



Article

# Genome-Wide Identification and Classification of Arabinogalactan Proteins Gene Family in *Gossypium* Species and *GhAGP50* Increases Numbers of Epidermal Hairs in *Arabidopsis*

Renhui Wei <sup>1,†</sup>, Ziru Guo <sup>1,†</sup>, Zheng Yang <sup>2</sup>, Yanpeng Zhao <sup>2</sup>, Haoliang Yan <sup>1</sup>, Muhammad Tehseen Azhar <sup>2,3</sup>, Yamin Zhang <sup>2</sup>, Gangling Li <sup>2</sup>, Jingtao Pan <sup>1</sup>, Aiying Liu <sup>1</sup>, Wankui Gong <sup>1</sup>, Qun Ge <sup>1</sup>, Juwu Gong <sup>1</sup>, Youlu Yuan <sup>1,2,\*</sup> and Haihong Shang <sup>1,2,\*</sup>

<sup>1</sup> State Key Laboratory of Cotton Bio-Breeding and Integrated Utilization, Institute of Cotton Research, Chinese Academy of Agricultural Sciences, Anyang 455000, China; wrh7117@163.com (R.W.); 18838990553@163.com (Z.G.); yanhl1989@163.com (H.Y.); panjingtao@caas.cn (J.P.); liuaiying@caas.cn (A.L.); gongwankui@caas.cn (W.G.); gequn@caas.cn (Q.G.); gongjuwu@caas.cn (J.G.)

<sup>2</sup> Zhengzhou Research Base, State Key Laboratory of Cotton Bio-Breeding and Integrated Utilization, School of Agricultural Sciences, Zhengzhou University, Zhengzhou 450001, China; yangz@zzu.edu.cn (Z.Y.); ypzha@zzu.edu.cn (Y.Z.); tehseenpb@uaf.edu.pk (M.T.A.); ymzhang2017@163.com (Y.Z.); ligangling@zzu.edu.cn (G.L.)

<sup>3</sup> Department of Plant Breeding and Genetics, University of Agriculture, Faisalabad 38040, Pakistan

\* Correspondence: yuanyoulu@caas.cn (Y.Y.); shanghaihong@caas.cn (H.S.)

† These authors contributed equally to this work.

**Abstract:** Arabinogalactan proteins (AGPs) constitute a diverse class of hydroxyproline-rich glycoproteins implicated in various aspects of plant growth and development. However, their functional characterization in cotton (*Gossypium* spp.) remains limited. As a globally significant economic crop, cotton serves as the primary source of natural fiber, making it essential to understand the genetic mechanisms underlying its growth and development. This study aims to perform a comprehensive genome-wide identification and characterization of the AGP gene family in *Gossypium* spp., with a particular focus on elucidating their structural features, evolutionary relationships, and functional roles. A genome-wide analysis was conducted to identify AGP genes in *Gossypium* spp., followed by classification into distinct subfamilies based on sequence characteristics. Protein motif composition, gene structure, and phylogenetic relationships were examined to infer potential functional diversification. Subcellular localization of a key candidate gene, *GhAGP50*, was determined using fluorescent protein tagging, while gene expression patterns were assessed through  $\beta$ -glucuronidase (GUS) reporter assays. Additionally, hormonal regulation of *GhAGP50* was investigated via treatments with methyl jasmonate (MeJA), abscisic acid (ABA), indole-3-acetic acid (IAA), and gibberellin (GA). A total of 220 AGP genes were identified in *Gossypium* spp., comprising 19 classical AGPs, 28 lysine-rich AGPs, 55 AG peptides, and 118 fasciclin-like AGPs (FLAs). Structural and functional analyses revealed significant variation in gene organization and conserved motifs across subfamilies. Functional characterization of *GhAGP50*, an ortholog of AGP18 in *Arabidopsis thaliana*, demonstrated its role in promoting epidermal hair formation in leaves and stalks. Subcellular localization studies indicated that *GhAGP50* is targeted to the nucleus and plasma membrane. GUS staining assays revealed broad expression across multiple tissues, including leaves, inflorescences, roots, and stems. Furthermore, hormonal treatment experiments showed that *GhAGP50* expression is modulated by MeJA, ABA, IAA, and GA, suggesting its involvement in hormone-mediated developmental processes. This study presents a comprehensive genome-wide analysis of the AGP gene family in cotton, providing new insights into their structural diversity and functional significance. The identification and characterization of



Academic Editors: María Serrano and Xujun Zhu

Received: 19 December 2024

Revised: 12 March 2025

Accepted: 2 April 2025

Published: 27 April 2025

**Citation:** Wei, R.; Guo, Z.; Yang, Z.; Zhao, Y.; Yan, H.; Azhar, M.T.; Zhang, Y.; Li, G.; Pan, J.; Liu, A.; et al.

Genome-Wide Identification and Classification of Arabinogalactan Proteins Gene Family in *Gossypium* Species and *GhAGP50* Increases Numbers of Epidermal Hairs in *Arabidopsis*. *Int. J. Mol. Sci.* **2025**, *26*, 4159. <https://doi.org/10.3390/ijms26094159>

**Copyright:** © 2025 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/>).

*GhAGP50* highlight its potential role in epidermal hair formation and hormonal regulation, contributing to a deeper understanding of AGP functions in cotton development. These findings offer a valuable genetic resource for future research aimed at improving cotton growth and fiber quality through targeted genetic manipulation.

**Keywords:** cotton; arabinogalactan protein; gene regulation; functional characterization

## 1. Introduction

Cotton is one of the important economic crops in tropical and sub-tropical regions of the world which provides natural fiber. The improvement of fiber quality is of great importance for the textile industry. Cotton fiber development is important for fiber quality. Different kinds of stresses influence cotton growth and development, such as drought, cold, and heat stresses [1]. Cotton fiber is a unicellular trichome differentiated from the ovule epidermis [2]. The thickness of the fiber cell is highly related to fiber quality, while excessive thickening of the cell wall leads to poor fiber quality [3,4]. Hence, further studies of the genes involved in the cell wall are crucial and are useable for improvement of these traits. It is found that several genes and hormones play roles in cell elongation and the thickness of the cell wall in cotton [5]. Fiber elongation and development are regulated by indole-3-acetic acid (IAA), abscisic acid (ABA), and methyl jasmonate (MeJA) [6]. Through controlling the cellulose synthesis induced by IAA, *GhGASA10-1* can advance fiber elongation [6], whereas fiber length is decreased due to ABA treatment and increased under ABAI treatment [7]. The up-regulation of gibberellin (GA) 20-oxidase homologous genes suggested that GA participates in the development of fibers [8]. However, there are a large number of related genes to be identified for function annotation in cotton.

AGPs are a large group of highly glycosylated proteins that are classified in the family of plant cell wall hydroxyproline-rich glycoproteins [9]. Previous research revealed that AGP genes are involved in the biogenesis of plant cell walls, namely in the synthesis and deposition of cellulose. AGPs are divided into various subfamilies, such as classical AGPs, AG-peptides, lysine (Lys)-rich AGPs, and FLAs by domain constitutions of core proteins [10]. AGP structure affects the synthesis and deposition of pectin and secretion of callose [11]. *AtFLA4* mutant exhibited aberrant cell growth, thinner walls, and lessening in cellulose [12]. AGPs are involved in regulation of cell wall signaling as well as structural integrity [13].

It has been determined and shown that the AGP genes are present in additional plants besides *Arabidopsis*. For instance, five *BrFLAs* are expressed in inflorescence in Chinese cabbage [14]. Some findings revealed that AGPs play an important role in supporting cell wall remodeling in cork oak [14]. In addition, fiber development and reactions to NaCl and phytohormones regulate expression of *GhFLA* [15], and 24 tomato *FLA* genes are responsive to ABA and MeJA [16]. PtFLAs perform various functions in poplar for wood formation which is regulated by GA signaling [17]. Previous studies suggested that *GhAGP31* could participate in the root development of cotton [18].

Hormones are important for the growth of various organs in plants [19,20], and AGPs can respond to various hormone signals. Previous studies suggested that *AtAGP30* functions in response to ABA and *AtFLA4* is regulated by ABA signaling [19]. However, GUS activity of *GbEXPA2* in trichomes of *Arabidopsis* is up-regulated by GA and down-regulated by ABA [20]. Functioning of AGP was monitored in various tissues and organs of plants [21]. Further investigation demonstrated that disruption of AGP21 in *Arabidopsis* leads to the formation of aberrant root hairs [22]. It was found that young tomato has higher

concentrations of Lys-rich LeAGP1 protein than roots, young stems, leaves, and flowers [23]. Interestingly, *AtAGP18* is modestly expressed in seedlings and rosettes but highly expressed in roots, shoots, and stems [24]. *AtAGP17* and *AtAGP19* exhibit the highest expression levels in roots and flowers, with moderate abundance in stems and seedlings [25]. Likewise, over-expression of *AtAGP18* leads to observations of reduced rosette size, heightened branching, and decreased seed viability [26]. Primarily, AGPs are pivotal contributors to various aspects of plant growth and development [27]. *AtFLA11* plays roles in the formation of the secondary cell wall [28], and *FLAs* regulate the deposition of cellulose [29]. *GhAGP4* plays a significant role in the development of the cell wall, which is important for the organization and function of the cell wall [30]. Suppressing the expression of *GhAGP4* through RNA interference hinders both the initiation and elongation phases of fiber development in cotton plants [30]. *GhFLA1* affects the composition of AGP, which is useable for the initiation and elongation of fiber [30]. Thus, it is important to investigate the function of AGPs for improvement of fiber quality in cotton.

Quantitative trait locus (QTL) analysis is an important tool for improving cotton fiber quality, and exploiting the functioning of candidate genes placed in QTL is also meaningful [31]. The gene *Gh\_A13G0395* was harbored in a QTL cluster related to fiber strength and fiber length [32]. Combining the transcript data of TM-1 [33] and quantitative Real-Time Polymerase Chain Reaction (qRT-PCR) analysis revealed that *Gh\_A13G0395* were highly expressed at 20, 25, and 30 DPA [32]. In the current study, *Gh\_A13G0395* is described as AGP18 in *Arabidopsis* and renamed as *GhAGP50*.

