



Genomic, Molecular Evolution, and Expression Analysis of Genes Encoding Putative Classical AGPs, Lysine-Rich AGPs, and AG Peptides in *Brassica rapa*

Tianyu Han^{1,2,3}, Heng Dong^{1,2,3}, Jie Cui^{1,2,3}, Ming Li^{1,2,3}, Sue Lin⁴, Jiashu Cao^{1,2,3} and Li Huang^{1,2,3*}

OPEN ACCESS

Edited by:

Jianjun Chen,
University of Florida, USA

Reviewed by:

Liezhao Liu,
Southwest University, China
Li-Song Chen,
Fujian Agriculture and Forestry
University, China
Taras P. Pasternak,
Albert Ludwig University of Freiburg,
Germany
David Domozych,
Skidmore College, USA

*Correspondence:

Li Huang
lihuang@zju.edu.cn

Specialty section:

This article was submitted to
Crop Science and Horticulture,
a section of the journal
Frontiers in Plant Science

Received: 28 September 2016

Accepted: 08 March 2017

Published: 29 March 2017

Citation:

Han T, Dong H, Cui J, Li M, Lin S,
Cao J and Huang L (2017) Genomic,
Molecular Evolution, and Expression
Analysis of Genes Encoding Putative
Classical AGPs, Lysine-Rich AGPs,
and AG Peptides in *Brassica rapa*.
Front. Plant Sci. 8:397.
doi: 10.3389/fpls.2017.00397

¹ Laboratory of Cell and Molecular Biology, Institute of Vegetable Science, Zhejiang University, Hangzhou, China, ² Key Laboratory of Horticultural Plant Growth, Development and Quality Improvement, Ministry of Agriculture, Hangzhou, China, ³ Zhejiang Provincial Key Laboratory of Horticultural Plant Integrative Biology, Hangzhou, China, ⁴ Institute of Vegetable Science, Wenzhou Vocational College of Science and Technology, Wenzhou, China

Arabinogalactan proteins (AGPs) belong to a class of Pro/Hyp-rich glycoproteins and are some of the most complex types of macromolecules found in plants. In the economically important plant species, *Brassica rapa*, only chimeric AGPs have been identified to date. This has significantly limited our understanding of the functional roles of AGPs in this plant. In this study, 64 AGPs were identified in the genome of *B. rapa*, including 33 classical AGPs, 28 AG peptides and three lys-rich AGPs. Syntenic gene analysis between *B. rapa* and *A. thaliana* suggested that the whole genome triplication event dominated the expansion of the AGP gene family in *B. rapa*. This resulted in a high retained proportion of the AGP family in the *B. rapa* genome, especially in the least fractionated subgenome. Phylogenetic and motif analysis classified the classical AGPs into six clades and three orphan genes, and the AG peptides into three clades and five orphan genes. Classical AGPs has a faster rate of molecular evolution than AG peptides revealed by estimation of molecular evolution rates. However, no significant differences were observed between classical AGPs and lys-rich AGPs. Under control conditions and in response to phytohormones treatment, a complete expression profiling experiment has identified five anther-specific AGPs and quite a number of AGPs responding to abscisic acid, methyl jasmonate and/or gibberellin. In this study, we presented a bioinformatics approach to identify important types of AGPs. Moreover, the association between their function and their protein structure, as well as the evolution and the expression of AGP genes were investigated, which might provide fundamental information for revealing the roles of AGPs in *B. rapa*.

Keywords: AGPs, arabinogalactan proteins, phytohormones, pollen, syntenic analysis

INTRODUCTION

Arabinogalactan proteins (AGPs), together with proline-rich proteins (PRPs) and extensins (EXTs), constitute the superfamily of proline/hydroxyproline-rich glycoproteins (P/HRGPs), which are abundant in the extracellular matrix throughout the plant kingdom (Gaspar et al., 2001; Showalter, 2001). AGPs function in diverse aspects of plant growth and development, such as microspore embryogenesis (Tang et al., 2006), cell wall strengthening (Hijazi et al., 2014), cell proliferation (Serpe and Nothnagel, 1994), cell expansion (Willats and Knox, 1996), programmed cell death (Pennell et al., 1991), pollen tube growth (Levitin et al., 2008; Lin et al., 2014; Pereira et al., 2016), cell separation in vestigial abscission (Stenvik et al., 2006), and signal recognition and transduction (Lee et al., 2008, 2009). In addition, small number of initial studies have illustrated that AGPs are involved in response to phytohormones such as abscisic acid (ABA), methyl jasmonate (MeJA), and gibberellin (GA). For example, *AGP30* specifically responds to the concentration of ABA and then affects seed germination in *Arabidopsis thaliana* (van Hengel and Roberts, 2003). *AGP31* is down-regulated in MeJA treated *A. thaliana* plants (Liu and Mehdy, 2007), and some AGPs affect the expressions of gibberellin-induced genes in barley (Mashiguchi et al., 2008).

The AGP family consists of different members that can be variable in specific plant species. Therefore, genome-wide screens are commonly used for the identification of AGPs. Knowledge of distinct characteristics shared across known members of the protein family enables their detection within the complete set of proteins in an organism. AGPs are rich in proline or hydroxyproline (Pro/Hyp), serine (Ser), threonine (Thr), and alanine (Ala), which comprise up to 99% of the molecular mass of AGP proteins (Ellis, 2010). According to differences in the composition of their protein backbone, AGPs are further classified into classical AGPs, arabinogalactan (AG) peptides, lysine (Lys)-rich AGPs, and chimeric AGPs (Schultz et al., 2002). Classical AGPs are defined by the core protein containing Hyp, Ala, Ser, Thr, and Glycine (Gly) as the major amino acid constituents, and their C terminus is glycosylphosphatidylinositol (GPI) anchored (Showalter et al., 2010). Lys-rich AGPs have a Lys-rich domain of approximately 16 amino acid residues that is flanked on both sides by AGP glycol modules. AG peptides are composed of only 10–13 amino acid residues and the putative cell adhesion molecules (Schultz et al., 2002). Most AGPs are characterized by the entire protein containing only P/HRGP modules, while chimeric AGPs are consisted of one or two known P/HRGP motifs and

additional unrelated motifs such as, fascilin-like domain for fasciclin-like AGPs (FLAs), early nodulin-like domain for eNod-like AGPs (ENODL) and non-specific lipid transfer protein-like domain for nsLTP-like AGPs (Schultz et al., 2002). Based on specific characteristics of the protein backbone and length, the presence of Ala-Pro, Pro-Ala, Ser-Pro, or Thr-Pro repeats, signal peptide and GPI anchor addition sequence, 22 classical AGPs, 16 AG peptides and three lys-rich AGPs were identified in the *A. thaliana* genome (Showalter et al., 2010) and 11 classical AGPs, 15 AG peptides and two lys-rich AGPs were found in the rice (*Oryza sativa*) genome (Ma and Zhao, 2010). In vegetable crops, chimeric AGPs like FLAs and ENODLs have been identified in *Brassica rapa* (Jun and Xiaoming, 2012; Li et al., 2013), which is the most important economic vegetable crop in East Asia (Wang et al., 2012). *B. rapa* shares a common ancestor with *A. thaliana*; their chromosomes were derived from the rearrangement of 24 conserved collinear blocks in ancestral karyotype of crucifer species (Cheng et al., 2013). A whole genome triplication (WGT) event, occurred 13–17 million years ago, has differentiated the genome of *B. rapa* from that of *A. thaliana*. This makes *B. rapa* an excellent system for studying the expansion of gene families (Wang et al., 2011). Chromosomal localization and gene duplication analysis illustrated that the expansion of FLAs and ENODLs in *B. rapa* depends on the WGT event, and several FLAs in *B. rapa* display similar expression patterns as their orthologs in *A. thaliana*. In our previous studies, a screening for differentially expressed genes between the male fertile and sterile plants of the *B. rapa* genic male sterility (GMS) A/B line has identified two classical AGPs encoding genes exclusively expressed in the fertile flower buds (Huang et al., 2008a). Further functional characterization revealed that these two AGPs might play important roles in pollen wall formation (Huang et al., 2008b; Lin et al., 2014). However, it is still unknown whether there are more classical AGPs function in pollen development in *B. rapa*. Also, there is limited information about the presence of other types of AGPs in *B. rapa*, which restricts a comprehensive understanding of the roles of AGP gene family in this crop.

In this study, in order to obtain more detailed information about the AGP family in *B. rapa*, Perl scripts were specifically written to screen for classical AGPs, AG peptides and lys-rich AGPs in the genome of *B. rapa*. Subsequently, the candidate *B. rapa* AGPs (*BrAGPs*) were annotated and their characteristics, including retained proportion, syntenic relationship, protein motif structure and molecular evolutionary rates, were analyzed. To understand the functions of *BrAGPs* and identify the candidate AGPs involved in pollen development, the expression of AGP genes in different tissues and different developmental stages of pollen were investigated. Furthermore, the response of *BrAGPs* to GA, MeJA and ABA treatments were also evaluated. This study may provide valuable insights and reveal some underlying mechanisms of the AGP gene family in *B. rapa*.

MATERIALS AND METHODS

Plant Materials

The “*Bcajh97-01A/B*” is a GMS A/B line (sister line) of *B. rapa*, with the progenies segregating into sterile and fertile types in a 1:1 ratio. The sterile “*Bcajh97-01A*” has been demonstrated

Abbreviations: AGPs, Arabinogalactan proteins; ABA, abscisic acid; AK, ancestral karyotype; AtAGPs, *A. thaliana* AGPs; BLAST, the basic local alignment search tool; BRAD, Brassica database; *BrAGPs*, *B. rapa* AGPs; ENODL, eNod-like AGPs; EXTs, extensions; FLAs, fasciclin-like AGPs; GA, gibberellin; GPI, glycosylphosphatidylinositol; HAP, hour after pollination; HAT, hours after treatment; HMM, hidden Markov model; Ka, nonsynonymous substitution rate; Ks, synonymous substitution rate; MeJA, methyl jasmonate; MEME, Multiple Em for Motif Elicitation; ML, maximum likelihood; NJ, neighbor-joining; P/HRGPs, proline/hydroxyproline-rich glycoproteins; PAST, Pro, Ala, Ser, Thr; Pro/Hyp, proline or hydroxyproline; PRPs, proline-rich proteins; qRT-PCR, quantitative real-time PCR; RT-PCR, reverse transcription polymerase chain reaction; TAIR, The Arabidopsis Information Resource; WGT, whole genome triplication.

as a mutant of male meiosis cytokinesis without mature pollen, which is the only difference detected between “*Bcajh97-01A*” and the fertile “*Bcajh97-01B*” (Huang et al., 2008a, 2009). This characteristic of “*Bcajh97-01A/B*” GMS line is ideal to study gene expression during pollen development.

“*Bcajh97-01A/B*” was planted in experimental green-house of Zhejiang University. At the flowering stage, siliques at 48 h after pollination (HAP), roots, stems, leaves, and inflorescences were collected from 15 individual plants of the fertile line for reverse transcription PCR. At the same time, according to the longitudinal diameters, the flower buds were divided into five stages corresponding to five distinct pollen developmental stages (Huang et al., 2008a). The pollen development in flower buds was confirmed by microscopic examination.

Pistils in the fertile plants were collected at 1, 3, and 10 HAP. One HAP pistils correspond to the period of pollen germination on the stigma, 3 HAP pistils represent pollen tube growth phase, and 10 HAP pistils correspond to the period when the pollen tube reaches the ovule and fertilization occurs (Jiang et al., 2013). The materials were frozen in liquid nitrogen immediately after harvested and stored at -75°C for subsequent analysis.

Phytohormone Treatments

For phytohormone treatments, one of the following solutions of 100 μM GA, 100 μM MeJA, and 100 μM ABA was sprayed on all leaves (Liu et al., 2014; Duan et al., 2015). Control plants were sprayed with distilled water. The concentration of phytohormone treatments chosen was based on the studies by Suzuki et al. (2002), Liu and Mehdy (2007), and van Hengel et al. (2004). Seedlings in three containers were used in each treatment, and each container had 12 seedlings. The treatments were repeated three times. Control and treated leaves were harvested at 0, 4, and 12 h after treatment (HAT). The samples were frozen in liquid nitrogen and stored at -80°C until use.

