



Article Genome-Wide Identification and Expression Analyses of the Cotton AGO Genes and Their Potential Roles in Fiber Development and Stress Response

Mingchuan Fu, Yizhen Chen, Hao Li, Liguo Wang, Renzhong Liu and Zhanji Liu *D

Key Laboratory of Cotton Breeding and Cultivation in Huang-Huai-Hai Plain, Ministry of Agriculture and Rural Affairs, Institute of Industrial Crops Shandong Academy of Agricultural Sciences, Jinan 250100, China * Correspondence: scrcliuzhanji@sina.com; Tel.: +86-531-6665-9992

Abstract: Argonaute proteins (AGOs) are indispensable components of RNA silencing. However, systematic characterization of the *AGO* genes have not been completed in cotton until now. In this study, cotton *AGO* genes were identified and analyzed with respect to their evolution and expression profile during biotic and abiotic stresses. We identified 14 *GaAGO*, 14 *GrAGO*, and 28 *GhAGO* genes in the genomes of *Gossypium arboreum*, *Gossypium raimondii*, and *Gossypium hirsutum*. Cotton AGO proteins were classified into four subgroups. Structural and functional conservation were observed in the same subgroups based on the analysis of the gene structure and conserved domains. Twenty-four duplicated gene pairs were identified in *GhAGO* genes, and all of them exhibited strong purifying selection during evolution. Moreover, RNA-seq analysis showed that most of the *GhAGO* genes exhibit high expression levels in the fiber initiation and elongation processes. Furthermore, the expression profiles of *GhAGO* genes tested by quantitative real-time polymerase chain reaction (qPCR) demonstrated that they were sensitive to Verticillium wilt infection and salt and drought stresses. Overall, our results will pave the way for further functional investigation of the cotton *AGO* gene family, which may be involved in fiber development and stress response.

Keywords: cotton; AGO gene family; expression pattern; Verticillium wilt; drought and salt stress

1. Introduction

Argonaute (AGO) proteins are core effectors of the RNA-induced silencing complex (RISC), which are well known to regulate gene expression in RNA interference (RNAi) or RNAi-independent pathways [1]. For instance, Arabidopsis AGO10 specifically binds to microRNA166/165 to regulate the development of the shoot apical meristem. A loss of function of *AGO10* resulted in the pinhead phenotype in the Arabidopsis ecotype Ler mutant plants [2]. Typically, AGO proteins contain four domains viz. N-terminal (Argo-N), Piwi Argonaute Zwille (PAZ), MID, and PIWI domains. The Argo-N domain is believed to participate in the separation of the small RNA:target duplex, while the PAZ domain contains a specific binding pocket that can anchor small RNAs. The MID domain can bind the 5' phosphates of small RNAs and anchors small RNAs onto the AGO proteins. The PIWI domain is functionally similar to RNase H with endonuclease activity and is in charge of the cleavage of target mRNA. In addition to these domains, the Argo-L1 and Argo-L2 domains are revealed in a number of AGO proteins [3].

The *AGO* gene family has been widely studied in a number of plant species. In general, different numbers of *AGO* genes are present in various plant species. Ten *AGO* genes have been revealed in *Arabidopsis thaliana* [4], 19 in *Oryza sativa* [5], 17 in *Zea mays* [6], 15 in *Solanum lycopersicum* [7], 13 in *Vitis vinifera* [8], 7 in *Cucumis sativus* [9], 27 in *Brassica napus* [10], 12 in *Capsicum annuum* [11], 13 in *Citrus sinensis* [12], 69 in *Triticum aestivum* [13], 13 in *Phaseolus vulgaris* [14], 14 in *Solanum tuberosum* [15], 18 in *Camellia*



Citation: Fu, M.; Chen, Y.; Li, H.; Wang, L.; Liu, R.; Liu, Z. Genome-Wide Identification and Expression Analyses of the Cotton *AGO* Genes and Their Potential Roles in Fiber Development and Stress Response. *Genes* 2022, *13*, 1492. https://doi.org/10.3390/ genes13081492

Academic Editors: Zengjian Jeffrey Chen and Youlu Yuan

Received: 18 July 2022 Accepted: 19 August 2022 Published: 20 August 2022

Publisher's Note: MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Copyright: © 2022 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (https:// creativecommons.org/licenses/by/ 4.0/). *sinensis* [3], and 13 in *Musa acuminata* [16]. However, no *AGO* genes in *Gossypium* species have been elucidated to date.

Gossypium hirsutum (AD₁) is an allotetraploid, which evolved from interspecific hybridization between an A-subgenome species and a D-subgenome species around 1-2 million years ago [17]. G. arboreum (A2) and a G. raimondii (D5) are considered as the putative donor diploid species for the A- and D-subgenomes, respectively. Cotton has been grown as the most important natural fiber crop and provides approximately 35% of the total fiber consumed worldwide [18]. It is also a vital oilseed and feed crop in the world owing to the high contents of oil and protein in cottonseeds [19]. However, cotton production is severely constrained by several biotic and abiotic stresses such as Verticillium wilt, salt, and drought, which result in decreased yields and inferior harvest quality. In this study, members of the AGO gene family were identified from systematic analyses of the allotetraploid G. hirsutum genome, as well as two diploid progenitors G. arboreum and *G. raimondii*. The gene structure and conserved domains of these genes were comprehensively analyzed, and the gene duplication events were identified. The expression patterns of GhAGO genes were examined in various organs using public RNA-Seq data and after treatments with Verticillium wilt, salt, and drought via quantitative RT-PCR (qPCR). This study will provide comprehensive information about the AGO genes and pave the way for further investigation of their function in cotton.

