Title page

Title:

Genome-wide characterization and expression analysis of PP2CA family members in response to ABA and osmotic stress in *Gossypium*

Authors:

Tingting Lu[1,2†], Gaofeng Zhang[1†], Yibin Wang[1], Shibin He[1], Lirong Sun[1*] and Fushun Hao[1*]

[1]State Key Laboratory of Cotton Biology, Henan Key Laboratory of Plant Stress Biology, School of Life Sciences, Henan University, Kaifeng 475004, China

[2]Henan University of Animal Husbandry and Economy, Zhengzhou 450046, China

[†]These authors contributed equally to this work

Tingting Lu: 4147505@qq.com

Gaofeng Zhang: 2604035350@qq.com

Yibin Wang: 1532231014@qq.com

Shibin He: Hsb108@163.com

[*]Corresponding author

Lirong Sun: sunlr9208@henu.edu.cn

Fushun Hao: haofsh@henu.edu.cn

Tel. 86-371-23881387; Fax 86-371-23881387

Genome-wide characterization and expression analysis of PP2CA family members in response to ABA and osmotic stress in *Gossypium*

Tingting Lu[1,2†], Gaofeng Zhang[1†], Yibin Wang[1], Shibin He[1], Lirong Sun[1*] and Fushun Hao[1*]

[1]State Key Laboratory of Cotton Biology, Henan Key Laboratory of Plant Stress Biology, College of Life Sciences, Henan University, Kaifeng 475004, China

[2]Henan University of Animal Husbandry and Economy, Zhengzhou 450046, China

ABSTRACT

Clade A type 2C protein phosphatases (PP2CAs), as central regulators of abscisic acid (ABA) signaling, negative control growth, development and responses to multiple stresses in plants. PP2CA gene families have been characterized at genome-wide levels in several diploid plants like *Arabidopsis* and rice. However, the information about genome organization, phylogenesis and putative functions of PP2CAs in *Gossypium* is lacking. Here, PP2CA family members were comprehensively analyzed in four *Gossypium* species including the diploid progenitor species *G. arboretum*, *G. raimondii* and the tetraploid species *G. hirsutum* and *G. barbadense*, and 14, 13, 27 and ~~29~~ 23 PP2CA genes were identified in the genomic~~es~~ sequences of these plants, respectively. Analysis results showed that most *Gossypium* PP2CAs were ~~localized in the nucleus, and the *PP2CAs* were~~ highly conserved in ~~physical properties,~~ chromosomal locations, structures and phylogeny among the four cotton species.

Formatted: Font: Not Italic

Segmental duplication might play~~ed~~ important roles in the formation of the PP2CAs, and most PP2CAs may be~~were~~ under purifying selection in *Gossypium* during evolution. ~~Moreover,~~The majority of the PP2CAs were expressed specifically in diverse tissues, and highly expressed in flowers in *G. hirsutum*. The *GhPP2CAs* displayed diverse expression patterns in responding to ABA and osmotic stress. ~~Besides, y~~Yeast-two hybrid assays revealed that many *GhPP2CAs* were capable of interaction with the cotton ABA receptors pyrabactin resistance1/PYR1-like/regulatory components of ABA receptors (PYR1/PYL/RCAR) *GhPYL2-2D* (*Gh_D08G2587*), *GhPYL6-2A* (*Gh_A06G1418*) and *GhPYL9-2A* (*Gh_A11G0870*) , key regulators of ABA signaling, in the presence and/or absence of ABA. These results gave a comprehensive view of the *Gossypium* PP2CAs~~,~~ and are valuable for further studying the functions of PP2CAs in *Gossypium*.

Keywords *Gossypium*; Clade A type 2C protein phosphatases (PP2CAs); gene family; phylogeny; protein interaction

INTRODUCTION

Protein phosphorylation and dephosphorylation, as two central mechanisms of cellular signal transduction, play pivotal roles in many biological processes including growth, development and adaptations to various environmental stimuli in plants (Schweighofer et al., 2004). They are catalyzed by protein kinases and phosphatases, respectively. Phosphatases are generally categorized into serine/threonine (Ser/Thr) phosphatases and tyrosine (Tyr) phosphatases according to the different amino acid residues they dephosphorylate. Based on biochemical and pharmacological properties,

Formatted: Font: Not Italic
Formatted: Font: Not Italic
Formatted: Font: Italic
Formatted: Not Highlight
Formatted: Not Highlight
Formatted: Font: Italic
Formatted: Font: Italic
Formatted: Font: Italic
Formatted: Font: Italic
Formatted: Font: Italic
Formatted: Font: Italic

Ser/Thr phosphatases can be further classified into three large families: phosphoprotein phosphatases (PPs), phosphoprotein metallophosphatases and aspartate-based protein phosphatases (Schweighofer et al., 2004; Kerk et al., 2007; Fuchs et al., 2013; Singh et al., 2015). The PPs includes PP1, PP2A, PP2B, PP4, PP5, PP6 and PP7, the phosphoprotein metallophosphatases consist of Mg^{2+}/Mn^{2+}-dependent type 2C protein phosphatases (PP2Cs) and other Mg^{2+}-dependent phosphatases (Schweighofer et al., 2004; Singh et al., 2010; Fuchs et al., 2013; Singh et al., 2015). PP2Cs, which play key roles in dephosphorylation events in plants, belong to a large subfamily, and can be further divided into 11 clades (A–K) in *Arabidopsis* and rice (Singh et al., 2010) and 12 clades (A-L) in *Brachypodium distachyon* (Cao et al., 2016). Among these, ~~clade A proteins (~~PP2CAs~~)~~ are the ones of well-studied PP2Cs in *Arabidopsis*, and they have been shown to have important roles in ~~are of most importance. They~~ controlling abscisic acid (ABA) signaling, and negatively regulate plant growth, development and response to various biotic and abiotic stresses in plants (Tähtiharju et al., 2001; Fuchs et al., 2013; Singh et al., 2015). In *Arabidopsis* genome, 9 *PP2CA* members have been identified. They are *ABI1* (*ABA insensitive 1*), *ABI2*, *HAB1* (*Homology to ABI1*), *HAB2*, *AHG1* (*ABA hypersensitive germination 1*), *HAI1* (*Highly ABA-induced PP2C ~~3~~1*), *HAI2*, *HAI3* and *AHG3*/*AtPP2CA* (Fuchs et al., 2013). These genes, particularly *ABI1*, *ABI2 and AHG3/PP2CA* alone or cooperatively control ABA-mediated transpiration, stomatal closure, seed germination and root growth, and are involved in the regulation of many abiotic stress responses like drought, high salinity, cold, heat and potassium

Formatted: Font: Times New Roman

Formatted: Font: Times New Roman, Italic

Formatted: Font: Times New Roman

Formatted: Font: Not Italic

Formatted: Font: Italic

deprivation (Schweighofer et al., 2004; Rubio et al., 2009; Singh et al., 2015). Some PP2CAs also play important roles in responses to pathogen attack (Schweighofer et al., 2004; Singh et al., 2015). PP2CAs are functionally redundant, and their expression is upregulated by high concentrations of ABA (Rubio et al., 2009; Singh et al., 2015). Moreover, PP2CAs physically interact with numerous cytosolic and nuclear localized proteins such as AtHB6 (Homeobox protein 6), CIPK8 (Calcineurin B–like protein-interacting protein kinase 8), CIPK24, and SnRK2s (Sucrose nonfermenting 1-related protein kinases subfamily 2 proteins) (Ohta et al., 2003; Fuchs et al., 2013; Singh et al., 2015). SnRK2s exert central and positive roles in ABA signal cascade in plants (Fujii et al., 2009a; Fujita et al., 2009).

Recently, ABA receptors pyrabactin resistance1/PYR1-like/regulatory components of ABA receptors (PYR1/PYL/RCAR) (named PYLs for simplicity) have been found (Ma et al., 2009; Park et al., 2009). This is a breathtaking discovery in plants. PP2CAs were identified as co-receptors, specifically interact with PYLs and control ABA signaling. In the presence of ABA, ABA binds to PYLs, further interacts with and inhibits the activities of PP2CAs; thereby releasing and activating SnRK2s. SnRK2s subsequently regulate multiple downstream transcriptional factors and other proteins to trigger ABA responses (Fujii et al., 2009b; Geiger et al., 2009~~Singh et al., 2015~~; Lee et al., 2009~~; Zhang et al., 2017b~~).