The findings from this study will provide valuable theoretical support for exploring the function of AGPs. Later, candidate gene expression was analyzed among various tissues for further experiment. The experiment demonstrated the roles of *GhAGP50* in root, stem, and leaves in *Arabidopsis*, which could reveal its biological function. Its subcellular localization, over-expression in *Arabidopsis*, and response to hormones can provide information to reveal its function and the mechanisms involved in growth and development in crop plants, particularly cotton. The improvement of cotton fiber quality is a major goal in cotton breeding, yet the genetic and molecular mechanisms underlying fiber development are not fully understood. AGPs, as key regulators of cell wall biogenesis and signaling, are likely to play important roles in fiber development. However, the specific functions of AGPs in cotton, particularly in fiber initiation and elongation, remain largely unexplored. This study focuses on *GhAGP50*, a candidate gene identified through QTL analysis, to elucidate its role in fiber development and its response to hormonal signals. By investigating the expression, subcellular localization, and functional characterization of *GhAGP50* in *Arabidopsis*, this study aims to provide insights into the mechanisms by which AGPs regulate fiber development and to identify potential targets for improving cotton fiber quality.

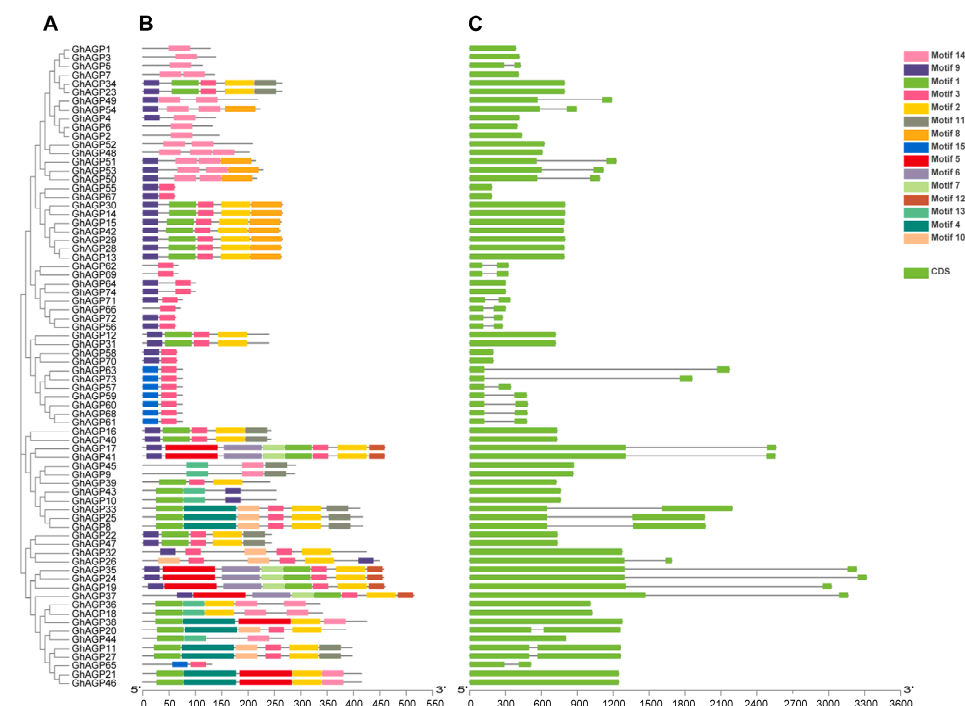
## 2. Results

### 2.1. Genome-Wide Identification and Sequence Analysis of AGP

Four subfamilies were detected, including AG peptides, FLAs, classical AGPs, and Lys-rich AGPs. Four subfamilies of proteins were discovered with the exception of the three subfamilies for *G. arboreum*. The genome of the four cotton species rendered 24, 44, 78, and 74 AGP proteins, accordingly (Tables S2 and S3). Likewise, *G. hirsutum* and *G. raimondii* had maximum molecular weights of 57.294 kDa. The results indicated the variation among features and in the cotton AGP gene family suggested a stronger necessity for these genes to operate.

## 2.2. Phylogenetic Tree and Genomic Architecture of AGP Gene Family

There were 220 AGP proteins in total, which included 74 AGPs of *G. hirsutum* (Figure 1), 24 AGPs of *G. arboreum* (Figure S1), 78 AGPs of *G. barbadense* (Figure S2), and 44 AGPs of *G. raimondii* (Figure S3). Group I had 21 AGPs, whereas Group IV had 27 AGPs. Moreover, 19, 15, and 20 *Gh*AGPs were grouped into groups I, II, and III, correspondingly. In addition, 11, 10, and 12 *Gr*AGPs were categorized into groups I, II, and IV, correspondingly. Furthermore, 15 unique motifs were discovered, and protein sequences of four cotton species with significant homology retained the same patterns (Figure S4). However, there were variations in motif compositions among different subfamilies. Consequently, unique protein motifs within the subfamily are crucial for each AGP protein.



**Figure 1.** Phylogenetic and Structural Analysis of AGP Genes in *G. hirsutum*. (A) Phylogenetic tree illustrating the evolutionary relationships among AGP genes in *G. hirsutum*, constructed using the Maximum Likelihood method with bootstrap values to indicate confidence levels. Different clades are highlighted in distinct colors to represent major evolutionary groups. (B) Conserved motif analysis displaying the distribution and arrangement of key motifs among AGP proteins. Motifs were identified using MEME, and their positions within each gene sequence are visualized, revealing similarities and differences in domain composition among AGP family members. (C) Gene structure analysis showing the exon-intron organization of AGP genes in *G. hirsutum*. Exons, introns, and untranslated regions (UTRs) are represented by different symbols and colors, providing insight into the structural diversity and evolutionary conservation of these genes.

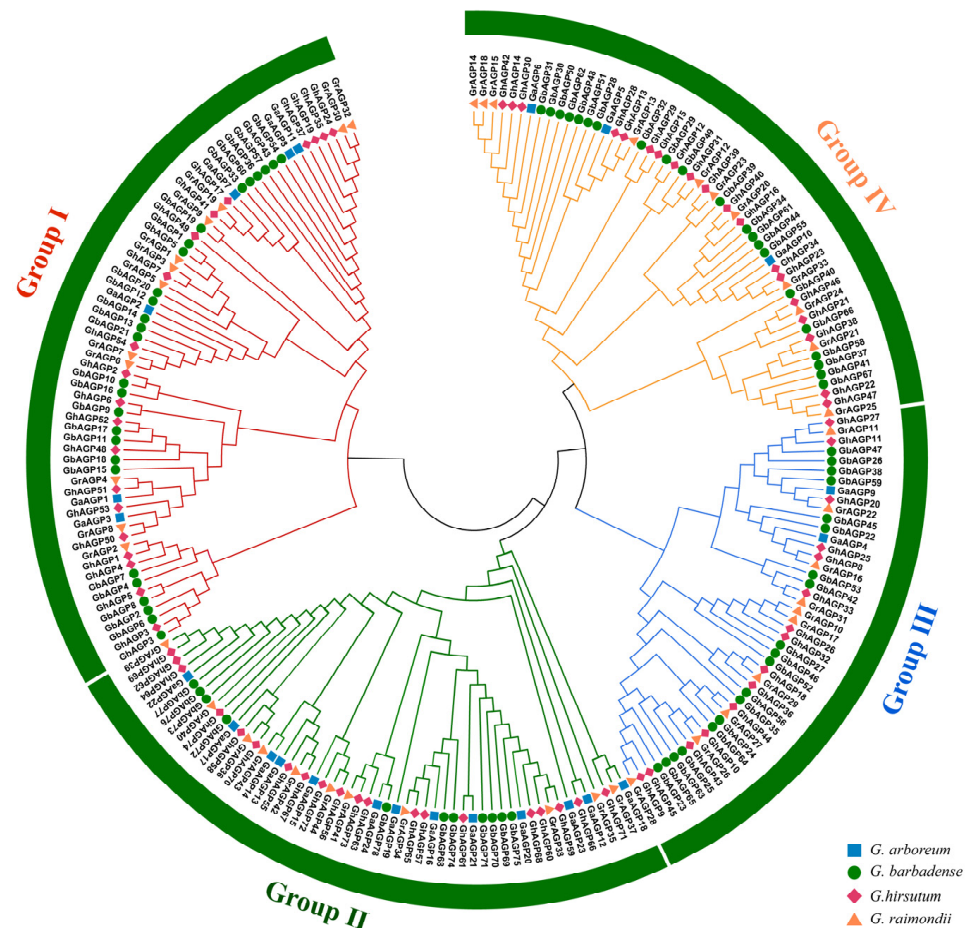
Surprisingly, genes that were closely related had stronger structural similarities despite differences in exon and intron length. Phylogeny and exon–intron structure showed a positive correlation. The results revealed the condition of exons and the gene length in the four cotton species in the AGP gene family.

## 2.3. Localization of AGP Genes on Chromosomes and Phylogenetic Analysis

The findings demonstrated that 220 cotton AGP genes were distributed broadly on chromosomes. There were significant differences in the number of AGP genes among each species. Simultaneously, AGP genes were found most on D07 in *G. hirsutum*, and on A07, A11, and D07 in *G. barbadense*. Among the four cotton species, the largest chromosome sizes



were as follows: 135.71 Mb, 70.71 Mb, 119.88 Mb, and 103.17 Mb (Figure S5). In order to gain a deeper understanding of the evolutionary relationships among AGP proteins, an unrooted phylogenetic tree was built. Four subfamilies were identified based on evolutionary analysis among the members of cotton AGP gene family. Compared to group III, which contained the fewest AGP gene family members, group I had the most (Figure 2). The phylogenetic tree demonstrates the evolutionary relationships of the genes. Therefore, subgroups recommended for interpretation exhibit higher conservation.



