3 Biotech (2018) 8:5
https://doi.org/10.1007/s13205-017-1031-z

ORIGINAL ARTICLE

Gene expression of an arabinogalactan lysine-rich protein *Ca*AGP18 during vegetative and reproductive development of bell pepper (*Capsicum annuum* L.)

Mercedes Verdugo-Perales[1,4] · Rosabel Velez-de la Rocha[1] · Josefina León-Félix[1] · Tomas Osuna-Enciso[1] · José B. Heredia[1] · Juan A. Osuna-Castro[2] · Maria A. Islas-Osuna[3] · J. Adriana Sañudo-Barajas[1]

Received: 5 July 2017 / Accepted: 4 December 2017 / Published online: 8 December 2017

Abstract

Lysine-rich (Lys-rich) proteins encoded by *AGP17*, *AGP18*, and *AGP19* genes are cell wall-associated glycopeptides related to sexual reproduction in flowering plants. This subclass belongs to classical arabinogalactan proteins (AGPs) widely studied in model plants like *Arabidopsis*. In this study, we identified the *CaAGP18* cDNA from bell pepper (*Capsicum annuum* L.), as well as its expression pattern during vegetative and reproductive development. The deduced amino acid sequence revealed a Lys-rich AGP18 protein of 238 amino acids residues in length with an estimated molecular mass of 22.85 kDa and an isoelectric point of 9.7. The protein is predicted as canonical AGP due to the presence of a small Lys-rich region and a C-terminal sequence essential for posttranslational modification with a glycosylphosphatidylinositol (GPI). Phylogenetic analysis showed that *CaAGP18* is clustered together with *NtAGP18*, *SpAGP18*, *StAGP18* and *NaAGP18* from *Solanaceae* species. *CaAGP18* expression through plant phenological stages had the highest transcription level in leaves at the seedling stage, whereas in reproductive organs there was a significant up-regulation in pistils during anthesis, also in petals 2 days post-anthesis (DPA), and in fruit at the expansion stage. Our results open future research for possible roles of *Ca*AGP18 in cell expansion as a wall-associated plasticizer and reproductive processes like pistil interactions and petal cell death.

Keywords *Solanaceae* · Seedling · Flowering · Fruit expansion · Pistil · Petal · Anthesis · Cell wall · Bell pepper

Introduction

Plant growth and development processes in crops of commercial interest are of great concern. Mutations and variability in genes related to agronomic traits have led to a vast and diverse genetic resource in *Solanaceae* family that comprises important edible species like *Capsicum annuum* (Wang et al. 2015). In the advent of genome sequencing and the omics, the knowledge of individual genes and the validation of expression patterns remain as an important component for the establishment of relationships between genes and important agronomical traits. Cell expansion, differentiation, and cell death, involve complex mechanisms of intercellular communication through activation of signaling pathways where arabinogalactan proteins (AGPs) have been associated (Ellis et al. 2010). AGPs belong to a superfamily of hydroxyproline-rich glycoproteins (HRGPs) associated with reproduction in flowering plants. One of the most distinctive characteristics is the hyper glycosyl structure decorating the protein sequence rich in proline (P), alanine (A),

✉ Maria A. Islas-Osuna
islasosu@ciad.mx

✉ J. Adriana Sañudo-Barajas
adriana@ciad.mx

[1] Laboratorio de Bioquímica Vegetal, Centro de Investigación en Alimentación y Desarrollo, A.C-Unidad Culiacán, Carret. a Eldorado km 5.5, Campo El Diez, 80110 Culiacán, Sinaloa, Mexico

[2] Facultad de Ciencias Biológicas y Agropecuarias, Universidad de Colima, Autopista Colima-Manzanillo km 40, 28100 Tecomán, Colima, Mexico

[3] Laboratorio de Genética y Biología Molecular de Plantas, Centro de Investigación en Alimentación y Desarrollo, A.C-Unidad Hermosillo, Carr. a La Victoria Km 0.6, Ejido La Victoria, 83304 Hermosillo, Sonora, Mexico

[4] Instituto de Investigación Lightbourn A.C. Carretera Las Pampas Km. 2.5 Col. Tierra y Libertad, Cd. Jiménez, 33980 Chihuahua, Mexico

serine (S) and threonine (T) residues (Showalter 2001; Ellis et al. 2010).

AGPs are classified based on their protein sequence into two subclasses: classical and chimeric. The classical AGP domain consists of an N-terminal signal peptide, a central domain rich in PAST and a C-terminal GPI anchor addition sequence which allows the AGPs location on the surface of the plasma membrane and cell wall (Schultz et al. 2002). Classical AGPs are sub-classified in arabinogalactan peptides (AG peptide) and Lys-rich AGPs (Yang et al. 2007; Seifert and Roberts 2007; Showalter et al. 2010). The AG peptides have a short protein backbone from 10 to 15 amino acids in length, differing to Lys-rich that contain 185–247 amino acids and a small Lys-rich subdomain (approximately twelve amino acids) in the PAST-rich central domain. Seven members have been identified as Lys-rich AGPs; *At*AGP17, 18 and 19 from *Arabidopsis thaliana* (Sun et al. 2005); *Cs*AGP1 from *Cucumis sativus* (Park et al. 2003); *Pta*AGP6 from *Pinus taeda* (Zhang et al. 2003); *Le*AGP-1 from *Solanum lycopersicon* (Gao and Showalter 1999) and *Na*AGP4 from *Nicotiana alata* (Gilson et al. 2001).

