



Genome-Wide Analysis of the GW2-Like Genes in *Gossypium* and Functional Characterization of the Seed Size Effect of *GhGW2-2D*

Li Huang^{1,2†}, Shuxian Yang^{2†}, Luyao Wu², Yue Xin², Jikun Song², Li Wang², Wenfeng Pei^{2,3}, Man Wu², Jiwen Yu², Xiaoyan Ma^{2*} and Shoulin Hu^{1*}

¹ College of Plant Sciences, Tarim University, Xinjiang, China, ² State Key Laboratory of Cotton Biology, Key Laboratory of Cotton Genetic Improvement, Ministry of Agriculture, Institute of Cotton Research of Chinese Academy of Agricultural Sciences, Anyang, China, ³ Western Agriculture Research Centre, Chinese Academy of Agricultural Sciences, Changji, China

OPEN ACCESS

Edited by:

Baohua Wang,
Nantong University, China

Reviewed by:

Jinfa Zhang,
New Mexico State University,
United States
Guanghui Xiao,
Shaanxi Normal University, China

*Correspondence:

Xiaoyan Ma
maxy_caas@126.com
Shoulin Hu
hushoulinghu@163.com

† These authors have contributed
equally to this work

Specialty section:

This article was submitted to
Plant Breeding,
a section of the journal
Frontiers in Plant Science

Received: 24 January 2022

Accepted: 15 February 2022

Published: 07 March 2022

Citation:

Huang L, Yang S, Wu L, Xin Y,
Song J, Wang L, Pei W, Wu M, Yu J,
Ma X and Hu S (2022) Genome-Wide
Analysis of the GW2-Like Genes
in *Gossypium* and Functional
Characterization of the Seed Size
Effect of *GhGW2-2D*.
Front. Plant Sci. 13:860922.
doi: 10.3389/fpls.2022.860922

Cotton is one of the most economically important crops worldwide. Seed size is a vital trait for plants connected with yield and germination. *GW2* encodes a RING_Ubox E3 ubiquitin ligase that controls seed development by affecting cell growth. Here, are few reports on *GW2*-like genes in cotton, and the function of *GW2* in cotton is poorly understood. In the present study, a genome-wide analysis identified 6 and 3 *GW2*-like genes in each of the two cultivated tetraploids (*Gossypium hirsutum* and *G. barbadense*) and each of their diploid ancestral species (*G. arboreum*, *G. raimondii*), respectively. *GhGW2-2D* has the same functional domain and high sequence similarity with *AtDA2* in *Arabidopsis*. Overexpression of *GhGW2-2D* in *Arabidopsis* significantly reduced seed and seedling size, suggesting *GhGW2-2D* is a potential target for regulating cotton seed size. These results provided information on the genetic and molecular basis of *GW2*-like genes in cotton, thus establishing a foundation for functional studies of cotton seeds.

Keywords: cotton, *GW2*-like, expression pattern, seed size, *GhGW2-2D*

INTRODUCTION

Cotton is one of the most valuable commercial crops worldwide. Cotton fiber is important for textile manufacturing, and cottonseed is a major source of nutrients in human and livestock feed (Bellaloui et al., 2015; Bellaloui et al., 2021). Despite cottonseeds being of high economic value, they have not received sufficient attention in the past years (Li et al., 2009). Understanding the characteristics of cottonseeds is important for optimizing cotton plant growth, for instance, large seeds usually have higher germination and seedling vigor indices (Bewley, 1997; Pahlavani et al., 2008; Zhao et al., 2019). Seed size is an important trait that is limited by both genetic and environmental (Hu et al., 2021). Numerous studies conducted in rice (*Oryza sativa*) and *Arabidopsis thaliana* have provided an insight into the molecular mechanisms controlling seed size (Xing and Zhang, 2010; Valluru et al., 2014; Li et al., 2019). These mechanisms include several signaling pathways that control seed size, such as G-protein signaling, the ubiquitin-proteasome pathway, mitogen-activated protein kinase (MAPK) signaling, the HAIKU (IKU) pathway, phytohormones, and some transcriptional regulators (Li and Li, 2016; Li et al., 2019). The ubiquitin proteasome pathway plays an important role in controlling seed size

(Li and Li, 2016; Li et al., 2019; Hao et al., 2021). Several RING-type E3 ubiquitin ligases have been reported to control seed size. For example, *Grain Width 2* (*GW2*), encoding RING_Ubox type E3 ubiquitin ligase, was found to affect grain size and improve grain yield potential (Song et al., 2007). *GW2* directly interacts with EXPANSIN-like1 (*EXPLA1*), a cell wall-loosening protein that increases cell growth, and inactivates it through ubiquitination, thereby reducing seed growth (Choi et al., 2018). The function of *GW2*-like genes in plants has been reported in many types of crops. In spikelet hulls, a loss-of-function mutation in *GW2* has been reported to increase cell proliferation, leading to wide and heavy grains with a high yield of rice (Yamaguchi et al., 2020). Seeds of *GW2*-knockout mutants have a high protein content and essential dietary minerals in the endosperm (Achary and Reddy, 2021). Reducing the abundance of *GW2* transcript with RNA interference in the durum wheat cultivar Svevo increased the grain starch content, width, and surface area (Sestili et al., 2019). In addition, *GW2* homologs were also reported to control grain size in maize and wheat (Li et al., 2010; Su et al., 2011). The gene *AtDA2* shares significant homology with *OsGW2* (Xia et al., 2013), which encodes a protein that can monoubiquitinate *DA1* to activate its peptidase activity, thereby causing the cleavage of *DA1* substrates (Dong et al., 2017). In *Arabidopsis*, *AtDA2* also interacts with *AtDA1* and *da2-1* has larger seeds and organs than wild type, suggesting a conserved function between *OsGW2* and *AtDA2* in seed size control (Xia et al., 2013).

Nonetheless, the function of *GW2* in cotton is still poorly understood. In recent years, complete sequencing of the cotton genome has facilitated the comprehensive identification and analysis of the cotton genes (Wang K. et al., 2012; Li et al., 2014, 2015; Zhu et al., 2017). In this study, we identified several *GW2*-like genes in *Gossypium hirsutum* (6), *G. barbadense* (6), *G. arboreum* (3), and *G. raimondii* (3). The sequence characteristics, chromosomal distribution, evolutionary relationship, and the expression patterns of *GW2*-like genes in cotton were analyzed. As the homologous gene of *AtDA2* and *OsGW2*, *GhGW2-2D* was also found to encode a RING_Ubox domain in cotton, which may negatively regulate seed size. Then we overexpressed *GhGW2-2D* into *Arabidopsis thaliana* ecotype Col-0 to further investigate the function of *GhGW2-2D*. The results of this study lay the foundation for future studies on *GW2*-like genes in the improvement of cotton seed size.

MATERIALS AND METHODS

Sequence Retrieval and Identification of *GW2*-Like Genes in Cotton

The four completed genome assemblies of *G. arboreum* (A2, CRI_V1.0) (Du et al., 2018), *G. raimondii* (D5, JGI v2_a2.1) (Paterson et al., 2012), *hirsutum* acc. TM-1 (AD1, ZJU) (Hu et al., 2019), and *G. barbadense* acc.H7124 (AD2, ZJU) (Hu et al., 2019) were downloaded from the CottonGen database¹ (Yu et al., 2014). The protein sequences of *AtDA2* and *OsGW2* were acquired from

¹<https://www.cottongen.org>

the *Arabidopsis* Information Resource² and the Rice Genome Annotation Project³, respectively. The protein sequences of *AtDA2* and *OsGW2* were used as queries to identify *GW2*-like genes against four genome databases of *Gossypium* with *e*-values of $1e^{-10}$. The candidate sequences were submitted to the NCBI batch CDD program⁴ and SMART database⁵ (Letunic et al., 2015) for further confirmation.