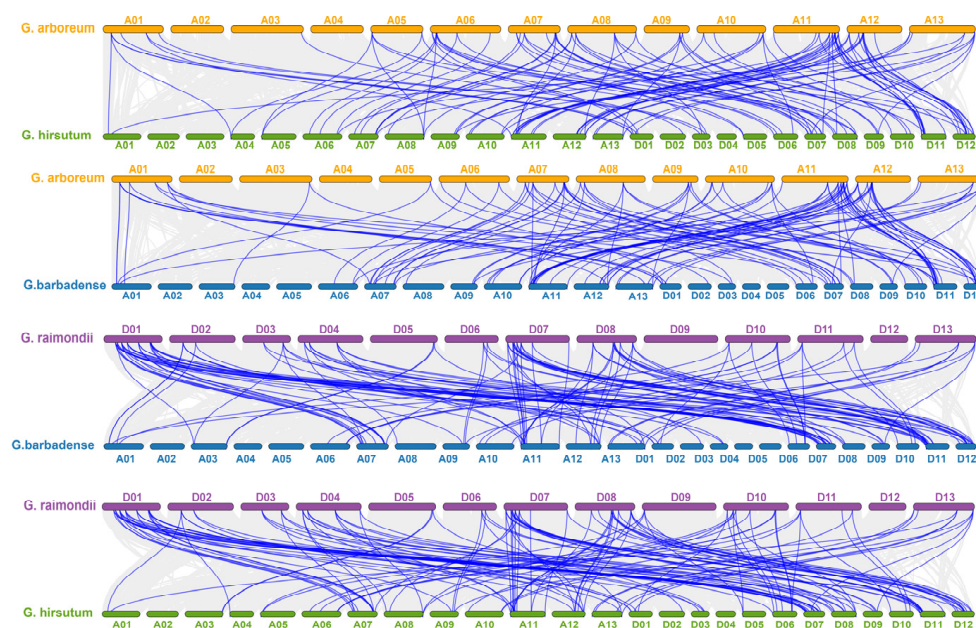
**Figure 2.** Phylogenetic analysis of AGP gene family among the four cotton species *G. hirsutum*, *G. arboreum*, *G. raimondii*, and *G. barbadense*.

#### 2.4. Collinearity Analysis of AGP Genes

In *G. hirsutum*, 51,575 genes with conserved collinearity and 3049 genes with tandem repeat were discovered (Table S4). In addition, the results also revealed that 15,904 genes with conserved collinearity were in *G. arboreum* (Table S5), 5202 collinear genes were in *G. barbadense* (Table S6), and 13,754 genes with conserved collinearity were discerned in *G. raimondii* (Table S7). By the collinearity analysis of AGP genes between *G. arboreum* and *G. hirsutum* (Table S8), *G. arboreum* and *G. barbadense* (Table S9), *G. barbadense* and *G. raimondii* (Table S10), *G. hirsutum* and *G. raimondii* (Table S11), the quantity and percentage of collinear genes were uncovered.

Accordingly, it indicated that *G. hirsutum* showed 12 dispersed, 25 segmental, and five tandem gene duplications (Table S12). *G. arboreum* shared the majority of genes with collinearity on chromosomes A05, A11, and A12 (Figure 3). The result of MCScanX analysis revealed the duplication in *G. raimondii* with tetraploid cotton. Correspondingly, 118 and

115 *AGP* homologous gene duplicates were identified between diploid and tetraploid cotton (Table S13).



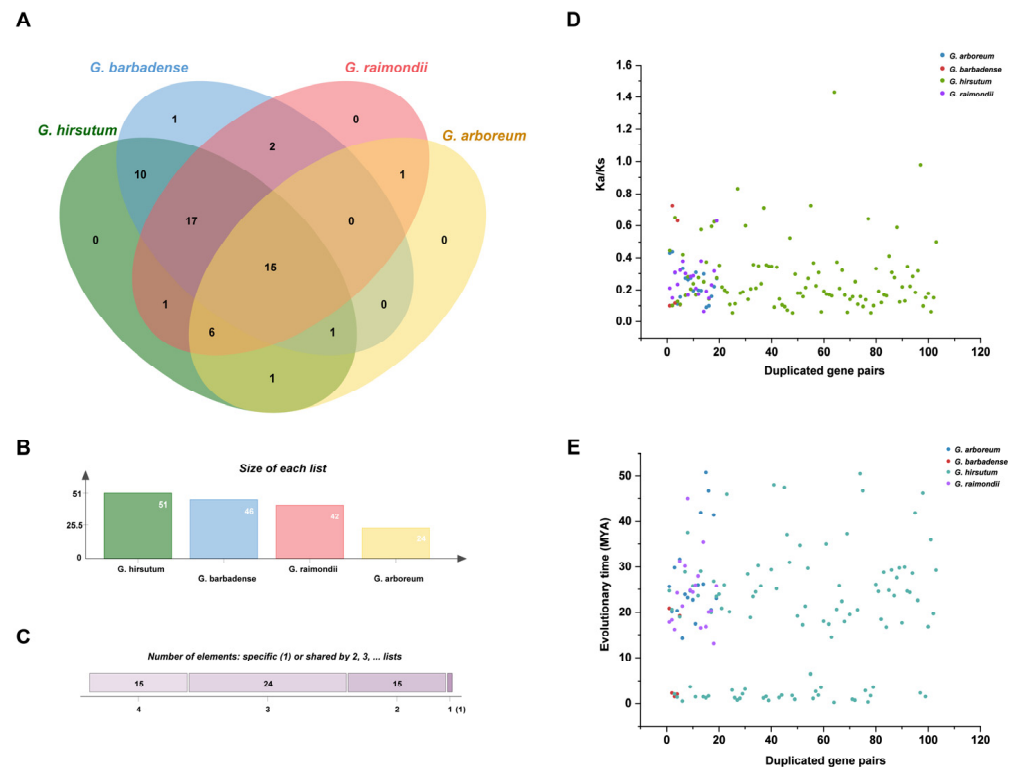
**Figure 3.** Duplication events analysis of *AGP* genes in cotton. Collinear blocks are symbolized by grey lines, while duplicated gene pairs are highlighted with blue lines.

### 2.5. Analysis of Orthologous Gene Clusters

Following relative assessment, the orthologous *AGP* gene clusters spanning the four cotton species were discovered. As a result, the cluster counts exhibited variations among the four cotton species, indicating differences in their genetic composition and diversity (Figure 4A–C). Moreover, 15 orthologous gene clusters were found. In addition, polyploidization was responsible for the emergence of unique gene clusters with particular orthologues, and orthologous gene clusters have been established. The results also revealed that in-paralogs were detected in three cotton species, the exception being *G. raimondii* (Table S14).

To explore the duplication patterns and evolutionary pressures of *AGP* paralogous gene pairs, synonymous and nonsynonymous rates were analyzed. The  $K_a/K_s$  ratio serves as a predictive measure of evolutionary history with a single value representing neutral selection. Conversely, a value less than one denotes purifying selection, while the value exceeding one signifies positive selection.

The  $K_a/K_s$  ratio in the current study was below 1 for both *AGP* orthologous and paralogous gene pairs. This suggests a scenario of purifying selection pressure accompanied by constrained functional divergence subsequent to segmental duplication. During the study, among the specific four cotton species, 146 duplicated gene pairs were identified (Figure 4D,E). Segmental duplication events involving *AGP* genes occurred in *G. arboreum* from approximately 14.15 to 50.78 Mya. Correspondingly, in *G. barbadense*, segmental duplication occurred from approximately 1.50–20.89 Mya. In addition, in *G. hirsutum* and *G. raimondii*, segmental duplication of *AGP* genes occurred between approximately 0.24–50.51 Mya and 13.02–44.96 Mya, respectively (Table S15).

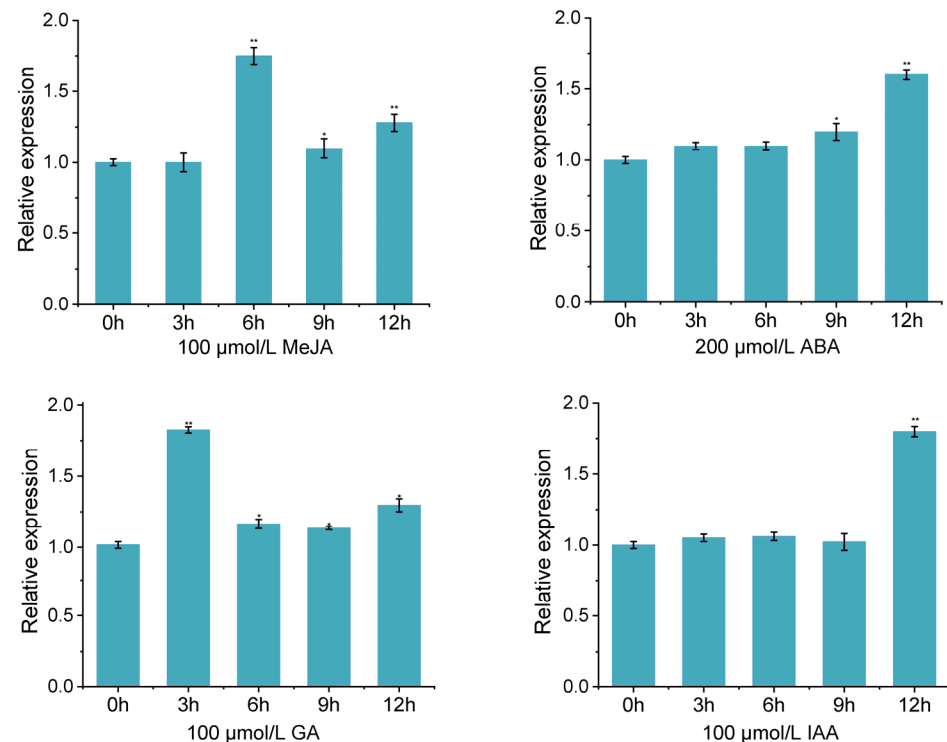


**Figure 4.** Clustering of orthologous genes and duplication events analysis of AGP genes in cotton. (A) The Venn diagram of syntenic AGP genes. (B) Number of orthologous genes. (C) Number of elements shared. (D) Synonymous and nonsynonymous ratios of AGP gene family. (E) The evolutionary time of 146 duplicated gene pairs. Ka, nonsynonymous substitution rate; Ks, synonymous substitution rate; Mya, million years ago.

## 2.6. cis Element Analysis of AGP Genes

The identified *cis*-regulatory elements exhibited a remarkably consistent pattern across four cotton species. ABRE, G-box, and GT1-motif were observed in comparatively greater quantities. Numerous *cis*-elements were associated with responsiveness to light and hormones (Table S16). Nine *cis*-regulatory elements were detected in *G. arboreum* (Figure S6a), *G. barbadense* (Figure S6b), and *G. raimondii* (Figure S6c) including one, two, three, two, and one elements related to ABA, GA, MeJA, auxin, and salicylic acid (SA), respectively. Besides SA, *G. hirsutum* also harbored *cis* elements associated with various other hormones (Figure S6d). Simultaneously, a variety of light-responsive *cis*-regulatory elements, i.e., GT1 motif, G-box, and AE-box, were discovered.