Advances in knowledge of Lys-rich AGPs sequences were obtained after completion of several plant genomes. *Na*AGP4 and *Le*AGP-1 were identified as members of the *Solanaceae* family, sharing about 78% sequence identity, close phylogenetic relationship, same spatial expression, and participation in tissue wounding repair and pathogen responses (Gilson et al. 2001). *Le*AGP-1 was confirmed anchored to the plasma membrane via GPI group, and it was differentially expressed in young stem and flowers, old internodes and green fruit (Sun et al. 2004). *Le*AGP-1 and *At*AGP18 share overexpression phenotypes and have been associated with cytokinin signal transduction pathway (Zhang et al. 2011). From this perspective, as cytokinin is involved in plant growth and development, AGPs might be part of processes such as differentiation, cell growth and division, and others affecting the plant physiology (Sun et al. 2004; Yang et al. 2007; Zhang et al. 2011).

A precise function of AGP18 has not been demonstrated; however, it is likely implicated in the initiation process during female gametogenesis (Acosta-García and Vielle-Calzada 2004) and may act regulating the selection of the functional megaspore (Demesa-Arevalo and Vielle-Calzada 2013). As far as we have investigated, there is no report about the expression patterns of Lys-rich AGPs pepper although its genome is available. *C. annuum* represents an excellent model for AGPs study as the well-identified high rate of growth during seedling, flowering, and fructification stages. In this work, a putative sequence encoding for AGP18 isolated from *C. annuum*-type bell pepper at seedling stage was cloned, phylogenetically compared and its expression pattern characterized during vegetative and reproductive development.

Materials and methods

Plant material and growth conditions

Bell pepper plants (*Capsicum annuum* L. cv. Cannon) were cultivated in a greenhouse at the Instituto Nacional de Investigaciones Forestales Agrícolas y Pecuarias (INIFAP) research station located in Culiacán, Sinaloa, Mexico (24°37′57.0″N and 107°26′27.3″W). Phenological stages of seedling, flowering and fruiting corresponded to 45, 75 and 115 days after the emergency (DAE), respectively. Sampling of leaf, stem, petal, and pistil [anthesis, 1, 2 and 3 days post-anthesis (DPA)] and fruit (10, 13, 16, 20, 23, 28, 30 DPA) was conducted in the early morning. Collected samples were washed with sterile water; excess was removed, flower organs exscinded and immediately frozen in liquid nitrogen for storage at 80 °C until used.

Cloning of *AGP18* cDNA from bell pepper

Total RNA was isolated using Tri-Reagent (Sigma Aldrich) according to the manufacturer's instructions, and resuspended in RNase-Free Water. First-strand cDNA was synthesized from 5 µg of DNA-free RNA using Super Script III reverse transcriptase (Invitrogen) according to the manufacturer's protocol. Primers AGP18F 5′-GCATTGTATTGG CAAGTGTCACGG-3′ and AGP18R 5′-AACTGGAGCGCT TACTGGCGTAA-3′ were designed after the alignment of the *Solanum tuberosum AGP18* (GenBank NM_001288489), *Solanum pennellii* (*Lycopersicon pennellii*) *AGP18* (XM_015211025) and *Nicotiana tomentosiformis AGP18* (XM_009618065) coding sequences to clone a fragment of the bell pepper cDNA. Rapid Amplification of cDNA Ends (RACE) was done using the SMARTer® RACE 5′/3′ Kit (TaKaRa, Japan) following the manufacturer's instructions to obtain the 5′ and 3′ cDNA fragments, with the gene-specific primers for AGP18 mentioned above. PCR conditions were 94 °C for 30 s (1 cycle), 68 °C for 30 s, and 72 °C for 3 min (24 cycles). The PCR products were purified with the NucleoSpin Gel and PCR Clean-Up Kit (Promega), according to the manufacturer's instructions. Fragments obtained were sequenced at the National Laboratory of Genomics for Biodiversity (Irapuato, México).

*Ca*AGP18 amino acid sequence and phylogenetic analysis

The full-length cDNA sequence of *CaAGP18* was analyzed using the algorithm Blastn, and the deduced amino acid sequence was tested for signal peptide using SignalP 4.1 (http://www.cbs.dtu.dk/services/SignalP/) (Petersen et al.

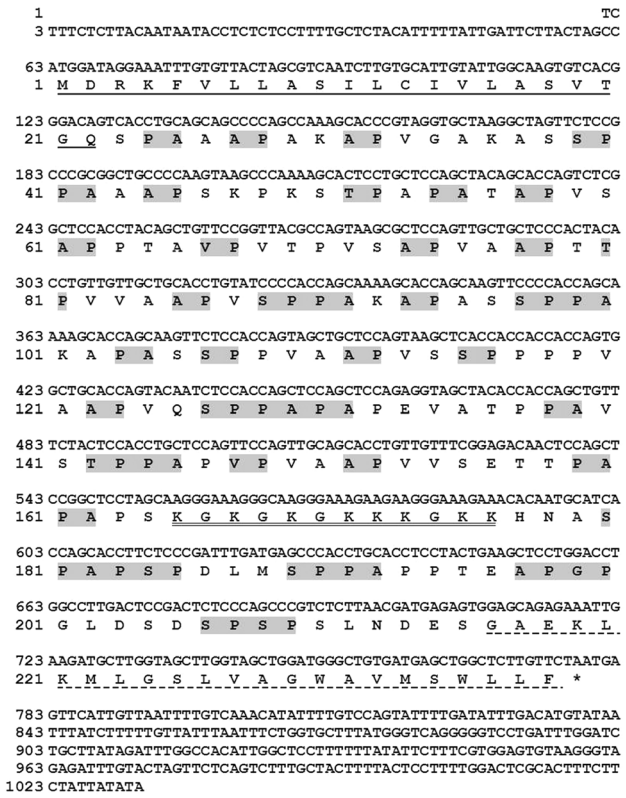

Fig. 1 *Ca*AGP18 cDNA and deduced amino acid sequence (GenBank MF278295 and ASB17094, respectively). 5′and 3′ untranslated (UTR) nucleotide sequences are shown (1–62 and 780–1032, respectively) and the coding sequence (63–779). Underlined residues at the N termini of the protein indicate the predicted signal peptide and the GPI anchor sequence is indicated with a dashed line; AP/PA/SP/TP repeats are shown in dark gray and Lys-rich region is indicated in parallel lines

2011). GPI modification site was predicted using the server big-PI Plant Predictor GPI Modification Site Prediction in Plants (http://mendel.imp.ac.at/gpi/plant_server.html) (Eisenhaber and Eisenhaber 2007). Proline-rich domains were identified via the Expert Protein Analysis System (http://www.expasy.org/). Domains of AGP18 were predicted using the simple modular architecture research tool (SMART) online server (http://smart.embl-heidelberg.de/).