Currently, PP2C gene families including PP2CAs have been analyzed at genome-wide levels in *Arabidopsis*, rice, maize and *Brachypodium distachyon* (Xue et al., 2008; Wei and Pan, 2014; Cao et al., 2016). The domain structure of PP2CAs

Formatted: Font: Not Italic

Formatted: Font: Not Italic

Formatted: Not Highlight

Formatted: Font: Not Bold

Formatted: Font: Not Italic

was also studied (Schweighofer et al., 2004). Moreover, the expression patterns of PP2CAs have been examined in response to ABA and multiple stresses in *Arabidopsis*, rice, maize and *Brachypodium distachyon* (Xue et al., 2008; Wei and Pan, 2014; Zhang~~Cao~~ et al., ~~2016~~2017a). However, knowledge about the genomic information and expression profiles of PP2CAs in cotton is unknown to date.

Cotton is the most important fiber crop, which provides the spinnable lint for the textile industry in the world. The yield and quality of cotton are adversely affected by many abiotic stresses such as drought and high salinity, which are governed by ABA signaling (Hauser et al., 2011; Liang et al., 2017; Ullah et al., 2017). Therefore, it is essential for us to uncover the functional mechanisms of PP2CAs in ABA signal transduction pathway in cotton. Here, we carried out a genome-wide identification of PP2CA gene family in diploid *G. arboretum* (A2) and *G. raimondii* (D5), and their decendant tetraploid species *G. hirsutum* (AD1) and *G. barbadense* (AD2). The evolutionary relationships of these PP2CAs were analyzed. Changes in the transcriptional levels of the PP2CAs were also investigated in diverse tissues and in response to ABA and osmotic stress in *G. hirsutum*. Furthermore, the interactions between *G. hirsutum* PP2CAs and several *GhPYLs* were detected by the yeast-two hybrid method. These results may be valuable for further functional characterization of cotton PP2CAs in ABA signaling in the future.

Formatted: Font: Not Italic

Formatted: Font: Times New Roman, 12 pt

Formatted: Font: Times New Roman, 12 pt

Formatted: Font: Times New Roman, 12 pt

Formatted: Font: Not Italic

Formatted: Font: Italic

MATERIALS AND METHODS

Analysis of the PP2C family in four *Gossypium* species

To explore all the members of the PP2C family in *Gossypium*, the protein sequences of 80 AtPP2Cs were initially applied as queries to ~~survey~~ search against the databases of *G. arboreum* (BGI-CGB v2.0 assembly genome), *G. raimondii* (JGI assembly v2.0 data.), *G. hirsutum* (NAU-NBI v1.1 assembly genome) (www.cottongen.org) and *G. barbadense* (http://database.chgc.sh.cn/cotton/index.html), respectively, using the BLAST program with default setting (E-value<e^{-10}) (Camacho et al., 2009). After removing the redundant sequences from the data set, the putative *Gossypium* PP2Cs were then characterized using the PP2C model (PF00481) (http://pfam.xfam.org/) by the Hmmer software (http://hmmer.org/), and the proteins without a PP2C catalytic domain were deleted. The molecular weight (MW) and the isoelectric point (pI) of *Gossypium* were predicted by the online tool ExPaSy (http://web.expasy.org/protparam/), which can give various physico-chemical properties of a protein based on its amino acid sequence (the extinction coefficient and the absorbance of a native protein in water at 280 nm were used). ~~was applied to analyze the properties of PP2CAs in *Gossypium*.The subcellular localizations of *Gossypium* PP2CAs were predicted using the program in the website (http://www.csbio.sjtu.edu.cn/bioinf/Cell-PLoc/).~~ The composition and position of exons and introns of the PP2CAs were obtained from the website (https://www.cottongen.org/) and characterized by the Gene Structure Display Server (GSDS) tools (http://gsds.cbi.pku.edu.cn/) (Hu et al., 2015). The conserved domains of PP2CAs were validated in NCBI (https://www.ncbi.nlm.nih.gov/Structure/cdd/wrpsb.cgi) using the automatic mode

Formatted: Font: Italic
Formatted: Font: Not Bold
Formatted: Font: Not Bold
Formatted: Not Highlight
Formatted: Not Highlight
Formatted: Not Highlight
Formatted: Not Highlight
Formatted: Font color: Blue

(Marchler-Bauer et al., 2017). ~~and SMART (http://smart.embl-heidelberg.de/).~~ The MEME program (meme-suite.org/tools/meme) was applied to determine the motifs of PP2CAs in *Gossypium* ('any number of repetitions' to be distributed in sequences was set). The locations of *Gossypium PP2CAs* in chromosomes were assessed using the MapInspect software (http://www.mybiosoftware.com/mapinspect-compare-display-linkage-maps.html).

Analysis of synteny and *Ka*/*Ks* of *PP2CAs*

The homologous regions of PP2CAs in *Gossypium* were identified by the MCScanx software (http://chibba.pgml.uga.edu/mcscan2/), and syntenic blocks were determined by the CIRCOS program (http://www.circos.ca/). The syntenic maps of the *PP2CAs* were obtained using the circos-0.69±3 software with default parameters (http://www.circos.ca/). Some genes located within the same or adjacent intergenic region were regarded as tandem duplications. The nucleotide substitution parameters *Ka* (non-synonymous) and *Ks* (synonymous) were assessed by the PAML program (http://abacus.gene.ucl.ac.uk/software/paml.html). Then, the ratio of *Ka*/*Ks* was calculated. *Ka*/*Ks*<1 means purifying selection; *Ka*/*Ks*=1 indicates neutral selection, while *Ka*/*Ks*>1 represents positive selection.

Phylogenetic analysis of *PP2CAs*

~~We downloaded t~~The PP2CA databases were downloaded for *Arabidopsis thaliana* (http://www.arabidopsis.org/), *Theobroma cacao* (http://cocoagendb.cirad.fr), *Ricinus communis* (http://castorbean.jcvi.org), *Populus trichocarpa*

Formatted: Font: Italic
Formatted: Font: Italic
Formatted: Font: Not Italic
Formatted: Font: Italic
Formatted: Font: Italic
Formatted: Font: Italic
Formatted: Font: Italic
Formatted: Font: Italic
Formatted: Font: Italic
Formatted: Font: Italic
Commented [JEF1]: Is there a reference for this?
Formatted: Font: Italic
Formatted: Font: Italic

(http://www.phytozome.net/poplar), *Glycine max* (http://www.phytozome.net/soybean), *Brachypodium distachyon* (http://plants.ensembl.org/Brachypodium_distachyon/Info/Index), *Oryza sativa* (http://rapdb.dna.affrc.go.jp), and the four *Gossypium* species mentioned above. The amino acid sequences of PP2CAs were aligned by the MUSCLE software (Edgar et al., 2004), and a phylogenetic tree of the PP2CAs was generated using the IQ-TREE server (http://www.iqtree.org/) following the ~~neighbor joining~~maximum likelihood ~~method~~ (~~Neighbor-Joining, NJ~~ML) method (Nguyen et al., 2015; Trifinopoulos et al., 2016). The ModelFinder, tree reconstruction and ultrafast bootstrap were used (Minh et al., 2013). The evolutionary tree was redrawn by the FigTree v1.4.4 software.~~One thousand bootstrap trials with the Clustal W tool (Larkin et al., 2007) and the MEGA 5.0 software (http://www.megasoftware.net/) were used.~~

Measurements of *GhPP2CAs* expression in tissues and in response to ABA or osmotic stress

For measuring the expression of *GhPP2CAs* in tissues in each experiment, about 2 g samples of roots, stems or leaves were collected from 10 *G. hirsutum* L. acc. Texas Marker-1 (TM-1) plants grown in soil for 21 d. About 20 flowers were ~~got~~ collected 1 d post anthesis, and about 5 g fibers were obtained from ovules 23 d post anthesis. For monitoring the expression of *GhPP2CAs* after ABA treatment or under osmotic stress, three-week-old TM-1 plants grown in liquid 1/2 MS medium (Murashige and Skoog, 1962) in a growth chamber (day/night temperature cycle of 28°C/26°C, 14 h light/10 h dark, and about 50% relative humidity) were sprayed with 100 μM ABA or treated

Formatted: Font: (Default) Times New Roman

Formatted: Font: (Default) Times New Roman

Formatted: Font: (Default) Times New Roman, No underline

Formatted: Font: (Default) Times New Roman, No underline

Formatted: Font: Not Bold, No underline

Formatted: Font: Not Bold, No underline

Formatted: Font: Not Bold, No underline

Formatted: Font: Italic

Formatted: Font: Italic

Formatted: Font: Italic

Commented [JEF2]: What is the estimated age of the plants when you started the experiment? It sounds like they are only 21 days old and normally they only have a few leaves and could not produce the necessary flowers. Were the seeds planted in soil and 21 days later you started collecting leaves, roots, stems? Once stem is collected that pretty much kills a 21 day old plant and can not produce flowers. Were seeds sown on MS medium and then 21 days later sprayed with ABA or treated with PEG? Please clarify more the details of the plant establishment and plant age of sampling.