2. Materials and Methods

2.1. Identification of Gossypium AGO Genes

A. thaliana AGO genes were retrieved from the Arabidopsis Information Resource (TAIR v10, https://www.arabidopsis.org/) (accessed on 19 November 2021) [20], while rice (*O. sativa* L., v7.0) *AGO* genes were obtained from the Phytozome v13 (https://phytozomenext.jgi.doe.gov/) (accessed on 19 November 2021) [21]. We used the HMMsearch program (http://hmmer.org/) (accessed on 22 November 2021) to screen the genome sequences of *G. hirsutum* (HAU, v1.1), *G. arboreum* (CRI, v1.0), and *G. raimondii* (JGI, v2.0) (https://cottonfgd.net/) (accessed on 22 November 2021) [22], employing the hidden Markov model (HMM) profiles of the PAZ domain (PF02170) and PIWI domain (PF02171) downloaded from the Pfam database (http://pfam.xfam.org/) (accessed on 22 November 2021) as queries [23]. The resulting sequences were then verified by the Pfam and Conserved Domain Database (CDD, https://www.ncbi.nlm.nih.gov/cdd) (accessed on 22 November 2019) [24]. The ExPASy ProtParam tool (https://web.expasy.org/protparam/) (accessed on 23 November 2021) was used to predict the molecular weight (MW) and theoretical isoelectric point (pI) of each AGO protein. Subcellular localization analysis was performed using the web server BUSCA (http://busca.biocomp.unibo.it/) (accessed on 23 November 2021) [25].

2.2. Phylogenetic, Gene Structure, and Conserved Domain Analyses

The phylogenetic tree was built by the neighbor-joining (NJ) method using MEGA 11 [26]. Bootstrap analysis was simulated with 1000 replicates. The exon–intron structures were visualized using the Gene Structure Display Server 2.0 (GSDS 2.0, http://gsds.cbi.pku.edu.cn/) (accessed on 22 March 2022) [27]. Conserved domain analysis of the AGO proteins was performed by the HMMER web server (https://www.ebi.ac.uk/Tools/hmmer/search/hmmscan) (accessed on 22 March 2022) using default parameters [28].

2.3. Chromosomal Mapping and Gene Duplication Analyses

The positions of the *GhAGO* genes were determined by mRNA location information retrieved from the GFF files of *G. hirsutum* genome. Additionally, the chromosomal distribution of *GhAGO* genes was illustrated by the MapChart (v2.32) program [29]. Gene duplication events were determined using methods described previously [30]. The Circos program was used to demonstrate the relationships of duplicated genes [31]. The value of nonsynonymous substitutions to synonymous substitutions (Ka/Ks) was calculated using the KaKs_Calculator package [32].

2.4. Cis-Acting Regulatory Element Analysis

We extracted the upstream 1500 bp DNA sequences of the *GhAGO* genes from the *G. hirsutum* genome sequences [33] and, then, submitted to the PlantCARE database (http://bioinformatics.psb.ugent.be/webtools/plantcare/html/) (accessed on 8 April 2022) [34]. The bed-files obtained from the above database were illustrated using GSDS 2.0 [27].

2.5. Transcriptome Data Analysis

The tissue-specific expression patterns of upland cotton *AGO* genes were measured using the transcriptome data of *G. hirsutum* TM-1, which were retrieved from NCBI (https://www.ncbi.nlm.nih.gov/bioproject/PRJNA248163) (accessed on 11 April 2022) [35]. The heatmap of the 28 *GhAGO* genes' expression profiles were performed using the TBtools package [36].

2.6. Plant Materials and Stress Treatments

Healthy seeds of *G. hirsutum* cv. Lumian418 were planted in sterilized soil at 28 °C with a photoperiod of 16 h light/8 h dark. Seedlings were grown for two weeks, then gently transferred to Hoagland solution for two days, and finally, transferred to Hoagland solution containing 200 mM NaCl or 15% (w/v) PEG6000. For Verticillium wilt infection, two-week-old seedlings were infected by the high-virulence VD8 strain of *Verticillium dahliae* (2 × 10⁷ spores/mL) by the root-dip method [37]. Roots from three biological replicates were harvested at 2, 6, and 12 h after treatment.

2.7. Quantitative Real-time PCR Analysis

RNA was isolated using Trizol reagent (Invitrogen, Carlsbad, CA, USA) and digested with RNase-free DNase I (Takara, Dalian, China) to eliminate trace genomic DNA. The quality of RNA samples was investigated by 1% agarose gel electrophoresis, while the concentration was examined by a spectrophotometer (NanoDrop 2000, Waltham, MA, USA). First-strand cDNA was synthesized from 1 μ g of RNA using a PrimerScriptTM 1st Strand cDNA Synthesis Kit (Takara, Dalian, China). The qPCR assay was performed in an ABI QuantStudio 5 Real-Time PCR System (Thermo Fisher Scientific, Waltham, MA, USA) using SYBR[®] Premix Ex TaqTM (Takara, Dalian, China) with three replicates. The qPCR procedure was set as follows: one cycle at 95 °C for 3 min, then 40 cycles at 95 °C for 15 s, and 60 °C for 15 s. The cotton *Histone3* (AF024716) was used as the internal reference gene [38]. The relative expression levels were calculated according to the 2^{- $\Delta\Delta$ Ct} method [39]. The qPCR primers are listed in Table S1.

2.8. Statistical Analysis

The qPCR data from three biological replicates were analyzed as the mean \pm the standard error. The difference between treatment and control was evaluated by Tukey's honestly significant difference tests. "*" and "**" indicate significant differences at $p \le 0.05$ and $p \le 0.01$, respectively.

3. Results

3.1. Identification and Phylogenetic Analysis of Gossypium AGO Genes

To determine the *AGO* genes in the genomes of *G. raimondii*, *G. arboretum*, and *G. hirsutum*, we carried out a genome-wide survey using HMMER search with the PAZ domain (PF02170) and PIWI domain (PF02171). After manual inspection and confirmation using the Pfam and CDD databases, 14 *G. raimondii AGO* (*GrAGO*), 14 *G. arboretum AGO* (*GaAGO*), and 28 *G. hirsutum AGO* (*GhAGO*) genes were identified in total (Table S2). The nomenclature of cotton *AGO* genes was defined based on the closest orthologs in Arabidopsis and rice. Comparison analysis suggested that the *G. hirsutum* genome harbored all copies of the *AGO* genes from the two diploid progenitor species. In addition, we found that the length of *Gossypium* AGO proteins ranged from 359 (GhAGO5c) to 1145 (GaAGO1b) amino acids. The predicted molecular weight ranged from 40.93 to 127.31 kDa, and the

calculated pI values ranged from 8.98 to 9.61 (Table S3). Additionally, all *Gossypium* AGO proteins were predicted to be nuclear proteins, which were highly consistent with AGO proteins from Arabidopsis and rice.