Chromosomal Location

The chromosomal distribution and the loci of all *GW2*-like genes were determined based on the results of identification genome annotation data. Molecular markers of reported QTLs were download from CottonQTLdb⁶ (Said et al., 2013, 2015a,b). Subsequently, the location images of *GW2*-like genes were drawn using the MapChart software (Voorrips, 2002).

Phylogenetic Analysis

The phylogenetic tree of *GW2*-like genes was analyzed and generated using multiple sequence alignments of four cultivated cotton species using ClusterW⁷ (Larkin et al., 2007). A phylogenetic tree was constructed using MEGA 7.0⁸ (Kumar et al., 2016) with pairwise distance and the neighbor-joining (NJ) method.

Gene Duplication and Synteny Analysis

Gene duplication events and genomic synteny were analyzed using MCScanX and visualized using the TBtools software (Wang Y. et al., 2012; Cheng et al., 2020). The synonymous (*Ks*) and non-synonymous (*Ka*) substitution ratios were calculated using TBtools software.

Gene Structure Analysis and Identified Motifs

The structural information of GFF3 files of four cotton species were download from CottonGen database (see text footnote 1). The protein length, molecular weight, and isoelectric point of the encoded proteins were predicted using the ProtParam tool⁹ (Gasteiger et al., 2003). To clarify the conserved motifs and protein domains, the online program MEME¹⁰ (Bailey et al., 2009) and NCBI database were used, respectively. The structure of *GW2*-like genes was predicted using TBtools software (Cheng et al., 2020).

Plant Material, RNA Extraction, and qRT-PCR Analysis

Wild-type cotton (TM-1) and *Arabidopsis* (Col-0) were planted in the experimental field and greenhouse of the Institute of Cotton

²<http://www.arabidopsis.org>

³<http://rice.uga.edu/expression.shtml>

⁴<https://www.ncbi.nlm.nih.gov/Structure/bwrpsb/bwrpsb.cgi>

⁵<http://smart.embl-heidelberg.de/>

⁶<http://www2.cottonqtl.org:8081/>

⁷<http://www.ebi.ac.uk/Tools/msa/clustalw2>

⁸<http://www.megasoftware.net>

⁹<https://web.expasy.org/>

¹⁰<http://meme-suite.org/>

Research of Chinese Academy of Agricultural Sciences under conventional field management conditions in Anyang, China. To detect the relative expression of GW2-like genes, different samples were collected from WT cotton at different stages of ovule development. All ovule samples were self-pollinated. Total RNA was isolated from samples using Fast Pure Plant Total RNA Isolation Kit (Vazyme, Nanjing, China). The qRT-PCR analysis primers were listed in **Supplementary Table 1**. Gh_D03G0370 (*GhActin3*) (Li et al., 2005) and AT3G18780 (*AtActin2*) (Wang et al., 2016) were used as internal controls for the qPCR experiments.

Constructs and Transformation

The 1,272 bp complete coding sequence (CDS) of *GhGW2-2D* was amplified using the primers 35S:*GhGW2-2D*-F and 35S:*GhGW2-2D*-R (**Supplementary Table 1**). The 35S:*GhGW2-2D* construct were cloned into a pBI121 vector digested with *Bam*HI and *Sac*I using a PCR-based homologous recombination system. The 35S:*GhGW2-2D* plasmid transferred into Col-0 plants using *Agrobacterium* GV3101 and the transformants were selected on medium supplemented with kanamycin (50 mg/L). Three homozygous T₃ generation lines were used for further analysis. The seed and cotyledon were conducted by scanning to generate a digital image and then measured by ImageJ software.

RESULTS

Identification of GW2-Like Genes in *Gossypium*

In order to identify all GW2-like proteins in tetraploid cotton (*G. hirsutum* and *G. barbadense*) and their diploid ancestors (*G. arboreum* and *G. raimondii*), the protein sequences of AtDA2 and OsGW2 in rice and *Arabidopsis* were used as queries.

We identified three, three, six, and six GW2-like genes in *G. arboreum*, *G. raimondii*, *G. barbadense*, and *G. hirsutum*, respectively. The length of these cotton GW2-like gene sequences ranged from 1,260 to 1,290 bp, encoding polypeptides from 419 to 429 amino acids long. The predicted isoelectric point of proteins varied from 4.70 to 4.93. According to the results of evolutionary analysis, all members were divided into three subgroups. More details of the gene names, locus IDs, and other features are shown in **Table 1**.

Chromosomal Location of GW2-Like Genes

All GW2-like genes were unevenly mapped onto chromosomes based on the genome annotation data of the four species (**Supplementary Figure 1** and **Table 1**). In the *G. arboreum* genome, the three identified *GaGW2s* were three located on chromosomes A01, A05, and A13. In *G. raimondii*, the three *GrGW2* genes were distributed on three chromosomes: chromosome 02 (D01), chromosome 09 (D05), and chromosome 13 (D13). In *G. hirsutum*, six *GhGW2* genes were mapped on three At subgenome chromosomes (A01, A05, and A13) and three Dt subgenome chromosomes (D01, D05, and D13). The chromosomal distribution of GW2-like genes in the *G. barbadense* genome was resemble to that of *GhGW2s*. Therefore, it could infer that GW2-like genes were highly conserved in the evolution from diploid to tetraploid in cotton.

Gene Structure and Protein Domain Analysis of GW2-Like Genes

An efficient way to predict gene function is to analyze its gene structure and the encoded protein domains. In this study, the GW2-like gene structures of four cotton species were

TABLE 1 | Information of the GW2-like genes in *Gossypium*.

Group name	Gene name	Gene identifier	Chromosomal localization	pI	MW (KD)	Size (AA)
Group 1	Ga01G0386	<i>GaGW2-1</i>	A01	4.90	47.58	429
	GB_A01G0356	<i>GbGW2-1A</i>	A01	4.93	47.45	428
	GB_D01G0377	<i>GbGW2-1D</i>	D01	4.83	47.5	429
	GH_A01G0369	<i>GhGW2-1A</i>	A01	4.93	47.52	429
	GH_D01G0356	<i>GhGW2-1D</i>	D01	4.82	47.52	429
	Gorai.002G038700	<i>GrGW2-1</i>	D01	4.83	47.53	429
Group 2	Ga05G3697	<i>GaGW2-2</i>	A05	4.84	46.97	423
	GB_A04G0530	<i>GbGW2-2A</i>	A04	4.85	47.08	423
	GB_D05G3554	<i>GbGW2-2D</i>	D05	4.82	46.97	423
	GH_A04G0522	<i>GhGW2-2A</i>	A04	4.85	46.91	423
	GH_D05G3532	<i>GhGW2-2D</i>	D05	4.79	46.61	419
	Gorai.009G359400	<i>GrGW2-2</i>	D05	4.82	46.96	423
Group 3	Ga13G1904	<i>GaGW2-3</i>	A13	4.85	47.02	426
	GB_A13G1898	<i>GbGW2-3A</i>	A13	4.88	46.75	423
	GB_D13G1839	<i>GbGW2-3D</i>	D13	4.73	47.25	428
	GH_A13G1787	<i>GhGW2-3A</i>	A13	4.85	46.73	423
	GH_D13G1740	<i>GhGW2-3D</i>	D13	4.73	47.31	429
	Gorai.013G177000	<i>GrGW2-3</i>	D13	4.70	47.26	429

visualized to further understand their differences and potential functional diversity.