Multiple sequence alignment was performed for AGP18 amino acid sequences from bell pepper, tomato, potato, tobacco, coyote tobacco, cotton, cacao, orange, *Jatropha*, eucalyptus, grape, *Arabidopsis*, cabbage, pigeon pea, soybean, wild soybean and spider flower using Clustal X2 (Thompson et al. 1997). The alignment was used to evaluate the evolutionary relationship of these proteins using the Neighbor-Joining method with AtFLA11 as an out-group, and 10,000 bootstrap replicates using Geneious R8. The tree model with genetic distance used was Jukes-Cantor.

Quantitative PCR (qPCR) analysis

Total RNA was isolated from pulverized frozen tissues using Tri-Reagent (Sigma Aldrich) according to the manufacturer's instructions and treated with RNase-free DNase I (Roche) to remove contaminant genomic DNA. RNA quantification was performed using a Nanodrop 2000c UV–VIS spectrophotometer at 260 nm (Thermo Scientific). RNA integrity was analyzed by 1% agarose gel electrophoresis under denaturing conditions and analyzed in a Chemi-Doc™ Imaging Systems (Bio-Rad). Five micrograms of total RNA were used for reverse transcription into complementary (cDNA) using Superscript III cDNA synthesis kit (Invitrogen). Quantitative real-time PCR (qPCR) assays were performed on a CFX96 Touch™ Real-Time PCR System (Bio-Rad). 40 ng of cDNA was used for reactions, in a 10 µL final reaction mixture containing cDNA, 5 µM of each forward and reverse primer (AGP18Fw and AGP18Rv), and 5 µL iTaq™ universal SYBR® Green supermix (Bio-Rad). β-tubulin (GenBank EF495259.1) was used as the reference gene (β-tubulin forward 5′ GAGGGTGAGTGAGCA GTTC-3′ and reverse 5′-CTTCATCGTCATCTGCTGTC-3′ primers). The PCR product was sequenced to validate the specificity of the primers. The qPCR run conditions were 95 °C for 6 min (1 cycle), 95 °C for 30 s, 60 °C for 30 s, 72 °C for 45 s (35 cycles). PCR products were melted by gradually increasing the temperature from 55 to 95 °C in 0.5 °C increments at every step, to confirm the specificity of the amplified products. Fluorescence intensity was measured after every cycle. Three biological replicates were done for cDNA obtained from plant leaves, stems, fruits, ovary and petals from different phenological stages. Each qPCR reaction was run in triplicate. The data were analyzed with the $2^{-\Delta\Delta Ct}$ method and gene expression was calculated as described by Livak and Schmittgen (2001).

Statistical analysis

A factorial design was conducted to analyze *CaAGP18* relative expression, with three biological replicates and three technical repetitions. Analysis of variance (ANOVA) and comparison of means (multiple range test of Tukey, a significance level of $P < 0.05$) were performed with the NCSS (2007) software.

Results and discussion

*Ca*AGP18 cDNA, deduced amino acid sequence and phylogeny

The complete nucleotide sequence of *CaAGP18* mRNA (GenBank MF278295) obtained was 1032 nt in length