Formatted: Not Highlight

with 10% PEG6000 (dissolved in medium) for 0, 3, 6, 12 and 24 h, respectively. Then, about 2 g roots were sampled, frozen in liquid nitrogen and stored at -70℃. Total RNA was extracted from some of the samples and cDNA was generated according to the method described previously (Ma et al., 2012; Zhang et al., 2017b).

Quantitative real-time RT-PCR (qRT-PCR) experiments were constructed in an ABI 7500 real-time PCR amplifier using the cDNA, SYBR Green Master mix, the specific primers of *GhPP2CA* genes (Table S1). *GhUBQ7* was used as an internal control (Lu et al., 2017). Experiments were independently repeated three times. The interval between two repeated experiments was 7-10 d.

Monitoring protein interaction by yeast two-hybrid method

The CDS sequences of *GhPYLs* (*GhPYL2-2D*, *GhPYL6-2A*, and *GhPYL9-2A*) and *GhPP2CAs* were amplified, and cloned into pGADT7 and pGBKT7 vectors respectively using gene specific primers (Table S2). After sequencing, the fused vectors were transformed into AH109. The cotransformants were plated on non-selective SD/-Leu/-Trp solid medium and selective SD/-Leu/-Trp/-His/-Ade solid medium as described previously (Lu et al., 2017; Zhang et al., 2017b).

RESULTS

Genome-wide analysis of *PP2CAs* in four *Gossypium* species

To identify the putative PP2CA family members in *Gossypium*, the amino acid sequences of 80 *Arabidopsis* PP2Cs (Xue et al., 2008) were used to survey the *Gossypium* dababases. Putative PP2CA gene function was assigned to a~~A~~ total of 114,

116, 239 and 232 genomic sequences that ~~PP2Cs~~ were retrieved from *G. arboretum*, *G. raimondii*, *G. hirsutum* and *G. barbadense*, respectively. They were individually denominated as *GaPP2Cs*, *GrPP2Cs*, *GhPP2Cs* and *GbPP2Cs* (Table S3). According to the phylogenetic relationships of PP2Cs between *Gossypium* and *Arabidopsis*, the *Gossypium* PP2Cs could be clustered into 12 clades (A-L) (Fig. S1-S4). The PP2CAs possessed 14 *GaPP2CAs*, 13 *GrPP2CAs*, 27 *GhPP2CAs* and 2~~3~~9 *GbPP2CAs*, respectively. They ~~GaPP2CAs~~ were named individually according to their gene identifiers ~~their similarity of amino acid sequences to the 9 AtPP2CAs. The nomenclature of GaPP2CAs was similar to that of *arboretum* CBL proteins (GaCBLs) (Lu et al., 2017). The GrPP2CAs, GhPP2CAs and GbPP2CAs were named based on their phylogenetic relationships with GaPP2CAs~~ (Table 1). In this report, we focused on the PP2CA family members in the four *Gossypium* species.

It was found that ~~Most PP2CAs from the four *Gossypium* species shared similar physical properties.~~the predicted coded amino acid lengths of *Gossypium* PP2CAs ranged from 118 to 593, with an average of ~~413~~420. These PP2CAs had molecular weights of 12.8 kDa (*Gh*PP2CA27~~HAI3-2~~) to 66 kDa (*GaPP2CA6*~~GaABI2-2~~). The mean theoretical pIs of PP2CAs was approximately 5.9 with a minimum of 4.65 (*GrPP2CA5*~~GrHAB1-2~~) and a maximum of 8.74 (*GhPP2CA22*~~GhABI2-3D~~) (Table 1).~~ The PP2CAs were predicted to locate in the nucleus except some members distributing in the chloroplast (GaABI2-2, GaHAB1-1, GrHAB1-1, GhHAB1-1A, GhHAB1-1D, GbHAB1-1D', GbHAB1-1D) and the mitochondrion (GhHAI3-1D, GhHAB1-1A).~~

Formatted: Font: Italic
Formatted: Font: Italic
Formatted: Font: Italic
Formatted: Font: Italic
Formatted: Font: Italic
Formatted: Font: Italic
Formatted: Font: Italic
Formatted: Font: Italic
Formatted: Font: Italic
Formatted: Font: Italic
Formatted: Font: Italic
Formatted: Font: Italic

Phylogenetic and structural analysis of PP2CAs in *Gossypium*

In order to understand the evolutionary relationship among *GaPP2CAs*, *GrPP2CAs*, *GhPP2CAs* and *GbPP2CAs*, we conducted a phylogenetic tree using the protein sequences of the *Gossypium* PP2CAs (Fig. 1A). As expected, most of *GaPP2CAs* were individually clustered closely with their corresponding orthologs of *GhPP2CAs* (*GhPP2CA1-13*) and *GbPP2CAs* (*GbPP2CA1-10*) in A genomes, and a majority of GrPP2CAs individually clustered closely with their homologs of *GhPP2CAs* (*GhPP2CA14-27*) and *GbPP2CAs* (*GbPP2CA11-23*) in D genomes. Noteworthily, *GaPP2CA10*~~AHG3-2~~ clustered together with *GaPP2CA11*~~AHG3-2'~~, and a similar case occurred between *GbPP2CA21*~~ABI2-1D and~~ /*GbPP2CA22*~~ABI2-1D'~~ and *GbPP2CA16/GbPP2CA23*. Moreover, homologues of 14~~3~~ *GaPP2CAs* ~~(except GaAHG3-2')~~ and of 13 *GrPP2CAs* were found in the *G. hirsutum* At and Dt subgenomes, respectively; and homologs of 10~~2~~ *GaPP2CAs* (except *GaPP2CA1*, *GaPP2CA3*, *GaPP2CA4* and ~~GaAHG3-2' and~~ *GaPP2CA14*~~ABI2-3~~) and 11~~2~~ *GrPP2CAs* (except *GrPP2CA6*~~ABI2-3~~ and *GrPP2CA9*) were detected in the *G. barbadense* At' or Dt' subgenomes, respectively. ~~One GhPP2CA (GhHAI3-2) and 6 GbPP2CAs (GbHAI3-2A, GbHAI3-2D, GbHAI3-3A', GbHAI3-3D, GbHAI3-3A, GbABI1-3D) had not orthologs in *G. arboretum* and *G. raimondii*.~~ Additionally, three pairs of paralogues with high sequence similarity in *GbPP2CAs* including *GbPP2CA8/GbPP2CA10*, *GbPP2CA16/GbPP2CA23*, *GbPP2CA21/GbPP2CA22* were clustered together. ~~were observed in GbPP2CAs. They were GbHAB1-1D/GbHAB1-1D', GbABI2-1D/GbABI2-1D', GbHAB1-2A/GbHAB1-2A',~~

Formatted: Font: Italic
Formatted: Font: Italic
Formatted: Font: Italic
Formatted: Font: Italic
Formatted: Font: Italic
Formatted: Font: Italic
Formatted: Font: Italic
Formatted: Font: Italic
Formatted: Font: Italic
Formatted: Font: Italic
Formatted: Font: Italic
Formatted: Font: Italic
Formatted: Font: Italic
Formatted: Font: Italic
Formatted: Font: Italic
Formatted: Font: Italic
Formatted: Font: Italic
Formatted: Font: Italic
Formatted: Font: Italic
Formatted: Font: Italic
Formatted: Font: Italic
Formatted: Font: Italic
Formatted: Font: Italic
Formatted: Font: Italic
Formatted: Font: Italic
Formatted: Font: Italic
Formatted: Font: Italic
Formatted: Font: Italic
Formatted: Font: Italic
Formatted: Font: Italic
Formatted: Font: Italic
Formatted: Font: Italic

GbHAI3-3A/GbHAI3-3A'; andThey were seemingly derived from GrHAB1-GaPP2CA81, GrPP2CA7 and *GrPP2CA1*ABI2-1, *GaHAB1-2 and GaHAI3*, respectively.