A phylogenetic tree of GW2-like genes in *G. hirsutum* and *G. barbadense* and their ancestral diploids was constructed using the NJ method (Figure 1A). All GW2-like proteins were divided into three subgroups. Each group contained six members, one from *G. arboreum*, one from *G. raimondii*, two from *G. hirsutum*, and two from *G. barbadense*. All GW2-like genes have eight exons that were highly conserved in the four species of cotton (Figure 1B).

The NCBI and SMART databases were used to predict the domains of the GW2-like proteins. All GW2-like proteins had a typical E3 ubiquitin ligase RING_Ubox domain (Figure 1C), which can bind to a specific E2 ubiquitin-binding enzyme and determines the specificity of the ubiquitinated substrate. The MEME online program was used to predict the motifs of the GW2-like family genes. The results showed that most members in the same subgroup had resemble gene structures and functional domains (Figure 1D).

Adaptive Evolution Analysis of the GW2-Like Genes

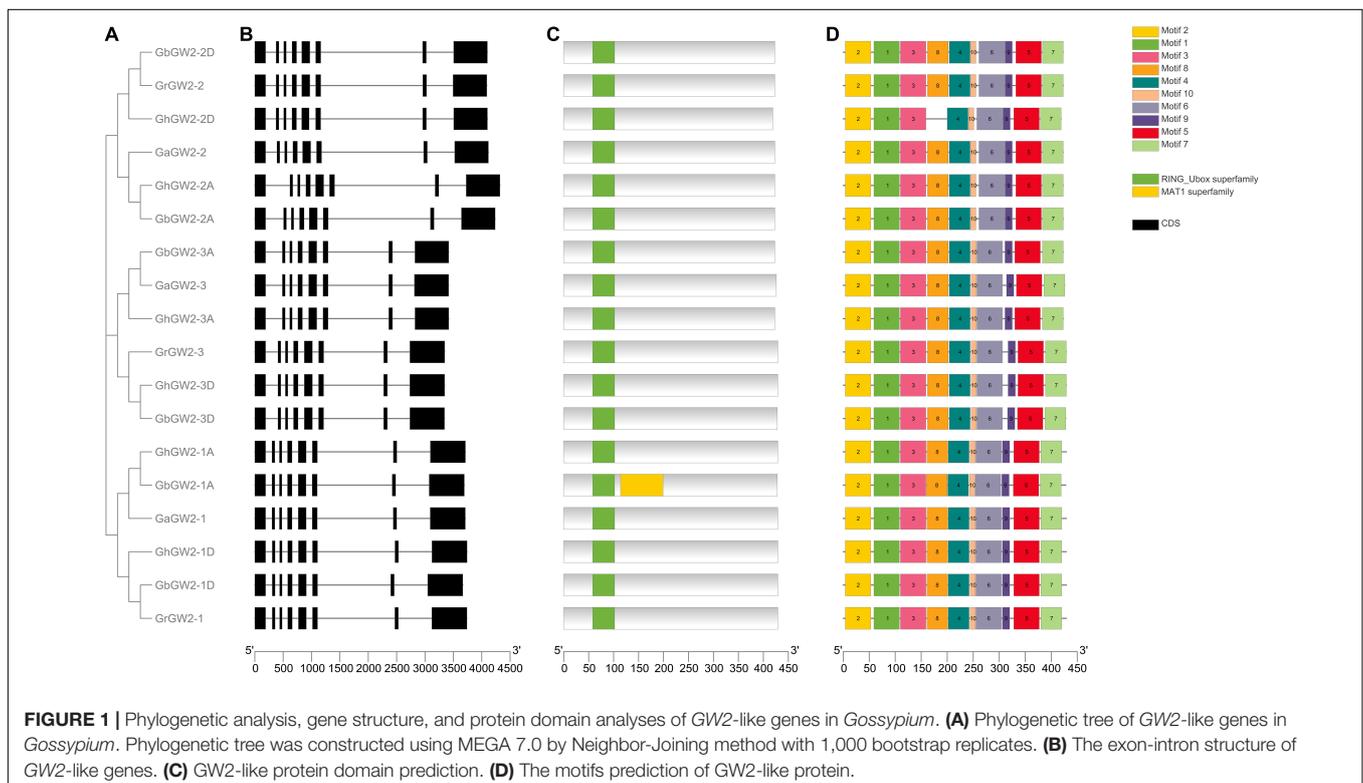
Gene replication and subsequent functionalization are important driving forces for genome and species evolution. In addition, gene duplication events are important for amplification of gene families. To clarify the amplification mechanism and evolution of GW2-like genes in the four cotton species, we conducted interspecific collinearity analysis and gene repeat event analysis (Figure 2 and Table 2).

Collinearity analysis was performed for *G. arboreum* and *G. hirsutum* (A subgenome), *G. arboreum* and *G. barbadense* (A subgenome), *G. raimondii* and *G. hirsutum* (D subgenome), *G. raimondii* and *G. barbadense* (D subgenome), respectively (Figure 2). A total of three, six, five, and seven pairs of collinear genes were found, respectively. The *Gb*GW2-like genes in *G. barbadense* had more collinear gene pairs with *G. arboreum* (A2) and *G. raimondii* (D5), while the *Gh*GW2-like genes in *G. hirsutum* had fewer collinear gene pairs with *G. arboreum* (A2) and *G. raimondii* (D5), indicating that GW2-like genes in *G. barbadense* were more related to two possible ancestral species of *G. arboreum* and *G. raimondii* than GW2-like genes in *G. hirsutum*.

In order to elucidating the evolutionary dynamics and selection pressure of protein coding, the non-synonymous (Ka) and synonymous (Ks) replacement rates were calculated using full-length sequences between each pair of linear homologous genes. Generally, a Ka/Ks ratio >1, =1 and <1 indicates diversified selection, neutral selection, and purification selection, respectively. The Ka/Ks ratio of all gene pairs was less than 1, suggesting that these genes had undergone purification selection (Table 2). Above results indicated that the GW2-like genes in *G. hirsutum* and *G. barbadense* were relatively conserved.