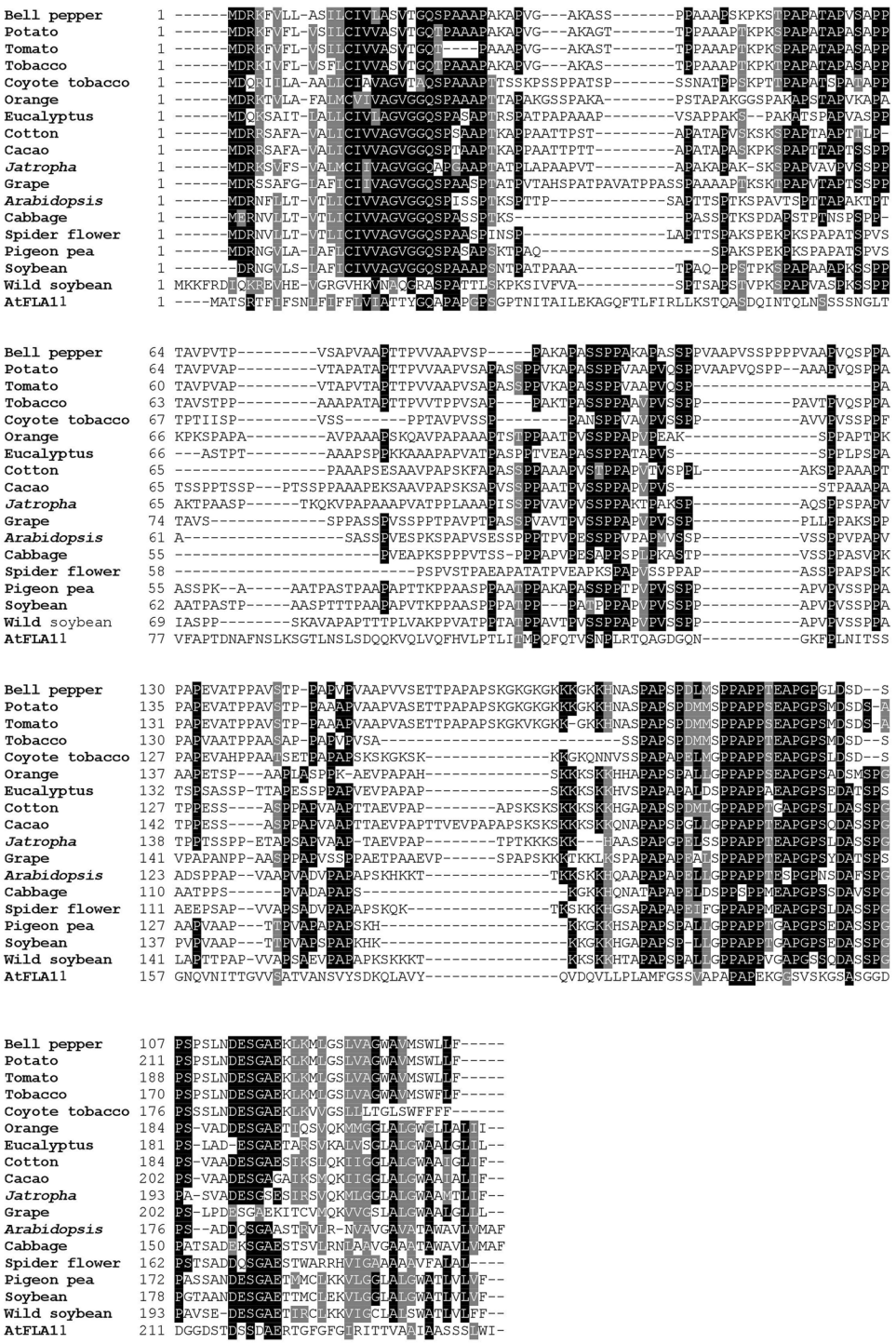

Fig. 2 Multiple sequence alignment of AGP18 from bell pepper, potato, tomato, tobacco, coyote tobacco, orange, eucalyptus, cotton, cacao, jatropha, grape, *Arabidopsis*, cabbage, spider flower, pigeon pea, soybean, wild soybean and AtFLA11 using Clustal X2 alignment tool. Identical amino acids are in a black background, while similar amino acids are in a gray background

and encodes a protein of 238 residues (Fig. 1, GenBank ASB17094). Predicted domains structure common to Lys-rich AGPs family were identified; the N-terminal signal peptide sequence, the basic central small Lys-rich domain, and a C-terminal GPI anchor addition sequence (Fig. 1). All these regions have been reported as highly conserved within the structure of the Lys-rich AGP subfamily in *Arabidopsis* (Nielsen et al. 1997; Yang and Showalter 2007). Moreover, the signal peptide in proteins has been reported as essential for glycosylation of AGPs during the biosynthesis, while GPI is required for plasma membrane anchoring, also, the AP/PA/SP/TP repeats (14/13/11/3) are characteristic of the central PAST (Pro, Ala, Ser, Thr) rich region.

The alignment of *Ca*AGP18 amino acid sequence with AGP18 from other plants species showed a high conservation of the central PAST-rich domain and the Lys-rich

