns in their gene structure, except for a few genes. Introns were observed in the gene structures of 36 genes. Moreover, 51 GhAAI genes encoded hydrophobic seed protein while only one gene (GhAAI4) encoded trypsin α-amylase protein. However, GhAAI genes encoding hydrophobic seed protein and trypsin α-amylase protein lacked introns. Collectively, from 122 GhAAI genes, three kinds of proteins are encoded and only 36 genes had introns in their structures.

Chromosomal location, gene duplication, and synteny analysis of GhAAI genes

We investigated the chromosomal location of 122 identified GhAAI genes on their corresponding chromosomes. Mapping results indicated that 51 GhAAI genes were located on At subgenome chromosomes while 58 GhAAI genes were on Dt subgenome chromosomes (Supplementary Fig. S5). Moreover, 13 GhAAI genes were present in different scaffolds. The highest number of genes were located on A11 (11 GhAAI genes) and its orthologous chromosome D11 (14 GhAAI genes). However, no GhAAI genes were located on A06, D03, or D06 chromosomes. Absence of GhAAI genes on these chromosomes, uneven distributions, as well as scaffold locations of GhAAI genes illustrated translocation during evolution and incomplete genome sequencing.

Phylogenetic analysis revealed the existence of numerous orthologous and paralogous gene pairs generated by gene duplication, so we further explored the locus relationships among and within At and Dt subgenomes as well as with A and D diploid cotton genomes. Based on syntenic relationships, there are 42 paralogous gene pairs identified as a result of segmental duplication (Fig. 2; Supplementary Table S3). Whole-genome duplication (WGD) resulted in 216 orthologous gene pairs and, of these, there were 32 pairs between the At subgenome and A genome, 36 pairs between the At subgenome and D genome, 39 pairs between the Dt subgenome and A genome, 38 pairs between the Dt subgenome and D genome, 23 pairs between orthologous chromosomes, and 48 orthologous gene pairs between non-orthologous chromosomes. However, no tandem duplication events were observed. Notably, most GhAAI genes formed orthologous or paralogous pairs with multiple genes. For instance, GhAAI66 and GhAAI18 accounted for 16 and 15 orthologous or paralogous pairs with other cotton AAI genes, respectively, demonstrating the significant contribution of gene duplication in cotton AAI gene family expansion. These results strengthen the idea that G. hirsutum was derived from hybridization of two diploids resembling G. arboreum and G. raimondii (Wendel and Cronn, 2003; Li et al., 2015). Here, we presumed that the high proportion of GhAAI genes from WGD as well as segmental duplication also contributed significantly in the expansion of the GhAAI gene family.

Next, we estimated the nature and extent of selection pressure on duplicated gene pairs as duplicated gene pairs underwent functional divergence and exhibited neo-functionalization, sub-functionalization, or non-functionalization during evolution (Prince and Pickett, 2002). Results of non-synonymous (Ka) and synonymous (Ks) values showed that 244 duplicated pairs had Ka/Ks <1.0 and, of these, 168 duplicated pairs exhibited a Ka/Ks value <0.5. However, 14 duplicated gene pairs showed Ka/Ks values >1.0 (Supplementary Table S3). Collectively, cotton AAI duplicated genes experienced strong purifying selection pressure as a high proportion of Ka/Ks values of duplicated gene pairs were <1.0, demonstrating positive selection pressure.

Tissue-specific expression pattern and responses to phytohormone treatment

In order to determine the potential functions of GhAAI genes, we first identified the expression pattern of all 122 GhAAI genes in data of 22 different tissues obtained from published transcriptomes downloaded from NCBI, and generated a heat map (Supplementary Fig. S6). The results indicated that all genes exhibited ubiquitous expression with no specific pattern, and genes depicting similar expression patterns were
closely clustered. However, only GhAAI6, GhAAI60, and GhAAI66 displayed significant expression values in all tissues. To validate these results, we selected 12 paralogous genes based on the fact that orthologous or paralogous genes usually exhibit similar functions (Altenhoff and Dessimoz, 2009).