RNA Isolation and Reverse Transcription PCR

Total RNA was extracted from tissues mentioned above using Trizol reagent (Invitrogen, Carlsbad, CA). The RNA integrity was analyzed on agarose gel. SuperScript IV (Invitrogen) and oligo (dT) primer was used to synthesize the cDNA. The expression of cDNA was detected using BiometraT Professional Thermo cycler and Taq (Dingguo, Shanghai). Reverse transcription polymerase chain reaction (RT-PCR) was employed to detect the expression of *BrAGPs* in different tissues of *B. rapa*, while quantitative real-time PCR (qRT-PCR) was carried out to test the expression variation of *BrAGPs* in treatment with endogenous hormones. RT-PCR and qRT-PCR were both carried out in triplicates using gene specific primers (Supplementary data 1: Tables S1, S2). *BrUBC10* was used as the reference gene. The results of qRT-PCR were calculated using the $2^{-\Delta\Delta\text{Ct}}$ method (Livak and Schmittgen, 2001) and normalized to the corresponding distilled water treatment. They were further gene-wise normalized, mean-centered and clustered hierarchically using the centroid linkage clustering method in Cluster 3.0 (<http://bonsai.hgc.jp/~mdehoon/software/cluster/index.html>).

Identification of AGPs, AG-Peptides and Lys-Rich AGPs by Calculating the Biased Amino Acid Composition and Length

The searching criteria for AGPs referred to the guidelines that are widely used (Schultz et al., 2002, 2004; Tan et al., 2003; Ma and Zhao, 2010). A Perl script (amino acid bias) was written to calculate the PAST (Pro, Ala, Ser, Thr) percentage for the candidate proteins (Supplementary data 1: Supplement Methods). All annotated proteins were downloaded from the Brassica database (BRAD, <http://brassicadb.org>, V1.2). Perl is a high-level, general-purpose, interpreted, dynamic programming language for UNIX, Windows, and OSX operating systems (<http://www.activeperl.com/>). For this study, the Perl language compiler ActivePerl-5.16.2.3010812913 was used.

The amino acid bias program was compiled with two different output files of the lists of proteins. The first output file named “AGPs” includes that all the proteins above 50% PAST threshold and >90 amino acid residues in length. Another output file named “AG-peptide” includes all proteins above 35% PAST and between 50 and 90 amino acid residues in length. The N-terminal and C-terminal signals of AG-peptide account for more than 50% of the deduced proteins and the length of AG-peptide is shorter than classical AGPs, so we reduced the PAST level and shortened the searching length of the sequences. Classical AGPs and lys-rich AGPs were extracted from file “AGPs,” in which lys-rich AGPs had a lys-rich domain of approximately 16 amino acid residues that was flanked on both sides by AGP glycomodules, and classical AGPs covered the other genes without the domains like fascilin-like domain and early nodulin-like domain. Pfam (<http://pfam.xfam.org/>) and SMART (<http://smart.embl-heidelberg.de/>) were used for analyzing the conserved domains. Moreover, orthologous genes of all the annotated classical AGPs, AG peptides and lys-rich AGPs in *A. thaliana* were screened in the proteome of *B. rapa*.

Signal Peptide and GPI Anchor Prediction

The signal peptide of proteins was predicted by the SignalP 4.1 Server (<http://www.cbs.dtu.dk/services/SignalP/>) (Petersen et al., 2011). In addition, proteins with predicted signal peptide cleavage site in the first 50 amino acids were considered as the presence of N-terminal signal sequences (Nielsen et al., 1997). To determine the presence of a C-terminal GPI anchor additional signal, all the proteins in the “AGPs” and “AG-peptide” files were subjected to the big-PI Plant Predictor (http://mendel.imp.ac.at/gpi/plant_server.html) and PSORT Prediction (<http://psort.hgc.jp/form.html>) (Nakai and Horton, 1999; Eisenhaber and Eisenhaber, 2007). Proteins having both signal peptide and GPI anchor were treated as “high confidence” AGPs.

The Analysis of Synteny and Retained Proportion

The analysis of synteny using information of 24 conserved collinear blocks of ancestral karyotype (AK) in *B. rapa* and *A. thaliana* was carried out according to the previous study (Cheng et al., 2012, 2013). Chromosomal locations of the syntenic genes of *A. thaliana* AGPs (*AtAGPs*) and *BrAGPs* were

gathered from The Arabidopsis Information Resource (TAIR, <http://arabidopsis.org>) and BRAD, respectively. The syntenic relationships between or inside the genomes were illustrated by Circos 5.05 (Krzywinski et al., 2009).

A set of 458 core eukaryotic genes and 458 random genes in *A. thaliana* (required for retained proportion analysis), were downloaded from CEGMA (<http://korflab.ucdavis.edu/Datasets/cegma>) (Parra et al., 2007) and were used to search for syntenic genes in the BRAD database. The orthologous genes of several *AtAGPs* in *A. lyrata*, *Capsella rubella*, *Camelina sativa*, *B. napus*, *B. oleracea*, *Thellungiella halophila*, *Thellungiella salsuginea*, *Sisymbrium irio*, *Schrenkiella parvula*, and *Aethionema arabicum* were detected through syntenic analysis in BRAD.

Phylogenetic and Protein Motifs Analysis

Multiple sequence alignment was performed by ClustalW (<http://www.ch.embnet.org/software/ClustalW.html>) in standard settings with some modification. Phylogenetic trees were constructed with both neighbor-joining (NJ) and maximum likelihood (ML) method with 1,000 bootstraps using MEGA6 software (Tamura et al., 2013). Poisson model were used in Model/Method and Uniform rates were chosen in Rates among sites. Treatment of Gaps/Missing Data was selected as Pairwise deletion. The clans consisting only one orthologous gene were treated as orphan genes, while other clans were named as clades in sequence. Motif analysis was performed in Multiple Em for Motif Elicitation (MEME) (<http://meme-suite.org/tools/meme>) (Bailey et al., 2009) for the identification of the short conserved motifs in the subfamilies of classical AGPs and AG peptides.

Molecular Evolutionary Rate Analysis

The molecular evolutionary rates between orthologous genes were estimated by calculating the ratio of nonsynonymous substitution rate (K_a) and synonymous substitution rate (K_s) between orthologous gene pairs. The protein sequences were aligned by ClustalW and the resulting alignment was used as a reference to align the nucleotide sequences. The K_a/K_s value was then determined using the YN86 module in PAML 4.7 (Yang, 2007). Significant differences were calculated by one-way ANOVA tests in SPSS 19 (IBM, USA).

RESULTS

Identification of AGPs in the *Brassica rapa* Genome

Initially, 76 proteins longer than 90 amino acids were retrieved in the *B. rapa* genome, with biased amino acid compositions of

at least 50% PAST. Similarly, 115 potential AG peptides were identified by searching for proteins between 50 and 90 amino acids in length with biased amino acid compositions of at least 35% PAST (Table 1).

Signal peptides are important for glycosylation of AGPs in the endoplasmic reticulum, while GPI anchor is required for anchoring AGPs to the plasma membrane. Therefore, these 191 proteins were further screened for the presence of a signal peptide and a GPI anchor. Fifty eight proteins longer than 90 amino acids and 24 proteins shorter than 90 amino acids were identified as they contain both the signal peptide and the GPI anchor (Table 1). By eliminating the proteins contained other domains such as fascilin-like domain and early nodulin-like domain, 22 classical AGPs, 24 AG peptides, and three lys-rich AGPs were identified as “high confidence” AGPs (Tables 1, 2). Additionally, the orthologous genes of all the annotated classical AGPs, AG peptides and lys-rich AGPs in *A. thaliana* were screened in the proteome of *B. rapa*. With this approach, 15 additional AGPs have been revealed as AGP candidates, including 11 classical AGPs, and four AG peptides (Table 2).

All AGPs in *B. rapa* were named after their orthologous proteins in *A. thaliana*. Bra038390 (81 amino acids in length) was identified as the ortholog of a classical AGP—*AtAGP10*. Hence, for convenience, Bra038390 was regarded as a classical AGP but not as an AG peptide. Moreover, sequence homology analysis did not result in any orthologous proteins in the annotated *AtAGPs* for three *BrAGPs*, Bra028633, Bra035584, and Bra036401, so they were designated as *BrAGP59*, *BrAGP60*, and *BrAGP61*, respectively. Overall, apart from the chimeric proteins, 64 AGPs were detected in *B. rapa*, including 33 classical AGPs, 28 AG peptides and three lys-rich AGPs (Table 2).

Retained Proportion Analysis

Most of the *BrAGPs* were distributed on chromosomes, with the exception of *BrAGP9.2*, *BrAGP11.2*, and *BrAGP12.2*, which were located on scaffolds. On average, the chromosomes contained four to nine *BrAGPs*, with chromosome A06 harboring only three and chromosome A03 having up to 11. The 24 conserved collinear blocks in ancestral karyotype were used to identify the syntenic relationship between *BrAGPs* and their corresponding orthologous genes in *A. thaliana*. The result illustrated that all *BrAGPs* are located in the same blocks as their *A. thaliana* orthologous genes (Figure 1), suggesting that the expansion of AGP gene family in *B. rapa* was depended on the WGT event.

No orthologous genes of *AtAGP5*, *AtAGP7*, *AtAGP19*, *AtAGP41*, *AtAGP51*, and *AtAGP56* were identified in *B. rapa*. To investigate whether these *AtAGPs* were newly acquired in

TABLE 1 | AGPs and AG peptides identified in the *Brassica rapa* genome based on amino acid composition, length and specific domains.

Search criteria	Total	SP ^a & GPI ^b	Classical AGPs	AG peptides	Lys-rich AGPs	Others
PAST ^c >50% & length >90	76	58	22	0	3	33
PAST >35% & 50 < length <90	115	24	0	24	0	0

^aSP, secretion peptide. N-terminal secretion signal peptides were predicted by SignalP; ^bGPI was predicted both by big-PI Plant Predictor and PSORT Prediction; ^cPAST, the proportion of Pro, Ala, Ser, and Thr.

TABLE 2 | Identification, characterization and classification of putative AGP encoding genes in the *Brassica rapa* genome.