Expression Profiles of GhGW2-Like Genes in Cotton

To investigate the tissue-specific expression profiles of GW2-like genes, published TM-1 expression data, including root, stem, leaf, torus, petal, stamen, pistil, calycle, fiber and ovule at various



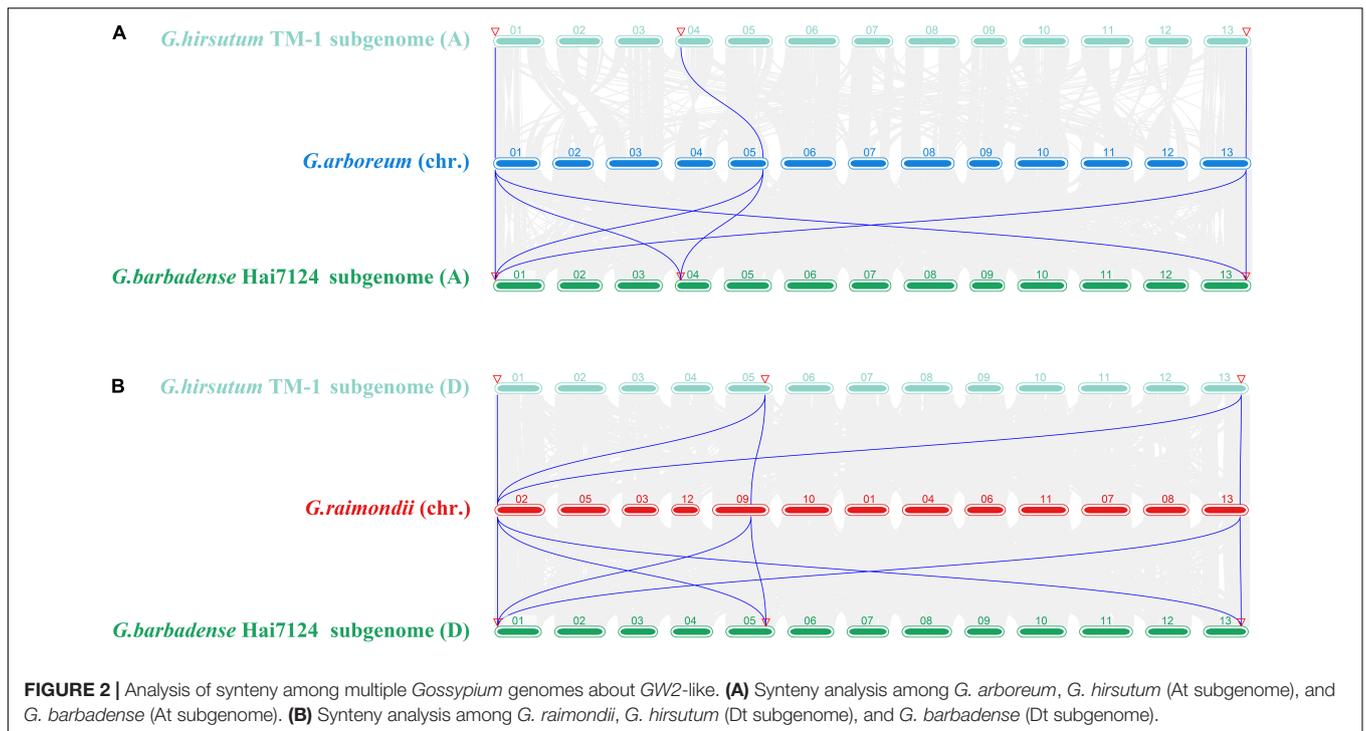


TABLE 2 | Ka and Ks calculations for GW2-like duplicated gene pairs.

Species	Gene 1	Gene 2	Ka	Ks	Ka/Ks	Duplicate type	
<i>G. hirsutum</i>	<i>GhGW2-1A</i>	<i>GhGW2-2A</i>	0.14	0.48	0.29	Segmental	
	<i>GhGW2-1A</i>	<i>GhGW2-3A</i>	0.15	0.47	0.32	Segmental	
	<i>GhGW2-1A</i>	<i>GhGW2-1D</i>	0.01	0.05	0.27	Segmental	
	<i>GhGW2-1A</i>	<i>GhGW2-2D</i>	0.14	0.46	0.29	Segmental	
	<i>GhGW2-1A</i>	<i>GhGW2-3D</i>	0.15	0.41	0.37	Segmental	
	<i>GhGW2-2A</i>	<i>GhGW2-1D</i>	0.15	0.44	0.33	Segmental	
	<i>GhGW2-2A</i>	<i>GhGW2-2D</i>	0.01	0.06	0.18	Segmental	
	<i>GhGW2-3A</i>	<i>GhGW2-3D</i>	0.01	0.07	0.16	Segmental	
	<i>GhGW2-1D</i>	<i>GhGW2-2D</i>	0.14	0.42	0.33	Segmental	
	<i>GhGW2-1D</i>	<i>GhGW2-3D</i>	0.15	0.37	0.41	Segmental	
	<i>G. barbadense</i>	<i>GbGW2-1A</i>	<i>GbGW2-2D</i>	0.14	0.47	0.29	Segmental
		<i>GbGW2-1A</i>	<i>GbGW2-2A</i>	0.14	0.49	0.29	Segmental
		<i>GbGW2-1A</i>	<i>GbGW2-3D</i>	0.15	0.41	0.38	Segmental
		<i>GbGW2-1A</i>	<i>GbGW2-1D</i>	0.02	0.05	0.33	Segmental
<i>GbGW2-2A</i>		<i>GbGW2-2D</i>	0.01	0.06	0.22	Segmental	
<i>GbGW2-2A</i>		<i>GbGW2-1D</i>	0.15	0.44	0.33	Segmental	
<i>GbGW2-3A</i>		<i>GbGW2-3D</i>	0.01	0.06	0.21	Segmental	
<i>GbGW2-3A</i>		<i>GbGW2-1D</i>	0.15	0.42	0.37	Segmental	
<i>GbGW2-1D</i>		<i>GbGW2-2D</i>	0.14	0.42	0.34	Segmental	
<i>GbGW2-1D</i>		<i>GbGW2-3D</i>	0.15	0.36	0.42	Segmental	

developmental stages were used. Most *GhGW2* genes were extensively expressed in many tissues (**Supplementary Figure 2**). *GhGW2-1A* and *GhGW2-1D* were relatively highly expressed in pollen, petal, and pistil. *GhGW2-2A* was highly expressed in 20 DPA fiber and roots. *GhGW2-2D* was highly expressed in the stem, root and early stages of ovule. *GhGW2-3A* and *GhGW2-3D* had similar expression trends with *GhGW2-1A* that highly

expressing in flowers and pistils (**Supplementary Figure 2A**). The expression patterns of *GbGW2*-like genes in *G. barbadense* and *G. hirsutum* shared alike (**Supplementary Figure 2B**), indicating that these genes may have similar functions. qRT-PCR of different ovule development stages showed that *GhGW2-1A*, *GhGW2-2A* and *GhGW2-1D* had a similar tendency with highly expressed in early stages and later stages, and less expressed in

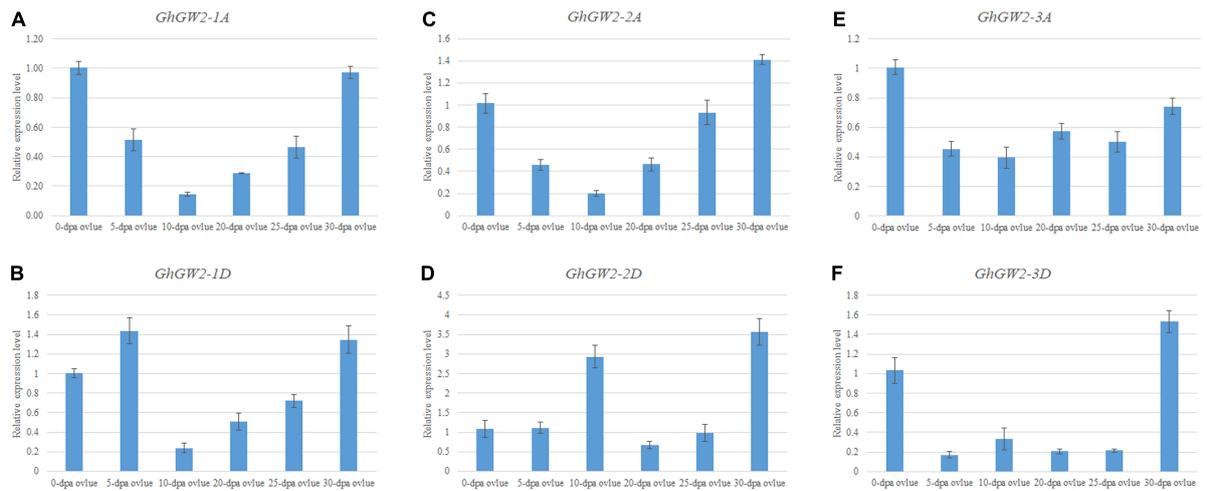


FIGURE 3 | qRT-PCR of GhGW2s in ovules. Relative expression level of *GhGW2-1A* (A), *GhGW2-1D* (B), *GhGW2-2A* (C), *GhGW2-2D* (D), *GhGW2-3A* (E), *GhGW2-3D* (F). GhActin3 was selected as internal and the 1 Ct value of 0-DPA-ovules was set as the control. The data presented are the means ± SD of three replicates.