We analyzed the tissue-specific expression level of these 12 GhAAI genes in 15 different tissues including root, stem, leaf, flower, ovule (1, 3, 5, 7, 10, 15, and 20 DPA) and fiber (10, 7, 15, and 20 DPA) using qRT–PCR (Fig. 3). The results were in accordance with transcriptomic data, and the transcript levels of most genes were high in different vegetative tissues, except that expression of GhAAI2, GhAAI8, GhAAI18, GhAAI65, and GhAAI81 was high at different stages of ovule development. Importantly, for GhAAI66, which had maximum gene duplication, its expression level was high in flower, with a 2-fold increase, indicating its potential function during flower development.

Next we investigated the responses of GhAAI genes under five different phytohormone treatments, namely BL, GA, IAA, SA, and MeJA, after 0.5, 1, 3, and 5 h using qRT–PCR (Fig. 4). GhAAI2, GhAAI18, and GhAAI66 were up-regulated across all hormonal treatments except at a few time points. Other observed genes showed ubiquitous responses without any specific pattern. Overall, all observed GhAAI genes were positively regulated at higher levels than the control following GA treatment, except at a few time points in some genes. Based on these findings, we assumed that GhAAI genes have potential functions in plant growth and development, and can be regulated by phytohormone treatments. Additionally, GhAAI66 exhibited higher transcript levels in flower tissues and under phytohormone responses, indicating that it might play an important role in plant growth, development, and phytohormonal response.

Ectopic expression of GhAAI66 regulates phase transition to induce early flowering

To determine the roles of GhAAI66 in flower development or regulation and transition from vegetative to reproductive growth, we overexpressed coding sequences under the control of the CaMV 35S promoter in Arabidopsis Col-0 plants. Compared with wild-type (WT) plants, transgenic plants overexpressing GhAAI66 showed early flowering (Fig. 5A). The statistical data indicated that OE-GhAAI66 lines opened their first flowers in 19.60±0.92 d as compared with WT plants where the first flower opened in 23.96±0.58 d (Fig. 5B),
indicating the transition from vegetative to reproductive phase. Previously, the vegetative to reproductive phase transition was considered to have three phases: V-phase (vegetative rosette), I1-phase (inflorescence with cauline leaves subtending axillary branches), and I2-phase (an inflorescence bearing flowers) (Ratcliffe et al., 1998). As in OE-GhAAI66 lines, inflorescences bearing flowers were observed (Fig. 5A), so here we speculated that ectopic expression of GhAAI66 regulates the vegetative to reproductive phase transition through the I2-phase, which in turn induces early flowering.
To identify the function of *GhAAI66* for flowering in cotton, we conducted a VIGS experiment. First, we monitored the relative expression of *GhAAI66* in CLCrV:00 and CLCrV:*GhAAI66* plants to confirm the silenced expression of *GhAAI66*. The transcript level indicated that *GhAAI66* expression was reduced in CLCrV:*GhAAI66* plants (Fig. 5C). CLCrV:*GhAAI66* plants had delayed flowering, while CLCrV:00 plants initiated flowering along with buds and developed cotton bolls; in addition, the plant height of CLCrV:*GhAAI66* plants was suppressed significantly as compared with control plants (Fig 5D). These results strengthen our hypothesis that *GhAAI66* regulates the vegetative to reproductive phase transition to induce early flowering.
GhAAI66 triggers early flowering signaling pathways