Subfamily	<i>Arabidopsis thaliana</i>			<i>Brassica rapa</i>									
	Gene name	Gene ID	Block	Gene name ^a	Gene ID	PAST (%)	SP ^b	GPI ^c	AP/PA/SP/TP repeats	Amino acids	Chromosome	Sub-genome	Block
Classical	<i>AtAGP1</i>	At5g64310	X	<i>BrAGP1.1</i>	Bra024284	59	Yes	Yes	10/16/20/4	124	A06	LF	X
				<i>BrAGP1.2</i>	Bra031924	56	Yes	Yes	8/14/24/2	130	A02	MF1	X
	<i>AtAGP2</i>	At2g22470	I	<i>BrAGP2.1</i>	Bra038521	62	Yes	Yes	14/6/16/4	132	A09	LF	I
				<i>BrAGP2.2</i>	Bra030234	61	Yes	Yes	16/10/14/4	132	A04	MF1	I
	<i>AtAGP3</i>	At4g40090	U	<i>BrAGP3.1</i>	Bra011806	61	Yes	Yes	14/14/18/8	143	A01	LF	U
				<i>BrAGP3.2</i>	Bra010639	60	Yes	Yes	14/12/18/10	143	A08	MF2	U
	<i>AtAGP4</i>	At5g10430	R	<i>BrAGP4.1</i>	Bra009032	71	Yes	Yes	18/24/6/20	143	A10	LF	R
				<i>BrAGP4.2</i>	Bra006060	72	Yes	Yes	14/20/8/18	141	A03	MF1	R
				<i>BrAGP4.3</i>	Bra028587	71	Yes	Yes	20/20/6/16	139	A02	MF2	R
	<i>AtAGP5</i>	At1g35230	B										
	<i>AtAGP6</i>	At5g14380	R	<i>BrAGP6</i>	Bra008762	66	Yes	Yes	14/8/14/6	152	A10	LF	R
	<i>AtAGP7</i>	At5g65390	X										
	<i>AtAGP9</i>	At2g14890	H	<i>BrAGP9.1</i>	Bra013116	70	Yes	Yes	14/22/14/18	179	A03	MF2	H
				<i>BrAGP9.2</i>	Bra039829	67	Yes	Yes	14/18/16/16	166	Scaffold000178		
	<i>AtAGP10</i>	At4g09030	P	<i>BrAGP10.1</i>	Bra038390	62	Yes	No	6/2/8/6	81	A09	LF	P
				<i>BrAGP10.2</i>	Bra000670	65	Yes	Yes	12/8/12/10	126	A03	MF1	P
	<i>AtAGP11</i>	At3g01700	H	<i>BrAGP11.1</i>	Bra000995	59	Yes	Yes	16/4/8/4	136	A03	MF2	H
				<i>BrAGP11.2</i>	Bra040548	63	Yes	Yes	18/6/10/4	138	Scaffold000217		
	<i>AtAGP25</i>	At5g18690	R	<i>BrAGP25</i>	Bra002180	50	Yes	Yes	8/4/14/2	135	A10	LF	R
	<i>AtAGP26</i>	At2g47930	J	<i>BrAGP26</i>	Bra021448	52	Yes	Yes	4/2/20/2	140	A04	MF1	J
	<i>AtAGP27</i>	At3g06360	F	<i>BrAGP27</i>	Bra040224	55	Yes	Yes	8/6/8/0	126	A01	MF1	F
	<i>AtAGP50</i>	At1g24520	B	<i>BrAGP50.1</i>	Bra032796	46	Yes	Yes	4/0/8/2	119	A09	LF	B
				<i>BrAGP50.2</i>	Bra010966	48	Yes	Yes	6/2/8/2	117	A08	MF1	B
				<i>BrAGP50.3</i>	Bra012505	47	Yes	Yes	4/2/10/0	119	A07	MF2	B
	<i>AtAGP51</i>	At1g31250	B										
	<i>AtAGP52</i>	At1g63530	D	<i>BrAGP52</i>	Bra027649	46	No	No	8/24/16/24	420	A09	MF2	D
	<i>AtAGP53</i>	At1g63540	D	<i>BrAGP53</i>	Bra027648	46	Yes	No	12/16/14/32	618	A09	MF2	D
	<i>AtAGP54</i>	At2g28440	I	<i>BrAGP54.1</i>	Bra011948	53	Yes	Yes	12/18/54/2	234	A07	LF	I
				<i>BrAGP54.2</i>	Bra035700	44	Yes	Yes	2/8/24/4	186	A04	MF1	I
	<i>AtAGP55</i>	At2g45000	J	<i>BrAGP55</i>	Bra004882	59	No	Yes	46/38/26/22	784	A05	LF	J
	<i>AtAGP56</i>	At3g22070	F										
	<i>AtAGP57</i>	At3g45230	M	<i>BrAGP57</i>	Bra038294	43	Yes	Yes	8/8/24/2	168	A06	LF	M
	<i>AtAGP58</i>	At4g16980	U	<i>BrAGP58.1</i>	Bra040103	56	No	No	8/4/16/8	135	A01	LF	U
<i>BrAGP58.2</i>				Bra021074	58	No	Yes	12/2/16/4	130	A08	MF2	U	
<i>BrAGP59</i>				Bra028633	53	Yes	Yes	6/4/14/6	115	A02	MF2	R	
<i>BrAGP60</i>				Bra035584	60	Yes	Yes	24/24/50/18	339	A02	MF2	W	
<i>BrAGP61</i>	Bra036401	70	Yes	Yes	14/16/18/14	171	A07	LF	H				
AG-peptide	<i>AtAGP12</i>	At3g13520	F	<i>BrAGP12.1</i>	Bra027425	44	Yes	Yes	2/4/4/0	61	A05	LF	F
				<i>BrAGP12.2</i>	Bra039397	47	Yes	Yes	2/4/4/0	62	Scaffold000164		
	<i>AtAGP13</i>	At4g26320	U	<i>BrAGP13.1</i>	Bra026448	46	Yes	Yes	4/4/2/0	59	A01	LF	U
				<i>BrAGP13.2</i>	Bra019114	42	Yes	Yes	4/2/2/0	59	A03	MF1	U
	<i>AtAGP14</i>	At5g56540	W	<i>BrAGP14</i>	Bra002808	43	Yes	Yes	4/2/2/0	60	A10	LF	W
	<i>AtAGP15</i>	At5g11740	R	<i>BrAGP15.1</i>	Bra008940	53	Yes	Yes	4/2/4/0	64	A10	LF	R
				<i>BrAGP15.2</i>	Bra006115	53	Yes	Yes	2/2/6/0	64	A03	MF1	R
				<i>BrAGP15.3</i>	Bra023339	52	Yes	Yes	4/2/4/0	64	A02	MF2	R
	<i>AtAGP16</i>	At2g46330	J	<i>BrAGP16.1</i>	Bra004546	42	Yes	Yes	6/4/0/0	73	A05	LF	J
				<i>BrAGP16.2</i>	Bra000419	42	Yes	Yes	6/4/0/0	73	A03	MF2	J
	<i>AtAGP20</i>	At3g61640	N	<i>BrAGP20</i>	Bra014427	40	Yes	Yes	4/2/4/0	68	A04	MF1	N

(Continued)

TABLE 2 | Continued

Subfamily	<i>Arabidopsis thaliana</i>			<i>Brassica rapa</i>									
	Gene name	Gene ID	Block	Gene name ^a	Gene ID	PAST (%)	SP ^b	GPI ^c	AP/PA/SP/TP repeats	Amino acids	Chromosome	Sub-genome	Block
	<i>AtAGP21</i>	At1g55330	C	<i>BrAGP21.1</i>	Bra037993	46	Yes	Yes	4/4/2/0	59	A06	LF	C
				<i>BrAGP21.2</i>	Bra011914	46	Yes	Yes	4/4/2/0	59	A07	MF2	C
				<i>BrAGP21.3</i>	Bra030868	47	Yes	Yes	4/4/2/0	58	A08	MF1	C
	<i>AtAGP22</i>	At5g53250	W	<i>BrAGP22.1</i>	Bra003071	38	Yes	Yes	4/4/4/0	63	A10	LF	W
				<i>BrAGP22.2</i>	Bra029086	38	Yes	Yes	4/4/4/0	63	A03	MF1	W
				<i>BrAGP22.3</i>	Bra022641	38	Yes	Yes	4/4/4/0	63	A02	MF2	W
	<i>AtAGP23</i>	At3g57690	N	<i>BrAGP23.1</i>	Bra007349	46	No	Yes	4/6/0/0	61	A09	LF	N
				<i>BrAGP23.2</i>	Bra014611	44	No	Yes	4/6/0/0	61	A04	MF1	N
				<i>BrAGP23.3</i>	Bra003296	43	Yes	Yes	4/6/0/0	61	A07	MF2	N
	<i>AtAGP24</i>	At5g40730	S	<i>BrAGP24</i>	Bra025551	37	Yes	Yes	4/6/2/0	67	A04	LF	S
	<i>AtAGP40</i>	At3g20865	F	<i>BrAGP40.1</i>	Bra031236	45	Yes	Yes	4/4/2/0	64	A05	LF	F
				<i>BrAGP40.2</i>	Bra023919	47	Yes	Yes	4/4/2/0	64	A01	MF1	F
	<i>AtAGP41</i>	At5g24105	Q										
	<i>AtAGP42</i>	At1g51915	C	<i>BrAGP42</i>	Bra030409	30	Yes	No	0/2/2/0	77	A05	MF2	C
	<i>AtAGP43</i>	At2g41905	J	<i>BrAGP43.1</i>	Bra016902	44	Yes	Yes	4/6/0/0	61	A04	MF1	J
				<i>BrAGP43.2</i>	Bra000244	44	Yes	Yes	4/6/0/0	61	A03	MF2	J
	<i>AtAGP44</i>	At3g01730	F	<i>BrAGP44</i>	Bra000992	39	Yes	Yes	4/2/4/0	90	A03	MF2	F
	<i>AtAGP45</i>	At5g12880	R	<i>BrAGP45</i>	Bra008874	46	Yes	No	2/0/4/0	76	A10	LF	R
Lys-rich	<i>AtAGP17</i>	At2g23130	I	<i>BrAGP17</i>	Bra039184	68	Yes	Yes	44/42/56/12	305	A09	LF	I
	<i>AtAGP18</i>	At4g37450	U	<i>BrAGP18.1</i>	Bra011767	62	Yes	Yes	20/14/26/4	178	A01	LF	U
				<i>BrAGP18.2</i>	Bra017823	65	Yes	Yes	30/30/28/10	202	A03	MF1	U
	<i>AtAGP19</i>	At1g68725	E										

^aBoldface indicates a protein that is treated as "high confidence" AGPs; ^bSignal peptide; ^cGPI was predicted both by big-PI Plant Predictor and PSORT Prediction.

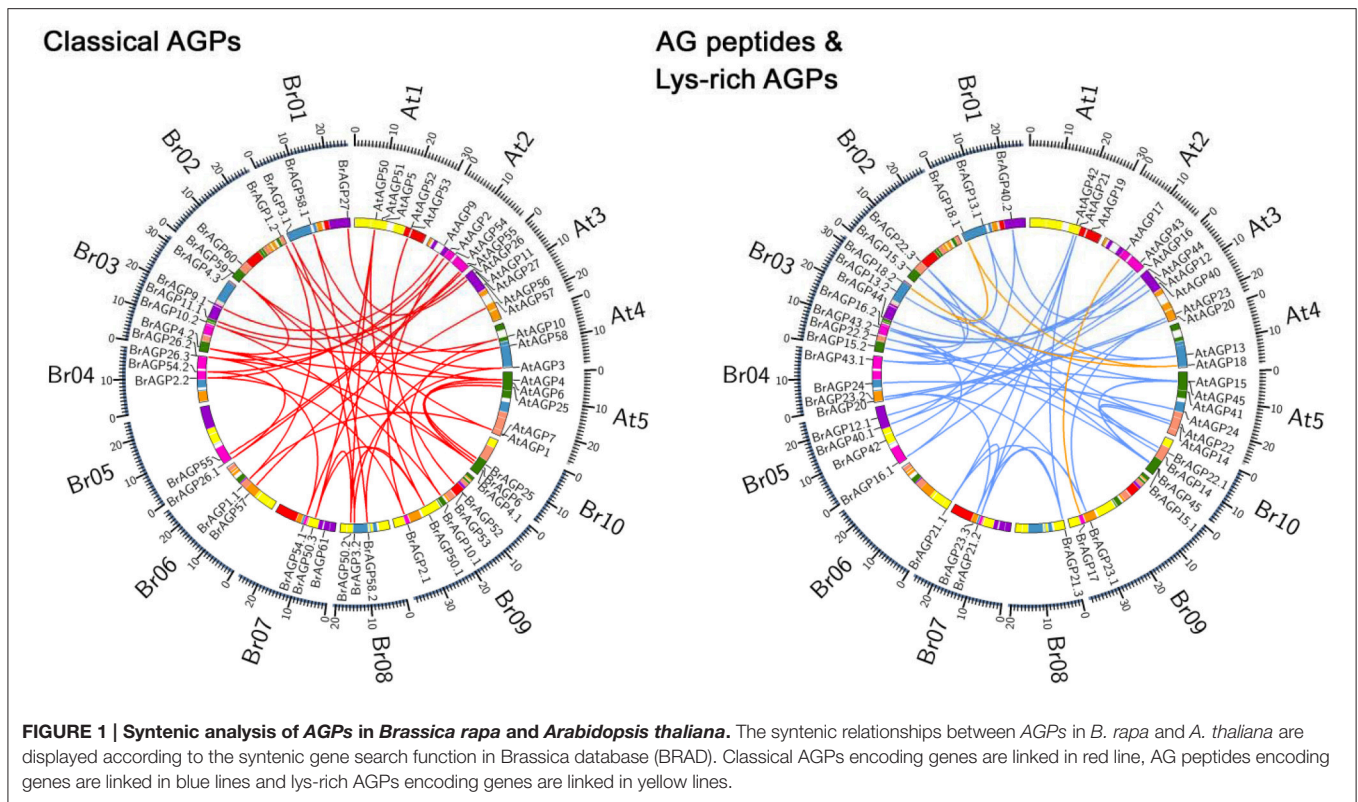
A. thaliana or were lost in *B. rapa*, their orthologous genes in other sequenced *Brassica* species were analyzed. *Brassica* species can be classified into three lineages by spectrum analysis, and most sequenced species were concentrated on Lineage I and Lineage II. Specifically, *A. thaliana*, *A. lyrata*, *C. rubella*, and *C. sativa* belonged to Lineage I, while *B. rapa*, *B. napus*, *B. oleracea*, *T. halophila*, *T. salsuginea*, *S. irio*, and *S. parvula* were in Lineage II (Koch and German, 2013). Analysis of these orthologous genes illustrated that *AtAGP5*, *AtAGP7*, *AtAGP19*, *AtAGP41*, and *AtAGP56* exist not only in the species belonging to Lineage I, but also in the species belonging to Lineage II as well as in the primitive *Brassica* species, *A. arabicum*. This finding indicated that these genes may be lost in *B. rapa* (Supplementary data 1: Figure S1). The orthologs of *AtAGP51* could only be detected in species in Lineage I, indicating that this gene arose after the separation of *A. thaliana* and *B. rapa*. Based on this result, the retained proportion of AGPs in *B. rapa* was calculated. The AGP gene family had a retained proportion of 50%, which was much higher than that of randomly selected genes (45%) and similar to that of core eukaryotic genes (52%) (Figure 2A). Particularly, classical AGPs, AG peptides, and lys-rich AGPs had a proportion of 48, 58, and 33%, respectively (Figure 2B).