To identify the evolutionary relationship of the *Gossypium* AGO proteins, we constructed a phylogenetic tree using the NJ method (Figure 1). The result indicated that the *Gossypium* AGO proteins clustered into four separate subgroups (i.e., AGO1, MEL1, AGO4, and ZIPPY), which was consistent with Arabidopsis, rice, and pepper [11]. Among the four subgroups, subgroup AGO1 contained the most AGO members with 23 *Gossypium* AGO proteins, while subgroup MEL1 contained the least AGO members with only 9 *Gossypium* AGO proteins. In addition, both AGO4 and ZIPPY subgroups contained 12 *Gossypium* AGO proteins. Furthermore, compared with those in Arabidopsis, *AGO1, AGO4, AGO5, AGO7*, and *AGO10* have greatly expanded in *Gossypium* species, while *AGO3, AGO8*, and *AGO9* were not detected in cotton.



Figure 1. Phylogenetic tree of the Gossypium AGO proteins. The AGO protein sequences of Gossypium, Arabidopsis, and rice were aligned using ClustalW. The MEGA 11 program was used to generate the NJ tree with 1000 bootstrap replicates. Different subgroups of AGO proteins are highlighted with various colors.

3.2. Genomic Localization and Gene Duplication Analysis of GhAGO Genes

GhAGO genes were anchored to their corresponding chromosomes (Figure 2). A total of 27 *GhAGO* genes unevenly distributed on 16 chromosomes and one (*GhAGO5e*) was mapped on a D-subgenome scaffold region (Scaffold635). Among the 16 chromosomes, 11 chromosomes (A06, A07, A08, A10, A12, D06, D07, D08, D10, D12, and D13) contained a single *GhAGO* gene. In addition, 2 chromosomes (A05 and A13) and 3 chromosomes (A09, D05, and D09) possessed two and four *GhAGO* genes, respectively (Figure 2).



Figure 2. Chromosomal localizations of the *GhAGO* genes. Green bars indicate the *G. hirsutum* chromosomes. The scale bar on the left denotes the chromosomal lengths (Mb).

To reveal *GhAGO* duplication events, we performed a genome-wide collinearity analysis using the MCScanX program [40]. As a result, 2159 collinear blocks were identified in the *G. hirsutum* (HAU) genome and 54,839 genes (78.12%) were collinear genes. In particular, 22 pairs of *GhAGO* genes were segmental duplications, which involved 22 *GhAGO* genes, whereas two pairs (*GhAGO5a/GhAGO5b* and *GhAGO5c/GhAGO5d*) were tandem duplications (Figure 3 and Table S4). The Ka/Ks values of all duplication pairs were less than 1, ranging from 0.063 to 0.859 (Table S4), indicating that the *AGO* gene family in *G. hirsutum* had undergone purifying selection during the long evolutionary period.



Figure 3. Collinearity analysis of *GhAGO* genes in *G. hirsutum*. The red lines highlight the 22 pairs of segmental duplications. The scale bar marked on the chromosome indicates chromosome lengths (Mb).

3.3. Gene Structure and Conserved Domain Analysis of GhAGO Genes

We analyzed the *GhAGO* gene structure and display it in Figure 4. The number of exons in the *GhAGO* genes varied from 3 in *GhAGO2d*, *GhAGO7a*, *GhAGO7b*, and *GhAGO7c* to 26 in *GhAGO5a*, which might be related to the diversification of their functions. Interestingly, the *GhAGO* genes from the same phylogenetic subgroup share highly similar gene structures. For example, in the subgroup AGO4, most *GhAGO* genes and the counterparts from *G. arboretum* and *G. raimondii* contained 21 introns and 22 exons. However, *GhAGO6d* had 22 introns and 23 exons. Additionally, *GhAGO5c* showed a distinct pattern of gene structure. It seemed to have lost part of the nucleotides from both the PAZ and Piwi domains.



Figure 4. Conserved domains and gene structures of the *GbNAC* genes. A: Phylogenetic tree of the *GhAGO* genes. Multiple sequence alignment of the GhAGO proteins was performed using ClustalW. The neighbor-joining (NJ) tree was constructed using MEGA 11 with 1000 bootstrap replicates. B: Conserved domain of the *GhAGO* genes. The conserved domain of the GhAGO proteins was detected by the HMMER web server (https://www.ebi.ac.uk/Tools/hmmer/search/hmmscan) (accessed on 22 March 2022) using default parameters. C: Gene structure of the *GhAGO* genes. The blue boxes and black lines represent exons and introns, respectively.

Seven conserved domains were identified among the 28 GhAGO proteins (Figure 4). As a result, with the exception of GhAGO5c, all GhAGO proteins shared four conserved domains, N-terminal ArgoN domain, PAZ, ArgoL1, and C-terminal Piwi domain, which is in line with known plant AGOs [41,42]. Notably, a Gly-rich Ago1 domain was revealed in front of the ArgoN domain in the GhAGO1a/b/c/d/e proteins. Additionally, previous efforts revealed that the Piwi domain exhibited substantial homology to RNase H and enabled some AGO proteins to cleave target RNAs pairing to the bound small RNAs [43]. Generally, the catalytic activity is associated with the conserved triad (aspartate-aspartatehistidine, DDH) and an additional conserved histidine at position 798 (H798) in Arabidopsis AGO1 [5,8]. In this study, we aligned the Piwi domains of all GhAGO proteins and 10 AtAGO proteins using ClustalX (http://www.clustal.org/, accessed on 22 November 2021). The result indicated that 17 GhAGO proteins contained the conserved DDH/H798 residues (Table S5). Among the other 11 GhAGO proteins, 6 GhAGOs, all from the AGO4 subgroup, contained the conserved DDH triad, but the histidine at the 798th position in AtAGO1 was replaced by a proline. In GhAGO2a and GhAGO2d, the histidine at the 986th position in AtAGO1 was replaced by an aspartate.

3.4. Cis-Acting Regulatory Elements in Promoter Region of GhAGO Genes

A number of *cis*-acting regulatory elements were revealed in the promoter regions of the 28 *GhAGO* genes. The essential regulatory elements such as TATA-box and CAAT-box were detected in all *GhAGO* genes. Other *cis*-acting elements identified in the *GhAGO* genes can be divided into four groups according to their functional properties, namely light response, stress response, plant growth, and hormone-responsive elements. The distribution of these elements identified in promoter regions of each *GhAGO* gene is illustrated in Figure 5. In particular, we identified eight elements associated with six hormone responses.