Most PP2CAs had 3-4 exons except that *GbAHG1D-GaPP2CA6*, *GrPP2CA9*, *GhPP2CA22*, *GbPP2CA13* possessed 5 exons, and *GhPP2CA27HAI3-2* and *GbABI1-1D* had 2 exons. Among the PP2CA genes, *GaPP2CA6ABI2-2*, *GrPP2CA9ABI2-3*, *GhPP2CA9ABI2-3A* and *GhPP2CA22ABI2-3D and GhHAI3-2* individually had a longer intron sequence than other genes did (Fig. 1B). These results indicate that the exon/intron structures of the *Gossypium* PP2CA genes were highly conserved.

The motif compositions of the PP2CA proteins were analyzed in the four *Gossypium* species. Twenty putative motifs named motif 1 to motif 20 were identified. Among those, motif 1, 2, 3, 4, 5, 6 and 67, 8 and 12 existed in every cluster and the majority of the PP2CA members. Moreover, most orthologous PP2CA proteins in the four *Gossypium* plants had the same or very similar compositions and distributions of motifs, suggesting that the PP2CA members in the same cluster likely share similar functions (Fig. 1C).

Chromosomal distributions of *PP2CAs* in *Gossypium*

To determine the putative evolutionary relationships of the *Gossypium* PP2CA genes, We analyzed the positions of the genes on chromosomes were analyzed. We found that the distributions of these PP2CAs were uneven. The 14 *GaPP2CAs*, 13 *GrPP2CAs*, 27 *GhPP2CAs* and 29 23 *GbPP2CAs* were distributed on 8, 9, 16 and 146

Formatted: Font: Not Italic, No underline
Formatted: Font: Not Italic, No underline
Formatted: No underline
Formatted: Font: Italic
Formatted: Font: Not Italic, No underline
Formatted: Font: Not Italic
Formatted: Font: Italic
Formatted: Font: Italic
Formatted: Highlight
Formatted: Not Highlight
Formatted: Not Highlight
Formatted: Font: Italic
Formatted: Not Highlight
Formatted: Font: Not Italic

chromosomes, respectively. Most of the chromosomes contained one PP2CA gene. By contrast, some chromosomes individually had two PP2CA genes. They were D11 and D13 in *G. raimondii*, At10, At13, Dt05, Dt10 and Dt13 in *G. hirsutum*, and Dt'07 and Dt'13 in *G. barbadense*. Besides, each of D09 and Dt'05 owned 3 PP2CAs. A13, At05 and At'05 separately possessed 4 *PP2CAs*. ~~and At'05 contained 5 *PP2CAs*.~~ In contrast, *GaPP2CA5*~~*ABI2-1*~~, *GaPP2CA8*~~*HAB1-2*~~, *GaPP2CA14*~~*ABI2-3*~~, *GhPP2CA18*~~*HAI3-2*~~, *GhPP2CA20*~~*HAI2D*~~, *GhPP2CA27*~~*ABI2-1D*~~, *GbPP2CA2*~~*AHG3-2A*~~ and *GbPP2CA12*~~*AHG3-1D*~~ were located on scaffolds, ~~rather than chromosomes~~in which contigs were not spliced into any chromosome in genomic mapping.

We compared the positions of the orthologs of *GaPP2CAs*, *GrPP2CAs* and *GhPP2CAs* or *GbPP2CAs* in chromosomes. As expected, most homologs of *GaPP2CAs* and *GrPP2CAs* in *G. hirsutum* were located in their corresponding At subgenomes and Dt subgenomes. The similar situation also occurred in *G. barbadense*. However, a few homologous genes of *GaPP2CAs* and *GrPP2CAs* were not ~~barely~~ located in their expected corresponding homoeologous chromosomes and collinear loci in *G. hirsutum* and *G. barbadense* (Fig. 2). These results imply that specific, unique, and complex variation events in PP2CA-contained homoeologous chromosomes may happen ~~among~~ within each of the two diploid and tetraploid species during genetic evolution.

Synteny analysis of *PP2CA* genes

During evolutionary processes, tandem and segmental duplications contribute to

Formatted: Font: Not Italic

Commented [JEF3]: It suggest to me that one could hope to focus more closely on the unique ones and simplify comparisons between species, or within a species for association with valuable responses or traits.

expanding gene family in plants (Cannon et al., 2004). We ~~therefore~~ examined the duplication relationship of the PP2CAs among *G. arboretum*, *G. raimondii* and *G. hirsutum* (the related database for *G. barbadense* was lacking). It was found that *GaPP2CA10*~~*AHG3-2*~~ and *GaPP2CA11*~~*AHG3-2'*~~ joined together, and ~~*GbPP2CA16GbABI1-1D*~~ and *GbPP2CA23*~~*ABI1-3D*~~ clustered together in the chromosome. There are less than 5 genes between each pair of the genes, suggesting that the two pairs of genes are tandemly duplicated.

The synteny relationship of gene pairs was also explored among *GaPP2CAs*, *GrPP2CAs* and *GhPP2CAs*. A total of 136 homologous gene pairs were observed in 133 collinearity blocks. Most of the blocks had one gene pair~~s~~. Some blocks owned two gene pairs (*GrPP2CA9*~~*ABI2-3*~~*/GhPP2CA22*~~*ABI2-3D*~~, *GrPP2CA10*~~*AHG3-2*~~*/GhPP2CA23*~~*AHG3-2D*~~) between chromosomal D11 and Dt10. Another block harbored three gene pairs (*GrPP2CA5*~~*HAB1-2*~~*/GhPP2CA2*~~*HAB1-2A*~~, *GrPP2CA6*~~*HAI2*~~*/GhPP2CA3*~~*HAI2A*~~, *GrPP2CA7*~~*HAB1-1*~~*/GhPP2CA4*~~*HAB1-1A*~~) between chromosomal D09 and At05 (Fig. 3). These findings imply that segmental duplication plays major roles in generating PP2CAs during evolution in *Gossypium*.

Analysis of *Ka*/*Ks* values of *PP2CAs*

To further understand the evolution processes among *Gossypium* PP2CAs, the effects of selection on duplication of PP2CA genes were determined. The non-synonymous (*Ka*) and synonymous (*Ks*) substitutions, and *Ka*/*Ks* values were calculated for the homologous gene pairs among *GaPP2CAs*, *GrPP2CAs* and *GhPP2CAs*. The mean values of *Ka*/*Ks* for these gene pairs between species Ga/Gh, Gr/Ga, Gr/Gh, Ga/Ga,

Formatted: Font: Not Italic

Formatted: Font: Italic

Formatted: Font: Not Italic

Formatted: Font: Italic

Formatted: Font: Italic

Formatted: Font: Italic

Formatted: Font: Italic

Formatted: Font: Italic

Formatted: Font: Italic

Formatted: Font: Italic, No underline, Font color: Auto

Formatted: Font: Italic, No underline, Font color: Auto

Formatted: No underline, Font color: Auto

Formatted: Font: Italic, No underline, Font color: Auto

Gr/Gr, Gh/Gh were 0.22, 0.21, 0.22, 0.21, 0.19 and 0.21, respectively. All of them were less than 1, indicating that the formation of these genes ~~were~~ was mainly under purifying selection during evolution. The *Ka*/*Ks* ratios for the two gene pairs *Gr*PP2CA11~~AHG3-1~~/*Gh*PP2CA15~~AHG3-1D~~ and *Gr*PP2CA3~~ABI2-2~~/*Gh*PP2CA21~~ABI2-2D~~ were higher than 1, suggesting that the two gene pairs ~~may~~ were generated under positive selection (Fig. 4).

Phylogenetic analysis of PP2CAs in *Gossypium* and other plants

We constructed a phylogenetic tree of PP2CA proteins in *G. arboretum*, *G. raimondii*, *G. hirsutum*, *G. barbadense*, *A. thaliana*, *T. cacao*, *R. communis*, *P. trichocarpa*, *G. max*, *B. distachyon* and *O. sativa* using the maximum likelihood (ML)~~neighbor-joining~~ method, and analyzed the evolutionary relationships of these PP2CAs. It was found that the PP2CAs included both dicotyledonous and monocotyledonous members (Fig. 5). This suggests that these PP2CA formed before the divergence of eudicots and monocots and are in general very highly conserved. Indeed, the PP2CAs from the eudicots *Gossypium*, cocao, poplar, castor, soybean and *Arabidopsis* clustered more closely, and those of the monocots rice and *distachyon* clustered together. Moreover, many PP2CAs from *Gossypium* clustered more closely with those from cocao than from poplar, castor, soybean and *Arabidopsis* (Fig. 5), indicating that PP2CAs of *Gossypium* had closer relationship with those of cacao than those of other plants. As expected, PP2CAs in the four *Gossypium* species always clustered together, in line with their homologous evolutionary relationships (Fig. 5).