As illustrated in the whole genome level, *BrAGPs* also displayed differentially-retained proportions among the three subgenomes, namely least fractionated subgenome (LF),

medium fractionated subgenome (MF1) and most fractionated subgenome (MF2) of *B. rapa* (Figure 2C). The AGP gene family contained more genes of LF subgenome (63%) than those of random selected genes (59%) and core eukaryotic genes (59%), while it had similar proportion of core eukaryotic genes in the MF1 and MF2 subgenomes for AGP genes and core eukaryotic genes, respectively. This indicated that the highly retained proportion of AGPs is mainly attributed to the gene reservation in LF.

Phylogenetic and Protein Structure Analysis

Full length protein sequences of all identified *BrAGPs* and their orthologs in *A. thaliana* were used to construct phylogenetic trees. The topologies of NJ and ML trees were mainly consistent, thus only the NJ trees are presented here (Figure 3). The 55 classical AGPs in *B. rapa* and *A. thaliana* were classified into six clades and three orphan genes with high bootstrap support (Figure 3). Motif analysis showed that the AGPs in Clade I mainly have two types of structural features. AGP2, AGP3, AGP4, and AGP7 contained motif 1, motif 4 and motif 3, while AGP1, AGP5, and AGP10 included motif 1 and motif 4. Clade II was constituted of AGP54 and AGP57 with motif 1 and motif 2. Clade III had three members, AGP6, AGP11, and AGP50, in which AGP6 and AGP11 had motif 1 and motif 4, similar to some of



the members in Clade I, while AGP50 only had motif 1. AGP9, AGP58, and AGP62 belonged to Clade IV with AGP58 having only motif 4, while AGP9 and AGP62 having motif 1 and motif 4. Clade V was composed of AGP25, AGP26, AGP27, which only hold motif 1. Tandem gene pair, AGP52 and AGP53 contained motif 4. Phylogenetic analysis indicated that the tandem repeat event occurred after the separation of *B. rapa* and *A. thaliana*, as the AGPs belonging to *B. rapa* had a closer resemblance to each other than to the orthologs in *A. thaliana*.

Phylogenetic analysis of AG peptides led to the division of three clades and five orphan genes, and almost all AG peptides had motif 6 and motif 7 (Figure 4). Four members of Clade I, AGP12, AGP13, AGP14, and AGP21, had motif 10 in addition to motif 6 and motif 7. AGP16, AGP20, AGP22, and AGP41 in Clade II all contained each motif 6, motif 7 and motif 8. Clade III included AGP23 and AGP43, which all had motif 8.

Estimation of the Molecular Evolutionary Rates of AGPs

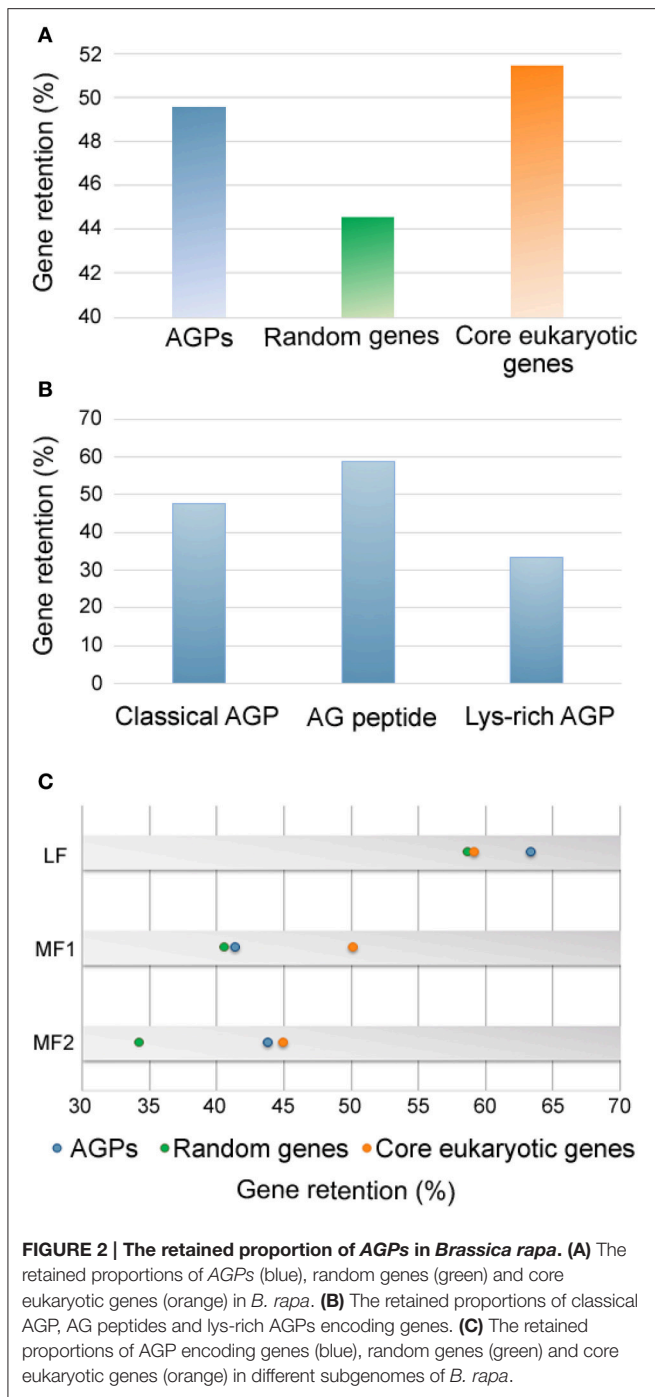
Estimation of molecular evolution rates forms an important basis for our understanding of the evolution processes in *B. rapa*. YN86 module was used for estimating the molecular evolutionary rates of the orthologous gene pairs among AGPs (Supplementary data 1: Table S3). All three types of AGPs encoding genes were evolved at a Ka/Ks value lower than 1, indicating that they all have evolved through purifying selection (Figure 5A). Classical AGPs showed an average Ka/Ks value of 0.30, AG peptides exhibited and Lys-rich AGPs were evolved at an average Ka/Ks value of

0.20 and 0.38, respectively. Statistical analysis illustrated that classical AGPs have evolved faster than AG peptides, but there is no significant differences existed between classical AGPs and lys-rich AGPs. Furthermore, the average molecular evolutionary rates of each clade were compared. In classical AGPs, Clade VI had a significantly higher Ka/Ks value than the other five clades (Figure 5B). In AG peptides, Clade I, Clade II, and Clade III shared a similar molecular evolutionary rate, while the orphan genes evolved at a faster rate of evolution (Figure 5C).

Differential Expression of *BrAGPs*

Half of the classical *BrAGPs* showed expression in all of the tissues detected, including root, stem, leaf, inflorescence and siliqua. Seven AGPs were expressed in more than one but not all tissues. *BrAGP6*, *BrAGP11.1*, *BrAGP11.2*, and *BrAGP54.1* expression was restricted in inflorescence (Figure 6A). Eleven of the 24 AG peptides encoding genes were expressed in all tissues with *BrAGP12.2*, *BrAGP13.2*, *BrAGP15.1*, *BrAGP15.2*, *BrAGP22.1*, and *BrAGP22.2* displaying similar expression levels among different tissues. Half of the genes encoding AG peptides were expressed in different tissues, of which *BrAGP15.3* was inflorescence-specific and *BrAGP14* was expressed exclusively in the stem. The expression of *BrAGP44* was not detected in any of the tissues (Figure 6A).

Among the three lys-rich AGPs encoding genes, *BrAGP18.1* and *BrAGP18.2* were expressed in all tissues, while *BrAGP17* was only detectable in stem, inflorescence and siliqua (Figure 6A). Interestingly, most paralogous gene pairs



displayed different expression patterns. Specifically, from the classical AGPs group, the gene pairs of *BrAGP1.1* and *BrAGP1.2*, *BrAGP2.1* and *BrAGP2.2*, *BrAGP3.1* and *BrAGP3.2*, *BrAGP4.1* and *BrAGP4.2/BrAGP4.3* had differential expression patterns. Likewise, the paralogous gene pairs of *BrAGP12.1* and *BrAGP12.2*, *BrAGP13.1* and *BrAGP13.2*, *BrAGP15.1*, *BrAGP15.2* and *BrAGP15.3*, *BrAGP16.1* and *BrAGP16.2*, *BrAGP21.1*, *BrAGP21.2* and *BrAGP21.3*, *BrAGP40.1* and *BrAGP40.2*, *BrAGP43.1* and *BrAGP43.2*,

encoding for AG peptides, showed differential expression patterns.

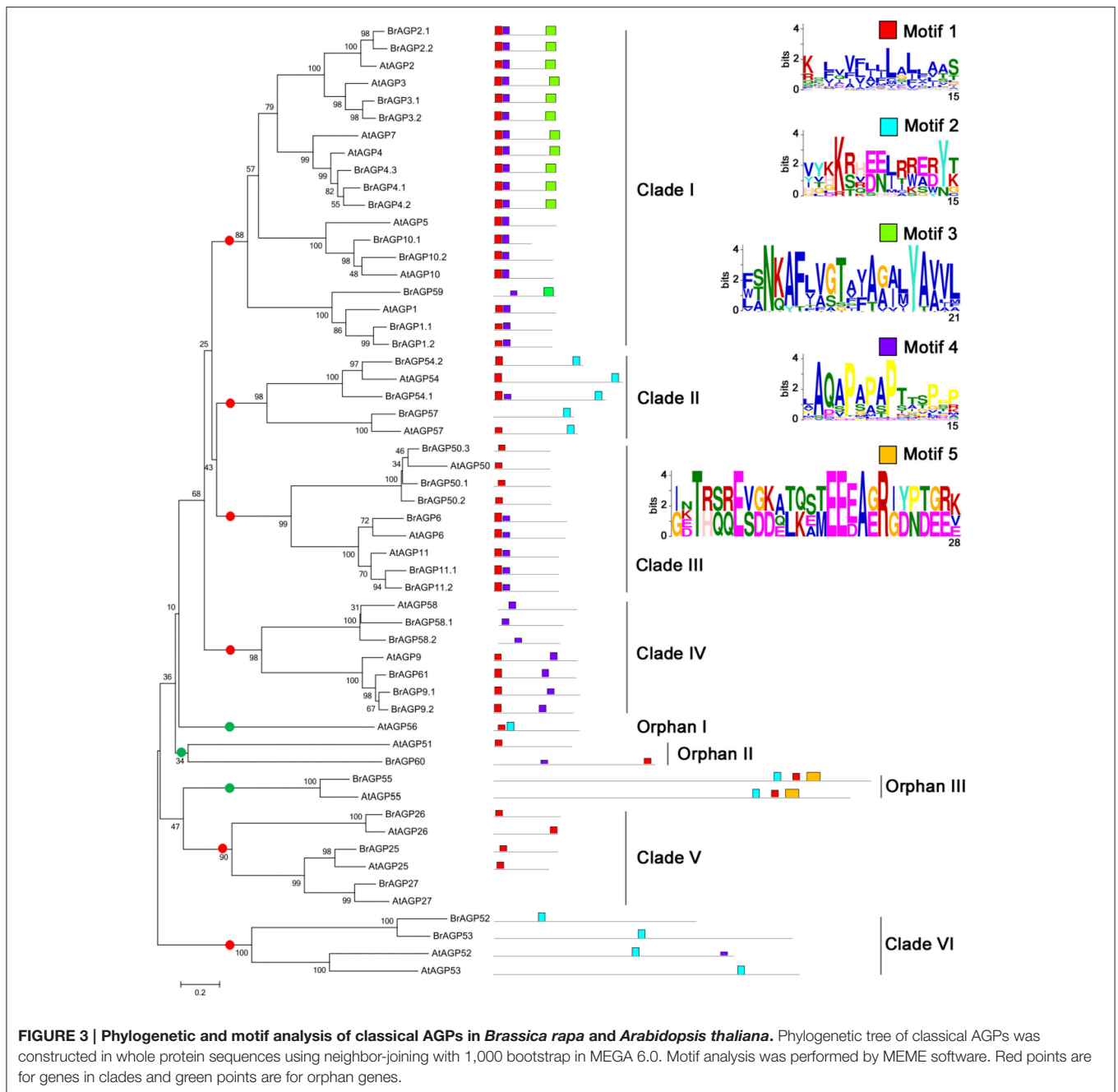
For the five inflorescence-specific genes, including four classical AGPs and one AG peptide encoding genes, their expression were further analyzed among different developmental stages of pollen in subdivided flower buds from the “*Bcaj97-01A/B*”GMS line (Figure 6B). These five *BrAGPs* were expressed in the fertile but not in sterile flower buds. *BrAGP6* and *BrAGP11.2* were only expressed at Stage I which corresponds to the mature pollen stage, while *BrAGP11.1*, *BrAGP15.3*, and *BrAGP54.1* were detected at Stage III to V, which means that they are expressed throughout the uninucleate pollen stage to the mature pollen stage. Since some genes expressed in mature pollen usually continue to be expressed in germinating pollen and pollen tubes, the expression levels of the five *BrAGPs* were investigated in pistils at three different time points after pollination in a fertile line. *BrAGP11.1* and *BrAGP54.1* were found to be expressed in pistils during the whole fertilization processes; however the other three genes were not detected in the pistils (Figure 6C).