To further investigate tissue-preferential and developmental regulated expression of *GhAGP50*, the 2000-bp 5'-flanking fragment upstream *GhAGP50* translation start codon (ATG) was isolated by genome walking PCR. Various putative *cis* acting regulatory elements were detected in *GhAGP50*, such as ABA responsiveness, MeJA responsiveness, auxin responsiveness, and light responsiveness (Table S16). After ABA or IAA treatment, the expression of *GhAGP50* showed the highest at 12 h (Figure 5). In addition, *GhAGP50* showed an increasing expression at 9 h and 12 h, while its expression showed highest at 6 h and 3 h after GA and MeJA treatment. These results indicated *GhAGP50* plays a role in hormone responses.



**Figure 5.** The relative expression levels of *GhAGP50* with MeJA, ABA, GA, and IAA treatment. \*  $p < 0.05$ ; \*\*  $p < 0.01$ .

## 2.7. Subcellular Localization and Expression of Promoter::GUS Fusion of *GhAGP50*

The gene *Gh\_A13G0395* was pinpointed through a comprehensive analysis of genes localized within stable QTLs associated with fiber quality and identification of AGP genes. Based on the results of coding sequence comparison, *Gh\_A13G0395* exhibits a high degree of homology to *AT4G37450*, recognized as AGP18 in *Arabidopsis*. After genome-wide identification and classification, *Gh\_A13G0395* underwent renaming to *GhAGP50*. The results showed that *GhAGP50* was mainly localized in the cell nucleus and membrane. In addition, subcellular localization of *GhAGP50* was investigated in onion epidermal cells. The findings showed that fluorescence signals of *GhAGP50*-GFP were observed in both the cell nucleus and membrane (Figures 6 and S7).

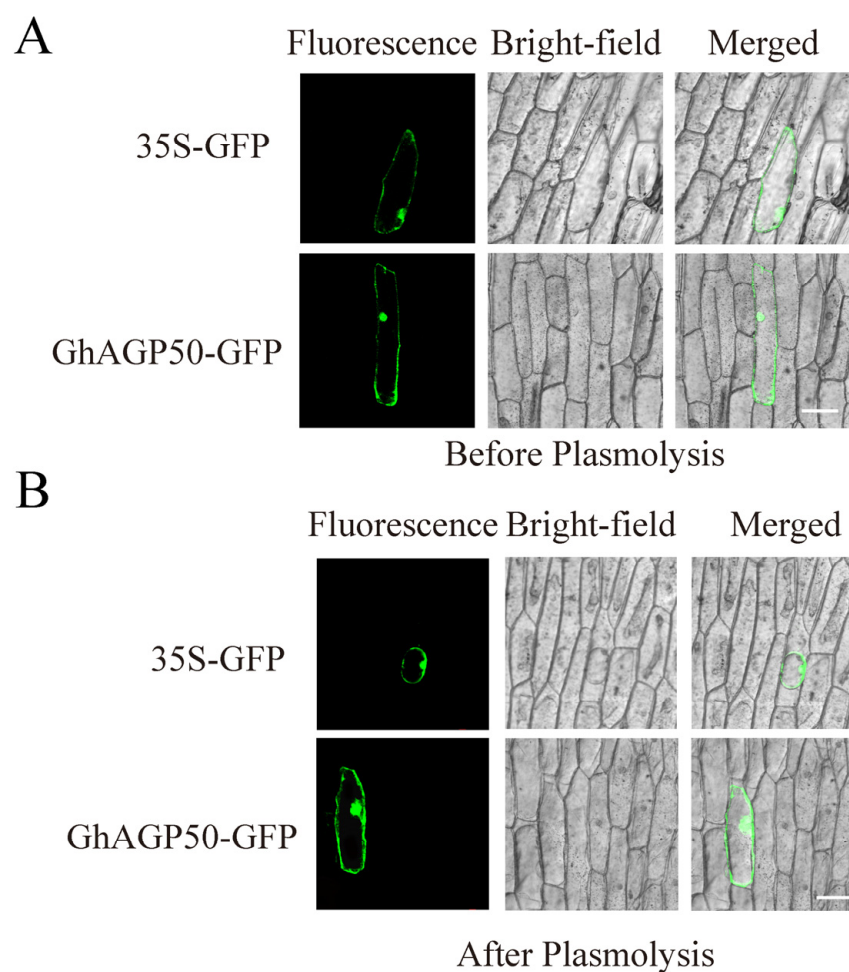
GUS activities were detected in multiple organs, including leaves, inflorescences, roots, and stems. Meanwhile, no signal was found in fruit pods and other tissues of transgenic *Arabidopsis* (Figure 7).

## 2.8. Overexpression of *GhAGP50* Improved the Growth of Root Hair and Epidermal Hairs in *Arabidopsis*

In addition, functioning of *GhAGP50* was explored by overexpression in *Arabidopsis*. The lengths of root hairs and numbers of epidermal hairs on leaves and stalks were measured in 7-day-old wild-type, *GhAGP50*-mutant, and *GhAGP50*-overexpressing plants. In comparison to wild-type plants, overexpressed plants have longer root hairs.

Moreover, the number of epidermal hairs on leaves and stalks was increased significantly among them. In comparison to the control group, the root length of OE-*GhAGP50* lines increased by 40.40%, while Mut-*GhAGP50* lines showed a decrease of 41.30%. In addition, the numbers of trichomes on leaves and stalks increased by 30.4% and 25% in OE-*GhAGP50* lines, respectively (Figure 8).



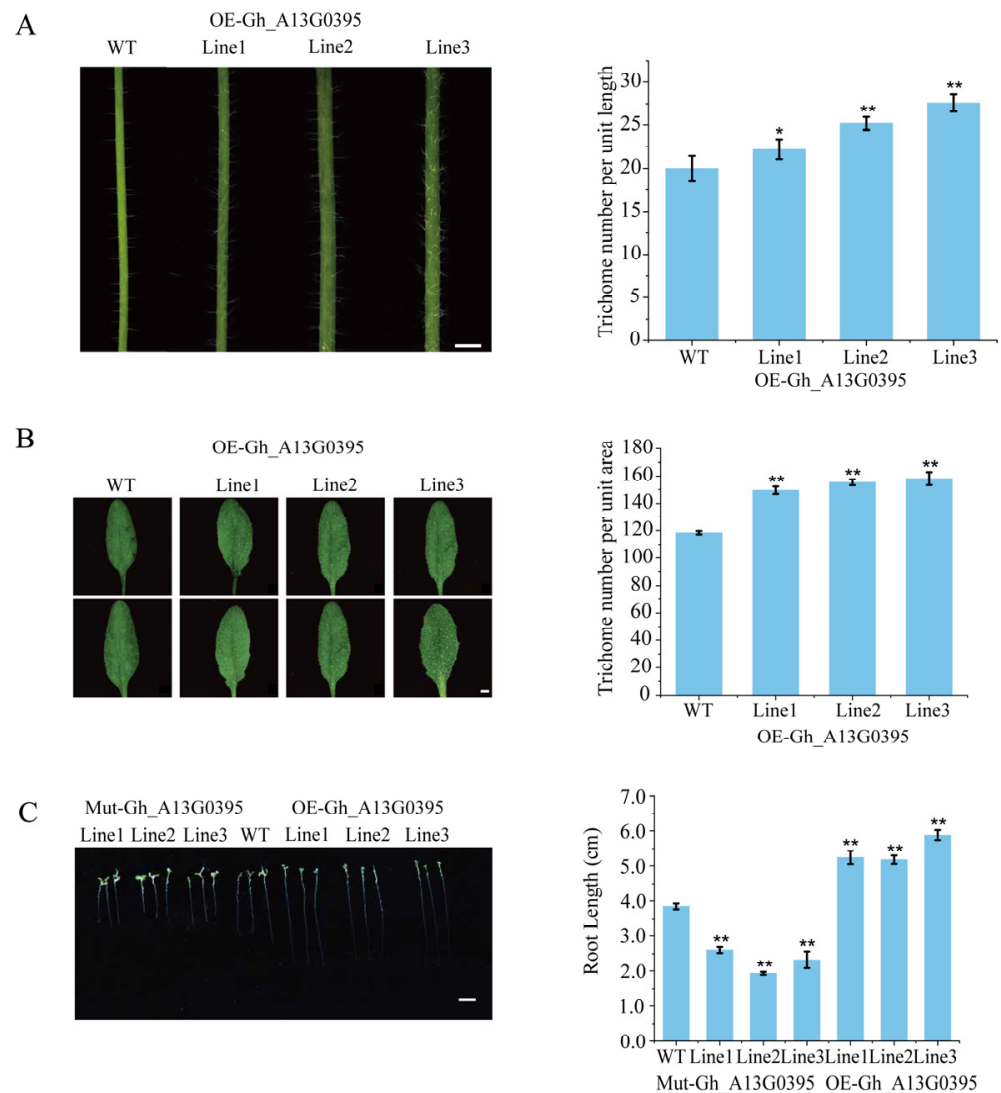


**Figure 6.** Subcellular localization of *GhAGP50*. (**A,B**) *GhAGP50*-GFP fusion protein was transiently observed in onion epidermal cells. Scale bar = 100  $\mu$ m.



**Figure 7.** Tissue-specific expression of *GhAGP50*-GUS transgenic plants. Scale bar = 1 mm.

Due to structural and genetic similarities between trichomes of *Arabidopsis* and cotton fibers, researchers have used these heterologous model species to track the activity of promoters particular to cotton fiber. *GhAGP50* overexpressed in *Arabidopsis* resulted in higher plant height than the wild type (Figure S8). Thus, *GhAGP50* can play roles in epidermal hairs of various organs in *Arabidopsis*.



**Figure 8.** Characterization of over-expressed *GhAGP50* in *Arabidopsis*, comparing trichomes on stalks (A) and trichomes on leaves (B) between wild-type and transgenic *Arabidopsis*. Scale bars = 2 cm. (C) Comparing root lengths among wild-type, mutant, and transgenic *Arabidopsis*. Scale bar = 1 cm. \*  $p < 0.05$ ; \*\*  $p < 0.01$ .