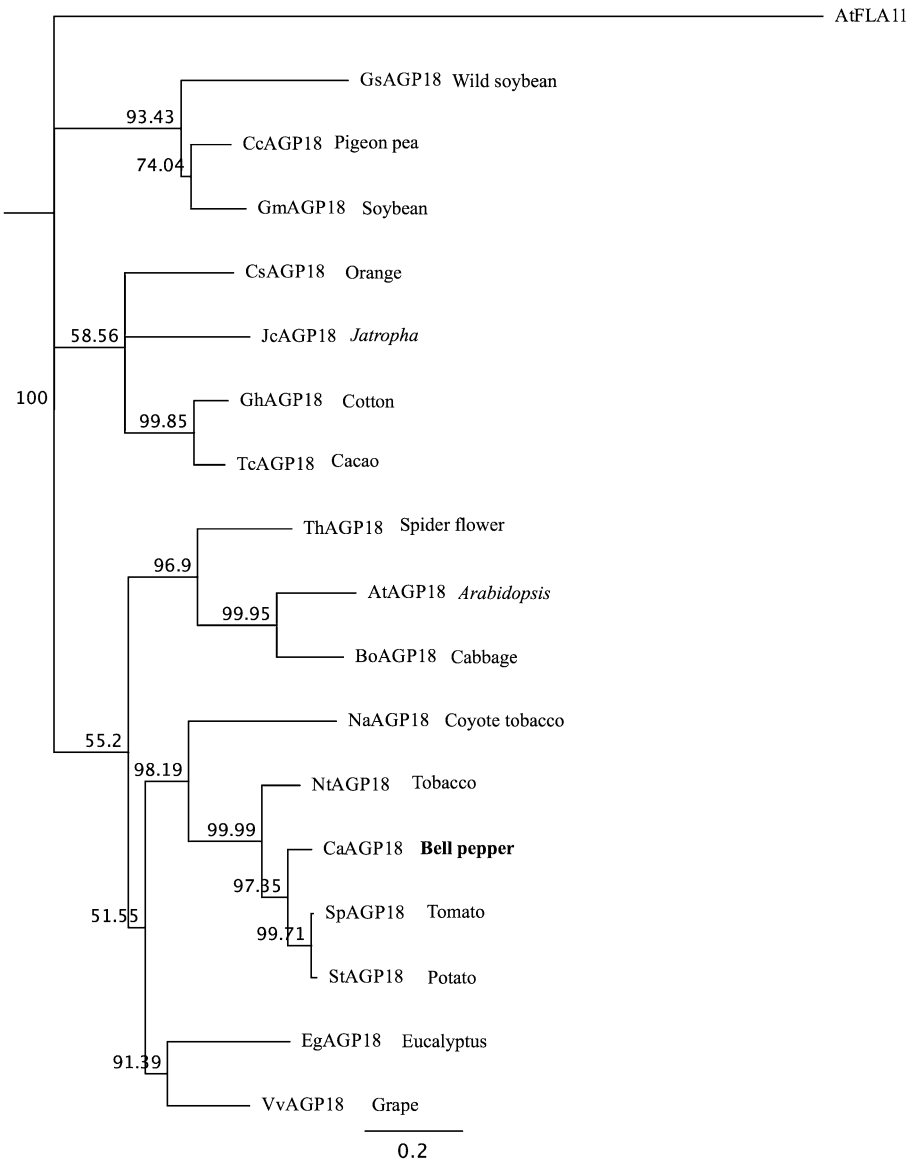

Fig. 3 Phylogeny tree of classical Lys-rich AGP18 s. Boot-strap values from 10,000 trials are indicated at the branch points. The scale indicates the branch length. AGP18 Protein identification numbers are: Bell pepper, MF278295; Tomato, XP_015066511.1; Potato, NP_001275418.1; Tobacco, XP_018630660.1; Coyote tobacco, XP_019245423.1; Cotton, XP_016681363.1; Cacao, XP_007047448.2; Orange, XP_015388701.1; *Jatropha*, XP_012079351.1; Eucalyptus, XP_010028135.1; Grape, XP_002284001.1; *Arabidopsis*, NP_568027.1; Cabbage, XP_013583391.1; Pigeon pea, XP_020240026.1; Soybean, XP_014625605.1; Wild soybean, KHN12969.1; Spider flower, XP_010548448 and *Arabidopsis* FASCICLIN-like arabinogalactan-protein 11 (AtFLA11), NP_195937 as an outgroup protein

region, suggesting a conserved structure between these proteins (Fig. 2). The phylogeny of the AGP family 18 is presented in Fig. 3. The legumes group in one clade; interestingly, soybean is closer to pigeon pea compared to wild soybean. The second clade is more diverse and includes orange, Jatropha, cotton, and cacao, dicots with important genomic features in common, revealing the existence of a common ancestor (Wang et al. 2016). Clade 3 is branched into the *Brassicaceae*, *Solanaceae* and a final group with two rosid eudicots: grape and eucalyptus. A recent study suggested an evolutionary modification of Lys-rich AGPs in monocots and eudicots, resulting in at least two ancient duplications. This evidence suggests the divergence and phylogenetic variation in gymnosperm and angiosperm plants that genomic sequencing projects have uncovered (Ma et al. 2017). Not all AGP18s deposited into the GenBank have been studied experimentally. For *Solanaceae*, the AGP18 function is unknown, and it could be related to another Lys-rich *Na*AGP4 protein, since both proteins are highly similar and share the same characteristic domains (Gilson et al. 2001). Differential expression pattern by single or large scale analysis such as RNA-Seq might confirm this hypothesis.

Relative expression analysis of *CaAGP18* gene in vegetative organs

The expression levels of *CaAGP18* in leaves were 96-fold higher during seedling compared to the flowering and fructification stages (Fig. 4). In contrast, expression was reduced in stems at all developmental stages. Authors have suggested a role of AGP18 in leave cell elongation (Park et al. 2003; Seifert and Roberts 2007), and the results in leaves and

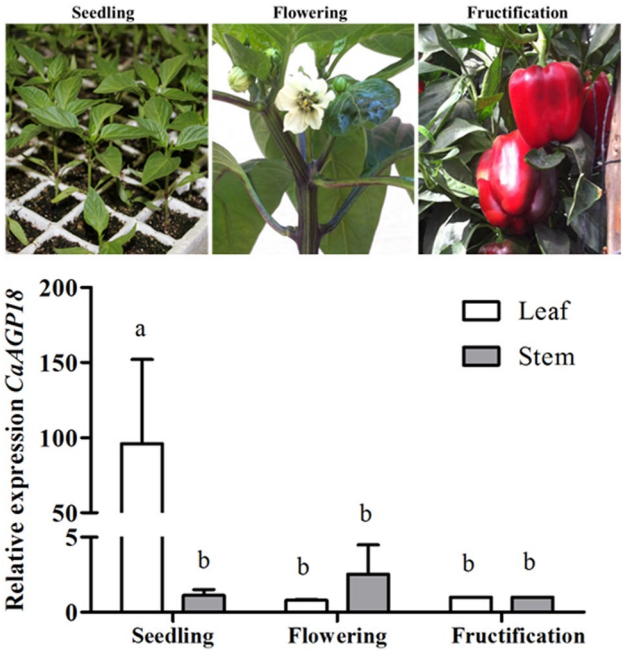

Fig. 4 Tissue-specific expression levels of *CaAGP18*. Total RNA was extracted from leaf and stem of bell pepper and cDNAs were synthesized. Quantitative real-time PCR was carried out from cDNAs of the organs collected at the Seedling, Flowering and Fructification stages. Expression levels were evaluated from three biological replicates of each organ and run in triplicates. Data were analyzed with the $2^{-\Delta\Delta Ct}$ method and expression levels of *CaAGP18* are relative to β-tubulin levels. The results were represented as mean ± standard deviation (*bars* in the graph) ($n = 9$). Expression levels of *CaAGP18* in leaf and stem from the fructification stage were chosen as calibrators (1) and the relative expression levels from the other stages (seedling and flowering) were represented as fold changes