These *cis*-acting elements include the AuxRR-core and TGA-element related to auxin response, the GARE-motif and P-box associated with gibberellin response, ABRE involved in abscisic acid (ABA) response, ERE related to ethylene-response, the CGTCA-motif involved in methyl jasmonate (MeJA) response, and the TCA-element associated with salicylic acid (SA) response. Additionally, the promoter sequences of some *GhAGO* genes also contain several elements involved in environmental stress responses, including pathogen defense (AT-rich and TC-rich), cold (DRE and LTR), wounding (WUN-motif), and anaerobic stress (ARE). Taken together, these results suggested that the *GhAGO* genes might play vital roles in regulating cotton response to hormone and environmental stresses.



Figure 5. Analysis of cis-acting regulatory elements of the *GhAGO* genes. The yellow lines represent the promoter regions of the *GhAGO* genes. The scale bar at the bottom denotes the length of the promoter sequence.

3.5. GhAGO Gene Expression Patterns in Diverse Cotton Tissues

The expression patterns of *GhAGO* genes in diverse tissues were investigated by using the method described by Zheng et al. [44]. The results indicated that the expression levels of the *GhAGO* genes were significantly different in diverse tissues (Figure 6). *GhPNH1a* and *GhPNH1b* were detected with high expression levels in the leaf and ovule 20 days post-anthesis (dpa). *GhAGO5c* showed a relatively distinctive expression pattern compared to all the other analyzed genes. This was due to the significant difference in its level of accumulation in the pistil. Additionally, all of the *GhAGO* genes showed extremely low expression levels in the calycle, petal, and stamen (Figure 6). In particular, most of the

GhAGO genes were expressed highly in the early (-3-5 dpa) and middle (3-20 dpa) stages of ovule development, suggesting that these genes may function in the fiber initiation and elongation processes.



Figure 6. Expression profiles of the *GhAGO* genes in various tissues. The expression levels are illustrated in different colors on the scale. Red represents high expression, and blue indicates low expression. dpa is an abbreviation for days post-anthesis.

3.6. GhAGO Genes Were Influenced by Verticillium Wilt Infection

To investigate whether the *GhAGO* genes play roles in pathogen defense in cotton, we analyzed the transcriptional patterns of the *GhAGO* genes in response to *V. dahliae* infection by qPCR at 2, 6, and 12 h post-inoculation (hpi). The results demonstrated that Verticillium wilt infection significantly affected the expression of the *GhAGO* genes (Figure 7). Half of the tested *GhAGO* genes were significantly suppressed by Verticillium wilt infection, while five *GhAGO* genes (*GhAGO2a, GhAGO4a, GhAGO7c, GhAGO7d,* and *GhPNH1d*) were remarkably induced by 2h treatment with Verticillium wilt infection. In addition, five members out of the tested *GhAGO* genes contained pathogen-responsive element AT-rich or TC-rich. Verticillium wilt infection significantly influenced the expression levels of *GhAGO2a, GhAGO7a, GhPNH1a,* and *GhPNH1d,* but had almost no effect on the expression of *GhAGO1a* (Figure 7). Interestingly, some *GhAGO* genes might have undergone neofunctionalization after duplication. For instance, the expression level of *GhAGO7a* was significantly downregulated by 2h treatment with *V. dahliae* infection, while its homeolog, *GhAGO7d,* was greatly upregulated (Figure 7).



Figure 7. Expression analyses of the *GhAGO* genes in roots under Verticillium wilt infection. qPCR was used to determine the expression profiles of 14 *GhAGO* genes under Verticillium wilt infection. The standard deviation is indicated by the error bars, and "*" (Tukey's HSD, $p \le 0.05$) and "**" ($p \le 0.01$) indicate significant differences between the treatment and control.

3.7. GhAGO Genes Were Modulated by Salt and Drought Stresses

Accumulating evidence indicates that RNA silencing exerts essential functions in plant resistance to abiotic stresses [11,13]. Therefore, the expression of the *GhAGO* genes was evaluated after salt and drought treatments. The qPCR assay showed that salt and drought stresses resulted in significant differences in the expression levels of the *GhAGO* genes (Figure 8). Surprisingly, all of the tested *GhAGO* genes were significantly downregulated by salt stress, indicating that these *GhAGO* genes might be important in response to salt stress. Similarly, the expression levels of seven *GhAGO* genes were downregulated by drought stress. In addition, the expression of *GhAGO1a* decreased significantly at 2 h after drought treatment, increased remarkably at 6 h, and then, decreased at 12 h. The expression level of *GhAGO10b* was greatly upregulated at 2 h and downregulated at 6 h. Conversely, the expression levels of *GhAGO2a*, *GhAGO2a*, and *GhAGO7c* were not modified under drought stress (Figure 8).



Figure 8. Expression analyses of the *GhAGO* genes in roots under salt and drought stresses. qPCR was used to determine the expression profiles of 12 *GhAGO* genes under salt and drought stresses. The standard deviation is indicated by the error bars, and "*" (Tukey's HSD, $p \le 0.05$) and "**" ($p \le 0.01$) indicate significant differences between the treatment and control.

4. Discussion

4.1. Characterization of Gossypium AGO Genes

In this study, we carried out a survey of the *Gossypium AGO* genes at the wholegenome scale to examine their potential functions in fiber development and stress response. Consequently, 28 *GhAGO* genes were identified in the *G. hirsutum* genome, which is twice as much as that found in *G. arboreum* and *G. raimondii*, presumably because *G. hirsutum* is a tetraploid species evolved from the hybridization between the diploid *G. raimondii* and *G. arboretum* [17]. Specifically, the respective number of *AGO2*, *AGO4*, *AGO6*, *AGO7*, *AGO10*, and *PNH1* genes in the tetraploid *G. hirsutum* is exactly double the numbers in diploid cotton, suggesting that no recent gene duplication or deletion occurred in these *AGO* genes after allotetraploid formation. However, gene deletion in *AGO1* and gene duplication in *AGO5* were observed because tetraploid upland cotton contains 5 *AGO1* and 5 *AGO5* genes, while each diploid cotton contains 3 *AGO1* and 2 *AGO5* genes, respectively (Figure 1).