Effects of Exogenous GA, MeJA, and ABA on the Expression of *BrAGPs* in the Leaf

Phytohormones serve as key signals for the plants to respond to environmental stimuli (Wania et al., 2016). In order to investigate whether *BrAGPs* were involved in phytohormone responses, we chose 15 classical AGPs, 13 AG peptides and two lys-rich AGPs, which according to our qRT-PCR displayed expression in the leaf. Three vital phytohormones, ABA, GA, and MeJA, involved in abiotic and biotic stress resistance, were used in this study to investigate the leaf *BrAGPs* expression in response to phytohormone treatments.

Following ABA treatment, 22 *BrAGPs* were up-regulated, including 13 classical AGPs, eight AG peptides and one lys-rich AGPs encoding genes (Figure 7A). Expression of 11 classical AGPs, four AG peptides, and a lys-rich AGPs encoding genes were markedly increased at 4 HAT and then recovered at 12 HAT; whereas, the expression levels of two classical AGPs and three AG peptides encoding genes were increased at 12 HAT. Eight *BrAGPs*, including two classical AGPs, five AG peptides, and one lys-rich AGPs encoding genes were down-regulated upon ABA treatment. Their expression levels decreased mainly at 4 HAT and didn't recover at 12 HAT, while *BrAGP22.3* decreased at 12 HAT. Except *BrAGP2.1* and *BrAGP2.2* that shared similar responses, most paralogous gene pairs displayed different expression patterns after ABA treatment.

Up-regulation of 12 classical AGPs, seven AG peptides, and one lys-rich AGPs genes was observed after GA treatment (Figure 7B). The expression levels of three classical AGPs encoding genes were increased at 4 HAT and subsequently at 12 HAT. Three classical AGPs and three AG peptides encoding genes were also up-regulated at 4 HAT but their transcript levels dropped back at 12 HAT. There was no effect on the expression of genes encoding for six classical AGPs, four AG peptides, and one lys-rich AGP, until 12 HAT. In contrast, the expression of encoding genes for three classical AGPs, six AG peptides, and one lys-rich AGP was down-regulated. The expression level of

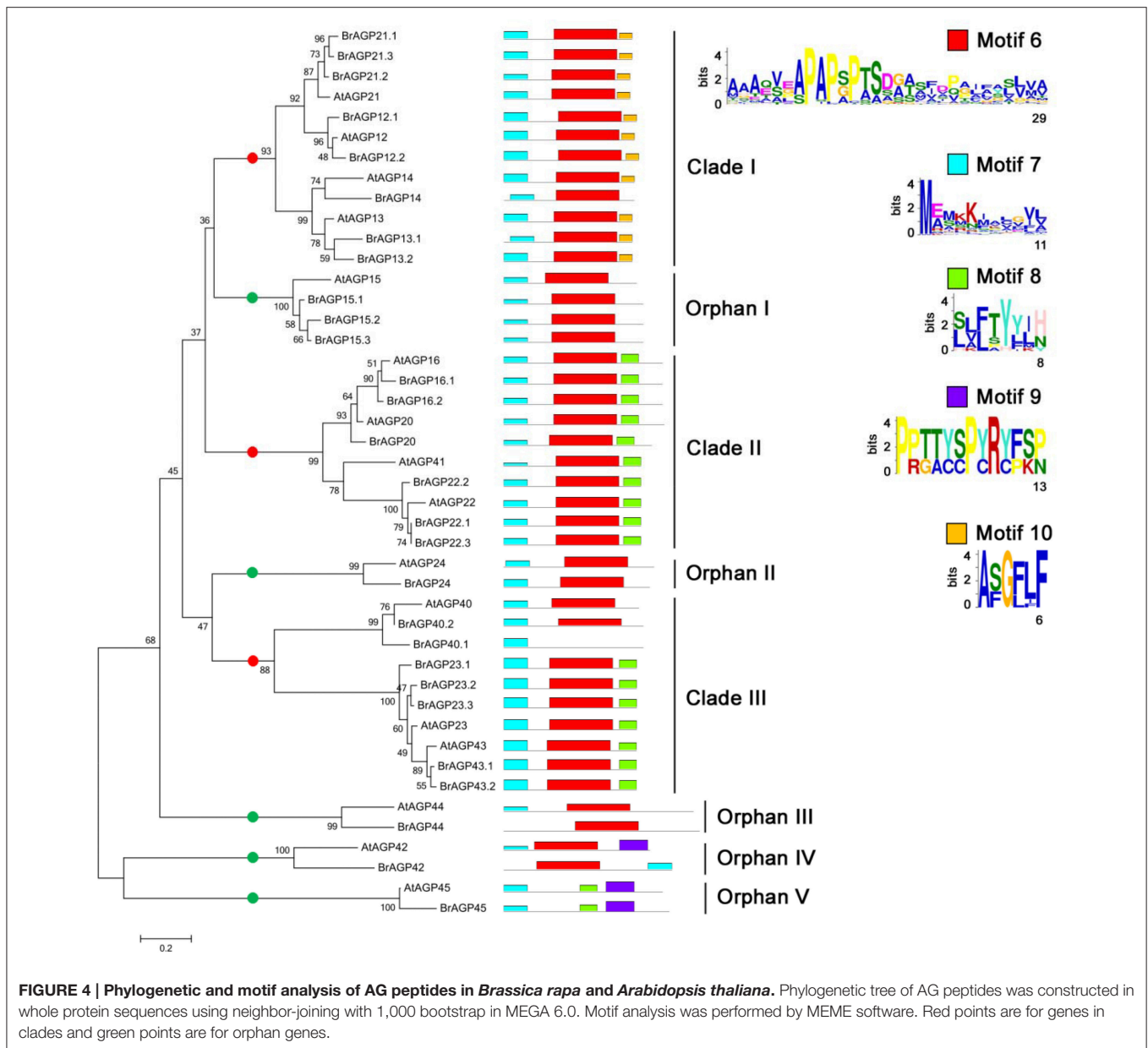


BrAGP2.2 decreased at 4 HAT and recovered at 12 HAT, while *BrAGP12.1* was down-regulated at 12 HAT. The remaining genes were down-regulated at 4 HAT and sustained low expression levels at 12 HAT. Although the expression of most paralogous gene pairs was different, the response pattern between *BrAGP40.1* and *BrAGP40.2*, and between *BrAGP22.2* and *BrAGP22.3*, was similar (Figure 7B).

Nineteen *BrAGPs* were up-regulated in MeJA treatment. Two classical AGPs and two AG peptides encoding genes were induced at 12 HAT, while 10 classical AGPs, four AG peptides and one lys-rich AGPs encoding genes, were up-regulated at 4

HAT and recovered at 12 HAT (Figure 7C). In addition, 11 AGPs were down-regulated, following MeJA treatment. One classical AGP and two AG peptides were down-regulated at 4 HAT and recovered at 12 HAT, while the expression levels of the encoding genes for two classical AGPs, five AG peptides, and one lys-rich AGP decreased at 4 HAT and didn't recover at 12 HAT. *BrAGP1.1* and *BrAGP1.2* showed the same response to ABA, but the rest of the paralogous gene pairs displayed different expression patterns.

Interestingly, the ABA-, GA-, and MeJA-induced expression of *BrAGPs* were largely overlapped. Seventeen *BrAGPs* were up-regulated by all three phytohormones, including genes encoding



10 classical AGPs, six AG peptides, and one lys-rich AGPs. In addition, *BrAGP2.2* and *BrAGP26* were up-regulated by both ABA and MeJA, and *BrAGP13.1* was up-regulated both by ABA and GA (Figure 7D). Five AG peptides encoding genes and one lys-rich AGP encoding gene were down-regulated by all three phytohormones. However, *BrAGP3.2* and *BrAGP60* were solely down-regulated by ABA and MeJA, the expression levels of *BrAGP10.1* and *BrAGP12.2* decreased only in response to GA and MeJA, respectively (Figure 7E).

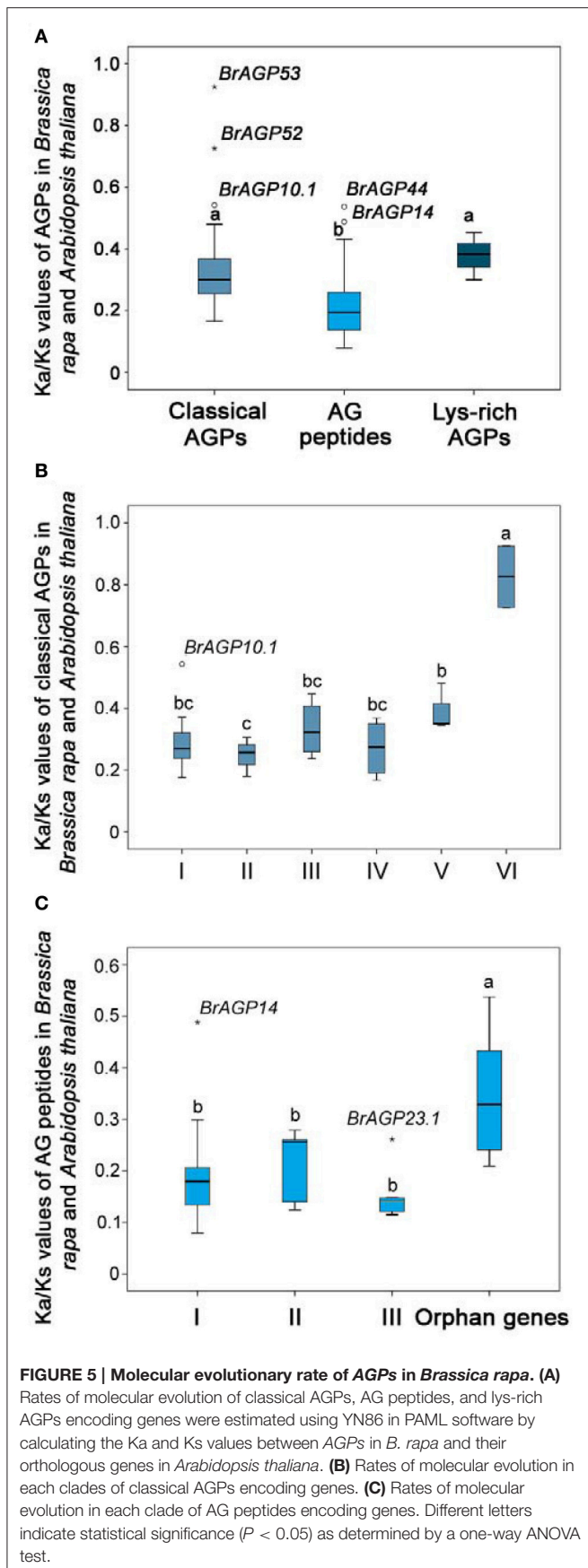
Comparison of the gene expression induction kinetics of different hormones revealed that the overlap between the response to ABA and MeJA is much bigger than that between ABA and GA, or MeJA and GA (Figure 7). In general, most of the transcriptional response of *BrAGPs* to ABA and MeJA treatments

was rapid and transient with expression changes detected at 4 HAT, while these of GA were relatively slow (expression changes were not detectable until 12 HAT).