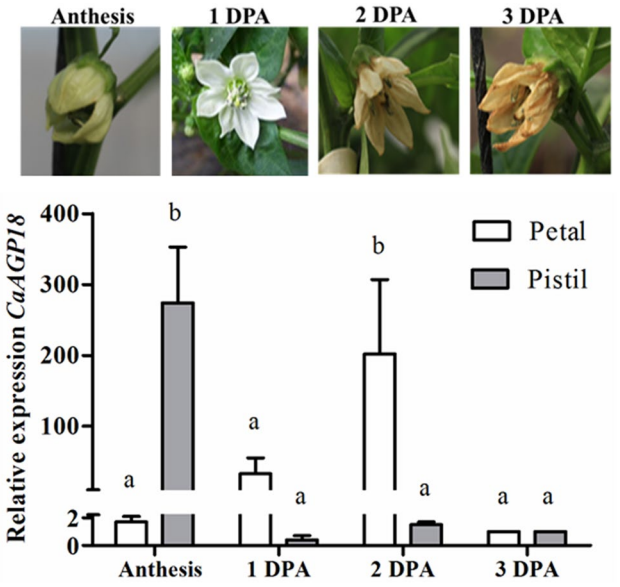

Fig. 5 Tissue-specific expression levels of *CaAGP18*. Total RNA was extracted from Petal and Pistil of bell pepper and cDNAs were synthesized. Quantitative real-time PCR was carried out from cDNAs of the organs collected at anthesis, 1 day post-anthesis (DPA), 2 DPA and 3 DPA. Expression levels were evaluated from three biological replicates of each organ and run in triplicates. Data were analyzed with the $2^{-\Delta\Delta Ct}$ method and expression levels of *CaAGP18* are relative to β-tubulin levels. The results were represented as mean ± standard deviation (*bars* in the graph) ($n = 9$). Expression levels of *CaAGP18* in petal and pistil from 3DPA were chosen as calibrators (1) and the relative expression levels from the other stages (anthesis, 1 DPA and 2 DPA) were represented as fold changes

seedlings support a potential role of *CaAGP18* as part of the biochemical and physiological processes that influence leaf growth in early development.

Adequate leaf area index (needed to sustain the root system and shoot growth) is determined by cytokinins; these plant hormones are involved in coordinating the directional growth of tissues (Sakakibara 2006). The differences in expression levels found in this work suggest a specific regulation of *CaAGP18* between organs and phenological stages. The relationship between gene expression and phenotype may not be intuitive, for example, overexpression of *At*AGP18 in *Arabidopsis* produced phenotypes with shorter stems similar to phenotypes of tobacco overproducing cytokinins (Zhang et al. 2011).

Expression analysis of *CaAGP18* gene in flower organs

The plant reproductive development is a complex process that involves male and female organogenesis. The expression pattern of *CaAGP18* in pistils and petals, at four different stages of development (Anthesis, 1, 2 and 3 DPA) suggest a potential role of AGP18 during pistil interactions and petal senescence. The transcript accumulated 276-fold in pistils from anthesis and 200-fold in petals from two DPA, when compared to 3 DPA (Fig. 5). The *CaAGP18* up-regulation in pistils could be related to female gametogenesis according to the previous report from Demesa-Arevalo and Vielle-Calzada (2013). In petals, AGP18 may associate with programed cell death due to ethylene-driven processes during fertilization and pollination-induced petal senescence (Kovaleva et al. 2013). The understanding of the mechanisms regulating stigma receptivity and pollen tube growth into the style are limited (Dresselhaus and Sprunk 2012) and the timing for maximum fertility is regulated by endogenous hormones (ethylene, gibberellins and auxins) and environmental factors (light, temperature and circadian clock rhythms) (Van Doorn and Kamdee 2014). At the transcript level, AGP18 has been detected at different expression levels in flowers, stems, seedlings and siliques (Acosta-García and Vielle-Calzada 2004). Thus, further studies are needed to test this

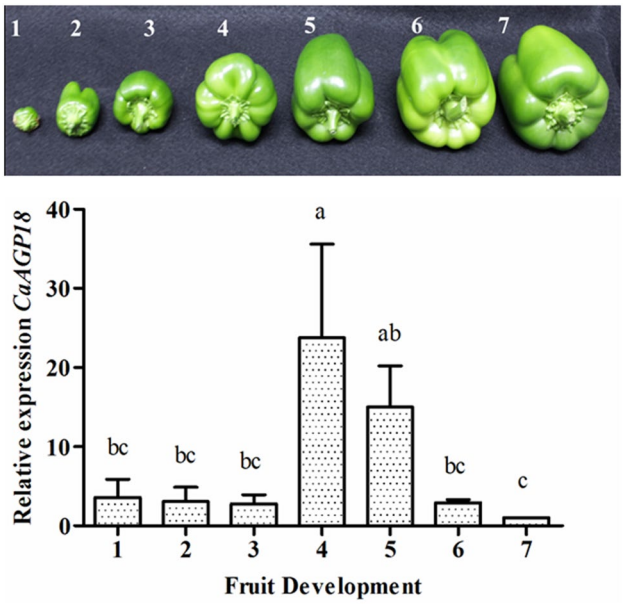

Fig. 6 Expression levels of *CaAGP18* in bell pepper fruits. Fruit samples were collected from seven different developmental stages (10, 13, 16, 20, 23, 28, 30 days post-anthesis) (Stages 1–9). (1) Fruit of 1.09 × 2.15 cm and 1.92 g of fresh weight (FW). (2) Fruit of 4.1 × 3.8 cm and 14.8 g of FW. (3) Fruit of 4.81 × 5.1 cm and 27.2 g of FW. (4) Fruit of 5.2 × 7 cm and 67.2 g of FW. (5) Fruit of 7.5 × 7.1 cm and 96.3 g of FW. (6) Fruit of 8.3 × 8.2 cm and 180.33 g of FW. (7) Fruit of 9.9 × 8.5 cm and 215.4 g of FW. Total RNA was extracted from bell pepper fruit. Quantitative real-time PCR was carried out from cDNAs synthesized from the biological replicates and run in triplicates. Data were analyzed with the $2^{-\Delta\Delta Ct}$ method and expression levels of *CaAGP18* are relative to β-tubulin levels. The results were represented as mean ± standard deviation (*bars* in the graph) ($n = 9$). Expression levels of *CaAGP18* in fruit from 30 DPA was chosen as calibrator (1) and the relative expression levels from the other stages (10, 13, 16, 20, 23, and 28 DPA) were represented as fold changes

hypothesis in *C. annuum*, because flowering stimuli are species dependent (Rogers 2006).