Cotton *AGO* genes were classified into four subgroups viz. AGO1, MEL1, AGO4, and ZIPPY, which is consistent with the previous results based on 206 *AGO* genes from 23 plant species [45]. The *AGO* genes, however, were divided into three subgroups (AGO1/5/10,

AGO4/6/8/9, and AGO2/3/7) in several plant species such as common bean [14] and potato [15]. Indeed, compared to three subgroups in common bean and potato, we split the subgroup AGO1/5/10 into two subgroups AGO1 and MEL1 (homolog of AGO5) with well-supported bootstrap values, which is consistent with the results reported for rice [5] and grapevine [8]. Extensive studies showed that several AGO subgroups have expanded through lineage-specific duplication [45]. For instance, compared with Arabidopsis, subgroups AGO1 and MEL1 (AGO5) expanded in a number of flowering species. More specifically, common bean harbors four AGO10 genes [14], whereas wheat contains 15 AGO5 paralogs [13]. In cotton, expansion of AGO5 and AGO10 has been also identified. The upland cotton contains five AGO5 and six AGO10 paralogs (Figure 1). In addition, AGO1 (five members), AGO4 (four members), and AGO7 (four members) have significantly expanded in the G. hirsutum genome. Meanwhile, we further investigated AGO gene duplication in the upland cotton genome and revealed 24 duplicated *GhAGO* gene pairs, including 22 segmental duplication pairs and 2 tandem duplication pairs (Figure 3). These results suggested that segmental duplication was the primary driving force for the expansion of the *GhAGO* genes. Similarly, the expansion of wheat *AGO* genes was dominated by segmental duplication [13]. Furthermore, all duplicated gene pairs had undergone strong purifying selection in the process of evolution (Table S4), indicating that purifying selection exerted an important role in the formation of the *GhAGO* gene functions. Previous studies have demonstrated that the cotton lineage experienced a fiveto sixfold ploidy increase approximately 57–70 million years ago (Mya) [46], while the A and D ancestor genome diverged around 6.2–7.1 Mya [17]. In this study, about half of the duplication events occurred after the cotton lineage ploidy increase. Additionally, one duplicate (GhAGO7a/GhAGO7d) might occur after the divergence of the two diploid progenitors (Table S4).

4.2. Differential Expression of GhAGO Genes in Response to Multiple Stresses

Extensive evidence shows that AGO2 and AGO4 play vital roles in the modulation of plant immunity [47–49]. In Arabidopsis, AtAGO2, which was highly induced by the bacterial pathogen *Pseudomonas syringae*, regulates innate immunity by binding miRNA393b* to orchestrate exocytosis of antimicrobial pathogenesis-related proteins [47]. AtAGO4 acts as a positive regulator of DNA methylation and mediates resistance to *P. syringae* [48]. In *Nicotiana attenuata*, among all 11 *NaAGO* genes, only transcripts of *NaAGO4a* and *NaAGO4b* were induced by hemibiotrophic pathogen *Fusarium brachygibbosum* infection. A loss of function of NaAGO4 confers mutant hypersusceptible to F. brachygibbosum [49]. Moreover, a strong upregulation of PvA_AGO2a and PvA_AGO4a expression was observed in P. vulgaris after inoculation with the fungus *Colletotrichum lindemuthianum* [14]. Additionally, StAGO15 in Solanum tuberosum belonging to the AGO4 clade was suppressed at 0-3 dpi and then significantly activated at 4–5 dpi when Phytophthora infestans had completed the transition from the biotrophic to necrotrophic stage [15]. In this study, to determine the contribution of some of the *GhAGO* genes involved in the defense response in upland cotton, we performed expression analysis in roots inoculated with the fungus V. dahliae based on the qPCR assay. The expression levels of GhAGO2a, GhAGO4a, GhAGO7c, GhAGO7d, and *GhPNH1d* were significantly upregulated at 2 hpi. On the contrary, the expression of GhAGO1b, GhAGO2d, GhAGO5b, GhAGO5d, GhAGO7a, GhAGO7b, and GhAGO10b was downregulated (Figure 7). Notably, the transcripts of most of the *GhAGO* genes containing pathogen-responsive element were significantly influenced by V. dahliae infection. These results suggest that these *GhAGO* genes may participate in the regulation of cotton defense.

Emerging evidence suggests that the *AGO* genes not only contribute to biotic stress response, but also modulate plant resistance to abiotic stresses such as salt, drought, cold, and heat stresses [8,13,50]. Arabidopsis *AtAGO2* was significantly induced by salt stress. Further analysis revealed that AtAGO2 improves Arabidopsis salt tolerance by interacting with an R3H-type RNA binding protein MUG13.4 and then influences the SOS signaling cascade at the transcription level [51]. In *Z. mays*, 17 *ZmAGO* genes have been reported, and all *ZmAGO* genes were induced under drought stress. Surprisingly, transcripts of *ZmAGO18a* and *ZmAGO18b* were 539.9-fold and 730.8-fold upregulated at 1h under drought stress in comparison to the control. The mutation of *ZmAGO18b* rendered plants hypersensitive to drought stress [50]. In *Setaria italica*, the mutation of *SiAGO1b* resulted in enhanced susceptibility to drought stress [52]. In addition, the expression levels of most of the *TaAGO* genes in *T. aestivum* were influenced by salt and drought stresses [13]. Similarly, most of the *VvAGO* genes in *V. vinifera* were downregulated under salt and drought stresses [8]. Salt and drought are considered the most destructive abiotic stresses to cotton. In this study, all of the tested *GhAGO* genes were significantly suppressed by salt stress. Two *GhAGO* genes (*GhAGO1a* and *GhAGO10b*) exhibited upregulation in response to drought stress at specific time points. This result indicated that these *GhAGO* genes may play important roles in plant adaptation to salt and drought stresses.