Expression patterns of *GhPP2CA* genes in different tissues

Formatted: Font: Italic, No underline

Formatted: No underline

Formatted: Font: Italic, No underline

Formatted: Font: Italic

Formatted: Font: Italic

Formatted: Font: Italic

Formatted: Font: Italic

Commented [JEF4]: More discussion could involve the Ka/Ks even if it is just speculation with limited results. I would expect there to be a difference in the ratios or in the terms purifying and positive selection when comparing wild versus cultivated cotton species. Cultivated plants are selected to yield in different environments or even to yield in spite of the environment which is contrary to wild species that are responding to favorable or unfavorable natural conditions to maximize survival and reproductive success with minimal fitness costs.

Commented [JEF5]: If one finds sequences characterized as PP2CA in such diverse plants then it plays a very important role in the plant and some sequences/functions are highly conserved.

The transcript abundances of 27 *GhPP2CA* in various tissues were measured by qRT-PCR to determine the putative functions of the PP2CAs in cotton. The results showed that all of the *GhPP2CAs* were highly expressed in flowers. *GhPP2CA11*~~*AHG1A*~~ and *GhPP2CA27*~~*HAI3-2*~~ were also preferentially expressed in roots. Moreover, the transcriptional levels of *GhPP2CA3*, *11*, *13*, *27*~~*AHG1A, GhHAI2A, GhABI1-2A and GhHAI3-2*~~ were high in fibers. The transcripts of *GhPP2CA4*, *16*, *22*~~*HAB1-1A, GhHAB1-2D and GhABI2-3D*~~ were abundant in stems. These results imply that most cotton PP2CA members may function in reproductive development, and some PP2CAs also play roles in some specific tissues like roots, fibers and stems (Fig. 6).

Transcriptional changes of *GhPP2CAs* in responses to ABA and osmotic stress

To gain insight into the roles of *GhPP2CAs* in ABA signaling, transcriptional abundances of *GhPP2CAs* in roots were detected after treatments with 100 μM ABA or 10% PEG6000 for indicated periods of time. We observed that the transcriptional levels of some *GhPP2CA* genes continually increased with the extension of ABA treatment time such as *GhPP2CA5*, *11*, *18*, *20*, *25*, *27*~~*.ABI2-1D, GhAHG1A, GhAHG3-1A, GhAHG3-3D, GhHAI2D and GhHAI3-2.*~~ In contrast, the expression levels of some members including *GhPP2CA2*~~-~~*4*, *8*, *10*, *24*~~*ABI2-2A, GhAHG1D, GhAHG3-2A, GhHAB1-1A, GhHAB1-2A and GhHAI2A*~~ had decreasing trend. The expression levels of some genes were decreased at 3 h or 6 h but increased at 12 h or 24 h. These genes included *GhPP2CA1*, *16*, *17*, *19*, *26*~~*.ABI1-1D, GhABI1-2D, GhHAB1-1D, GhHAB1-2D, GhHAI3-1A, GhABI2-1D, GhAHG1A, GhAHG3-1A,*~~

Formatted: Font: Not Italic
Formatted: Font: Not Italic
Formatted: Font: Not Italic
Formatted: Font: Italic
Formatted: Font: Italic
Formatted: Font: Not Italic
Formatted: Font: Not Italic
Formatted: Font: Not Italic
Formatted: Font: Not Italic
Formatted: Font: Not Italic
Formatted: Font: Not Italic
Formatted: Font: Not Italic
Formatted: Font: Not Italic
Formatted: Font: Italic

~~*GhAHG3-3D, GhHAI2D,* and *GhHAI3-2.*~~ The expression levels of several genes increased at 3 h or 6 h but decreased at 12 h or 24 h. These genes were *GhPP2CA6*, *7*, *9*, *12-15*, *21-23*~~*ABI1-1A, GhABI1-2A, GhABI2-1A, GhABI2-2D, GhABI2-3A, GhABI2-3D, GhAHG3-1D, GhAHG3-2D, GhAHG3-3A* and *GhHAI3-1D*~~ (Fig. 7). Treatment of cotton seedlings with PEG6000 also altered the expression of most *GhPP2CA* genes (Fig. 8). The majority of *GhPP2CAs* were upregulated after treatments with PEG for a short time period, and downregulated afterwards. For example, the transcriptional levels of *GhPP2CA8*~~*ABI2-2A*~~ and *GhPP2CA21*~~*ABI2-2D*~~ were prominently enhanced at 3 h, and then reduced at 6 h, 12 h and 24 h, while those of *GhPP2CA5*, *6*, *17*, *23*, *25*, *26*~~*ABI1-1A, GhABI1-2D, GhAHG3-1A, GhAHG3-2D, GhAHG3-3D* and *GhHAB1-1D*~~ were pronouncedly increased at 6 h and decreased at 12 h and 24 h. By contrast, the expression of some genes was significantly elevated at 12 or 24 h post PEG treatment. These genes included *GhPP2CA1*, *3*, *4*, *7*, *9*, *12-14*, *18*, *20*, *24*, *27*~~*ABI1-2A, GhABI2-1A, GhABI2-1D, GhABI2-3A, GhAHG1D, GhAHG3-3A, GhHAB1-1A, GhHAI2A, GhHAI2D, GhHAI3-1A, GhHAI3-1D,* and *GhHAI3-2.*~~ The overall ~~changed~~ trend of expression of *GhPP2CA10*~~*AHG3-1D*~~ and *GhPP2CA15*~~*AHG3-2A*~~ was increased while that of *GhPP2CA2*, *11*, *16*~~*AHG1A, GhHAB1-2A,* and *GhHAB1-2D*~~ was decreased under osmotic stress (Fig. 8). Together, these data suggest that *GhPP2CAs* exhibit diverse expression patterns in responses to ABA and osmotic stress.

Many GhPP2CAs interact with GhPYL2-2D, GhPYL6-2A and GhPYL9-2A

PP2CAs have been documented to interact with ABA receptor PYLs in ABA signal

Formatted: Font: Not Italic
Formatted: Font: Not Italic
Formatted: Font: Not Italic
Formatted: Font: Not Italic
Formatted: Font: Italic
Formatted: Font: Not Italic
Formatted: Font: Italic
Formatted: Font: Italic
Formatted: Font: Italic
Formatted: Font: Not Italic
Formatted: Font: Not Italic
Formatted: Font: Not Italic
Formatted: Font: Not Italic
Formatted: Font: Not Italic
Formatted: Font: Not Italic
Formatted: Font: Not Italic
Formatted: Font: Not Italic
Formatted: Font: Not Italic
Formatted: Font: Not Italic

pathway (~~Fujii~~Ma et al., 2009; ~~Singh~~Park et al; 2009~~15~~). Accordingly, we investigated the interactions between *GhPP2CAs* and *GhYPLs* in the absence or presence of ABA by yeast-two hybrid method. A total of 11 *GhPP2CAs* were cloned, and three *GhPYLs GhPYL2-2D* (Gh_D08G2587), *GhPYL6-2A* (Gh_A06G1418) and *GhPYL9-2A* (Gh_A11G0870) were randomly selected and cloned. These genes were fused into yeast vectors, and yeast-two hybrid experiments were performed. In the absence of ABA, GhPP2CA2~~AHG3-3D~~ and GhPP2CA25~~HAB1-2A~~ respectively interacted with *GhPYL2-2D* while multiple *GhPP2CAs* like *GhPP2CA2*, 3, 6, 10, 13, 15, 18, 19, 25~~ABI1-1A, GhABI1-1D, GhABI1-2A, GhAHG3-1D, GhAHG3-2A, GhAHG3-3D, GhHAB1-2A, GhHAI2A, and GhHAI2D~~ individually interplayed with *GhPYL2-2D* in the presence of ABA (Fig. 9). In contrast, several *GhPP2CAs* could respectively interact with *GhPYL6-2A* or *GhPYL9-2A* either with or without ABA. These *GhPP2CAs* included *GhPP2CA2*, 6, 10, 13, 15, 18, 19, 24, 25.~~ABI1-1A, GhABI1-1D, GhABI1-2A, GhAHG1D, GhAHG3-1D, GhAHG3-2A, GhAHG3-3D, GhHAB1-2A and GhHAI2D.~~ Besides, *GhPP2CA3*~~HAI2A~~ interact with *GhPYL9-2A* but not with *GhPYL6-2A* either in the presence or absence of ABA (Fig.10, Fig. 11). These results imply that *GhPP2CAs* differentially interact with *GhPYLs* in responding to ABA in cotton.