Localization patterns of AGP18 proteins are also necessary. For example, in the model plant *Arabidopsis*, the AGP18 protein has been found critical for the initiation of female gametogenesis. Using immunochemical methods, AGPs have been shown to localize to the pistil and are present during the gametophytic cell differentiation (Coimbra et al. 2007). In *Arabidopsis* and tobacco, the localization of AGPs in the stigma exudates and style transmitting tissues has been related to functions of pollen-specific interactions in the stigmatic surface (Lee et al. 2009).

Expression analysis of *CaAGP18* gene during fruit growth

The transcript levels of *CaAGP18* also changed during fruit development (Fig. 6). Higher levels of the transcript accumulated at stage four (average fruit size 5.2 × 7 cm and 67.2 g) and five (average fruit size 7.5 × 7.1 cm and 96.3 g) whereas the lowest levels were found in stage 7, when fruit growth stopped. These results suggest that *CaAGP18* is more significant during cell elongation, fruit growth and sugar accumulation (Fig. 6). Several researchers have previously reported that this gene favors tissue growth acting as cell shape regulator and cell wall plasticizer as proposed for GPI-anchored AGPs (Seifert and Roberts 2007). Fruit development is a complex process, in the first stages of fruit enlargement there is an increment of cell number, followed by cell expansion and finally the physiological ripening with important changes in metabolite content besides capsaicin accumulation (Lamport et al. 2006; Ellis et al. 2010; Chaki et al. 2015). Different authors agree that the AGP-encoding genes participate during leaf, steam, fruit and flower development (Zhang et al. 2011; Fragkostefanakis et al. 2012). Future research in bell pepper using mutational studies would clarify the potential role of AGP18 as well as other members of Lys-rich AGPs.

Conclusions

This study reveals new findings about the potential roles of AGP18 during *Capsicum annuum* growth and development. The gene *CaAGP18* is predicted as a Lys-rich classical AGP protein in vegetative and reproductive organs of bell pepper at seedling, flowering, and fruiting stages. The expression patterns suggest a potential role during leaf growth, petal senescence, pistil interaction and fruit expansion. However, a better understanding through gene screening of other class members and functional mutation analysis may reveal the importance of Lys-rich proteins in this crop. Also, further research is yet to be undertaken to demonstrate posttranslational regulatory mechanisms and protein–protein interactions that would provide concluding insights of protein functionality.

Acknowledgements MVP work was supported by a PhD scholarship from the National Council for Science and Technology (CONACyT Mexico), No. 273500. MAIO thanks CONACyT for grant CB2012-01-178296. We thank Dr. Carmen Contreras, Nancy Bojórquez and Pedro Rojas for contributions in figure edition, field sampling and culture establishment. Thanks to Dr. Rogerio Sotelo-Mundo and Daniel Lira for critical reading of the manuscript.

Author contributions Conceived and designed the experiments: JASB, MVP, MAIO, TOE. Performed the experiments: MVP. Analyzed the data: MVP, JASB, MAIO. Wrote the paper: MVP, JASB, MAIO. Reviewed the paper: JLF, TOE, JBH, JAOC, RVR. Final approval of the version to be published: JASB, MAIO.

Compliance with ethical standards

Conflict of interest The authors declare that they have no competing interest.

References

Acosta-García G, Vielle-Calzada J-P (2004) A classical arabinogalactan protein is essential for the initiation of female gametogenesis in *Arabidopsis*. Plant Cell 16:2614–2628. https://doi.org/10.1105/tpc.104.024588

Chaki M, Álvarez de Morales P, Ruiz C, Begara-Morales JC, Barroso JB, Corpas FJ, Palma JM (2015) Ripening of pepper (*Capsicum annuum*) fruit is characterized by an enhancement of protein tyrosine nitration. Ann Bot Lond 116:637–647. https://doi.org/10.1093/aob/mcv016

Coimbra S, Almeida J, Junqueira V, Costa ML, Pereira LG (2007) Arabinogalactan proteins as molecular markers in *Arabidopsis thaliana* sexual reproduction. J Exp Bot 58:4027–4035. https://doi.org/10.1093/jxb/erm259

Demesa-Arevalo E, Vielle-Calzada J-P (2013) The classical arabinogalactan protein AGP18 mediates megaspore selection in *Arabidopsis*. Plant Cell 25:1274–1287. https://doi.org/10.1105/tpc.112.106237

Dresselhaus T, Sprunk S (2012) Plant fertilization: maximizing reproductive success. Curr Biol 22:487–489. https://doi.org/10.1016/j.cub.2012.04.048

Eisenhaber B, Eisenhaber F (2007) Posttranslational modifications and subcellular localization signals: indicators of sequence regions without inherent 3D structure? Curr Protein Pept Sci 8:197–203. https://doi.org/10.2174/138920307780363424

Ellis M, Egelund J, Schultz CJ, Bacic A (2010) Arabinogalactan-proteins: key regulators at the cell surface? Plant Physiol 153:403–419. https://doi.org/10.1104/pp.110.156000

Fragkostefanakis S, Dandachi F, Kalaitzis P (2012) Expression of arabinogalactan proteins during tomato fruit ripening and in response to mechanical wounding, hypoxia and anoxia. Plant Physiol Biochem 52:112–118. https://doi.org/10.1016/j.plaphy.2011.12.001

Gao M, Showalter AM (1999) Yariv reagent treatment induces programmed cell death in *Arabidopsis* cell cultures and implicates arabinogalactan protein involvement. Plant J 19:321–331. https://doi.org/10.1046/j.1365-313X.1999.00544.x

Gilson P, Gaspar YM, Oxley D, Youl JJ, Bacic A (2001) $NaAGP_4$ is an arabinogalactan protein whose expression is suppressed by wounding and fungal infection in *Nicotiana alata*. Protoplasma 215:128–139. https://doi.org/10.1007/BF01280309