DISCUSSION

The High Retained Proportions of AGPs in *B. rapa* May Be Attributed to Subfunctionalization

AGPs are characterized by the repetitive nature of their protein backbones with less than 40% similarity between them (Schultz et al., 2002). The low similarity makes it difficult to be identified using the basic local alignment search tool (BLAST) or



profile hidden Markov model (HMM) methods. “Hyp contiguity hypothesis” has provided a guidance to identify AGPs by Perl scripts (Kieliszewski and Lamport, 1994). However, it may lead to variable results if different programs and standards are employed (Schultz et al., 2002; Showalter et al., 2010). Here, based on the threshold values set in Showalter et al. (2010), a whole genome wide screening of classical AGPs, AG peptides, and lys-rich AGPs was performed in *B. rapa*, resulting in the identification of 64 *BrAGPs* including 49 with high confidence.

B. rapa contains more AGPs than *A. thaliana*, especially classical AGPs and AG peptides that is mainly attributed to the WGT event. However, we could not identify orthologs for three *BrAGPs* in *A. thaliana*, and orthologs for six *AtAGPs* in *B. rapa*. Ortholog comparison with closely related *Brassica* species revealed that five *AtAGPs* are lost in *B. rapa*, while one *AtAGP* and three *BrAGPs* were independently evolved after the separation of *A. thaliana* and *B. rapa*. These results indicate that besides the WGT event, other processes, for instance, dosage effect, might have also contributed to the evolution of the AGP family. In this study we have shown that classical AGPs, AG peptides and lys-rich AGPs are highly represented in the genome of *B. rapa*. Similar results were also reported by two previous studies where 33 FLAs and 52 ENODLs were identified in *B. rapa*. The corresponding retained proportions for these two types of chimeric AGPs were described to be 52 and 67%, respectively (Jun and Xiaoming, 2012; Li et al., 2013).

According to previous study, the over retained genes or gene families were always dosage sensitive or usually undergone subfunctionalization and/or neofunctionalization (Kim et al., 2014). In our investigation of the expression of *BrAGPs*, we revealed that 4/5 “high confidence” classical AGP and 6/7 AG peptide paralogous gene pairs had distinct expression patterns in different tissues. Such results indicated that the high retained proportions of classical AGPs and AG peptides are probably due to subfunctionalization and/or neofunctionalization, but not due to the dosage effect, which may yield increased expression of a gene (Edger and Pires, 2009). However, the result of molecular evolutionary rate analysis indicated that these AGPs were all under purifying selection, illustrating that neofunctionalization may have hardly occurred in these proteins. We thus suggest that subfunctionalization is the most likely mechanism that facilitated the retention of duplicated genes in classical AGPs and AG peptides. Unfortunately, in the previous studies on FLAs and ENODLs, the paralogous gene pairs in *B. rapa* and *A. thaliana* were treated as one gene for their gene expression pattern (Jun and Xiaoming, 2012; Li et al., 2013). Therefore, we were unable to verify whether the high retained proportion of the chimeric AGPs are also depended on subfunctionalization. Relevant research work should be focused on this issue in the future.

Implication of a Subset of AGPs in Sexual Reproduction

Comprehensive studies on gene expression can provide useful information for predicting gene function. In this study, we first investigated the expression of *BrAGPs* in five different tissues and found tissue-specific and non-tissue specific genes. To identify candidate AGPs that may be involved in pollen

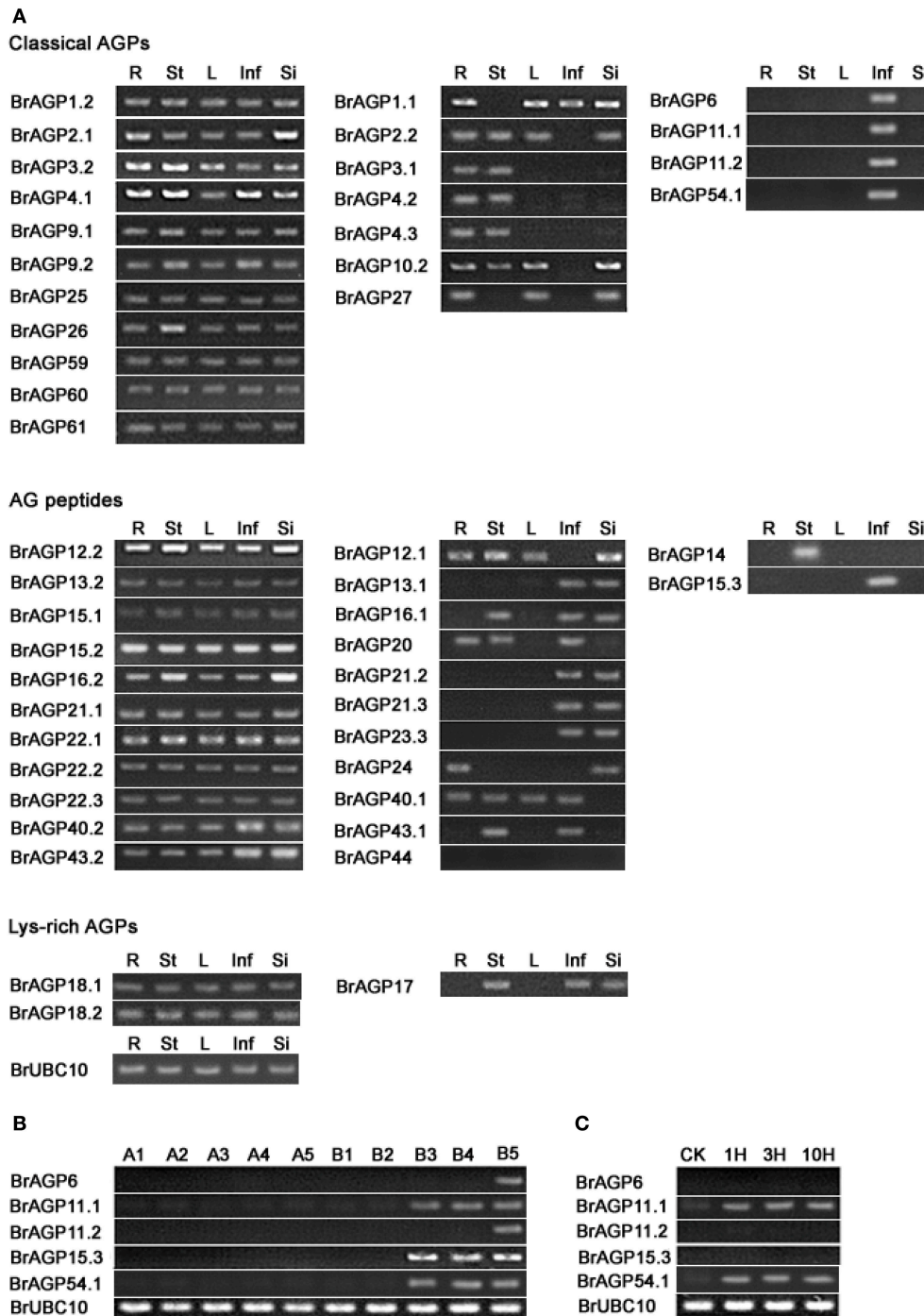
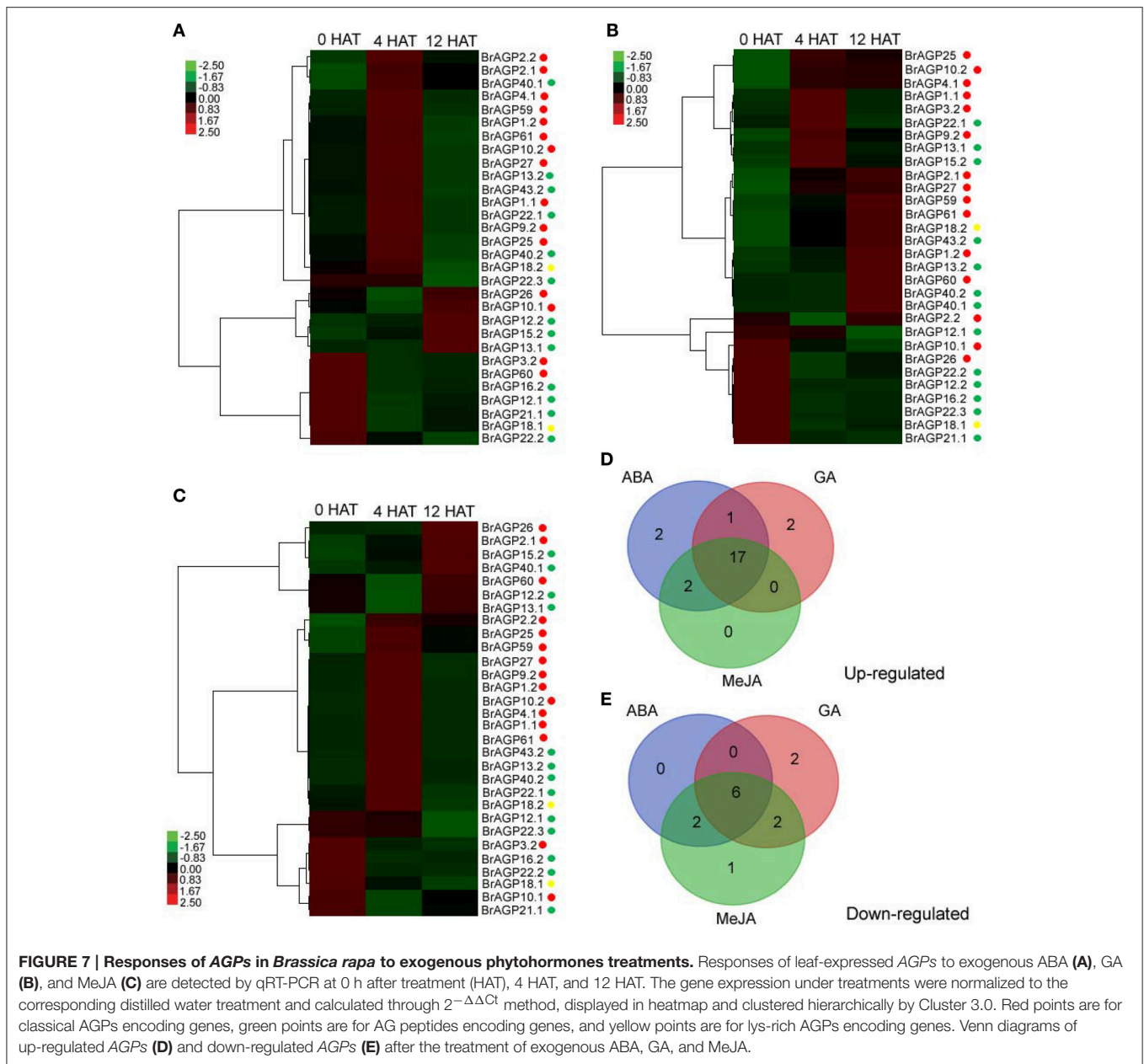


FIGURE 6 | Differential expression of AGPs in *Brassica rapa*. (A) Expression levels of AGPs in root (R), stem (St), leaf (L), inflorescence (Inf), and silique (Si) are detected by RT-PCR. (B) Expression of inflorescence-specific AGPs in flower buds of fertile line "*Bcajh97-01B*" and sterile line "*Bcajh97-01A*." B1–B5 are five stages of flower buds in "*Bcajh97-01B*," which are corresponding to five developmental stages of pollen; A1–A5 indicate five stages of flower buds in "*Bcajh97-01A*," which have the similar longitudinal diameter to the corresponding flower buds in "*Bcajh97-01B*." (C) Expression of inflorescence-specific AGPs in unpollinated pistils (CK) and pollinated pistils at 1 h after pollination (1 HAP), 3 HAP, and 10 HAP.