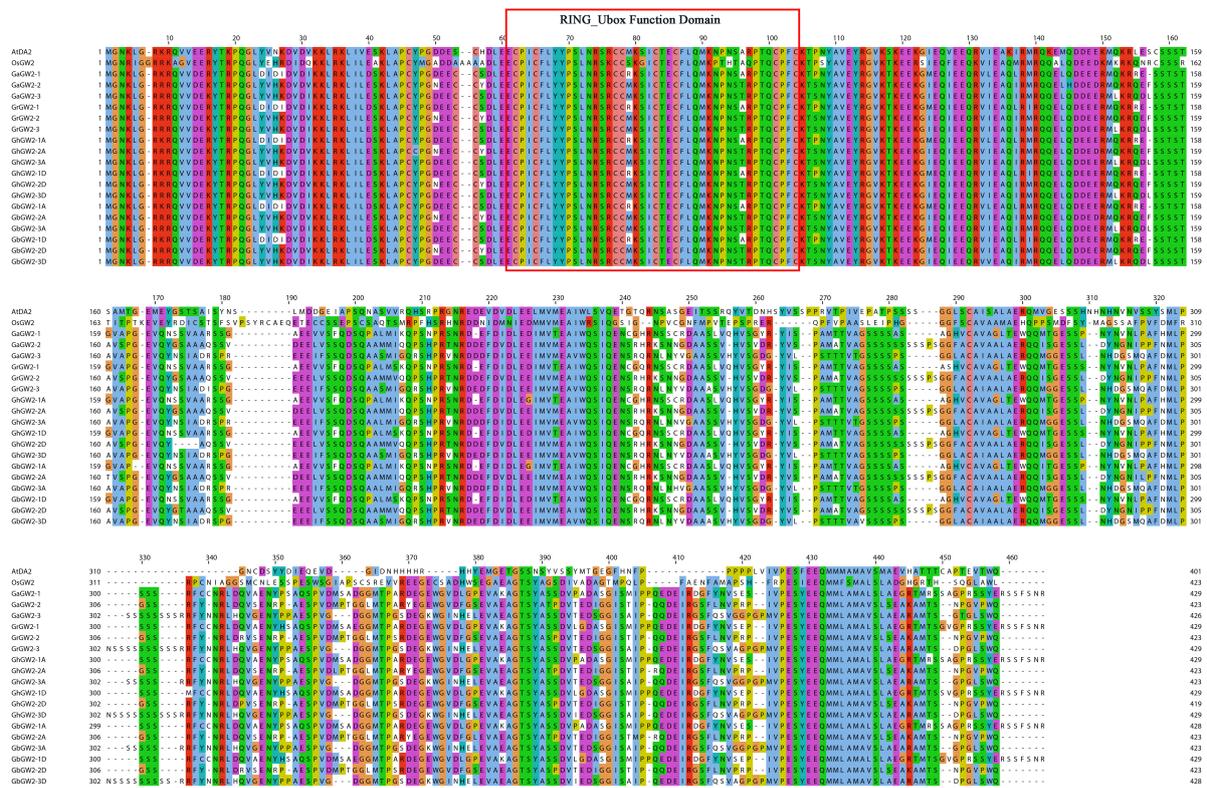
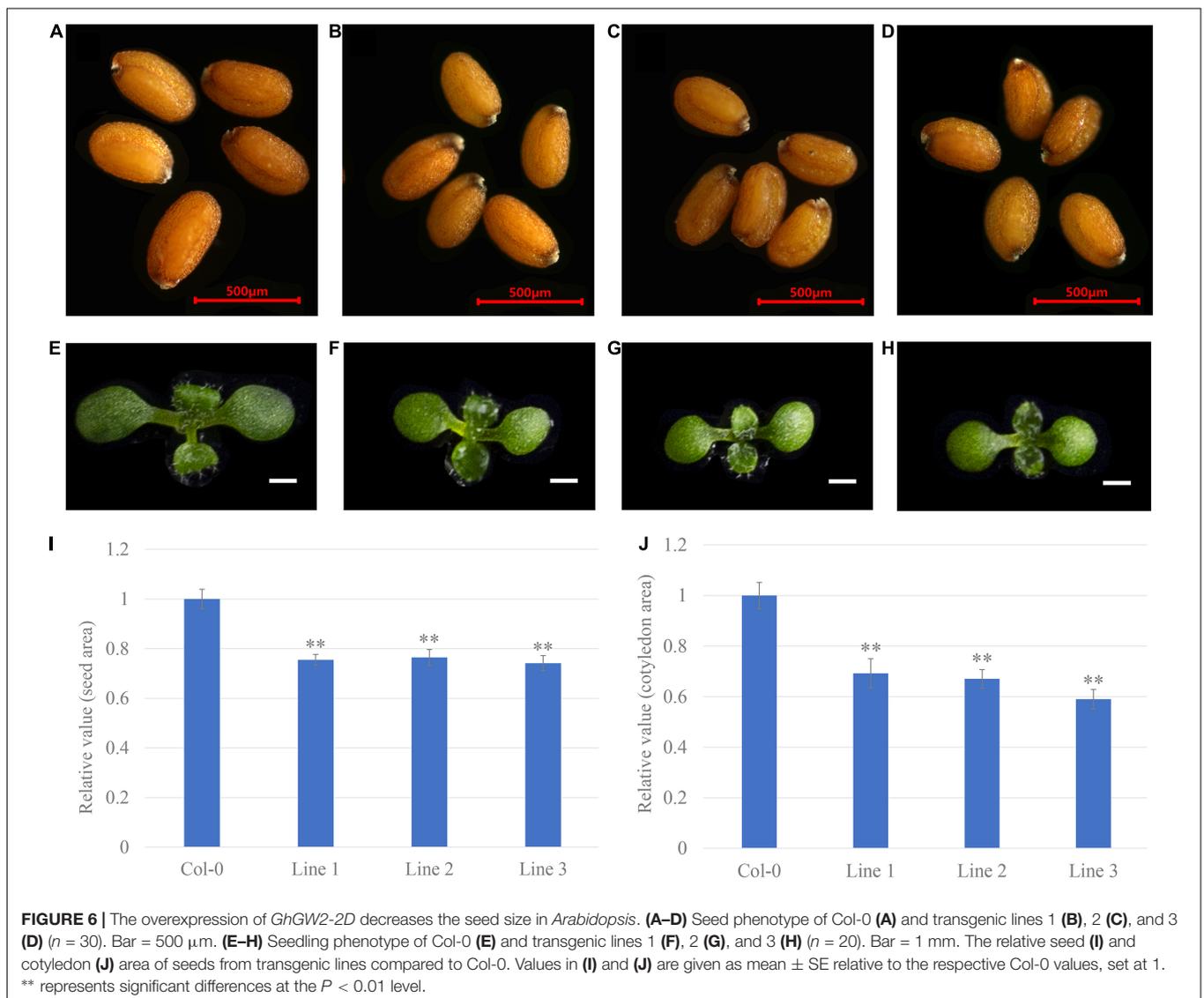
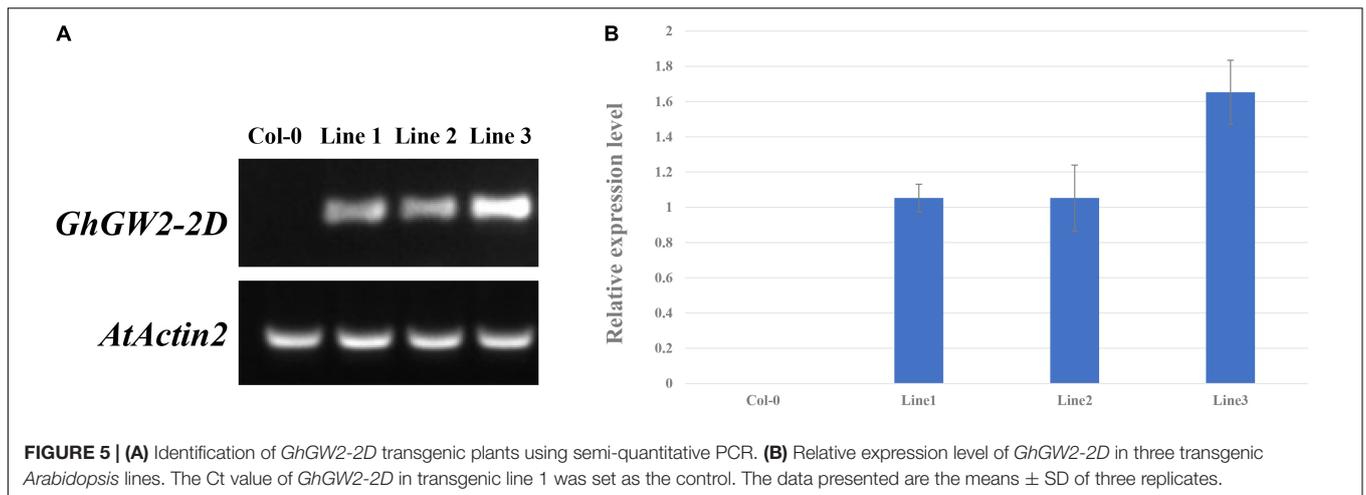