Next, to explore the mechanism by which ectopic expression of GhAAI66 regulates the vegetative to reproductive phase transition and induces early flowering, we conducted RNA-seq analysis of WT and OE-GhAAI66 lines. Among 27,337 expressed genes, ~3,891 genes (14.23% of all expressed genes) were differentially expressed, comprising 2,410 that were up-regulated (61.93% genes of all DEGs) and 1,481 that were down-regulated (38.76% genes of all DEGs) in OE-GhAAI66 lines (Fig. 6A; Supplementary Table S4, S5). KEGG analysis of RNA-seq data indicated that genes for plant hormone signal transduction, starch and sucrose metabolism, and valine, leucine, and isoleucine (BCAA; branched chain amino acids) degradation were up-regulated (Fig. 6B). In contrast, genes for ribosomes, biosynthesis of amino acids, ribosome biogenesis, as well as valine, leucine, and isoleucine (BCAA) biosynthesis and many others were down-regulated (Fig. 6C). The up-regulation of genes for plant hormone signal transduction illustrated the involvement of hormone signaling pathways during the phase transition and early flowering induction. Findings during the past decade have elaborated that BCAA degradation provides energy for the early phase of germination and plays critical roles in maintaining amino acid homeostasis as well as normal seed development (Lu et al., 2011; Ding et al., 2012; Angelovici et al., 2013; Peng et al., 2015; Gipson et al., 2017). Here, genes for valine, leucine, and isoleucine degradation were up-regulated, while genes involved in their biosynthesis were down-regulated, demonstrating that BCAA homeostasis in OE-GhAAI66 lines might be an important contributor in the phase transition and early flowering. Further, GO enrichment showed that genes up-regulated in OE-GhAAI66 lines were mainly enriched in plant developmental processes as well as in cell wall metabolism, biosynthesis, and biogenesis processes (Fig. 6D). In contrast, down-regulated genes were involved in cell death and responses to chitin, JA, wounding, and SA (Fig. 6E).

Deeper investigation indicated that ELF4 (EARLY FLOWERING 4), ELF4-L2 (ELF4-LIKE 2), ELF4-L3 (ELF4-LIKE 3), ELF4-L4 (ELF4-LIKE 4), GI (GIGANTEA), and FT (FLOWERING LOCUS T) were up-regulated in RNA-seq data (Fig. 6D). In contrast, the JA receptor COI1 (CORONATINE INSENSITIVE 1), biosynthesis gene AOS (ALLENE OXIDE SYNTHASE), and the three flowering repressors TEM1 (TEMPRANILLO 1), TEM2 (TEMPRANILLO 2), and FLC (FLOWERING LOCUS C) were down-regulated. These results were further validated by qRT–PCR analysis (Fig. 6F). As GA plays a positive role in flowering induction (Sun and Gubler, 2004) and because in our previous results most GhAAI genes including GhAAI66 had a positive response during GA treatment, we further investigated
Fig. 6. Gene Ontology (GO) analysis of RNA-seq data of OE-GhAAI66 lines with respect to Col-0 Arabidopsis plants. (A) The expression map of the differentially expressed genes (DEGs). The bar graph shows the number of DEGs between Col-0 and OE-GhAAI66 plants. ‘Up’ and ‘down’ are the up-regulated and down-regulated genes in OE-GhAAI66 plants, respectively. (B) KEGG analysis of up-regulated genes from RNA-seq data. (C) KEGG analysis of down-regulated genes from RNA-seq data. Each experiment was conducted in three biological repeats. (D and E) Gene ontology enrichment of up- and down-regulated genes in OE-GhAAI66 plants. The lengths of the bars indicate the $-\log_{10}$-transformed $P$-values. (F, G) Relative expression pattern analysis of floral integrators, JA and GA receptors, biosynthesis, and catabolic genes by qRT–PCR analysis in order to validate RNA-seq data. (F) Up- and down-regulated floral integrators and JA receptor and biosynthesis genes. (G) Up- and down-regulated GA receptor and biosynthesis, responsive, and catabolic genes. Data are the log2 fold change. Each experiment was conducted with three biological repeats. The error bars show the SD of three independent biological repeats. (This figure is available in color at JXB online.)
the mechanism to determine whether GA is also playing a part in the GhAAI66-induced phase transition and early flowering in OE-GhAAI66 lines as GA seemed to be acting upstream of AAI genes during GA treatment. We found that gibberellin receptor GID1A (GIBBERELLIN INSENSITIVE DWARF 1A), GID1C (GIBBERELLIN INSENSITIVE DWARF 1C), GID1L1 (GID1-LIKE1), GID1L2 (GID1-LIKE2), CXE17 (CARBOXYESTERASE 17), CXE18 (CARBOXYESTERASE 18), gibberellin biosynthetic GASA4 (GIBBERELLIC ACID STIMULATED ARABIDOPSIS 4), GASA10 (GIBBERELLIC ACID STIMULATED ARABIDOPSIS 10), and SOC1 (SUPPRESSOR OF OVEREXPRESSION CONSTANS1) were down-regulated in RNA-seq.