development, those inflorescence-specific *BrAGPs* including four classical AGPs encoding gene *BrAGP6*, *BrAGP11.1*, *BrAGP11.2*, *BrAGP54.1*, and one AG peptide encoding gene *BrAGP15.3*,

were further investigated for their in male fertile/sterile flower buds at five different developmental stages. All these genes were specifically expressed in fertile flower buds at the late



developmental stages, indicating a role in pollen formation or pollen function. Some genes expressed in mature pollen usually continued to be expressed in germinating pollen and pollen tubes, showing that they are specifically involved in the regulation of the subsequent pollination and fertilization processes (Pina et al., 2005; Becker and Feijo, 2007; Wang et al., 2008). We also detected the expression of these genes in pistils at three different time points after pollination. *BrAGP11.1* and *BrAGP54.1* were shown to be expressed in pistils during the whole fertilization processes, which indicated that they might also function in subsequent sexual reproductive events. A previous study has demonstrated that *AtAGP6* and *AtAGP11* are the only two pollen-specific classical AGPs and

are essential for pollen and pollen tube function in *A. thaliana* (Pereira et al., 2006; Levitin et al., 2008). Recent studies using qRT-PCR and *in situ* hybridization in our laboratory also revealed that the homologous gene *BrAGP6* (formerly named as *BcMF18*) was specifically expressed in pollen (unpublished data), and *BrAGP11.1* (or *BcMF8*) was expressed in the developing pollen and the pollen tube (Huang et al., 2008b; Lin et al., 2014). The functional disruption of *BrAGP6* by antisense RNA technology caused collapse of the pollen due to the absence of cellular content and nucleus, and failure of intine layer formation due to lack of cellulose deposition (unpublished data). The inhibition of *BrAGP11.1* resulted in slipper-shaped and bilaterally sunken pollen with abnormal intine development

and aperture formation, an arrest of pollen germination and unstable pollen tube formation (Lin et al., 2014). Although the functions of *BrAGP15.3* and *BrAGP54.1* are not fully understood, their expression patterns are similar to the known pollen-related AGP genes. Therefore, we suggest that these two genes may play important roles in pollen development and/or fertilization. Our results also revealed that only a small number of *BrAGPs* was inflorescence-specific. This is in agreement with a previous expression analysis of the FLAs and ENODLs in *B. rapa*, which also showed that only two *BrENODL* genes and five *FLA* genes were specifically expressed in inflorescence (Jun and Xiaoming, 2012; Li et al., 2013). Similarly, in *A. thaliana*, only a subset of *AGP* genes is expressed in pollen grain and pollen tubes (Pereira et al., 2006). Therefore, we speculate that, despite of the diversity of *AGPs*, only a small number of *AGPs* are endowed with a specific and defined role in male sexual reproduction in *B. rapa* as those in *A. thaliana*.

Novel Roles of *BrAGPs* in Phytohormone Signaling

AGPs are also implicated in signal recognition and transduction (Suzuki et al., 2002; van Hengel and Roberts, 2003), although the exact mechanisms are still elusive. The mRNA level of *AtAGP31* decreases at 8 h after ABA, MeJA, and wounding treatment (Liu and Mehdy, 2007), while *AtAGP30* specifically responds to ABA treatment (van Hengel and Roberts, 2003). Moreover, a mutation of this gene showed suppression of the ABA-induced delay in germination and alteration of the expression of some ABA-regulated genes (van Hengel and Roberts, 2003). In rice, two classical *AGPs*, *OsAGP1*, and *OsAGP15* were significantly up-regulated under the ABA treatment, indicating that they may response to ABA (Ma and Zhao, 2010). Moreover, *AGPs* are also involved in response to GA, for example, *AGPs* are suggested to play important roles in GA-induced α -amylase production in barley aleurone cells (Suzuki et al., 2002). The ATH1 microarray analysis results (GEO accession number: GSE39384) also revealed that many *AGP*-encoding genes are regulated by ABA, GA, and MeJA in *A. thaliana* (Supplementary data 2: Table S4). A similar trend was observed for some homologous *AGPs* between *B. rapa* and *A. thaliana*. For example, *AGP1* and *AGP2* were up-regulated in *A. thaliana* after 3 h treated by ABA, while their orthologs in *B. rapa* were also up-regulated at 4 HAT by ABA (Figure 7). The conserved response to the phytohormone treatments in different plant species indicates that *AGPs* may be important functional genes for phytohormone signaling pathways. However, many homologous *BrAGP* genes show diverse change trends under phytohormone treatments with their *A. thaliana* homologs. This result demonstrated that the differentiation of the *AGP* genes might be intended to acclimatize the plant to specific environmental changes. In our study, 17 *BrAGPs* were up-regulated and six were down-regulated by all three types of phytohormones, ABA, GA, and MeJA treatments. These genes comprise the majority (approximately 76%) of *AGPs* which are expressed in the leaf of *B. rapa*. These findings indicate that quite a few *AGPs* might be directly

associated or indirectly interact with the hormone signaling network.

The molecular mechanism by which *AGPs* are involved in phytohormone signaling remains to be elucidated. The relative rapid expression of *AGPs* under exogenous phytohormone treatments suggested that they may function in maintaining the organism's homeostasis rather than solely being structural components. In plants, sugar levels reflect the plant's physiological conditions and sugar signaling is closely associated with multiple phytohormone signaling cascades and abiotic stress signaling (Smeekens, 2000; Gibson, 2005; Rolland et al., 2006). Interestingly, sugar can induce ABA in the sugar-ABA signaling cascade (Laby et al., 2000). Changes in the expression of each *BrAGP* gene by exogenous ABA treatment indicates that the *AGPs* may participate in the signaling cascade by the dynamic release and uptake of glycosyl radicals. It was clear that there is complex cross-talk among plant hormones, which are important in plant growth regulation and defense responses (Kohli et al., 2013; Wania et al., 2016). In this study, we have shown that there is a large overlap of the *BrAGPs* induced or suppressed by ABA, GA, and MeJA treatments. It indicates that these *BrAGPs* might participate in different signaling pathways or in overlapping processes that are controlled by phytohormonal interactions.

CONCLUSION

We identified, for the first time, 33 classical *AGPs*, 28 *AG* peptides, and three lys-rich *AGPs* in the genome of *B. rapa*. We also elucidated their genomic characteristics, protein structures, duplication status, molecular evolutionary rates and expression characteristics in different tissues as well as under phytohormone treatments. High retained proportions of *AGPs* were observed in *B. rapa*, which may be resulted from subfunctionalization. Similar expression patterns to the known pollen-related *BrAGPs*, were found in the remaining two genes with unknown function, *BrAGP15.3* and *BrAGP54.1*, indicating they may also be involved in pollen development and/or fertilization. Large numbers of *BrAGPs* displayed responses under the treatment of ABA, MeJA, and/or GA, suggesting that they may participate in the response to phytohormone signaling independently and/or cooperatively. In summary, this study has provided fundamental information for revealing the roles of the classical *AGPs*, *AG* peptides, and lys-rich *AGPs*, three important types of *AGPs* in *B. rapa*.

AUTHOR CONTRIBUTIONS

TH performed the identification of *AGPs*, expression analysis, and participated in drafting the manuscript. HD performed the protein structure and evolution analysis, and participated in manuscript preparation. JCu, ML, SL, and JCa participated in phytohormone treatments, expression analysis, identification of *AGPs*, research design, respectively. LH conceived the project, designed research, and revised the manuscript. All authors read and approved the final manuscript.

ACKNOWLEDGMENTS

We thank Associate Professor Zhong-Hua Chen and David Randall (Western Sydney University) for their comments on this manuscript. This work was supported by the National Natural Science Foundation of China (No. 31372078, No. 31572126), the Key Technology Innovation Team of Zhejiang Province (No. 2013TD05), and the Grand Science and

Technology Special Project of Zhejiang Province (No. 2016C02051-6-1).

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: <http://journal.frontiersin.org/article/10.3389/fpls.2017.00397/full#supplementary-material>

REFERENCES

- 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
- Becker, J. D., and Feijo, J. A. (2007). How many genes are needed to make a pollen tube? lessons from transcriptomics. *Ann. Bot.* 100, 1117–1123. doi: 10.1093/aob/mcm208
- Cheng, F., Mandakova, T., Wu, J., Xie, Q., Lysak, M. A., and Wang, X. (2013). Deciphering the diploid ancestral genome of the Mesohexaploid *Brassica rapa*. *Plant Cell* 25, 1541–1554. doi: 10.1105/tpc.113.110486
- Cheng, F., Wu, J., Fang, L., and Wang, X. (2012). Syntenic gene analysis between *Brassica rapa* and other Brassicaceae species. *Front. Plant Sci.* 3:198. doi: 10.3389/fpls.2012.00198
- Duan, W., Song, X., Liu, T., Huang, Z., Ren, J., Hou, X., et al. (2015). Genome-wide analysis of the MADS-box gene family in *Brassica rapa* (Chinese cabbage). *Mol. Genet. Genomics* 290, 239–255. doi: 10.1007/s00438-014-0912-7
- Edger, P. P., and Pires, J. C. (2009). Gene and genome duplications: the impact of dosage-sensitivity on the fate of nuclear genes. *Chromosome Res.* 17:699. doi: 10.1007/s10577-009-9055-9
- Eisenhaber, B., and Eisenhaber, F. (2007). Posttranslational modifications and subcellular localization signals: indicators of sequence regions without inherent 3D structure? *Curr. Protein Pept. Sci.* 8, 197–203. doi: 10.2174/138920307780363424
- Ellis, M., Egelund, J., Schultz, C. J., and Bacic, A. (2010). Arabinogalactan-proteins: key regulators at the cell surface? *Plant Physiol.* 153, 403–419. doi: 10.1104/pp.110.156000
- Gaspar, Y., Johnson, K. L., McKenna, J. A., Bacic, A., and Schultz, C. J. (2001). The complex structures of arabinogalactan-proteins and the journey towards understanding function. *Plant Mol. Biol.* 47, 161–176. doi: 10.1023/A:1010683432529
- Gibson, S. (2005). *The Role of CIL1 in Brassica Carinatalateral Meristem Development*. Dissertation/master's thesis, University of Saskatchewan, Saskatoon, SK.
- Hijazi, M., Roujol, D., Nguyen-Kim, H., Del Rocio Cisneros Castillo, L., Saland, E., Jamet, E., et al. (2014). Arabinogalactan protein 31 (AGP31), a putative network-forming protein in *Arabidopsis thaliana* cell walls? *Ann. Bot.* 114, 1087–1097. doi: 10.1093/aob/mcu038
- Huang, L., Cao, J. S., Zhang, A. H., and Ye, Y. Q. (2008b). Characterization of a putative pollen-specific arabinogalactan protein gene, BcMF8, from *Brassica campestris* ssp. *chinensis*. *Mol. Biol. Rep.* 35, 631–639. doi: 10.1007/s11033-007-9133-z
- Huang, L., Cao, J., Ye, W., Liu, T., Jiang, L., and Ye, Y. (2008a). Transcriptional differences between the male-sterile mutant bcms and wild-type *Brassica campestris* ssp. *chinensis* reveal genes related to pollen development. *Plant Biol.* 10, 342–355. doi: 10.1111/j.1438-8677.2008.00039.x
- Huang, L., Ye, W. Z., Liu, T. T., and Cao, J. S. (2009). Characterization of the male-sterile line Bcajh97-01A/B and identification of candidate genes for genic male sterility in Chinese cabbage-pak-choi. *J. Am. Soc. Horticult. Sci.* 134, 632–640.
- Jiang, J., Jiang, J., Qiu, L., Miao, Y., Yao, L., and Cao, J. (2013). Identification of gene expression profile during fertilization in *Brassica campestris* subsp. *chinensis*. *Genome* 56, 39–48. doi: 10.1139/gen-2012-0088
- Jun, L., and Xiaoming, W. (2012). Genome-wide identification, classification and expression analysis of genes encoding putative fasciclin-like arabinogalactan proteins in Chinese cabbage (*Brassica rapa* L.). *Mol. Biol. Rep.* 39, 10541–10555. doi: 10.1007/s11033-012-1940-1
- Kieliszewski, M. J., and Lamport, D. T. (1994). Extensin: repetitive motifs, functional sites, post-translational codes, and phylogeny. *Plant J.* 5, 157–172. doi: 10.1046/j.1365-313X.1994.05020157.x
- Kim, J., Lee, J., Choi, J. P., Park, I., Yang, K., Kim, M. K., et al. (2014). Functional innovations of three chronological mesohexaploid *Brassica rapa* genomes. *BMC Genomics* 15:606. doi: 10.1186/1471-2164-15-606
- Koch, M. A., and German, D. A. (2013). Taxonomy and systematics are key to biological information: *Arabidopsis*, *Eutrema* (*Thellungiella*), *Noccaea* and *Schrenkiella* (Brassicaceae) as examples. *Front. Plant Sci.* 4:267. doi: 10.3389/fpls.2013.00267
- Kohli, A., Sreenivasulu, N., Lakshmanan, P., and Kumar, P. P. (2013). The phytohormone crosstalk paradigm takes center stage in understanding how plants respond to abiotic stresses. *Plant Cell Rep.* 32, 945–957. doi: 10.1007/s00299-013-1461-y
- Krzywinski, M., Schein, J., Birol, I., Connors, J., Gascoyne, R., Horsman, D., et al. (2009). Circos: an information aesthetic for comparative genomics. *Genome Res.* 19, 1639–1645. doi: 10.1101/gr.092759.109
- Laby, R. J., Kincaid, M. S., Kim, D., and Gibson, S. I. (2000). The *Arabidopsis* sugar-insensitive mutants *sis4* and *sis5* are defective in abscisic acid synthesis and response. *Plant J.* 23, 587–596. doi: 10.1046/j.1365-313x.2000.00833.x
- Lee, C. B., Kim, S., and McClure, B. (2009). A pollen protein, NaPCCP, that binds pistil arabinogalactan proteins also binds phosphatidylinositol 3-phosphate and associates with the pollen tube endomembrane system. *Plant Physiol.* 149, 791–802. doi: 10.1104/pp.108.127936
- Lee, C. B., Swatek, K. N., and McClure, B. (2008). Pollen proteins bind to the C-terminal domain of *Nicotiana glauca* pistil arabinogalactan proteins. *J. Biol. Chem.* 283, 26965–26973. doi: 10.1074/jbc.M804410200
- Levitin, B., Richter, D., Markovich, I., and Zik, M. (2008). Arabinogalactan proteins 6 and 11 are required for stamen and pollen function in *Arabidopsis*. *Plant J.* 56, 351–363. doi: 10.1111/j.1365-313X.2008.03607.x
- Li, J., Gao, G., Zhang, T., and Wu, X. (2013). The putative phytoalexin genes in Chinese cabbage (*Brassica rapa* L.): genome-wide identification, classification and expression analysis. *Mol. Genet. Genomics* 288, 1–20. doi: 10.1007/s00438-012-0726-4
- Lin, S., Dong, H., Zhang, F., Qiu, L., Wang, F., Cao, J., et al. (2014). BcMF8, a putative arabinogalactan protein-encoding gene, contributes to pollen wall development, aperture formation and pollen tube growth in *Brassica campestris*. *Ann. Bot.* 113, 777–788. doi: 10.1093/aob/mct315
- Liu, C., and Mehdy, M. C. (2007). A nonclassical arabinogalactan protein gene highly expressed in vascular tissues, AGP31, is transcriptionally repressed by methyl jasmonic acid in *Arabidopsis*. *Plant Physiol.* 145, 863–874. doi: 10.1104/pp.107.102657
- Liu, Z., Zhang, M., Kong, L., Lv, Y., Zou, M., Lu, G., et al. (2014). Genome-wide identification, phylogeny, duplication, and expression analyses of two-component system genes in Chinese cabbage (*Brassica rapa* ssp. *pekinensis*). *DNA Res.* 21, 379–396. doi: 10.1093/dnares/dsu004
- Livak, K. J., and Schmittgen, T. D. (2001). Analysis of relative gene expression data using real-time quantitative PCR and the $2^{-\Delta\Delta CT}$ Method. *Methods* 25, 402–408. doi: 10.1006/meth.2001.1262
- Ma, H., and Zhao, J. (2010). Genome-wide identification, classification, and expression analysis of the arabinogalactan protein gene family in rice (*Oryza sativa* L.). *J. Exp. Bot.* 61, 2647–2668. doi: 10.1093/jxb/erq104