### 3. Discussion

#### 3.1. Evolution Analysis Revealed Purifying Selection

In the current study, 24, 78, 74, and 44 AGP genes were found in the four cotton species. The hybridization of *G. arboreum* and *G. raimondii* approximately 1–2 Mya may be the cause of the existence of most AGP genes in tetraploid cotton species [34,35]. The preservation of conserved domains throughout cotton evolution was demonstrated by comparative analysis of AGP genes. In this study, seven classical AGPs, 40 FLAs, seven Lys-rich AGPs, and 20 AG-peptides were identified in *G. hirsutum*. The results also revealed three classical AGPs, 24 FLAs, six Lys-rich AGPs, and 11 AG-peptides in *G. raimondii*, while *G. barbadense* showed nine classical AGPs, 46 FLAs, and 12 Lys-rich AGPs and 11 AG-peptides. During cotton fiber growth and development, many AGP carbohydrate epitopes were discovered from *G. hirsutum* L. [15,30]. The amplification of the AGP gene was primarily caused by segmental duplications. Compared to the other two cotton species, *G. arboreum* and *G. hirsutum* possessed a higher quantity of collinear blocks. Tandem and segmental duplication generally have little effect on the evolution of cotton AGP genes. On the other

hand, the regulation of genomic diversity was dependent on purifying selection that was applied to pairs of duplicated AGP genes in wild populations [36]. Moreover, purifying selection has been demonstrated in many plants, namely rice [37], soybean [38], wheat [39], cucumber [40], cocoa [41], and cotton [42]. Despite the fact that *G. hirsutum* (AD1) and *G. barbadense* (AD2) were the initial species to undergo transoceanic hybridization, this evolutionary process commenced about 1–2 million years ago [43].

### 3.2. The Evolutionary Events of AGPs Influence Functional Diversification

The AGP gene family has been expanded through gene duplication and differentiation events with differentiating functionally and adapting to the different physiological and ecological needs of plants. This gene family expansion and functional differentiation is an important mechanism for adaption to various environmental stresses [44]. AGP family genes contain functional domains, such as the fasciclin domain and the arabinogalactan domain, with different structures and functions. The diversity of these functional domains allows AGP proteins to perform multiple functions in processes such as cell wall construction, cell signaling, and cell–cell interactions [45]. The functional variety of AGP genes is further increased by variations in their expression patterns throughout different developmental stages and tissues. Plant growth and environmental adaptability are significantly influenced by certain AGP genes, which are strongly expressed during particular developmental stages or in response to particular environmental stimuli [44].

### 3.3. Regulation of AGPs by Various Hormones During Plant Growth

AGPs are not only significant cell wall components but also important nutritional or signaling molecules [46]. Although the precise processes are still unclear, AGPs have also been associated with signal transduction and recognition [47]. ABA is an important hormone, while several AGPs are responsive to ABA and salt stresses in rice [48]. Furthermore, AGPs are important in  $\alpha$ -amylase synthesis that GA induced in wheat cells [48]. In *Brassica rapa*, ABA, GA, and MeJA treatments resulted in up-regulation of 17 *BrAGPs* and down-regulation of six *BrAGPs* [49]. *AmAGP* is expressed high with IAA treatment in *Amorphophallus muelleri* [50]. IAA and MeJA promote fiber elongation, while ABA inhibits fiber elongation. Likewise, GA could promote cotton fiber elongation and ovule expansion. In this study, expression of *GhAGP50* was regulated by various concentrations of GA, IAA, ABA, and MeJA. The interaction of GA with auxins orchestrates the differentiation and growth of epidermal hairs, ensuring proper development and functionality [21]. ABA inhibits root hair formation in epidermal cells [51]. MeJA plays a significant role in the initiation and elongation of root hairs [11]. *GhAGP50* might be influenced by these hormones to regulate epidermal hair formation and development. By participating in cell expansion, differentiation, and interaction with other cellular components, AGPs play a crucial role in regulating the formation of root and shoot epidermal hairs. AGPs that are recognized by specific antibodies are essential for these developmental processes [52].

### 3.4. AGPs Play Roles in the Growth and Development of Plants

AGP is one of the most intricate kinds of macromolecules [53]. In addition, AGPs are complex hydroxyproline-rich glycoproteins, which affect the structure and composition of the cell wall [10]. AGPs are involved in diverse processes of plant development, including cellular differentiation, signaling, and interaction between microbes and plants [54]. In addition, *AtAGP19* promoter-controlled GUS activity was identified in the veins of leaves, roots, stems, styles, and siliques [55]. In this study, *GhAGP50* was found to be expressed in leaves, inflorescences, roots, and stems, suggesting its potential function in controlling growth and development of plants. AGPs play crucial roles in root growth and regeneration, root hair formation, and seedling growth [56]. In this study, the numbers of epidermal

hairs on the stems and leaves of transgenic *Arabidopsis* were greater than those of wild-type *Arabidopsis*, demonstrating that *GhAGP50* promotes the formation of epidermal hairs. Thus, it is important for further exploration to investigate the function of *AGP*.

### 3.5. AGPs Regulate Cotton Fiber Development

In *Arabidopsis*, traditional AGP18 actively regulates survival and megaspore selection [57]. AGPs are located on plasma membranes and AGPs play important roles in various biological processes [14]. Fiber elongation to transitional cell-wall remodeling is important for the development of cotton fiber [58]. AGPs are highly correlated with cell wall components, which play roles in the fiber initial differentiation and elongation developmental stages [30]. Previous researchers reported that AGPs are involved in the elongation of fiber, which plays a role in improving cotton fiber quality [59]. The gene *GhGalT1* controls glycosylation of AGP to regulate cotton fiber elongation [60]. The AG glycan on AGP plays a role in controlling fiber development, which shows the importance of discovering genes related to specific glycans to improve fiber quality [60]. In this study, several identified AGP genes were significant for further study for fiber development in cotton.

## 4. Materials and Methods

### 4.1. Identification and Characterization of AGP Genes

The genomic sequences of *G. arboreum* (CRI), *G. raimondii* (JGI), *G. barbadense* (ZJU) and *G. hirsutum* (NAU) were from the CottonFGD database and CottonGen database [61]. HMMER software v3.4 with an E value threshold of  $1 \times 10^{-10}$  utilized Hidden Markov Models (HMMs) for the detection of potential AGP sequences [62]. The HMM files (PF02469 and PF06376) were used to identify FLAs and AG-peptides. Protein domains (IPR044959 and IPR044981) were used to detect classical and Lys-rich AGPs. In addition, the Pfam (<https://pfam.xfam.org/>) and InterPro (<https://www.ebi.ac.uk/interpro/>) databases were utilized to verify the accuracy of the identified candidate protein sequences. The biophysical properties of AGP proteins underwent examination with an online tool, i.e., ExPASy [63]. Bologna Unified Subcellular Component Annotator (BUSCA) was used to predict subcellular localization of proteins [64].

### 4.2. Phylogenetic Investigation and Gene Structure of AGP Genes

The AGP sequences from four cotton species were under alignment using Clustal Omega (version 1.2.4) with standard parameters [65]. The generated multiple sequence alignments (MSAs) were constructed in the MEGA 7 software and then visualized for analysis [66]. The identification of conserved motifs was carried out using MEME [67], while the MAST algorithm was used to examine the protein database for 15 motifs [68]. Gene structure was determined by evaluating both coding sequences and genomic sequences. TBtools was employed to visualize the gene structure of AGPs [69]. To construct a phylogenetic tree, AGP proteins were analyzed in MEGA 7 [66].

### 4.3. Chromosomal Locations and Gene Collinearity Analysis of AGP Genes

Chromosome locations of AGP genes were drawn with TBtools software (Version 2.027) [69]. AGP genes were given systematic nomenclature depending on the location of their chromosomes. The assessment of gene duplication occurrences involved the utilization of the MCScanX toolkit, which performed multiple collinearity scans for analysis [70]. The visualization of duplicated regions involved illustrating the syntenic relationships among AGP genes using Circos. Tandem duplicates are gene pairs that are duplicates of each other, originating from same genome, and located on same chromosome [71]. TBtools facilitated the computation of selection pressure by evaluating the ratio of nonsynonymous



(Ka) to synonymous (Ks) substitutions. The duplication time (T) in evolutionary analysis was computed using the formula  $T = Ks / 2\lambda \times 10^{-6}$  (Mya), with a  $\lambda$  value of  $1.5 \times 10^{-8}$  [72].

#### 4.4. Identifying of cis Regulatory Element

In the cotton AGP promoter region, the 1 kb promoter was to pinpoint the important cis-regulatory components. The sequence was supplied from the CottonFGD database. PlantCARE (accessed on 1 April 2025) was used to determine cis-regulatory elements [73].

#### 4.5. Subcellular Localization

The XbaI and SmaI restriction sites on the pCAMBIA2300 vector were analyzed to build the translation fusion structure. As part of the experimental procedure, the recombinant plasmid experienced transformation into *Agrobacterium tumefaciens* GV3101. *Nicotiana benthamiana* (tobacco) leaves were concurrently transformed with the pCAMBIA2300 vector to serve as the control group under identical conditions. The observation of the transformed gene expression was conducted using a CCD optical microscope (Germany's Leica Microsystems, Wetzlar, Germany).

To investigate subcellular localization of GhAGP50, the PDS-1000/He system (Bio-Rad, Hercules, CA, USA) facilitated the transformation of onion epidermal cells with both the empty vector and recombinant plasmids containing GFP fusions. Following a 24–36 h incubation period on MS media, the fluorescence of GFP was detected using confocal microscopy. The 30% sucrose solution was applied to the cells to perform the plasmolysis experiment.

#### 4.6. GhAGP50 Promoter Expression in Transgenic Arabidopsis Plants

A GUS reporter cassette was utilized to assess the promoter activity. It contained the GhAGP50 promoter inserted into the SalI/BamHI sites of the pBI121 vector. The GhAGP50::GUS construct was introduced into *A. tumefaciens* GV3101 using normal transformation techniques. The selection of positive plants was carried out using MS plates supplemented with kanamycin at the concentration of 50 g/mL. Using the GUS Stain Kit (Solarbio, Beijing, China, G3061), histochemical tests were performed to determine the GUS activity in transgenic *Arabidopsis*. A Leica stereomicroscope (Leica MZ16f) was utilized for examining and photographing the stained tissue.