FIGURE 4 | Multiple sequence alignments of GW2-like amino acid sequences.

10–20 DPA ovules (Figures 3A–C). *GhGW2-2D* expression was highest in 10 DPA ovule and 30 DPA ovule, suggesting it may involve in cotton ovules development (Figure 3D). *GhGW2-3A* revealed an extensive expression in all ovule development stages (Figure 3E). *GhGW2-3D* was expressed at very low

levels in all ovule development stages except in 30 DPA ovule (Figure 3F).

GhGW2-like genes had distinct expression patterns in different tissues of TM-1, suggesting that *GhGW2*-like genes may have different functions in cotton growth.



Generation of *GhGW2-2D*-Overexpressing *Arabidopsis* Lines

Based on the sequence alignment and homology analysis, *GhGW2-2D* was selected for further study because of its similarity to *OsGW2* and *AtDA2* (Figure 4). Moreover, the expression of *GhGW2-2D* was relatively high in the stage of ovule development. Plants overexpressing *GhGW2-2D* were generated and identified using semi-quantitative PCR (Figure 5A), and a total of three transgenic lines with high expression of *GhGW2-2D* were identified. qRT-PCR was used to further assess the relative expression levels of the three diverse transgenic lines (Figure 5B).

Overexpression of *GhGW2-2D* Decreases Seed Size

To evaluate the applicability of transgenic breeding to alter seed size, different developmental stages of transgenic *Arabidopsis* were characterized. Overexpression of *GhGW2-2D* reduced seed size (Figures 6A–D). The seed areas of lines 1, 2, and 3 decreased by 25, 24, and 26%, respectively, compared with that of the wild type Col-0 (Figure 6I). The growth of nine-day-old seedlings was measured after germinating. The transgenic lines had smaller cotyledons than Col-0 (Figures 6E–H). Compared with that of wild type Col-0, the cotyledon area of lines 1, 2, and 3 decreased by 31, 33, and 42%, respectively (Figure 6J). These results further support the role of *GhGW2-2D* in limiting seed and organ growth.

DISCUSSION

Seed size is a vital characteristic of flowering plants (Orsi and Tanksley, 2009). Cottonseed kernels are regarded as the best source of vegetable protein and oil and are associated with seed size (Poehleman, 1994; Pahlavani et al., 2009). Advances have revealed that plants with large seeds accumulate more nutrients, which may affect seed germination, plant growth, and development (Wang, 2008). Thus, understanding the mechanisms controlling seed size is essential for agricultural production and germplasm improvement.

Recent studies have shown that *GW2*-like genes have a far-reaching impact on seed size and weight. *GW2*, which encodes an E3 ubiquitin ligase involved in the regulation of cell division, was found to affect grain size and improve grain yield potential (Song et al., 2007). The seeds of the offspring of mutant *da2-1* plants were bigger than those of the control, and the 1,000 grain weight was increased (Xia et al., 2013). Moreover, the loss of the function of this gene has been reported to increase the number of cells and the grain size of rice (Yamaguchi et al., 2020). Altogether, these studies highlight that the *GW2* gene has an important value in breeding and plays an important regulatory role in controlling seed development.

We identified 3, 3, 6, and 6 *GW2*-like genes in *G. raimondii*, *G. arboreum*, *G. barbadense*, and *G. hirsutum*, respectively. The chromosomal distribution, evolutionary relationship, and expression patterns of *GW2*-like genes in cotton were analyzed.

It has been found that all *GW2*-like genes have a typical RING_U-box domain, which can interact with a specific E2 ubiquitin binding enzyme and ubiquitinate specific substrates (Song et al., 2007). Gene structure analysis revealed that all *GW2*-like genes have eight exons, which indicates that the gene may be functionally conserved during evolution. According to the selection pressure analysis, the Ka/Ks ratio of *GW2*-like was less than 1, further supporting the evolutionary conservation of these genes. The *GhGW2-2D* was co-localized with quantitative trait loci (QTL) for seed size (boll weight, BW), suggesting that its sequence variations may be genetically associated with the natural variation in seed size (Supplementary Figure 3). Processing and analysis of published RNA-Seq data, we found that *GhGW2-2D* was highly expressed in roots, stems, and during early ovule development. Moreover, *GhGW2-2D* is most similar to *AtDA2* and *OsGW2* and has the same protein conserved functional domain. Therefore, this gene was selected for the subsequent analysis.