Discussion

AAI genes form a large family in many plant species. For example, there are 68 AAIIs in Arabidopsis and 122 in allotetraploid cotton G. hirsutum. The members of this gene family encode three domains, namely the LTP2 domain, the hydrophobic seed domain, and the trypsin α-amylase domain. Despite this, no systematic study on the AAI gene family in any species had been conducted until now. There are only a few studies about AAI genes containing the LTP2 domain (Hsu et al., 2005; Chae et al., 2009; Guo et al., 2013; Deng et al., 2016; Jacq et al., 2017); however, there are no studies about AAI genes carrying the hydrophobic seed domain or trypsin α-amylase domain. Cotton is an important fiber crop and is the main source of fiber for the textile industry (Bao et al., 2011). Advances in cotton genomics and genetics allowed us to perform a systematic study on cotton AAI genes and to investigate their potential functions. This study will provide basic information for further investigation of cotton AAI gene functions.

AAI genes were highly conserved during evolution

In the current study, we classified 336 AAI genes from different dicotyledons (Arabidopsis, G. hirsutum, G. arboreum, G. raimondii, and T. cacao), monocotyledons (O. sativa and Z. mays), moss (P. patens), and fern (S. moellendorffii) into five major groups. The results of phylogenetic analysis using NJ and ME methods were consistent and validated our findings. With the exception of group AAI-e, all AAI groups showed advanced evolution as groups AAI-a and AAI-b lack fern and moss AAI genes and groups AAI-c and AAI-d lack AAI genes from monocots, highlighting their evolution after the separation of fern and moss or monocots, respectively. Group AAI-e members evolved after the separation of moss, fern, monocot, and dicot species. WOX and YABBY gene families exhibited conserved amino acid residues and were found to be evolutionarily conserved (Yang et al., 2017, 2018). Similarly, in our study, AAI domain sequences were highly conserved among dicot (Arabidopsis and G. hirsutum), monocot (O. sativa), moss (P. patens), and fern (S. moellendorffii).

Ten different motifs were distributed in all GhAAI proteins, ranging from three to six motifs in each protein, demonstrating that GhAAI genes displayed a conserved motif distribution pattern. Tandem duplication has been reported always to result in more introns and new genes (Iwamoto et al., 1998). In our study, no tandem duplication events were observed, while 36 out of 122 GhAAI genes had one to multiple introns and, surprisingly, all 36 genes encoded LTP2 protein. Introns might play essential roles during the evolution of different species (Roy and Gilbert, 2006). It has been reported that genes contained more introns at early stages of expansion, with introns being lost with passage of time (Roy and Penny, 2007), suggesting that more advanced families had fewer introns in their genome (Roy and Gilbert, 2005). Despite the fact that conserved exon/intron motifs are functionally important even with low sequence conservation, exon/intron patterns of many gene families are conserved (Frugoli et al., 1998). We thus speculated that these introns were not lost during evolution, but diverged at early expansion stages of evolution, while other genes lost their introns over evolutionary time.