5. Conclusions

This study performed a systematic analysis of the *AGO* gene family in three *Gossypium* species. A total of 14 *GrAGO*, 14 *GaAGO*, and 28 *GhAGO* genes were revealed in the genomes of *G. raimondii*, *G. arboretum*, and *G. hirsutum*, respectively. The *Gossypium AGO* genes were divided into four distinct subgroups. Duplication analysis demonstrated that the *GhAGO* genes experienced segmental and tandem duplication events during evolution. Furthermore, the predicted *cis*-acting regulatory elements of the *GhAGO* genes suggested their functional association with growth, development, hormone response, and environmental stress response. Tissue-specific expression analysis indicated that most of the identified *GhAGO* genes may play pivotal roles in the fiber initiation and elongation processes. Our qPCR analyses revealed that a number of *GhAGO* genes were involved in the response to *V. dahliae* infection and salt and drought stresses. Overall, our results will provide a solid foundation for further functional characterization of the *GhAGO* genes in response to biotic and abiotic stresses.

Supplementary Materials: The following Supporting Information can be downloaded at: https://www.mdpi.com/article/10.3390/genes13081492/s1, Table S1: The primers used for qPCR. Table S2: The gene ID of AGO genes identified in the genomes of Arabidopsis, rice, *Gossypium hirsutum*, *G. arboreum*, and *G. raimondii*. Table S3: The detail properties of cotton *AGO* genes. Table S4: Segmental and tandem duplications of *GhAGO* gene pairs in upland cotton and inference of duplication time. Table S5: The conserved DDH/H798 residues in upland cotton and Arabidopsis.

Author Contributions: Conceptualization, Z.L.; data curation, M.F., Y.C. and H.L.; formal analysis, M.F. and Y.C.; investigation, Y.C. and L.W.; methodology, Z.L.; software, M.F.; visualization, L.W.; writing—original draft, M.F. and Z.L.; writing—review and editing, R.L. and Z.L. All authors have read and agreed to the published version of the manuscript.

Funding: This study was financially supported by grants from the Seed-Industrialized Development Program in Shandong Province (2020LZGC002), and the Innovation Project of Agricultural Science and Technology of Shandong Academy of Agricultural Sciences (CXGC2021A46).

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: Not applicable.

Conflicts of Interest: The authors declare no conflict of interest.

References

- 1. Liu, C.; Xin, Y.; Xu, L.; Cai, Z.; Xue, Y.; Liu, Y.; Xie, D.; Liu, Y.; Qi, Y. Arabidopsis ARGONAUTE 1 binds chromatin to promote gene transcription in response to hormones and stresses. *Dev. Cell* **2018**, *44*, 348–361. [CrossRef] [PubMed]
- Zhu, H.; Hu, F.; Wang, R.; Zhou, X.; Sze, S.H.; Liou, L.W.; Barefoot, A.; Dickman, M.; Zhang, X. Arabidopsis Argonaute 10 specifically sequesters miR166/165 to regulate shoot apical meristem development. *Cell* 2011, 145, 242–256. [CrossRef] [PubMed]