- Mashiguchi, K., Urakami, E., Hasegawa, M., Sanmiya, K., Matsumoto, I., Yamaguchi, I., et al. (2008). Defense-related signaling by interaction of arabinogalactan proteins and beta-glucosyl Yariv reagent inhibits gibberellin signaling in barley aleurone cells. *Plant Cell Physiol.* 49, 178–190. doi: 10.1093/pcp/pcm175
- Nakai, K., and Horton, P. (1999). PSORT: a program for detecting sorting signals in proteins and predicting their subcellular localization. *Trends Biochem. Sci.* 24, 34–36. doi: 10.1016/S0968-0004(98)01336-X
- Nielsen, H., Engelbrecht, J., Brunak, S., and von Heijne, G. (1997). Identification of prokaryotic and eukaryotic signal peptides and prediction of their cleavage sites. *Protein Eng.* 10, 1–6. doi: 10.1093/protein/10.1.1
- Parra, G., Bradnam, K., and Korf, I. (2007). CEGMA: a pipeline to accurately annotate core genes in eukaryotic genomes. *Bioinformatics* 23, 1061–1067. doi: 10.1093/bioinformatics/btm071
- Pennell, R. I., Janniche, L., Kjellbom, P., Scofield, G. N., Peart, J. M., and Roberts, K. (1991). Developmental regulation of a plasma membrane arabinogalactan protein epitope in oilseed rape flowers. *Plant Cell* 3, 1317–1326. doi: 10.1105/tpc.3.12.1317
- Pereira, A. M., Nobre, M. S., Pinto, S. C., Lopes, A. L., Costa, M. L., Masiero, S., et al. (2016). “Love is strong, and you’re so sweet”: JAGGER is essential for persistent synergid degeneration and polytubey block in *Arabidopsis thaliana*. *Mol. Plant* 9, 601–614. doi: 10.1016/j.molp.2016.01.002
- Pereira, L. G., Coimbra, S., Oliveira, H., Monteiro, L., and Sottomayor, M. (2006). Expression of arabinogalactan protein genes in pollen tubes of *Arabidopsis thaliana*. *Planta* 223, 374–380. doi: 10.1007/s00425-005-0137-4
- Petersen, T. N., Brunak, S., von Heijne, G., and Nielsen, H. (2011). SignalP 4.0: discriminating signal peptides from transmembrane regions. *Nat. Methods* 8, 785–786. doi: 10.1038/nmeth.1701
- Pina, C., Pinto, F., Feijo, J. A., and Becker, J. D. (2005). Gene family analysis of the *Arabidopsis* pollen transcriptome reveals biological implications for cell growth, division control, and gene expression regulation. *Plant Physiol.* 138, 744–756. doi: 10.1104/pp.104.057935
- Rolland, F., Baena-Gonzalez, E., and Sheen, J. (2006). Sugar sensing and signaling in plants: conserved and novel mechanisms. *Annu. Rev. Plant Biol.* 57, 675–709. doi: 10.1146/annurev.arplant.57.032905.105441
- Schultz, C. J., Ferguson, K. L., Lahnstein, J., and Bacic, A. (2004). Post-translational modifications of arabinogalactan-peptides of *Arabidopsis thaliana*. endoplasmic reticulum and glycosylphosphatidylinositol-anchor signal cleavage sites and hydroxylation of proline. *J. Biol. Chem.* 279, 45503–45511. doi: 10.1074/jbc.M407594200
- Schultz, C. J., Rumsewicz, M. P., Johnson, K. L., Jones, B. J., Gaspar, Y. M., and Bacic, A. (2002). Using genomic resources to guide research directions. the arabinogalactan protein gene family as a test case. *Plant Physiol.* 129, 1448–1463. doi: 10.1104/pp.003459
- Serpe, M. D., and Nothnagel, E. A. (1994). Effects of Yariv phenylglycosides on Rosa cell suspensions: evidence for the involvement of arabinogalactan-proteins in cell proliferation. *Planta* 193, 542–550. doi: 10.1007/BF02411560
- Showalter, A. M. (2001). Arabinogalactan-proteins: structure, expression and function. *Cell. Mol. Life Sci.* 58, 1399–1417. doi: 10.1007/PL00000784
- Showalter, A. M., Keppler, B., Lichtenberg, J., Gu, D., and Welch, L. R. (2010). A bioinformatics approach to the identification, classification, and analysis of hydroxyproline-rich glycoproteins. *Plant Physiol.* 153, 485–513. doi: 10.1104/pp.110.156554
- Smeeckens, S. (2000). Sugar-induced signal transduction in plants. *Annu. Rev. Plant Physiol. Plant Mol. Biol.* 51, 49–81. doi: 10.1146/annurev.arplant.51.1.49
- Stenvik, G. E., Butenko, M. A., Urbanowicz, B. R., Rose, J. K., and Aalen, R. B. (2006). Overexpression of INFLORESCENCE DEFICIENT 1 IN ABSCISSION activates cell separation in vestigial abscission zones in *Arabidopsis*. *Plant Cell* 18, 1467–1476. doi: 10.1105/tpc.106.042036
- Suzuki, Y., Kitagawa, M., Knox, J. P., and Yamaguchi, I. (2002). A role for arabinogalactan proteins in gibberellin-induced α -amylase production in barley aleurone cells. *Plant J.* 29, 733–741. doi: 10.1046/j.1365-313X.2002.01259.x
- Tamura, K., Stecher, G., Peterson, D., Filipiński, A., and Kumar, S. (2013). MEGA6: molecular evolutionary genetics analysis version 6.0. *Mol. Biol. Evol.* 30, 2725–2729. doi: 10.1093/molbev/mst197
- Tan, L., Leykam, J. F., and Kieliszewski, M. J. (2003). Glycosylation motifs that direct arabinogalactan addition to arabinogalactan-proteins. *Plant Physiol.* 132, 1362–1369. doi: 10.1104/pp.103.021766
- Tang, X. C., He, Y. Q., Wang, Y., and Sun, M. X. (2006). The role of arabinogalactan proteins binding to Yariv reagents in the initiation, cell developmental fate, and maintenance of microspore embryogenesis in *Brassica napus* L. cv. Topas. *J. Exp. Bot.* 57, 2639–2650. doi: 10.1093/jxb/erl027
- van Hengel, A. J., Barber, C., and Roberts, K. (2004). The expression patterns of arabinogalactan-protein AtAGP30 and GLABRA2 reveal a role for abscisic acid in the early stages of root epidermal patterning. *Plant J.* 39, 70–83. doi: 10.1111/j.1365-313X.2004.02104.x
- van Hengel, A. J., and Roberts, K. (2003). AtAGP30, an arabinogalactan-protein in the cell walls of the primary root, plays a role in root regeneration and seed germination. *Plant J.* 36, 256–270. doi: 10.1046/j.1365-313X.2003.01874.x
- Wang, F., Li, L., Li, H., Liu, L., Zhang, Y., Gao, J., et al. (2012). Transcriptome analysis of rosette and folding leaves in Chinese cabbage using high-throughput RNA sequencing. *Genomics* 99, 222–229. doi: 10.1016/j.ygeno.2012.02.005
- Wang, X., Wang, H., Wang, J., Sun, R., Wu, J., Liu, S., et al. (2011). The genome of the mesopolyploid crop species *Brassica rapa*. *Nat. Genet.* 43, 1035–1039. doi: 10.1038/ng.919
- Wang, Y., Zhang, W.-Z., Song, L.-F., Zou, J.-J., Su, Z., and Wu, W.-H. (2008). Transcriptome analyses show changes in gene expression to accompany pollen germination and tube growth in *Arabidopsis*. *Plant Physiol.* 148, 1201–1211. doi: 10.1104/pp.108.126375
- Wania, S. H., Kumarb, V., Shriramc, V., and Sah, S. K. (2016). Phytohormones and their metabolic engineering for abiotic stress tolerance in crop plants. *Crop J.* 4, 162–176. doi: 10.1007/978-94-017-7758-2_10
- Willats, W. G., and Knox, J. P. (1996). A role for arabinogalactan-proteins in plant cell expansion: evidence from studies on the interaction of beta-glucosyl Yariv reagent with seedlings of *Arabidopsis thaliana*. *Plant J.* 9, 919–925. doi: 10.1046/j.1365-313X.1996.9060919.x
- Yang, Z. (2007). PAML 4: phylogenetic analysis by maximum likelihood. *Mol. Biol. Evol.* 24, 1586–1591. doi: 10.1093/molbev/msm088

Conflict of Interest Statement: 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.

Copyright © 2017 Han, Dong, Cui, Li, Lin, Cao and Huang. 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) or licensor 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.