To date, only few studies have investigated *GW2*-like genes in cotton. In the present study, a comprehensive analysis of *GW2*-like genes in the four sequenced cotton species was performed, which provided more details of the cotton *GW2*-like genes. Overexpression of *GhGW2-2D* in *Arabidopsis* decreased seed size, indicating that the product encoded by this gene may play a negative regulatory role in seed development. *DA1* interacts with the E3 ubiquitin ligase *DA2* to regulate seed and organ size in *Arabidopsis*, *Triticum aestivum*, and *G. hirsutum* (Xia et al., 2013; Liu et al., 2020; Yang et al., 2021). It may be interesting to further explore the molecular network controlling seed size in cotton. These findings could provide an important theoretical and experimental basis for growth and regulating the development of seeds and organs in cotton, which has great significance in creating excellent germplasm.

DATA AVAILABILITY STATEMENT

The original contributions presented in the study are included in the article/Supplementary Material, further inquiries can be directed to the corresponding authors.

AUTHOR CONTRIBUTIONS

LH and SY planned the experiments and wrote the manuscript. LWu and YX participated in the study. JS, MW, WP, and JY provided advice for experiments and manuscript writing. XM and SH conceived and designed the research and manuscript revision. All authors read and approved the final manuscript.

FUNDING

This present study was supported by the Natural Science Foundation of Xinjiang Uygur Autonomous Region of China (Grant Nos. 2021D01B113 and 2020D01A135), the National Natural Science Foundation of China (Grant Nos. 32060449 and 31660410), and Agricultural Science and

Technology Innovation Program of Chinese Academy of Agricultural Sciences.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: <https://www.frontiersin.org/articles/10.3389/fpls.2022.860922/full#supplementary-material>

REFERENCES

- Achary, V. M. M., and Reddy, M. K. (2021). CRISPR-Cas9 mediated mutation in GRAIN WIDTH and WEIGHT2 (GW2) locus improves aleurone layer and grain nutritional quality in rice. *Sci. Rep.* 11:21941. doi: 10.1038/s41598-021-00828-z
- Bailey, T. L., Boden, M., Buske, F. A., Frith, M., Grant, C. E., Clementi, L., et al. (2009). MEME SUITE: tools for motif discovery and searching. *Nucleic Acids Res.* 37, W202–W208. doi: 10.1093/nar/gkp335
- Bellaloui, N., Stetina, S. R., and Turley, R. B. (2015). Cottonseed protein, oil, and mineral status in near-isogenic *Gossypium hirsutum* cotton lines expressing fuzzy/linted and fuzzless/linted seed phenotypes under field conditions. *Front. Plant Sci.* 6:137. doi: 10.3389/fpls.2015.00137
- Bellaloui, N., Turley, R. B., and Stetina, S. R. (2021). Cottonseed protein, oil, and minerals in cotton (*Gossypium hirsutum* L.) lines differing in curly leaf morphology. *Plants* 10:525. doi: 10.3390/plants10030525
- Bewley, J. D. (1997). Seed germination and dormancy. *Plant Cell* 9, 1055–1066. doi: 10.1105/tpc.9.7.1055
- Cheng, J., Chen, Y., Hu, Y., Zhou, Z., Hu, F., Dong, J., et al. (2020). Fine mapping of restorer-of-fertility gene based on high-density genetic mapping and collinearity analysis in pepper (*Capsicum annuum* L.). *Theor. Appl. Genet.* 133, 889–902. doi: 10.1007/s00122-019-03513-y
- Choi, B. S., Kim, Y. J., Markkandan, K., Koo, Y. J., Song, J. T., and Seo, H. S. (2018). GW2 functions as an E3 ubiquitin ligase for rice expansin-like 1. *Int. J. Mol. Sci.* 19:1904. doi: 10.3390/ijms19071904
- Dong, H., Dumenil, J., Lu, F. H., Na, L., Vanhaeren, H., Naumann, C., et al. (2017). Ubiquitylation activates a peptidase that promotes cleavage and destabilization of its activating E3 ligases and diverse growth regulatory proteins to limit cell proliferation in *Arabidopsis*. *Genes Dev.* 31, 197–208. doi: 10.1101/gad.292235.116
- Du, X., Huang, G., He, S., Yang, Z., Sun, G., Ma, X., et al. (2018). Resequencing of 243 diploid cotton accessions based on an updated A genome identifies the genetic basis of key agronomic traits. *Nat. Genet.* 50, 796–802. doi: 10.1038/s41588-018-0116-x
- Gasteiger, E., Gattiker, A., Hoogland, C., Ivanyi, I., Appel, R. D., and Bairoch, A. (2003). ExPASy: the proteomics server for in-depth protein knowledge and analysis. *Nucleic Acids Res.* 31, 3784–3788. doi: 10.1093/nar/kg563
- Hao, J., Wang, D., Wu, Y., Huang, K., Duan, P., Li, N., et al. (2021). The GW2-GW1-OsbZIP47 pathway controls grain size and weight in rice. *Mol. Plant* 14, 1266–1280. doi: 10.1016/j.molp.2021.04.011
- Hu, S., Yang, H., Gao, H., Yan, J., and Xie, D. (2021). Control of seed size by jasmonate. *Sci. China Life Sci.* 64, 1215–1226. doi: 10.1007/s11427-020-1899-8
- Hu, Y., Chen, J., Fang, L., Zhang, Z., Ma, W., Niu, Y., et al. (2019). *Gossypium barbadense* and *Gossypium hirsutum* genomes provide insights into the origin and evolution of allotetraploid cotton. *Nat. Genet.* 51, 739–748. doi: 10.1038/s41588-019-0371-5
- Kumar, S., Stecher, G., and Tamura, K. (2016). MEGA7: molecular evolutionary genetics analysis version 7.0 for bigger datasets. *Mol. Biol. Evol.* 33, 1870–1874. doi: 10.1093/molbev/msw054
- Larkin, M. A., Blackshields, G., Brown, N. P., Chenna, R., McGettigan, P. A., McWilliam, H., et al. (2007). Clustal W and clustal X version 2.0. *Bioinformatics* 23, 2947–2948. doi: 10.1093/bioinformatics/btm404
- Letunic, I., Doerks, T., and Bork, P. (2015). SMART: recent updates, new developments and status in 2015. *Nucleic Acids Res.* 43, D257–D260. doi: 10.1093/nar/gku949
- Li, F., Fan, G., Lu, C., Xiao, G., Zou, C., Kohel, R. J., et al. (2015). Genome sequence of cultivated upland cotton (*Gossypium hirsutum* TM-1) provides insights into genome evolution. *Nat. Biotechnol.* 33, 524–530. doi: 10.1038/nbt.3208
- Li, F., Fan, G., Wang, K., Sun, F., Yuan, Y., Song, G., et al. (2014). Genome sequence of the cultivated cotton *Gossypium arboreum*. *Nat. Genet.* 46, 567–572. doi: 10.1038/ng.2987
- Li, N., and Li, Y. (2016). Signaling pathways of seed size control in plants. *Curr. Opin. Plant Biol.* 33, 23–32. doi: 10.1016/j.pbi.2016.05.008
- Li, N., Xu, R., and Li, Y. (2019). Molecular networks of seed size control in plants. *Annu. Rev. Plant Biol.* 70, 435–463. doi: 10.1146/annurev-arplant-050718-095851
- Li, Q., Li, L., Yang, X., Warburton, M. L., Bai, G., Dai, J., et al. (2010). Relationship, evolutionary fate and function of two maize co-orthologs of rice GW2 associated with kernel size and weight. *BMC Plant Biol.* 10:143. doi: 10.1186/1471-2229-10-143
- Li, W., Zhou, Z., Meng, Y., Xu, N., and Fok, M. (2009). Modeling boll maturation period, seed growth, protein, and oil content of cotton (*Gossypium hirsutum* L.) in China. *Field Crops Res.* 112, 131–140. doi: 10.1016/j.fcr.2009.02.009
- Li, X. B., Fan, X. P., Wang, X. L., Cai, L., and Yang, W. C. (2005). The cotton ACTIN1 gene is functionally expressed in fibers and participates in fiber elongation. *Plant Cell* 17, 859–875. doi: 10.1105/tpc.104.029629
- Liu, H., Li, H., Hao, C., Wang, K., Wang, Y., Qin, L., et al. (2020). TaDA1, a conserved negative regulator of kernel size, has an additive effect with TaGW2 in common wheat (*Triticum aestivum* L.). *Plant Biotechnol. J.* 18, 1330–1342. doi: 10.1111/pbi.13298
- Orsi, C. H., and Tanksley, S. D. (2009). Natural variation in an ABC transporter gene associated with seed size evolution in tomato species. *PLoS Genet.* 5:e1000347. doi: 10.1371/journal.pgen.1000347
- Pahlavani, M., Miri, A., and Kazemi, G. (2009). Response of oil and protein content to seed size in cotton (*Gossypium hirsutum* L., cv. Sahel). *Plant Breed. Seed Sci.* 59, 53–64. doi: 10.2478/v10129-009-0004-8
- Pahlavani, M. H., Miri, A. A., and Kazemi, G. (2008). Response of oil and protein content to seed size in cotton. *Int. J. Agric. Biol.* 10, 643–647.
- Paterson, A. H., Wendel, J. F., Gundlach, H., Guo, H., Jenkins, J., Jin, D., et al. (2012). Repeated polyploidization of *Gossypium* genomes and the evolution of spinnable cotton fibres. *Nature* 492, 423–427. doi: 10.1038/nature11798
- Poehlman, J. M. (1994). *Breeding Field Crops*, 3rd Edn. Ames, IA: Iowa State University Press.
- Said, J. I., Knapka, J. A., Song, M., and Zhang, J. (2015a). Cotton QTLdb: a cotton QTL database for QTL analysis, visualization, and comparison between *Gossypium hirsutum* and *G. hirsutum* × *G. barbadense* populations. *Mol. Genet. Genomics* 290, 1615–1625. doi: 10.1007/s00438-015-1021-y
- Said, J. I., Song, M., Wang, H., Lin, Z., Zhang, X., Fang, D. D., et al. (2015b). A comparative meta-analysis of QTL between intraspecific *Gossypium hirsutum* and interspecific *G. hirsutum* × *G. barbadense* populations. *Mol. Genet. Genomics* 290, 1003–1025. doi: 10.1007/s00438-014-0963-9
- Said, J. I., Lin, Z., Zhang, X., Song, M., and Zhang, J. (2013). A comprehensive meta QTL analysis for fiber quality, yield, yield related and morphological traits, drought tolerance, and disease resistance in tetraploid cotton. *BMC Genomics* 14:776. doi: 10.1186/1471-2164-14-776