- Krishnatreya, D.B.; Baruah, P.M.; Dowarah, B.; Chowrasia, S.; Mondal, T.K.; Agarwala, N. Genome-wide identification, evolutionary relationship and expression analysis of AGO, DCL and RDR family genes in tea. *Sci. Rep.* 2021, *11*, 8679. [CrossRef] [PubMed]
- 4. Baumberger, N.; Baulcombe, D.C. Arabidopsis ARGONAUTE1 is an RNA Slicer that selectively recruits mRNAs and siRNAs. *Proc. Natl. Acad. Sci. USA* 2005, *102*, 11928–11933. [CrossRef] [PubMed]
- Kapoor, M.; Arora, R.; Lama, T.; Nijhawan, A.; Khurana, J.P.; Tyagi, A.K.; Kapoor, S. Genome-wide identification, organization and phylogenetic analysis of Dicer-like, Argonaute and RNA-dependent RNA Polymerase gene families and their expression analysis during reproductive development and stress in rice. *BMC Genom.* 2008, *9*, 451. [CrossRef]
- 6. Qian, Y.; Cheng, Y.; Cheng, X.; Jiang, H.; Zhu, S.; Cheng, B. Identification and characterization of Dicer-like, Argonaute and RNA-dependent RNA polymerase gene families in maize. *Plant Cell Rep.* **2011**, *30*, 1347–1363. [CrossRef]
- Bai, M.; Yang, G.S.; Chen, W.T.; Mao, Z.C.; Kang, H.X.; Chen, G.H.; Yang, Y.H.; Xie, B.Y. Genome-wide identification of Dicer-like, Argonaute and RNA-dependent RNA polymerase gene families and their expression analyses in response to viral infection and abiotic stresses in *Solanum lycopersicum*. *Gene* 2012, 501, 52–62. [CrossRef]
- Zhao, H.; Zhao, K.; Wang, J.; Chen, X.; Chen, Z.; Cai, R.; Xiang, Y. Comprehensive analysis of Dicer-like, Argonaute and RNA-dependent RNA polymerase gene families in grapevine (*Vitis vinifera*). J. Plant Growth Regul. 2015, 34, 108–121. [CrossRef]
- Gan, D.; Liang, D.; Wu, J.; Zhan, M.; Yang, F.; Xu, W.; Zhu, S.; Shi, J. Genome-wide identification of the Dicer-like, Argonaute, and RNA-dependent RNA polymerase gene families in cucumber (*Cucumis sativus* L.). J. Plant Growth Regul. 2016, 35, 135–150. [CrossRef]
- Cao, J.Y.; Xu, Y.P.; Li, W.; Li, S.S.; Rahman, H.; Cai, X.Z. Genome-wide identification of Dicer-like, Argonaute, and RNA-dependent RNA polymerase gene families in *Brassica* Species and functional analyses of their Arabidopsis homologs in resistance to *Sclerotinia sclerotiorum*. Front. Plant Sci. 2016, 7, 1614. [CrossRef]
- 11. Qin, L.; Mo, N.; Muhammad, T.; Liang, Y. Genome-wide analysis of *DCL*, *AGO*, and *RDR* gene families in pepper (*Capsicum annuum* L.). *Int. J. Mol. Sci.* **2018**, *19*, 1038. [CrossRef] [PubMed]
- Sabbione, A.; Daurelio, L.; Vegetti, A.; Talon, M.; Tadeo, F.; Dotto, M. Genome-wide analysis of AGO, DCL and RDR gene families reveals RNA-directed DNA methylation is involved in fruit abscission in *Citrus sinensis*. *BMC Plant Biol.* 2019, 19, 401. [CrossRef] [PubMed]
- Liu, Y.F.; Wang, L.M.; Zhao, L.Z.; Wang, W.; Zhang, H.X. Genome-wide identification and evolutionary analysis of Argonaute genes in hexaploidy bread wheat. *Biomed. Res. Int.* 2021, 2021, 9983858.
- Alvarez-Diaz, J.C.; Richard, M.M.S.; Thareau, V.; Teano, G.; Paysant-Le-Roux, C.; Rigaill, G.; Pflieger, S.; Gratias, A.; Geffroy, V. Genome-Wide Identification of Key Components of RNA Silencing in Two *Phaseolus vulgaris* Genotypes of Contrasting Origin and Their Expression Analyses in Response to Fungal Infection. *Genes* 2022, 13, 64. [CrossRef]
- 15. Liao, Z.; Hoden, K.P.; Singh, R.K.; Dixelius, C. Genome-wide identification of Argonautes in Solanaceae with emphasis on potato. *Sci. Rep.* **2020**, *10*, 20577. [CrossRef] [PubMed]
- Ahmed, F.F.; Hossen, M.I.; Sarkar, M.A.R.; Konak, J.N.; Zohra, F.T.; Shoyeb, M.; Mondal, S. Genome-wide identification of DCL, AGO, and RDR gene families and their associated functional regulatory elements analyses in banana (*Musa acuminata*). *PLoS* ONE 2021, 16, e0256873. [CrossRef] [PubMed]
- 17. Hu, Y.; Chen, J.; Fang, L.; Zhang, Z.; Ma, W.; Niu, Y.; Ju, L.; Deng, J.; Zhao, T.; Lian, J.; et al. *Gossypium barbadense* and *Gossypium hirsutum* genomes provide insights into the origin and evolution of allotetraploid cotton. *Nat. Genet.* **2019**, *51*, 739–748. [CrossRef]
- Abdelraheem, A.; Esmaeili, N.; O'Connell, M.; Zhang, J. Progress and perspective on drought and salt stress tolerance in cotton. *Ind. Crop. Prod.* 2019, 130, 118–129. [CrossRef]
- 19. Song, J.; Pei, W.; Wang, N.; Ma, J.; Xin, Y.; Yang, S.; Wang, W.; Chen, Q.; Zhang, J.; Yu, J.; et al. Transcriptome analysis and identification of genes associated with oil accumulation in upland cotton. *Physiol. Plant.* **2022**, 174, e13701. [CrossRef]
- Reiser, L.; Subramaniam, S.; Li, D.; Huala, E. Using the Arabidopsis information resource (TAIR) to find information about Arabidopsis genes. *Curr. Protoc. Bioinf.* 2017, 60, 1.11.1–1.11.45. [CrossRef]
- Ouyang, S.; Zhu, W.; Hamilton, J.; Lin, H.; Campbell, M.; Childs, K.; Thibaud-Nissen, F.; Malek, R.L.; Lee, Y.; Zheng, L.; et al. The TIGR rice genome resource: Improvements and new features. *Nucl. Acids Res.* 2007, 35, D883–D887. [CrossRef] [PubMed]
- Zhu, T.; Liang, C.; Meng, Z.; Sun, G.; Meng, Z.; Guo, S.; Zhang, R. CottonFGD: An integrated functional genomics database for cotton. BMC Plant Biol. 2017, 17, 101. [CrossRef] [PubMed]
- Finn, R.D.; Coggill, P.; Eberhardt, R.Y.; Eddy, S.R.; Mistry, J.; Mitchell, A.L.; Potter, S.C.; Punta, M.; Qureshi, M.; Sangrador-Vegas, A.; et al. The Pfam protein families database: Towards a more sustainable future. *Nucleic Acids Res.* 2016, 44, D279–D285. [CrossRef] [PubMed]
- 24. Marchler-Bauer, A.; Derbyshire, M.K.; Gonzales, N.R.; Lu, S.; Chitsaz, F.; Geer, L.Y.; Geer, R.C.; He, J.; Gwadz, M.; Hurwitz, D.I.; et al. CDD: NCBI's conserved domain database. *Nucl. Acids Res.* **2015**, *43*, D222–D226. [CrossRef]
- 25. Savojardo, C.; Martelli, P.L.; Fariselli, P.; Profiti, G.; Casadio, R. BUSCA: An integrative web server to predict subcellular localization of proteins. *Nucleic Acids Res.* 2018, *46*, W459–W466. [CrossRef]
- Tamura, K.; Stecher, G.; Kumar, S. MEGA 11: Molecular evolutionary genetics analysis version 11. Mol. Biol. Evol. 2021, 38, 3022–3027. [CrossRef]
- Hu, B.; Jin, J.; Guo, A.Y.; Zhang, H.; Luo, J.; Gao, G. GSDS 2.0: An upgraded gene feature visualization server. *Bioinformatics* 2015, 31, 1296–1297. [CrossRef]