#### 4.7. Transformation in Arabidopsis

Wild-type *Arabidopsis* type Columbia Zero and *Arabidopsis* mutant (SALK\_117268) were acquired from Nottingham *Arabidopsis* Stock Centre (NASC). *Arabidopsis* was transformed using the floral dip approach after the pBI121 vector was introduced into *A. tumefaciens* GV3101. The GhAGP50 ORF was integrated into the pBI121 vector under the direction of cauliflower mosaic virus (CaMV) 35S promoter. The experiments utilized the T3 generation for further investigations. To determine the positivity of the offspring seedlings, PCR amplification was used to detect the target gene. *Arabidopsis* seeds of the wild type, genetically modified to include pBI121–GhAGP50, underwent sterilization before being placed onto MS medium for growth. The phenotypes of the planted *Arabidopsis* were evaluated and analyzed when reaching maturity.

#### 4.8. Exogenous Hormone Treatment

Upon reaching the three-leaf stage, a portion of the cotton plants were designated as the control group, while the remaining plants were treated with 100  $\mu$ mol/L MeJA, 200  $\mu$ mol/L ABA, 100  $\mu$ mol/L GA, and 100  $\mu$ mol/L IAA. Each treatment was replicated three times biologically. Leaf tissue samples were obtained at 3 h intervals following treatment, spanning 3 h, 6 h, 9 h, and 12 h. RNA extraction was then performed on these

samples to assess gene expression levels. The cotton material treated was 0–153, with high-quality fiber.

#### 4.9. RNA Extraction and qRT-PCR Analysis

The material used for RNA extraction was 0–153. Leaf tissue RNA extraction was conducted using an RNA extraction kit, FastPure Universal Plant Total RNA Isolation Kit (Vazyme Biotech Co., Ltd., Nanjing, China). To validate the expression pattern, qRT-PCR was conducted. HiScript<sup>®</sup> II Q RT SuperMix for qPCR (+gDNA wiper) kit (Vazyme, R223-01) was used to reverse transcribe 1 µg RNA of each sample into cDNA. Specific primers were designed using Primer Premier 5 and the housekeeping gene *GhHis3* was a reference gene in this experiment (Table S1) [74]. The relative expression levels of candidate genes were computed by the  $2^{-\Delta\Delta CT}$  method and the bar graph was plotted using OriginLab [75]. The statistical significance was analyzed by using two-tailed tests.

## 5. Conclusions

In order to highlight the potential functional diversity, whole-genome identification and evolutionary analysis were performed. In addition, phylogenetic relationships and expression patterns were analyzed for the cotton *AGP* gene family. These results could contribute to understanding the functional characteristics of cotton *AGPs*. *GhAGP50* was harbored in a fiber length and fiber strength QTL, and identified as AGP18, which is an important protein composing the plant cell wall. In this study, overexpression of *GhAGP50* promoted epidermal hair of leaves and stems, which means it contributes to regulating cotton fiber development. This study contributes to the understanding of the roles of *AGPs* in improving cotton fiber quality and the processes of plant growth and development.

**Supplementary Materials:** The supporting information can be downloaded at: <https://www.mdpi.com/article/10.3390/ijms26094159/s1>.

**Author Contributions:** Conceptualization, H.S. and Y.Y.; writing—original draft preparation, R.W.; methodology, Y.Z. (Yanpeng Zhao), Z.G. and J.G.; data curation, Y.Z. (Yamin Zhang), W.G. and Q.G.; writing—review and editing, Z.Y., G.L. and H.Y.; supervision, Z.Y. and G.L.; project administration, A.L., J.P. and M.T.A.; funding acquisition, H.S. and Y.Y. All authors have read and agreed to the published version of the manuscript.

**Funding:** This work was funded by grants from Key R&D Program of Henan Province (No. 241111114200), Kashgar Regional Science and Technology Plan (KS2024014), “Tianchi Talent” in Xinjiang Uygur Autonomous Region in 2023 (Yuan Youlu), Natural Science Foundation of Henan (242300421592), Independent Fund of State Key Laboratory of Cotton Bio-breeding and Integrated Utilization (CBIU2023ZZ05, CBIUZ2024002), Biological Breeding-Major Projects (2023ZD04062), National Natural Science Foundation of China (No. 32101799), the Central Public-interest Scientific Institution Basal Research Fund (1610162023002, 1610162023053), China Agriculture Research System—Cotton (CARS-15-02) and Agricultural Science and Technology Innovation Program of Chinese Academy of Agricultural Sciences (CAAS-ASTIP-2016-ICR).

**Institutional Review Board Statement:** Not applicable.

**Informed Consent Statement:** Not applicable.

**Data Availability Statement:** All data generated or analyzed during this study are included in this manuscript and its Supplementary Information Files.

**Acknowledgments:** This work was supported by National Supercomputing Center in Zhengzhou.

**Conflicts of Interest:** The authors declare no conflicts of interest.

## References

1. Sarwar, M.; Saleem, M.F.; Ali, B.; Saleem, M.H.; Rizwan, M.; Usman, K.; Keblawy, A.E.; Ali, A.; Afzal, M.; Sheteiwy, M.S. Application of Potassium, Zinc and Boron as Potential Plant Growth Modulators in *Gossypium hirsutum* L. under Heat Stress. *Turk. J. Agric. For.* **2022**, *46*, 567–584. [\[CrossRef\]](#)
2. Qin, Y.-M.; Zhu, Y.-X. How Cotton Fibers Elongate: A Tale of Linear Cell-Growth Mode. *Curr. Opin. Plant Biol.* **2011**, *14*, 106–111. [\[CrossRef\]](#) [\[PubMed\]](#)
3. Haigler, C.H.; Betancur, L.; Stiff, M.R.; Tuttle, J.R. Cotton Fiber: A Powerful Single-Cell Model for Cell Wall and Cellulose Research. *Front. Plant Sci.* **2012**, *3*, 104. [\[CrossRef\]](#) [\[PubMed\]](#)
4. Jareczek, J.J.; Grover, C.E.; Wendel, J.F. Cotton Fiber as a Model for Understanding Shifts in Cell Development under Domestication. *Front. Plant Sci.* **2023**, *14*, 1146802. [\[CrossRef\]](#)
5. Cao, J.-F.; Zhao, B.; Huang, C.-C.; Chen, Z.-W.; Zhao, T.; Liu, H.-R.; Hu, G.-J.; Shangguan, X.-X.; Shan, C.-M.; Wang, L.-J.; et al. The miR319-Targeted GhTCP4 Promotes the Transition from Cell Elongation to Wall Thickening in Cotton Fiber. *Mol. Plant* **2020**, *13*, 1063–1077. [\[CrossRef\]](#)
6. Chen, Y.; Zhang, J.-B.; Wei, N.; Liu, Z.-H.; Li, Y.; Zheng, Y.; Li, X.-B. A Type-2C Protein Phosphatase (GhDRP1) Participates in Cotton (*Gossypium hirsutum*) Response to Drought Stress. *Plant Mol. Biol.* **2021**, *107*, 499–517. [\[CrossRef\]](#)
7. Yang, Y.; Lai, W.; Long, L.; Gao, W.; Xu, F.; Li, P.; Zhou, S.; Ding, Y.; Hu, H. Comparative Proteomic Analysis Identified Proteins and the Phenylpropanoid Biosynthesis Pathway Involved in the Response to ABA Treatment in Cotton Fiber Development. *Sci. Rep.* **2023**, *13*, 1488. [\[CrossRef\]](#)
8. Xiao, Y.-H.; Li, D.-M.; Yin, M.-H.; Li, X.-B.; Zhang, M.; Wang, Y.-J.; Dong, J.; Zhao, J.; Luo, M.; Luo, X.-Y. Gibberellin 20-Oxidase Promotes Initiation and Elongation of Cotton Fibers by Regulating Gibberellin Synthesis. *J. Plant Physiol.* **2010**, *167*, 829–837. [\[CrossRef\]](#)
9. Paunović, D.M.; Ćuković, K.B.; Bogdanović, M.D.; Todorović, S.I.; Trifunović-Momčilov, M.M.; Subotić, A.R.; Simonović, A.D.; Dragičević, M.B. The Arabinogalactan Protein Family of Centaurea Erythraea Rafn. *Plants* **2021**, *10*, 1870. [\[CrossRef\]](#)
10. Ma, Y.; Yan, C.; Li, H.; Wu, W.; Liu, Y.; Wang, Y.; Chen, Q.; Ma, H. Bioinformatics Prediction and Evolution Analysis of Arabinogalactan Proteins in the Plant Kingdom. *Front. Plant Sci.* **2017**, *8*, 66. [\[CrossRef\]](#)
11. Lin, S.; Miao, Y.; Huang, H.; Zhang, Y.; Huang, L.; Cao, J. Arabinogalactan Proteins: Focus on the Role in Cellulose Synthesis and Deposition during Plant Cell Wall Biogenesis. *Int. J. Mol. Sci.* **2022**, *23*, 6578. [\[CrossRef\]](#) [\[PubMed\]](#)
12. Shi, H.; Kim, Y.; Guo, Y.; Stevenson, B.; Zhu, J.-K. The Arabidopsis SOS5 Locus Encodes a Putative Cell Surface Adhesion Protein and Is Required for Normal Cell Expansion. *Plant Cell* **2003**, *15*, 19–32. [\[CrossRef\]](#) [\[PubMed\]](#)
13. Ma, Y.; Johnson, K. Arabinogalactan Proteins—Multifunctional Glycoproteins of the Plant Cell Wall. *Cell Surf.* **2023**, *9*, 100102. [\[CrossRef\]](#) [\[PubMed\]](#)
14. Pérez-Pérez, Y.; Carneros, E.; Berenguer, E.; Solís, M.-T.; Bárány, I.; Pintos, B.; Gómez-Garay, A.; Risueño, M.C.; Testillano, P.S. Pectin De-Methylesterification and AGP Increase Promote Cell Wall Remodeling and Are Required During Somatic Embryogenesis of Quercus Suber. *Front. Plant Sci.* **2019**, *9*, 1915. [\[CrossRef\]](#)
15. Huang, G.-Q.; Xu, W.-L.; Gong, S.-Y.; Li, B.; Wang, X.-L.; Xu, D.; Li, X.-B. Characterization of 19 Novel Cotton FLA Genes and Their Expression Profiling in Fiber Development and in Response to Phytohormones and Salt Stress. *Physiol. Plant.* **2008**, *134*, 348–359. [\[CrossRef\]](#)
16. Yao, K.; Yao, Y.; Ding, Z.; Pan, X.; Zheng, Y.; Huang, Y.; Zhang, Z.; Li, A.; Wang, C.; Li, C.; et al. Characterization of the FLA Gene Family in Tomato (*Solanum lycopersicum* L.) and the Expression Analysis of SIFLAs in Response to Hormone and Abiotic Stresses. *Int. J. Mol. Sci.* **2023**, *24*, 16063. [\[CrossRef\]](#)
17. Wang, H.; Jin, Y.; Wang, C.; Li, B.; Jiang, C.; Sun, Z.; Zhang, Z.; Kong, F.; Zhang, H. Fasciclin-like Arabinogalactan Proteins, PtFLAs, Play Important Roles in GA-Mediated Tension Wood Formation in Populus. *Sci. Rep.* **2017**, *7*, 6182. [\[CrossRef\]](#)
18. Gong, S.-Y.; Huang, G.-Q.; Sun, X.; Li, P.; Zhao, L.-L.; Zhang, D.-J.; Li, X.-B. GhAGP31, a Cotton Non-classical Arabinogalactan Protein, Is Involved in Response to Cold Stress during Early Seedling Development. *Plant Biol.* **2012**, *14*, 447–457. [\[CrossRef\]](#)
19. Xue, H.; Seifert, G.J. FASCICLIN LIKE ARABINO GALACTAN PROTEIN 4 and RESPIRATORY BURST OXIDASE HOMOLOG D and F Independently Modulate Abscissic Acid Signaling. *Plant Signal. Behav.* **2015**, *10*, e989064. [\[CrossRef\]](#)
20. Li, F.; Fan, G.; Lu, C.; Xiao, G.; Zou, C.; Kohel, R.J.; Ma, Z.; Shang, H.; Ma, X.; Wu, J.; et al. Genome Sequence of Cultivated Upland Cotton (*Gossypium hirsutum* TM-1) Provides Insights into Genome Evolution. *Nat. Biotechnol.* **2015**, *33*, 524–530. [\[CrossRef\]](#)
21. Li, W.-J.; Wu, N.; Chen, C.; Zhao, Y.-P.; Hou, Y.-X. Identification and Expression Analysis of Arabinogalactan Protein Genes in Cotton Reveal the Function of GhAGP15 in Verticillium Dahliae Resistance. *Gene* **2022**, *822*, 146336. [\[CrossRef\]](#) [\[PubMed\]](#)
22. Borassi, C.; Gloazzo Dorosz, J.; Ricardi, M.M.; Carignani Sardoy, M.; Pol Fachin, L.; Marzol, E.; Mangano, S.; Rodríguez García, D.R.; Martínez Pacheco, J.; Rondón Guerrero, Y.D.C.; et al. A Cell Surface Arabinogalactan-peptide Influences Root Hair Cell Fate. *New Phytol.* **2020**, *227*, 732–743. [\[CrossRef\]](#) [\[PubMed\]](#)
23. Gao, M.; Showalter, A.M. Immunolocalization of LeAGP-1, a Modular Arabinogalactan-Protein, Reveals Its Developmentally Regulated Expression in Tomato. *Planta* **2000**, *210*, 865–874. [\[CrossRef\]](#) [\[PubMed\]](#)