Previous findings documented many gene families lacking, or with fewer, introns in their genes (Serrano et al., 2006; Qambar et al., 2018; Zhang et al., 2018). Insertion/deletion events contribute to exon/intron structural differences that might be helpful to estimate evolutionary mechanisms (Lecharny et al., 2003). Intron are under weak selection pressure, and genes with no introns might evolve at a rapid rate, while genes with larger or more introns contributed to gain of function in evolution. Further, gene loss/addition by segmental or WGD as well as incomplete sequencing of genomes are the main reasons for uneven distributions of GhAAI genes in At and Dt chromosomes of allotetraploid cotton.

Expansion and duplication of the GhAAI gene family during evolution

Allotetraploid cotton (G. hirsutum) evolved as a result of hybridization between A (G. arboreum) and D (G. raimondii) cotton genomes and subsequent polyploidization ~5–10 million years ago (mya) (Li et al., 2015) and provides the best model to study polyploidy (Wendel and Cronn, 2003). Generally, functional divergence is contributed by gene duplication that is important for environmental adaptability and evolutionary mechanisms (Conant and Wolfe, 2008). Segmental and salicoid duplications enlarged many gene families in ancestral plants ~65 mya (Barakat et al., 2009; Wang et al., 2013). Two large segmental and small-scale tandem duplications generated novel genes during evolution that accounted for genomic complexities in the plant kingdom (Cannon et al., 2004).

We identified 122 GhAAI genes, which is more than double the number of AAI genes in G. arboreum and G. raimondii, respectively, illustrating the effect of polyploidy which resulted in more gene duplication in the GhAAI gene family as compared with AAI genes in other cotton species (G. arboreum and
G. raimondii). The dramatic increase in GhAAI family members can be evaluated from AA1 genes in G. arboreum and G. raimondii. Although GhAAI genes were increased, gene loss after hybridization during genomic arrangements and chromosome doubling occurred (Paterson et al., 2012). Additionally, the cotton genome underwent fewer arrangements with respect to paleopolyploid maize and Brassica (Gaeta et al., 2007; Woodhouse et al., 2010). Our study identified 42 paralogous and 216 orthologous gene pairs as a result of segmental duplication and WGD, respectively. Interestingly, GhAAI66 and GhAAI18 accounted for 16 and 15 orthologous/paralogous pairs, demonstrating a significant contribution of gene duplication in cotton AA1 gene family expansion.

Segmental duplication is important during evolution, as many plant species have multiple duplicated chromosomal blocks (Cannon et al., 2004). Further, it has been reported that Arabidopsis experienced WGD twice (Wang et al., 2011). Moreover, malvids (cotton and cacao) displayed a common ancestor and underwent ancient duplication ~18–58 mya (Li et al., 2014). Many Arabidopsis gene families also show gene family expansion (Baumberger et al., 2003; Wang et al., 2008). Additionally, cotton RH2FE3, YABBY, WOX, GRAS, and MIK-C-Type MADS-Box, sesame heat shock proteins, and soybean WRKY show enlargement as the result of segmental duplication and WGD (Yin et al., 2013; Dossa et al., 2016; Ren et al., 2017; Yang et al., 2017, 2018; Qambar et al., 2018; Zhang et al., 2018). Gene duplication analysis of AA1 genes revealed a common ancestor between the A-genome and At subgenome similar to D-genome and Dt subgenome. We concluded that segmental duplication and WGD are attributes of expansion in the GhAAI gene family. These findings will provide understanding of chromosomal interactions, genetic evolution, and intergenomic hereditary information transfer.