Formatted: Font: Italic
Formatted: Font: Italic
Formatted: Font: Italic
Formatted: Font: Italic
Formatted: Font: Italic
Formatted: Font: Italic
Formatted: Font: Italic
Formatted: Font: Italic
Formatted: Font: Italic
Formatted: Font: Italic
Formatted: Font: Italic
Formatted: Font: Italic
Formatted: Font: Italic
Formatted: Font: Italic

DISCUSSION

PP2CAs are central components of ABA signal transduction pathway, and negatively control ABA and stress responses in plants (Fuchs et al., 2013; Singh et al., 2015).

They have been identified in several plants including *Arabidopsis*, rice, maize and *Brachypodium distachyon* in recent years (Xue et al., 2008; Wei and Pan, 2014; Cao et al., 2016). However, phylogenesis and putative functions of PP2CAs in *Gossypium* remain elusive. In the present study, 14, 13, 27 and ~~29~~ 23 PP2CA genes were characterized in genomes of *G. arboretum*, *G. raimondii*, *G. hirsutum* and *G. barbadense*, respectively (Table 1). Compared to the number of PP2CAs in *Arabidopsis* (10), rice (10), maize (16) and *Brachypodium distachyon* (8), that in *G. hirsutum* and *G. barbadense* was great (Xue et al., 2008; Wei and Pan, 2014; Cao et al., 2016). This suggests that more complex and elaborate ABA signaling mechanisms modulated by PP2CAs may exist in the upland and island cotton species. Conceivably, the high number of PP2CAs of the two species is related to their tetraploid nature. The two plants retain most PP2CA homologs of both *G. ar~~a~~boreum* and *G. raimondii* but not a copy of either one progenitor during evolution. This may be due to long-term human selection ~~for~~within these two tetraploid ~~the~~ cotton species for higher yields, growth in ~~that grow better in~~ hotter dr~~y~~ier regions, day neutral flowering, and adaptation to agronomic areas far outside their original habitats. ~~with high yields.~~ These ~~advantageous~~altered characteristics may be associated with more PP2CA proteins and complex ABA signal mechanisms in the cultivated cotton plants than in wild plants and in the greater opportunities to accumulate sequences, amenable to mutation and selection, in a tetraploid genome than in a diploid genome.

~~Analysis results revealed that the physical properties of PP2CAs are similar among the four cotton species (Table 1), pointing to the conservative functions of the PP2CA~~

~~members in *Gossypium*. Moreover, the predicted subcellular localizations of most *Gossypium* PP2CAs were the nucleus, consistent with those of PP2CAs in *Arabidopsis*, rice, maize and *Brachypodium distachyon* (Xue et al., 2008; Wei and Pan, 2014; Cao et al., 2016). These data imply that the functions of PP2CAs are conserved among different plants, reflecting the importance of PP2CAs in regulating gene expression in plants.~~

We noticed that 27 *GhPP2CAs* and 23 *GbPP2CAs* individually had their corresponding orthologs in *G. arboretum* or *G. raimondii* (Fig. 1), indicating that those PP2CAs in *G. hirsutum* and *G. barbadense* are ancestrally related to those~~directly descended from those~~ PP2CAs in the two diploid species. ~~However, no orthologs of GhHAI3-2, GbABI1-3D, GbHAI3-2A, GbHAI3-2D, GbHAI3-3A, GbHAI3-3A' and GbHAI3-3D were found in the diploid cotton genomes (Fig. 1), implying that these PP2CAs likely generated through gene duplication.~~ Additionally, no orthologous genes of *GaPP2CA1, 3, 4, 14~~*AHG3-3*~~, *GrPP2CA6~~*ABI1-2*~~ and *GrPP2CA9~~*HAI2*~~ were observed in *G. barbadense* (Fig. 1). This hints that these genes are possibly lost, or these genes arose after the tetraploid species appeared and separated from the diploid species during the evolutionary processes.

The structures and the numbers of introns and exons in PP2CAs were similar among the 4 *Gossypium* species as well as *Arabidopsis*, rice, maize and *Brachypodium distachyon* (Xue et al., 2008; Wei and Pan, 2014; Cao et al., 2016), suggesting that the PP2CAs undergo conserved evolutionary processes even after the divergence of monocotyledons and dicotyledons. Colinearity results showed that 136

Formatted: Not Highlight

Formatted: No underline, Font color: Auto

Formatted: Font: Italic

Formatted: Font: Italic

Formatted: Font: Not Italic

Formatted: Font: Not Italic

homologous gene pairs existed among *GaPP2CAs*, *GrPP2CAs* and *GhPP2CAs* (Fig. 3), indicating that PP2CA genes expand primarily through segmental duplication of DNA. Segmental duplicates may be more often maintained through subsequent gene subfunctionalization compared to tandem duplicates (Lynch and Conery, 2000). Accordingly, these PP2CAs ~~most~~ probably had diverse functions in *Gossypium*. Moreover, in agreement with our results, *Arabidopsis* phosphatase family genes also showed ~~to have high~~ segmental duplication ~~property~~ (Cannon et al., 2004), suggesting the evolutionary mechanism of PP2CAs may be conserved in plants. The mean values of *Ka*/*Ks* for majority of *PP2CA* homologous gene pairs were less than 1 (Fig. 4), implying that the homologous genes between any two of *G. arboretum*, *G. raimondii* and *G. hirsutum* were under purifying selection during evolution.

Phylogenetic results showed that the PP2CA members from monocotyledonous plants clustered together, and similar results occurred in dicotyledonous PP2CAs (Fig. 5). This suggests that great changes in DNA sequences of the PP2CAs have taken place after isolation of monocotyledons and dicotyledons although these genes shared a common ancestor. PP2CAs in *Gossypium* always clustered together with those in *T. cacao* rather than with those in *A. thaliana*, *R. communis*, *P. trichocarpa*, *G. max*, *B. distachyon* and *O. sativa* (Fig. 5), pointing to the closer evolutionary relationship of *Gossypium* with *T. cacao*. That is, most of homologous PP2CA members in *Gossypium* and *cacao* ~~may~~ were generated before separation of the two ~~plants~~genera from ~~a close~~the common ancestor~~s~~. The common PP2CAs across all plants show that they still have core functions essential to basic plant survival and functions.

Formatted: Font: Times New Roman, 12 pt

Formatted: Font: Times New Roman, 12 pt

Formatted: Font: Italic

Formatted: Font: Italic

Formatted: Font: Italic

Differences in PP2CAs that follow differentiation of different genera and even species show that they are still diverse and can accommodate functions specific to the survival of species and even in response to selection by man. 'Housekeeping' PP2CAs could probably be subtracted from the picture to illuminate the more unique ones to better understand functions of individual PP2CAs and their roles in specific species, traits, or even agronomic performance of specific cultivars.

Transcript abundance analysis indicated that the majority of the *GhPP2CAs* was predominantly expressed in flowers (Fig. 6), suggesting that *GhPP2CA*-mediated ABA signaling may be of great importance in flower development of cotton. High expression of *GhPP2CAs* in flowers was likely due to the importance of timing flowering to environmental conditions of native *Gossypium* plants. Because evolution of some species is tied to long-term human selection of cotton plants with high yields of fib~~r~~ers and good adaptations to hot and dry growth conditions. Cotton yields are closely associated with flowering in agronomic conditions created by man that often are much different than the natural habitats of wild or ancestral *Gossypium* species (e.g. cultivation only in summers of temperate-tropical latitudes instead of perennial growth in tropical, latitudes closer to the equator). Drought and hot stresses should limit flower development. PP2CAs are negative regulators of the adverse stress~~or~~es; and therefore, may facilitate flowering of cotton in these newer environments. The expression of most *GhPP2CAs* was upregulated after treatment with ABA or PEG6000 (Fig. 7, Fig. 8), in good agreement with the results from *AtPP2CAs*, *OsPP2CAs* and *BdPP2CAs* (Xue et al. 2008; Cao et al. 2016). These

Formatted: Font: Italic

Formatted: Font: Italic

Formatted: Font: Italic

Formatted: Font: Italic

findings imply that PP2CAs are essential for plant response to ABA and osmotic stress.