24. Yang, J.; Showalter, A.M. Expression and Localization of AtAGP18, a Lysine-Rich Arabinogalactan-Protein in Arabidopsis. *Planta* **2007**, *226*, 169–179. [[CrossRef](#)]
25. Yang, J.; Zhang, Y.; Liang, Y.; Showalter, A.M. Expression Analyses of AtAGP17 and AtAGP19, Two Lysine-rich Arabinogalactan Proteins, in Arabidopsis. *Plant Biol.* **2011**, *13*, 431–438. [[CrossRef](#)]
26. Zhang, Y.; Yang, J.; Showalter, A.M. AtAGP18 Is Localized at the Plasma Membrane and Functions in Plant Growth and Development. *Planta* **2011**, *233*, 675–683. [[CrossRef](#)]
27. Kaur, D.; Held, M.A.; Smith, M.R.; Showalter, A.M. Functional Characterization of Hydroxyproline-O-Galactosyltransferases for Arabidopsis Arabinogalactan-Protein Synthesis. *BMC Plant Biol.* **2021**, *21*, 590. [[CrossRef](#)]
28. Ito, S.; Suzuki, Y.; Miyamoto, K.; Ueda, J.; Yamaguchi, I. AtFLA11, a Fasciclin-Like Arabinogalactan-Protein, Specifically Localized in Sclerenchyma Cells. *Biosci. Biotechnol. Biochem.* **2005**, *69*, 1963–1969. [[CrossRef](#)]
29. MacMillan, C.P.; Mansfield, S.D.; Stachurski, Z.H.; Evans, R.; Southerton, S.G. Fasciclin-like Arabinogalactan Proteins: Specialization for Stem Biomechanics and Cell Wall Architecture in Arabidopsis and Eucalyptus: FLAs Specialized for Stem Biomechanics and Cell Walls. *Plant J.* **2010**, *62*, 689–703. [[CrossRef](#)]
30. Huang, G.-Q.; Gong, S.-Y.; Xu, W.-L.; Li, W.; Li, P.; Zhang, C.-J.; Li, D.-D.; Zheng, Y.; Li, F.-G.; Li, X.-B. A Fasciclin-Like Arabinogalactan Protein, GhFLA1, Is Involved in Fiber Initiation and Elongation of Cotton. *Plant Physiol.* **2013**, *161*, 1278–1290. [[CrossRef](#)]
31. Li, Y.; Si, Z.; Wang, G.; Shi, Z.; Chen, J.; Qi, G.; Jin, S.; Han, Z.; Gao, W.; Tian, Y. Genomic Insights into the Genetic Basis of Cotton Breeding in China. *Mol. Plant* **2023**, *16*, 662–677. [[CrossRef](#)] [[PubMed](#)]
32. Zhang, Z.; Li, J.; Jamshed, M.; Shi, Y.; Liu, A.; Gong, J.; Wang, S.; Zhang, J.; Sun, F.; Jia, F.; et al. Genome-wide Quantitative Trait Loci Reveal the Genetic Basis of Cotton Fibre Quality and Yield-related Traits in a *Gossypium hirsutum* Recombinant Inbred Line Population. *Plant Biotechnol. J.* **2020**, *18*, 239–253. [[CrossRef](#)] [[PubMed](#)]
33. Zhang, T.; Hu, Y.; Jiang, W.; Fang, L.; Guan, X.; Chen, J.; Zhang, J.; Saski, C.A.; Scheffler, B.E.; Stelly, D.M. Sequencing of Allotetraploid Cotton (*Gossypium hirsutum* L. Acc. TM-1) Provides a Resource for Fiber Improvement. *Nat. Biotechnol.* **2015**, *33*, 531–537. [[CrossRef](#)]
34. Hu, Y.; Chen, J.; Fang, L.; Zhang, Z.; Ma, W.; Niu, Y.; Ju, L.; Deng, J.; Zhao, T.; Lian, J. *Gossypium barbadense* and *Gossypium hirsutum* Genomes Provide Insights into the Origin and Evolution of Allotetraploid Cotton. *Nat. Genet.* **2019**, *51*, 739–748. [[CrossRef](#)]
35. Huang, G.; Wu, Z.; Percy, R.G.; Bai, M.; Li, Y.; Frelichowski, J.E.; Hu, J.; Wang, K.; Yu, J.Z.; Zhu, Y. Genome Sequence of *Gossypium herbaceum* and Genome Updates of *Gossypium arboreum* and *Gossypium hirsutum* Provide Insights into Cotton A-Genome Evolution. *Nat. Genet.* **2020**, *52*, 516–524. [[CrossRef](#)]
36. Cvijović, I.; Good, B.H.; Desai, M.M. The Effect of Strong Purifying Selection on Genetic Diversity. *Genetics* **2018**, *209*, 1235–1278. [[CrossRef](#)]
37. Ke, S.; Jiang, Y.; Zhou, M.; Li, Y. Genome-Wide Identification, Evolution, and Expression Analysis of the WD40 Subfamily in *Oryza* Genus. *Int. J. Mol. Sci.* **2023**, *24*, 15776. [[CrossRef](#)]
38. Jia, Q.; Song, J.; Zheng, C.; Fu, J.; Qin, B.; Zhang, Y.; Liu, Z.; Jia, K.; Liang, K.; Lin, W. Genome-Wide Analysis of Cation/Proton Antiporter Family in Soybean (*Glycine max*) and Functional Analysis of GmCHX20a on Salt Response. *Int. J. Mol. Sci.* **2023**, *24*, 16560. [[CrossRef](#)]
39. Zhang, S.; Yan, C.; Lu, T.; Fan, Y.; Ren, Y.; Zhao, J.; Shan, X.; Guan, Y.; Song, P.; Li, D. New Insights into Molecular Features of the Genome-Wide AOX Family and Their Responses to Various Stresses in Common Wheat (*Triticum aestivum* L.). *Gene* **2023**, *888*, 147756. [[CrossRef](#)]
40. Akram, J.; Siddique, R.; Shafiq, M.; Tabassum, B.; Manzoor, M.T.; Javed, M.A.; Anwar, S.; Nisa, B.U.; Saleem, M.H.; Javed, B. Genome-Wide Identification of CCO Gene Family in Cucumber (*Cucumis sativus*) and Its Comparative Analysis with *A. thaliana*. *BMC Plant Biol.* **2023**, *23*, 640. [[CrossRef](#)]
41. Pacheco, D.D.R.; Santana, B.C.G.; Pirovani, C.P.; De Almeida, A.-A.F. Zinc/Iron-Regulated Transporter-like Protein Gene Family in *Theobroma cacao* L: Characteristics, Evolution, Function and 3D Structure Analysis. *Front. Plant Sci.* **2023**, *14*, 1098401. [[CrossRef](#)] [[PubMed](#)]
42. You, J.; Liu, Z.; Qi, Z.; Ma, Y.; Sun, M.; Su, L.; Niu, H.; Peng, Y.; Luo, X.; Zhu, M.; et al. Regulatory Controls of Duplicated Gene Expression during Fiber Development in Allotetraploid Cotton. *Nat. Genet.* **2023**, *55*, 1987–1997. [[CrossRef](#)] [[PubMed](#)]
43. Wang, M.; Tu, L.; Lin, M.; Lin, Z.; Wang, P.; Yang, Q.; Ye, Z.; Shen, C.; Li, J.; Zhang, L.; et al. Asymmetric Subgenome Selection and Cis-Regulatory Divergence during Cotton Domestication. *Nat. Genet.* **2017**, *49*, 579–587. [[CrossRef](#)]
44. He, J.; Zhao, H.; Cheng, Z.; Ke, Y.; Liu, J.; Ma, H. Evolution Analysis of the Fasciclin-Like Arabinogalactan Proteins in Plants Shows Variable Fasciclin-AGP Domain Constitutions. *Int. J. Mol. Sci.* **2019**, *20*, 1945. [[CrossRef](#)]
45. Mao, Y.; Dai, F.; Si, Z.; Fang, L.; Zhang, T. Duplicate Mutations of GhCYP450 Lead to the Production of Ms5m6 Male Sterile Line in Cotton. *Theor. Appl. Genet.* **2023**, *136*, 2. [[CrossRef](#)]