- Sestili, F., Pagliarello, R., Zega, A., Saletti, R., Pucci, A., Botticella, E., et al. (2019). Enhancing grain size in durum wheat using RNAi to knockdown GW2 genes. *Theor. Appl. Genet.* 132, 419–429. doi: 10.1007/s00122-018-3229-9
- Song, X. J., Huang, W., Shi, M., Zhu, M. Z., and Lin, H. X. (2007). A QTL for rice grain width and weight encodes a previously unknown RING-type E3 ubiquitin ligase. *Nat. Genet.* 39, 623–630. doi: 10.1038/ng2014
- Su, Z., Hao, C., Wang, L., Dong, Y., and Zhang, X. (2011). Identification and development of a functional marker of TaGW2 associated with grain weight in bread wheat (*Triticum aestivum* L.). *Theor. Appl. Genet.* 122, 211–223. doi: 10.1007/s00122-010-1437-z
- Valluru, R., Reynolds, M. P., and Salse, J. (2014). Genetic and molecular bases of yield-associated traits: a translational biology approach between rice and wheat. *Theor. Appl. Genet.* 127, 1463–1489. doi: 10.1007/s00122-014-2332-9
- Voorrips, R. E. (2002). MapChart: software for the graphical presentation of linkage maps and QTLs. *J. Hered.* 93, 77–78. doi: 10.1093/jhered/93.1.77
- Wang, K., Wang, Z., Li, F., Ye, W., Wang, J., Song, G., et al. (2012). The draft genome of a diploid cotton *Gossypium raimondii*. *Nat. Genet.* 44, 1098–1103. doi: 10.1038/ng.2371
- Wang, X. Y. (2008). Effect of seed size on the plant growth at seedling stage in cotton. *J. Anhui Agric. Sci.* 36, 9939–9940.
- Wang, Y., Tang, H., Debarry, J. D., Tan, X., Li, J., Wang, X., et al. (2012). MCScanX: a toolkit for detection and evolutionary analysis of gene synteny and collinearity. *Nucleic Acids Res.* 40:e49. doi: 10.1093/nar/gkr1293
- Wang, Z., Li, N., Jiang, S., Gonzalez, N., Huang, X., Wang, Y., et al. (2016). SCF(SAP) controls organ size by targeting PPD proteins for degradation in *Arabidopsis thaliana*. *Nat. Commun.* 7:11192. doi: 10.1038/ncomms11192
- Xia, T., Li, N., Dumenil, J., Li, J., Kamenski, A., Bevan, M. W., et al. (2013). The ubiquitin receptor DA1 interacts with the E3 ubiquitin ligase DA2 to regulate seed and organ size in *Arabidopsis*. *Plant Cell* 25, 3347–3359. doi: 10.1105/tpc.113.115063
- Xing, Y., and Zhang, Q. (2010). Genetic and molecular bases of rice yield. *Annu. Rev. Plant Biol.* 61, 421–442. doi: 10.1146/annurev-arplant-042809-112209
- Yamaguchi, K., Yamamoto, T., Segami, S., Horikawa, M., Chaya, G., Kitano, H., et al. (2020). gw2 mutation increases grain width and culm thickness in rice (*Oryza sativa* L.). *Breed. Sci.* 70, 456–461. doi: 10.1270/jsbbs.20018
- Yang, S., Huang, L., Song, J., Liu, L., Bian, Y., Jia, B., et al. (2021). Genome-wide analysis of DA1-like genes in *Gossypium* and functional characterization of GhDA1-1A controlling seed size. *Front. Plant Sci.* 12:647091. doi: 10.3389/fpls.2021.647091
- Yu, J., Jung, S., Cheng, C. H., Ficklin, S. P., Lee, T., Zheng, P., et al. (2014). CottonGen: a genomics, genetics and breeding database for cotton research. *Nucleic Acids Res.* 42, D1229–D1236. doi: 10.1093/nar/gkt1064
- Zhao, W., Yan, Q., Yang, H., Yang, X., Wang, L., Chen, B., et al. (2019). Effects of mepiquat chloride on yield and main properties of cottonseed under different plant densities. *J. Cotton Res.* 2:10. doi: 10.1186/s42397-019-0026-1
- Zhu, T., Liang, C., Meng, Z., Sun, G., Meng, Z., Guo, S., et al. (2017). CottonFGD: an integrated functional genomics database for cotton. *BMC Plant Biol.* 17:101. doi: 10.1186/s12870-017-1039-x

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's Note: All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

Copyright © 2022 Huang, Yang, Wu, Xin, Song, Wang, Pei, Wu, Yu, Ma and Hu. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.