The interactions between 11 *GhPP2CA* proteins and 3 *GhPYLs* were examined. The results revealed that most *GhPP2CAs* can individually interact with the three *GhPYLs* in the absence or presence of ABA (Fig. 9-11). *GhPYLs* are homologs of AtPYLs and some *GhPYLs* are functional ABA receptors (Liang et al., 2017; Zhang et al., 2017b). These data indicate that a large number of *GhPP2CAs* ~~are functional phosphatases, and~~ may play roles via interactions with *GhPYLs* in ABA-dependent or ABA-independent manner in cotton. The detailed mechanisms of *GhPP2CAs* in ABA signaling will be further explored in the future.

CONCLUSIONS

In ~~T~~total~~ly~~, 14, 13, 27 and ~~29~~ 23 PP2CA genes were characterized from *G. arboretum*, *G. raimondii*, *G. hirsutum* and *G. barbadense*, respectively. These genes shared high similarity in ~~physical properties,~~ chromosomal locations, structures and phylogeny among the species. Most of them might be~~were~~ under purifying selection during evolution. Moreover, PP2CAs displayed specific expression patterns in tissues and diverse expression profiles in response to ABA and osmotic stress in *G. hirsutum*. Yeast-two hybrid experiments indicated that most *GhPP2CAs* interacted with *GhPYL2-2D*, *GhPYL6-2A* and *GhPYL9-2A* with or without ABA. These findings provide essential information for in-depth investigations of the functions of PP2CAs in *Gossypium* in the future.

Formatted: Font: Italic
Formatted: Font: Italic
Formatted: Font: Italic
Formatted: Font: Italic
Formatted: Font: Italic
Formatted: Font: Italic
Formatted: Font: Italic
Formatted: Font: Italic
Commented [JEF6]: Be more consistent with the italics and remove them when they are not needed.
Formatted: Font: Not Italic
Formatted: Font: Italic
Formatted: Font: Italic
Formatted: Font: Italic
Formatted: Font: Italic

ADDITIONAL INFORMATION AND DECLARATIONS

Funding

This work was supported by Foundation of Program for young backbone teachers in universities of Henan Province [2016GGJS-024], the scientific research and innovation team in Henan University of Animal Husbandry and Economy (2018KYTD18), and the "111" Project. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Grant Disclosures

The following grant information was disclosed by the authors:

Program for young backbone teachers in universities of Henan Province: 2016GGJS-024.

Foundation of scientific research and innovation team in Henan University of Animal Husbandry and Economy: (2018KYTD18).

"111" Project.

Competing Interests

The authors declare there are no competing interests.

Author Contributions

- Tingting Lu and Gaofeng Zhang analyzed the data and performed the experiments,
- Yibin Wang and Shibin He contributed to prepare various reagents/materials/analysis tools.
- Lirong Sun and Fushun Hao conceived and designed the experiments, and wrote the manuscript.

REFERENCES

Camacho C, Coulouris G, Avagyan V, Ma N, Papadopoulos J, Bealer K, Madden LT, 2009. Blast+: architecture and applications. *BMC Bioinformatics* **10:** 421.

Cannon SB, Mitra A, Baumgarten A, Young ND, May G. 2004. The roles of segmental and tandem gene duplication in the evolution of large gene families in *Arabidopsis thaliana*. BMC *Plant Biol* **4**:10.

Cao J, Jiang M, Li P, Chu Z. 2016. Genome-wide identification and evolutionary analyses of the PP2C gene family with their expression profiling in response to multiple stresses in *Brachypodium distachyon*. BMC *Genomics* **17**:175.

Edgar RC, 2004. Muscle: multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Res* **32:**1792–1797.

Fuchs S, Grill E, Meskiene I, Schweighofer A. 2013. Type 2C protein phosphatases in plants. *FEBS Journal* 280:681–693.

Fujii H, Zhu JK. 2009a. Arabidopsis mutant deficient in 3 abscisic acid-activated protein kinases reveals critical roles in growth, reproduction, and stress. *Proceedings of the National Academy of Sciences* **106:**8380–8385.

Fujii H, Chinnusamy V, Rodrigues A, Rubio S, Antoni R, Park SY, Cutler SR, Sheen J, Rodriguez PL, Zhu JK. 2009b. In vitro reconstitution of an abscisic acid signalling pathway. *Nature* **462**:660–664.

Fujita Y , Nakashima K , Yoshida T , Katagiri T, Kidokoro S, Kanamori N, Umezawa T, Fujita M, Maruyama K, Ishiyama K, Kobayashi M, Nakasone

Formatted: Font: Times New Roman, 12 pt, Bold, No underline

Formatted: Font: Times New Roman, 12 pt, Bold, No underline

Formatted: Indent: Left: 0", Hanging: 2 ch

Formatted: Font: Times New Roman, 12 pt, Bold, No underline

Formatted: Font: Times New Roman, 12 pt, No underline

Formatted: Font: Times New Roman, 12 pt, Bold, Not Italic, No underline

Formatted: Font: Bold, No underline

Formatted: Font: Times New Roman, 12 pt, No underline

Formatted: Font: 16 pt

Formatted: Font: (Default) Times New Roman, 12 pt, Bold, No underline

Formatted: Indent: Left: 0", Hanging: 2 ch

Formatted: Font: Bold, No underline

Formatted: Font: Times New Roman, 12 pt, Not Italic

Formatted: Font: (Default) Times New Roman, 12 pt, Not Italic

Formatted: Font: (Default) Times New Roman, 12 pt

Formatted: Font: (Default) Times New Roman, 12 pt, Not Italic

Formatted: Font: (Default) Times New Roman, 12 pt, Bold, Not Italic

Formatted: Font: Bold, Not Italic

Formatted: Font: (Default) Times New Roman, 12 pt, Not Italic

Formatted: Font: Times New Roman, 12 pt, Bold, Not Italic

Formatted: Indent: Left: 0", Hanging: 2 ch

Formatted: Font: Bold, Not Italic

S, Yamada K, Ito T, Shinozaki K, Yamaguchi-Shinozake K, 2009. Three SnRK2 protein kinases are the main positive regulators of abscisic acid signaling in response to water stress in *Arabidopsis*. *Plant and Cell Physiology*, **50:**2123–2132.

Geiger D , Scherzer S , Mumm P , Stange A, Marten I, Bauer H, Ache P, Matschi S, Liese A, AI-Rasheid KA, Romeis T, Hedrich R, 2009. Activity of guard cell anion channel SLAC1 is controlled by drought-stress signaling kinase-phosphatase pair. *Proceedings of the National Academy of Sciences*, **106:**21425–21430.

Hauser F, Waadt R, Schroeder J. 2011. Evolution of Abscisic Acid Synthesis and Signaling Mechanisms. *Current Biology*,**21:**346–355.

~~**Himmelbach A, Hoffmann T, Leube M, Höhener B, Grill E. 2002.** Homeodomain protein ATHB6 is a target of the protein phosphatase ABI1 and regulates hormone responses in *Arabidopsis*. *EMBO J* **21**:3029–3038.~~

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

Kerk D, Templeton G, Moorhead GBG. 2007. Evolutionary radiation pattern of novel protein phosphatases revealed by analysis of protein data from the completely sequenced genomes of humans, green algae, and higher plants. *Plant Physiology* **146**:351–367.

~~**Larkin MA, Blackshields G, Brown NP, Chenna R, McGettigan PA, McWilliam H, Valentin F, Wallace IM, Wilm A, Lopez R, Thompson JD, Gibson TJ,**~~

Formatted ...
Formatted: Font: Times New Roman, 12 pt
Formatted ...
Formatted: Font: Times New Roman, 12 pt, Not Italic
Formatted ...
Formatted ...
Formatted ...
Formatted: Font: Times New Roman, 12 pt, Not Italic
Formatted: Font: Times New Roman, 12 pt, Bold
Formatted: Indent: Left: 0", Hanging: 2 ch
Formatted ...
Formatted ...

~~**Higgins DG. 2007.** Clustal W and Clustal X version 2.0. *Bioinformatics* **23**:2947–2948.~~

Lee SC, Lan W, Buchanan BB, Luan S. 2009. A protein kinase-phosphatase pair interacts with an ion channel to regulate ABA signaling in plant guard cells. *Proceedings of the National Academy of Sciences* **106**:21419–21424.

Liang C, Liu Y, Li Y, Meng Z, Yan R, Zhu T, Wang Y, Kang S, Abid MA, Malik W, Sun G, Guo S, Zhang R. 2017. Activation of ABA receptors gene GhPYL9-11A is positively correlated with cotton drought tolerance in transgenic Arabidopsis. *Frontiers in Plant Science* **8**:1453.