46. Lamport, D.T.A.; Tan, L.; Held, M.A.; Kieliszewski, M.J. Pollen Tube Growth and Guidance: Occam's Razor Sharpened on a Molecular Arabinogalactan Glycoprotein Rosetta Stone. *New Phytol.* **2018**, *217*, 491–500. [\[CrossRef\]](#)
47. Van Hengel, A.J.; Barber, C.; Roberts, K. The Expression Patterns of Arabinogalactan-protein AtAGP30 and GLABRA2 Reveal a Role for Absciscic Acid in the Early Stages of Root Epidermal Patterning. *Plant J.* **2004**, *39*, 70–83. [\[CrossRef\]](#)
48. Suzuki, Y.; Kitagawa, M.; Knox, J.P.; Yamaguchi, I. A Role for Arabinogalactan Proteins in Gibberellin-induced  $\alpha$ -amylase Production in Barley Aleurone Cells. *Plant J.* **2002**, *29*, 733–741. [\[CrossRef\]](#)
49. Han, T.; Dong, H.; Cui, J.; Li, M.; Lin, S.; Cao, J.; Huang, L. Genomic, Molecular Evolution, and Expression Analysis of Genes Encoding Putative Classical AGPs, Lysine-Rich AGPs, and AG Peptides in *Brassica rapa*. *Front. Plant Sci.* **2017**, *8*, 397. [\[CrossRef\]](#)
50. Shi, H.-D.; Zhang, W.-Q.; Lu, H.-Y.; Zhang, W.-Q.; Ye, H.; Liu, D.-D. Functional Characterization of a Starch Synthesis-Related Gene AmAGP in *Amorphophallus muelleri*. *Plant Signal. Behav.* **2020**, *15*, 1805903. [\[CrossRef\]](#)
51. Shibata, M.; Sugimoto, K. A Gene Regulatory Network for Root Hair Development. *J. Plant Res.* **2019**, *132*, 301–309. [\[CrossRef\]](#) [\[PubMed\]](#)
52. Marzec, M.; Szarejko, I.; Melzer, M. Arabinogalactan Proteins Are Involved in Root Hair Development in Barley. *J. Exp. Bot.* **2015**, *66*, 1245–1257. [\[CrossRef\]](#) [\[PubMed\]](#)
53. Su, S.; Higashiyama, T. Arabinogalactan Proteins and Their Sugar Chains: Functions in Plant Reproduction, Research Methods, and Biosynthesis. *Plant Reprod.* **2018**, *31*, 67–75. [\[CrossRef\]](#) [\[PubMed\]](#)
54. Villa-Rivera, M.G.; Cano-Camacho, H.; López-Romero, E.; Zavala-Páramo, M.G. The Role of Arabinogalactan Type II Degradation in Plant-Microbe Interactions. *Front. Microbiol.* **2021**, *12*, 730543. [\[CrossRef\]](#)
55. Yang, J.; Sardar, H.S.; McGovern, K.R.; Zhang, Y.; Showalter, A.M. A Lysine-rich Arabinogalactan Protein in Arabidopsis Is Essential for Plant Growth and Development, Including Cell Division and Expansion. *Plant J.* **2007**, *49*, 629–640. [\[CrossRef\]](#)
56. Leszczuk, A.; Zajac, A.; Kurzyna-Szklarek, M.; Cybulska, J.; Zdunek, A. Investigations of Changes in the Arabinogalactan Proteins (AGPs) Structure, Size and Composition during the Fruit Ripening Process. *Sci. Rep.* **2020**, *10*, 20621. [\[CrossRef\]](#)
57. Demesa-Arévalo, E.; Vielle-Calzada, J.-P. The Classical Arabinogalactan Protein AGP18 Mediates Megaspore Selection in Arabidopsis. *Plant Cell* **2013**, *25*, 1274–1287. [\[CrossRef\]](#)
58. Zhou, X.; Hu, W.; Li, B.; Yang, Y.; Zhang, Y.; Thow, K.; Fan, L.; Qu, Y. Proteomic Profiling of Cotton Fiber Developmental Transition from Cell Elongation to Secondary Wall Deposition. *ABBS* **2019**, *51*, 1168–1177. [\[CrossRef\]](#)
59. Hossain, M.S.; Ahmed, B.; Ullah, M.W.; Aktar, N.; Haque, M.S.; Islam, M.S. Genome-Wide Identification of Fasciclin-like Arabinogalactan Proteins in Jute and Their Expression Pattern during Fiber Formation. *Mol. Biol. Rep.* **2020**, *47*, 7815–7829. [\[CrossRef\]](#)
60. Qin, L.-X.; Chen, Y.; Zeng, W.; Li, Y.; Gao, L.; Li, D.-D.; Basic, A.; Xu, W.-L.; Li, X.-B. The Cotton Beta-Galactosyltransferase 1 (GalT1) That Galactosylates Arabinogalactan Proteins Participates in Controlling Fiber Development. *Plant J.* **2017**, *89*, 957–971. [\[CrossRef\]](#)
61. Yu, J.; Jung, S.; Cheng, C.-H.; Lee, T.; Zheng, P.; Buble, K.; Crabb, J.; Humann, J.; Hough, H.; Jones, D.; et al. CottonGen: The Community Database for Cotton Genomics, Genetics, and Breeding Research. *Plants* **2021**, *10*, 2805. [\[CrossRef\]](#) [\[PubMed\]](#)
62. Finn, R.D.; Clements, J.; Eddy, S.R. HMMER Web Server: Interactive Sequence Similarity Searching. *Nucleic Acids Res.* **2011**, *39*, W29–W37. [\[CrossRef\]](#) [\[PubMed\]](#)
63. Duvaud, S.; Gabella, C.; Lisacek, F.; Stockinger, H.; Ioannidis, V.; Durinx, C. Expasy, the Swiss Bioinformatics Resource Portal, as Designed by Its Users. *Nucleic Acids Res.* **2021**, *49*, W216–W227. [\[CrossRef\]](#) [\[PubMed\]](#)
64. 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. [\[CrossRef\]](#)
65. Sievers, F.; Wilm, A.; Dineen, D.; Gibson, T.J.; Karplus, K.; Li, W.; Lopez, R.; McWilliam, H.; Remmert, M.; Söding, J.; et al. Fast, Scalable Generation of High-quality Protein Multiple Sequence Alignments Using Clustal Omega. *Mol. Syst. Biol.* **2011**, *7*, 539. [\[CrossRef\]](#)
66. Kumar, S.; Stecher, G.; Tamura, K. MEGA7: Molecular Evolutionary Genetics Analysis Version 7.0 for Bigger Datasets. *Mol. Biol. Evol.* **2016**, *33*, 1870–1874. [\[CrossRef\]](#)
67. Bailey, T.L.; Johnson, J.; Grant, C.E.; Noble, W.S. The MEME Suite. *Nucleic Acids Res.* **2015**, *43*, W39–W49. [\[CrossRef\]](#)
68. Bailey, T.L.; Gribskov, M. Concerning the Accuracy of MAST E-Values. *Bioinformatics* **2000**, *16*, 488–489. [\[CrossRef\]](#)
69. 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. [\[CrossRef\]](#)
70. 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. *Nucleic Acids Res.* **2012**, *40*, e49. [\[CrossRef\]](#)
71. Afzal Malik, W.; Afzal, M.; Chen, X.; Cui, R.; Lu, X.; Wang, S.; Wang, J.; Mahmood, I.; Ye, W. Systematic Analysis and Comparison of ABC Proteins Superfamily Confer Structural, Functional and Evolutionary Insights into Four Cotton Species. *Ind. Crops Prod.* **2022**, *177*, 114433. [\[CrossRef\]](#)

72. Wang, K.; Wang, Z.; Li, F.; Ye, W.; Wang, J.; Song, G.; Yue, Z.; Cong, L.; Shang, H.; Zhu, S.; et al. The Draft Genome of a Diploid Cotton *Gossypium raimondii*. *Nat. Genet.* **2012**, *44*, 1098–1103. [[CrossRef](#)] [[PubMed](#)]
73. Lescot, M. PlantCARE, a Database of Plant Cis-Acting Regulatory Elements and a Portal to Tools for in Silico Analysis of Promoter Sequences. *Nucleic Acids Res.* **2002**, *30*, 325–327. [[CrossRef](#)] [[PubMed](#)]
74. Tu, L.; Zhang, X.; Liu, D.; Jin, S.; Cao, J.; Zhu, L.; Deng, F.; Tan, J.; Zhang, C. Suitable Internal Control Genes for qRT-PCR Normalization in Cotton Fiber Development and Somatic Embryogenesis. *Chin. Sci. Bull.* **2007**, *52*, 3110–3117. [[CrossRef](#)]
75. Livak, K.J.; Schmittgen, T.D. Analysis of Relative Gene Expression Data Using Real-Time Quantitative PCR and the  $2^{-\Delta\Delta CT}$  Method. *Methods* **2001**, *25*, 402–408. [[CrossRef](#)]

**Disclaimer/Publisher’s Note:** The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.