- Potter, S.C.; Luciani, A.; Eddy, S.R.; Park, Y.; Lopez, R.; Finn, R.D. HMMER web server: 2018 update. *Nucl. Acids Res.* 2018, 46, W200–W204. [CrossRef]
- 29. Voorrips, R. MapChart: Software for the graphical presentation of linkage maps and QTLs. J. Hered. 2002, 93, 77–78. [CrossRef]
- Liu, Z.; Fu, M.; Li, H.; Chen, Y.; Wang, L.; Liu, R. Systematic analysis of NAC transcription factors in Gossypium barbadense uncovers their roles in response to Verticillium wilt. *PeerJ* 2019, 7, e7995. [CrossRef]
- Krzywinski, M.; Schein, J.E.; Birol, I.; Connors, J.; Gascoyne, R.; Horsman, D.; Jones, S.J.; Marra, M.A. Circos: An information aesthetic for comparative genomics. *Genome Res.* 2009, 19, 1639–1645. [CrossRef] [PubMed]
- Zhang, Z.; Li, J.; Zhao, X.Q.; Wang, J.; Wong, G.K.; Yu, J. KaKs_Calculator: Calculating Ka and Ks through model selection and model averaging. *Genom. Proteom. Bioinf.* 2006, 4, 259–263. [CrossRef]
- Wang, M.; Tu, L.; Yuan, D.; Zhu, D.; Shen, C.; Li, J.; Liu, F.; Pei, L.; Wang, P.; Zhao, G.; et al. Reference genome sequences of two cultivated allotetraploid cottons, *Gossypium hirsutum* and *Gossypium barbadense*. *Nat. Genet.* 2019, *51*, 224–229. [CrossRef] [PubMed]
- Lescot, M.; Dehais, P.; Thijs, G.; Marchal, K.; Moreau, Y.; Van de Peer, Y.; Rouze, P.; Rombauts, S. PlantCARE, a database of plant cis-acting regulatory elements and a portal to tools for in silico analysis of promoter sequences. *Nucl. Acids Res.* 2002, 30, 325–327. [CrossRef]
- Zhang, T.; Hu, Y.; Jiang, W.; Fang, L.; Guan, X.; Chen, J.; Zhang, J.; Saski, C.A.; Scheffler, B.E.; Stelly, D.M.; et al. Sequencing of allotetraploid cotton (*Gossypium hirsutum* L. acc. TM-1) provides a resource for fiber improvement. *Nat. Biotechnol.* 2015, 33, 531–537. [CrossRef]
- Chen, C.; Chen, H.; Zhang, Y.; Thomas, H.R.; Frank, M.H.; He, Y.; Xia, R. TBtools: An integrative toolkit developed for interactive analyses of big biological data. *Mol. Plant* 2020, *13*, 1194–1202. [CrossRef]
- Chen, J.Y.; Huang, J.Q.; Li, N.Y.; Ma, X.F.; Wang, J.L.; Liu, C.; Liu, Y.F.; Liang, Y.; Bao, Y.M.; Dai, X.F. Genome-wide analysis of the gene families of resistance gene analogues in cotton and their response to verticillium wilt. *BMC Plant Biol.* 2015, 15, 148. [CrossRef]
- Li, W.; Zhang, D.; Zhu, G.; Mi, X.; Guo, W. Combining genome-wide and transcriptome-wide analyses reveal the evolutionary conservation and functional diversity of aquaporins in cotton. *BMC Genom.* 2019, 20, 538. [CrossRef]
- Livak, K.J.; Schmittgen, T.D. Analysis of relative gene expression data using real-time quantitative PCR and the 2^{-ΔΔCT} method. *Methods* 2001, 25, 402–408. [CrossRef]
- 40. Wang, Y.; Tang, H.; DeBarry, J.D.; Tan, X.; Li, J.; Wang, X.; Lee, T.H.; Jin, H.; Marler, B.; Guo, H.; et al. MCScanX: A toolkit for detection and evolutionary analysis of gene synteny and collinearity. *Nucl. Acids Res.* **2012**, *40*, e49. [CrossRef]
- 41. Fang, X.; Qi, Y. RNAi in plants: An Argonaute-centered view. Plant Cell 2016, 28, 272–285. [CrossRef] [PubMed]
- Li, Z.; Li, W.; Guo, M.; Liu, S.; Liu, L.; Yu, Y.; Mo, B.; Chen, X.; Gao, L. Origin, evolution and diversification of plant ARGONAUTE proteins. *Plant J.* 2022, 109, 1086–1097. [CrossRef] [PubMed]
- 43. Hock, J.; Meister, G. The Argonaute protein family. Genome Biol. 2008, 210, 210. [CrossRef] [PubMed]
- Zheng, L.; Wu, H.; Qanmber, G.; Ali, F.; Wang, L.; Liu, Z.; Yu, D.; Wang, Q.; Xu, A.; Yang, Z. Genome-wide study of the GATL gene family in *Gossypium hirsutum* L. reveals that *GhGATL* genes act on pectin synthesis to regulate plant growth and fiber elongation. *Genes* 2020, 11, 64. [CrossRef]
- 45. Rodriguez-Leal, D.; Castillo-Cobian, A.; Rodriguez-Arevalo, I.; Vielle-Calzada, J.P. A primary sequence analysis of the ARG-ONAUTE protein family in plants. *Front. Plant Sci.* **2016**, *7*, 1347. [CrossRef]
- Paterson, A.H.; Wendel, J.F.; Gundlach, H.; Guo, H.; Jenkins, J.; Jin, D.; Llewellyn, D.; Showmaker, K.C.; Shu, S.; Udall, J.; et al. Repeated polyploidization of *Gossypium* genomes and the evolution of spinnable cotton fibres. *Nature* 2012, 492, 423–428. [CrossRef]
- Zhang, X.; Zhao, H.; Gao, S.; Wang, W.C.; Katiyar-Agarwal, S.; Huang, H.D.; Raikhel, N.; Jin, H. Arabidopsis Argonaute 2 regulates innate immunity via miRNA393^{*}-mediated silencing of a Golgi-localized SNARE gene *MEMB12*. *Mol. Cell* 2011, 42, 356–366. [CrossRef]
- Agorio, A.; Vera, P. ARGONAUTE4 is required for resistance to Pseudomonas syringe in Arabidopsis. *Plant Cell* 2007, 19, 3778–3790. [CrossRef]
- 49. Pradhan, M.; Pandey, P.; Baldwin, I.T.; Pandey, S.P. Argonaute4 modulates resistance to *Fusarium brachygibbosum* infection by regulating jasmonic acid signaling. *Plant Physiol.* **2020**, *184*, 1128–1152. [CrossRef]
- 50. Zhai, L.; Teng, F.; Zheng, K.; Xiao, J.; Deng, W.; Sun, W. Expression analysis of Argonaute genes in maize (*Zea mays* L.) in response to abiotic stress. *Hereditas* **2019**, *156*, 27. [CrossRef]
- 51. Wang, H.; Liu, C.; Ren, Y.; Wu, M.; Wu, Z.; Chen, Y.; He, L.; Tang, B.; Huang, X.; Shabala, S.; et al. An RNA-binding protein MUG13.4 interacts with AtAGO2 to modulate salinity tolerance in Arabidopsis. *Plant Sci.* **2019**, *288*, 110218. [CrossRef] [PubMed]
- Liu, X.; Tang, S.; Jia, G.; Schnable, J.C.; Su, H.; Tang, C.; Zhi, H.; Diao, X. The C-terminal motif of SiAGO1b is required for the regulation of growth, development and stress responses in foxtail millet (*Setaria italica* (L.) P. Beauv). *J. Exp. Bot.* 2016, 67, 3237–3249. [CrossRef] [PubMed]