Lu TT, Zhang GF, Sun LR, Wang J, Hao FS. 2017. Genome-wide identification of CBL family and expression analysis of CBLs in response to potassium deficiency in cotton. *PeerJ* **5**:e3653.

Lynch M, Conery JS. 2000. The evolutionary fate and consequences of duplicate genes. *Science* **290**:1151–1155.

Ma LY, Zhang H, Sun LR, Jiao YH, Zhang GZ, Miao C, Hao FS. 2012. NADPH oxidase AtrbohD and AtrbohF function in ROS-dependent regulation of Na^+/K^+ homeostasis in *Arabidopsis* under salt stress. *Journal of Experimental Botany* **63**:305–317.

Marchler-Bauer A, Bo Y, Han L, He J, Lanczycki CJ, Lu S, Chitsaz F, Derbyshire MK, Geer RC, Gonzales NR, Gwadz M, Hurwitz DI, Lu F, Marchler GH, Song JS, Thanki N, Wang Z, Yamashita RA, Zhang D, Zheng C, Geer LY, Bryant SH. 2017. CDD/SPARCLE: functional

Formatted: Font: Times New Roman, 12 pt, Bold, Not Italic
Formatted: Font: Times New Roman, 12 pt, Bold, Not Italic
Formatted: Indent: Left: 0", Hanging: 2 ch
Formatted: Font: Bold, Not Italic
Formatted: Font: Bold, Not Italic
Formatted: Font: Times New Roman, 12 pt, Not Italic
Formatted: Font: Times New Roman, 12 pt
Formatted: Font: Times New Roman, 12 pt, Bold, Not Italic
Formatted: Font: Times New Roman, 12 pt, Not Italic
Formatted: Font: Times New Roman, 12 pt, Not Italic
Formatted: Font: Times New Roman, 12 pt, Not Italic
Formatted: Font: Bold
Formatted: Indent: Left: 0", Hanging: 2 ch
Formatted: Font: Italic
Formatted: Font: Bold

classification of proteins via subfamily domain architectures. *Nucleic Acids Research* **45**: D200-D203.

Ma Y, Szostkiewicz I, Korte A, Moes A, Yang Y, Christmann A, Grill E. 2009. Regulators of PP2C phosphatase activity function as abscisic acid sensors. *Science* **324**:1064–1068.

Minh BQ, Nguyen MA, von Haeseler A. 2013. Ultrafast approximation for phylogenetic bootstrap. *Molecular & Biology & Evolution* **30**:1188–1195.

Murashige T, Skoog F. 1962. A revised medium for rapid growth and bio-assays with tobacco tissue culture. *Physiologia Plantarum* **15**:473–497.

Nguyen LT, Schmidt HA, Von Haeseler A, Minh BQ. 2015. IQ-TREE: a fast and effective stochastic algorithm for estimating maximum-likelihood phylogenies. *Molecular Biology and Evolution* **32**:268–274.

Ohta M, Guo Y, Halfter U, Zhu J. 2003. A novel domain in the protein kinase SOS2 mediates interaction with the protein phosphatase 2C ABI2. *Proceedings of the National Academy of Science* **100**:11771–11776.

Park SY, Fung P, Nishimura N, Jensen DR, Fujii H, Zhao Y, Lumba S, Santiago J, Rodrigues A, Chow TF, Alfred SE, Bonetta D, Finkelstein R, Provart NJ,Desveaux D, Rodriguez PL, McCourt P, Zhu JK, Schroeder JI, Volkman BF, Cutler SR. 2009. Abscisic acid inhibits type 2C protein phosphatases via the PYR/PYL family of START proteins. *Science* **324**:1068–1071.

Rubio S, Rodrigues A, Saez A, Dizon MB, Galle A, Kim TH, Santiago J, Flexas J, Schroeder JI, Rodriguez PL. 2009. Triple loss of function of protein

phosphatases type 2C leads to partial constitutive response to endogenous abscisic acid. *Plant Physiol* **150:**1345–1355.

~~**Ohta M, Guo Y, Halfter U, Zhu J. 2003.** A novel domain in the protein kinase SOS2 mediates interaction with the protein phosphatase 2C ABI2. *Proc Natl Acad Sci USA* **100**:11771–11776.~~

Schweighofer A, Hirt H, Meskiene I. 2004. Plant PP2C phosphatases: emerging functions in stress signaling. *Trends in Plant Science* **9**:236–243.

Singh A, Giri J, Kapoor S, Tyagi AK, Pandey GK. 2010. Protein phosphatase complement in rice: genome-wide identification and transcriptional analysis under abiotic stress conditions and reproductive development. *BMC Genomics* **11**:435.

Singh A, Pandey A, Srivastava AK, Tran LS, Pandey GK. 2015. Plant protein phosphatases 2C: from genomic diversity to functional multiplicity and importance in stress management. *Critical Reviews in Biotechnology* **36**:1023–1035.

Tähtiharju S, Palva T. 2001. Antisense inhibition of protein phosphatase 2C accelerates cold acclimation in *Arabidopsis thaliana*. *The Plant Journal* **26:** 461–470.

Trifinopoulos J, Nguyen LT, Von Haeseler A, Minh BQ. 2016. W-IQ-TREE: a fast online phylogenetic tool for maximum likelihood analysis. *Nucleic Acids Research* **44**:232–235.

Ullah A, Sun H, Yang X, Zhang X. 2017. Drought coping strategies in cotton:

Formatted: Font: Times New Roman, 12 pt, Not Italic
Formatted: Font: Times New Roman, 12 pt, Not Italic
Formatted: Font: Times New Roman, 12 pt, Not Italic
Formatted: Font: Times New Roman, 12 pt, Not Italic
Formatted: Font: Times New Roman, 12 pt, Not Italic
Formatted: Font: Times New Roman, 12 pt, Not Italic
Formatted: Font: Times New Roman, 12 pt, Not Italic
Formatted: Font: Times New Roman, 12 pt, Not Italic
Formatted: Font: Times New Roman, 12 pt, Not Italic
Formatted: Font: Times New Roman, 12 pt
Formatted: Font: Times New Roman, 12 pt, Bold, Not Italic
Formatted: Font: Times New Roman, 12 pt, Not Italic
Formatted: Font: Times New Roman, 12 pt, Not Italic
Formatted: Font: 16 pt
Formatted: Font: Italic
Formatted: Font: Times New Roman, 12 pt, Bold, Not Italic
Formatted: Indent: Left: 0", Hanging: 2 ch
Formatted: Font: Bold, Not Italic
Formatted: Font: Times New Roman, 12 pt, Bold, Not Italic
Formatted: Font: Bold, Not Italic
Formatted: Font: Times New Roman, 12 pt, Bold, Not Italic
Formatted: Font: Times New Roman, 12 pt, Not Italic
Formatted: Font: Times New Roman, 12 pt, Not Italic
Formatted: Font: Times New Roman, 12 pt, Not Italic
Formatted: Font: Times New Roman, 12 pt
Formatted: Font: Times New Roman, 12 pt, Not Italic
Formatted: Font: Bold, Not Italic
Formatted: Font: Times New Roman, 12 pt, Not Italic
Formatted: Font: Times New Roman, 12 pt, Bold
Formatted: Font: Times New Roman, 12 pt

increased crop per drop. *Plant Biotechnology Journal* 15: 271–284.

Wei K, Pan S. 2014. Maize protein phosphatase gene family: identification and molecular characterization. *BMC Genomics* **15**:773.

Xue T, Wang D, Zhang S, Ehlting J, Ni F, Jakab S, Zheng C, Zhong Y. 2008. Genome-wide and expression analysis of protein phosphatase 2C in rice and *Arabidopsis*. *BMC Genomics* **9**:550.

Zhang F, Wei Q, Shi J, Jin X, He Y, Zhang Y, Luo Q, Wang Y, Chang J, Yang G, He G. 2017a. *Brachypodium distachyon* BdPP2CA6 interacts with BdPYLs and BdSnRK2 and positively regulates salt tolerance in transgenic *Arabidopsis*. *Frontiers in Plant Science* **8**:264.

Zhang GF, Lu TT, Miao WW, Sun LR, Tian M, Wang J, Hao FS. 2017b. Genome-wide identification of ABA receptor PYL family and expression analysis of PYLs in response to ABA and osmotic stress in *Gossypium*. *PeerJ* **5**:e4126.

Formatted: Font: Times New Roman, 12 pt, Not Italic

Formatted: Font: Times New Roman, 12 pt

Formatted: Font: Times New Roman, 12 pt

Formatted: Font: 20 pt