**GhAAI genes have ubiquitous expression in tissues and regulated hormone treatments**

Few investigations have been conducted to explore biological and physiological functions of AA1 genes in plant biology. For instance, LTP2 is involved in cuticle–cell wall interface integrity and in permeability of the etiolated hypocotyl (Jaqc et al., 2017). LTP3 was positively regulated by MYB96 as it directly binds to the LTP3 promoter and is involved in plant tolerance to freezing and drought stress in Arabidopsis (Guo et al., 2013). Further, GhMYB7/9 mediates transcriptional regulation of the LTP3 gene during fiber development in cotton (Hsu et al., 2005). GhLTPG1 was located on the cell membrane and was highly expressed in elongating fibers and the outer integument of cotton ovules (Deng et al., 2016). However, no studies have been conducted on AA1 genes encoding the hydrophobic seed domain or trypsin α-amylase domain.

We determined that GhAAI gene expression was ubiquitous in different tissues. However, the transcript level of most genes was high in different vegetative tissues, except that in GhAAI2, GhAAI18, GhAAI18, GhAAI65, and GhAAI81, expression was high at different stages of ovule development. Importantly, the GhAAI66 expression level was high in flower, indicating its potential function during flower development. Our results from qRT–PCR analysis and previously published RNA–seq and transcriptomic data results were reasonable and validated our findings.

Additionally, all GhAAI genes exhibited up- or down-regulation when treated with various phytohormones. For instance, GhAAI2, GhAAI8, and GhAAI66 were up-regulated under all hormonal treatments with few exceptions. Overall, all estimated GhAAI genes were positively regulated following GA exposure except for a few time points with some genes. These findings strengthen our hypothesis that GhAAI genes might play important roles in signaling of different hormones in addition to the fact that GhAAI genes can be regulated by phytohormone treatments. Taken together, we speculated that GhAAI genes play diverse roles in plant growth and development under different hormonal treatments.

**GhAAI66 regulates early flowering in Arabidopsis**

In our study, ectopic expression of GhAAI66 regulated the phase transition from vegetative to reproductive growth to induce early flowering as OE-GhAAI66 lines opened their first flowers in fewer days as compared with WT plants that flowered 4 d later, indicating the transition from vegetative to reproductive phase. We deduced that ectopic expression of GhAAI66 regulates the vegetative to reproductive phase transition through the 12-phase, which in turn induces early flowering. Moreover, silenced CLCrV-GhAAI66 plants showed delayed flowering resulting from late phase transition from vegetative to reproductive growth, supporting our findings that GhAAI66 regulates the vegetative to reproductive phase transition to induce early flowering.

Decades of genetic research have revealed complex genetic networks for floral transition in Arabidopsis regulated by four genetic pathways: photoperiod, vernalization, autonomous, and GA-induced pathways (Simpson and Dean, 2002; Boss et al., 2004; Sung and Amasino, 2004; Baurle and Dean, 2006). The floral induction signals from these pathways are delivered to CONSTANS (CO) and FLC that antagonistically regulate flowering in Arabidopsis (Putterill et al., 1995; Samach et al., 2000). Here, GhAAI66 was shown to integrate multiple flower signaling pathways to induce early flowering in Arabidopsis, and we proposed a working model for GhAAI66 to induce flowering (Fig. 7). We found that expression of various GA receptor genes (GID1A, GID1C, GID1L1, GID1L2, CXE17, and CXE18), biosynthesis genes (GA5, GA20OX3, KSI, and SOCI), and enzymes was induced, while the expression of the GA catalytic gene GA20OX6 and JA receptor and biosynthesis genes (COI1 and AOS, respectively) was repressed. Previously, GA and JA were reported to antagonistically regulate flowering in Arabidopsis. JA promotes COI1-dependent degradation of JAZs, which in turn liberates transcriptional functions of the TOEs (TARGET of EAT1 and 2) to repress FT expression and mediate signaling cascades to delay flowering (Zhai et al., 2015). SOCI is integrated with the GA pathway: for instance, the slo1 null mutant had reduced sensitivity to GA, and further SOCI overexpression rescued the non-flowering phenotype of ga1-3 in Arabidopsis (Blazquez et al., 1998; Gocal et al., 2001). Even the genes responsive to SA were down-regulated in this study; it has previously been shown that SA regulates...
flowering time through photoperiod and autonomous pathways; however, it does not require functional flowering genes such as CO, FCA, and FLC (Martinez et al., 2004).