Kovaleva L, Timofeeva G, Rodionova G, Zakharova E, Rakitin V (2013) Role of ethylene in the control of gametophyte–sporophyte interactions in the course of the progamic phase of fertilization. Russ J Dev Biol 44:69–77

Lamport DTA, Kieliszewski MJ, Showalter AM (2006) Salt stress upregulates periplasmic arabinogalactan proteins: using salt stress to analyse AGP function. N Phytol 169:479–492. https://doi.org/10.1111/j.1469-8137.2005.01591.x

Lee CB, Kim S, 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. https://doi.org/10.1104/pp.108.127936

Livak KJ, Schmittgen TD (2001) Analysis of relative gene expression data using real-time quantitative PCR and the $2^{-\Delta\Delta Ct}$ method. Methods 25:402–408. https://doi.org/10.1006/meth.2001.1262

Ma Y, Yan C, Li H, Wu W, Liu Y, Wang Y, Chen Q, Ma H (2017) Bioinformatics prediction and evolution analysis of arabinogalactan proteins in the plant kingdom. Front Plant Sci 8:66. https://doi.org/10.3389/fpls.2017.00066

Nielsen H, Engelbrecht J, Brunak S, Von Heijne G (1997) Identification of prokaryotic and eukaryotic signal peptides and prediction of their cleavage sites. Protein Eng 10:1–6. https://doi.org/10.1093/protein/10.1.1

Park MH, Suzuki Y, Chono M, Knox JP, Yamaguchi I (2003) *CsAGP1*, a gibberellin-responsive gene from cucumber hypocotyls, encodes a classical arabinogalactan protein and is involved in stem elongation. Plant Physiol 131:1450–1459. https://doi.org/10.1104/pp.015628

Petersen TN, Brunak S, Von Hejine G, Nielsen H (2011) Signal P 4.0: discriminating signal peptides from transmembrane regions. Nat Methods 8:785–786. https://doi.org/10.1038/nmeth.1701

Rogers HJ (2006) Programmed cell death in floral organs: how and why do flowers die? Ann Bot Lond 97:09–315. https://doi.org/10.1093/aob/mcj051

Sakakibara H (2006) Cytokinins: activity, biosynthesis, and translocation. Ann Rev Plant Biol 57:431–449. https://doi.org/10.1146/annurev.arplant.57.032905.105231

Schultz CJ, Rumsewicz M, Johnson KL, Jones BJ, Gaspar YM, Bacic A (2002) Using genomic resources to guide research directions. The arabinogalactan protein gene family as a test case. Plant Physiol 129:1448–1463. https://doi.org/10.1104/pp.003459

Seifert GJ, Roberts K (2007) The biology of arabinogalactan proteins. Ann Rev Plant Biol 58:137–161. https://doi.org/10.1146/annurev.arplant.58.032806.103801

Showalter A (2001) Arabinogalactan-proteins: structure, expression and function. CMLS Cell Mol Life Sci 58:1399–1417. https://doi.org/10.1007/PL00000784

Showalter AM, Keppler B, Lichtenberg J, Gu D, Welch LR (2010) A bioinformatics approach to the identification, classification, and analysis of hydroxyproline-rich glycoproteins. Plant Physiol 153:485–513. https://doi.org/10.1104/pp.110.156554

Sun W, Kieliszewski MJ, Showalter AM (2004) Overexpression of tomato LeAGP-1 arabinogalactan-protein promotes lateral branching and hampers reproductive development. Plant J 40:870–881. https://doi.org/10.1111/j.1365-313X.2004.02274.x

Sun W, Xu J, Yang J, Kieliszewski MJ, Showalter AM (2005) The lysine-rich arabinogalactan-protein subfamily in *Arabidopsis*: gene expression, glycoprotein purification and biochemical characterization. Plant Cell Physiol 46:975–984. https://doi.org/10.1093/pcp/pci106

Thompson JD, Gibson TJ, Plewniak F, Jeanmougin F, Higgins DG (1997) The CLUSTAL_X windows interface: flexible strategies for multiple sequence alignment aided by quality analysis tools. Nucleic Acids Res 25:4876–4882

Van Doorn WG, Kamdee C (2014) Flower opening and closure: an update. J Exp Bot 65:5749–5757. https://doi.org/10.1093/jxb/eru327

Wang L, Li J, Zhao J, He C (2015) Evolutionary developmental genetics of fruit morphological variation within the Solanaceae. Front Plant Sci 6:248. https://doi.org/10.3389/fpls.2015.00248

Wang X, Guo H, Wang J, Lei T, Liu T, Wang Z, Li Y, Lee TH, Li J, Tang H, Jin D, Paterson AH (2016) Comparative genomic de-convolution of the cotton genome revealed a decaploid ancestor and widespread chromosomal fractionation. N Phytol 209:1252–1263. https://doi.org/10.1111/nph.13689

Yang J, Showalter AM (2007) Expression and localization of AtAGP18, a lysine-rich arabinogalactan-protein in *Arabidopsis*. Planta 226:169–179. https://doi.org/10.1007/s00425-007-0478-2

Yang J, Sardar HS, McGovern KR, Zhang Y, Showalter AM (2007) A lysine-rich arabinogalactan protein in *Arabidopsis* is essential for plant growth and development, including

cell division and expansion. Plant J 49:629–640. https://doi.org/10.1111/j.1365-313X.2006.02985.x

Zhang Y, Brown G, Whetten R, Loopstra CA, Neale D, Kieliszewski MJ, Sederoff RR (2003) An arabinogalactan protein associated with secondary cell wall formation in differentiating xylem of loblolly pine. Plant Mol Biol 52:91–102. https://doi.org/10.1023/A:1023978210001

Zhang Y, Yang J, Showalter AM (2011) AtAGP18, a lysine-rich arabinogalactan protein in *Arabidopsis thaliana*, functions in plant growth and development as a putative co-receptor for signal transduction. Plant Signal Behav 6:855–857. https://doi.org/10.4161/psb.6.6.15204