The transcripts of ELF4 and its orthologous genes, GI, FT, and SOC1, were up-regulated in OE-GhAAI66 lines, whereas, previously, physical interaction proved ELF4 to be a regulator of GI nuclear distribution where GI binds the promoter of CO and directly activates the expression of FT (Sawa and Kay, 2011; Kim et al., 2013). Moreover, important flowering repressors including TEM1/2 and FLC were repressed in our transcriptomic data, and qRT–PCR analysis supported the previous findings that FLC protein interacts directly in vivo with SOC1 and FT (Helliwell et al., 2006), and prevents them from triggering a signal cascade of flower induction. For instance, overexpression of FLC completely blocked the activation of SOC by CO overexpression (Hepworth et al., 2002), whereas photoperiod deficiency in co and gi mutants was rescued when combined with the increased level of FLC generated by fa (Koornneef et al., 1998). Moreover, FT is negatively regulated by TEM1/2 and FLC (Zhai et al., 2015), and reports demonstrated that FT is the output of CO and SOC1 regulated through FT (Wigge et al., 2005; Yoo et al., 2005).

Taken together, we concluded that GhAAI66 integrates multiple flower signaling pathways to induce early flowering in Arabidopsis.

### Supplementary data

Supplementary data are available at JXB online.

**Table S1.** Proposed names of putative AAI genes and their gene locus ID in *A. thaliana*, *G. arboreum*, *G. hirsutum*, *G. raimondii*, *O. sativa*, *T. cacao*, *P. patens*, *S. moellendorffii*, and *Z. mays*.  
**Table S2.** Basic properties of GhAAI genes with locus ID, start and end point, strand, CDS, protein length, MW (molecular weight), pl (isoelectric point), GRAVY, as well as predicted subcellular localization.  
**Table S3.** Gene duplication analysis in cotton.  
**Table S4.** RNA-seq data of up-regulated genes in OE-GhAAI66 of three independent transgenic lines.  
**Table S5.** RNA-seq data of down-regulated genes in OE-GhAAI66 of three independent transgenic lines.  
**Table S6.** List of all primers used in this study for gene cloning and qPCR analysis.

Fig. S1. Conserved amino acid residue analysis among (A) Arabidopsis, (B) *O. sativa*, (C) *G. hirsutum*, (D) moss, and (E) fern AAI genes.  
Fig. S2. Phylogenetic tree constructed using ME.  
Fig. S3. The phylogenetic tree was generated using the NJ method for cotton including *G. arboreum*, *G. hirsutum*, and *G. raimondii* AAI genes in order to estimate the common ancestor hypothesis.  
Fig. S4. Protein motif distribution and gene structure (exon/intron) analysis of GhAAI genes along with the phylogenetic tree inferred using the NJ method.  
Fig. S5. Chromosomal location of GhAAI genes on different chromosomes.  
Fig. S6. Expression pattern of GhAAI genes in different cotton tissues

### Author contributions

ZY and FL designed the experiments and coordinated the project; GQ, LL, and ZL performed the experiments and analyzed the data; DY, KZ, and PH prepared the samples; and GQ, ZY, and FL wrote and revised the manuscript.

### Acknowledgements

This project is supported by the Funds for Creative Research Groups of China (grant no. 31621005) and the Agricultural Science and Technology Innovation Program Cooperation and Innovation Mission (CAAS-XTCX2016). The authors declare that they have no competing interests.

### References

Qanmber et al.


Yin GJ, Xu HL, Xiao SY, Qin YJ, Li YX, Yan YM, Hu YK. 2013. The large soybean (Glycine max) WRKY TF family expanded by segmental duplication events and subsequent divergent selection among subgroups. BMC Plant Biology 13.


