**A peer-reviewed version of this preprint was published in PeerJ on 29 May 2018.**



Liu B, Sun Y, Xue J, Jia X, Li R. 2018. Genome-wide characterization and expression analysis of GRAS gene family in pepper (*Capsicum annuum* L.) PeerJ 6:e4796 https://doi.org/10.7717/peerj.4796

# Genome-wide characterization and expression analysis of GRAS gene family in pepper (*Capsicum annuum* L.)

**Liu Baoling** [1] , **Sun Yan** [1] , **Xue Jinai** [1] , **Li Runzhi** Corresp. [1]

[1] Institute of Molecular Agriculture and Bioenergy, Shanxi Agricultural University, Jinzhong City, China

Corresponding Author: Li Runzhi
Email address: rli2001@126.com

Plant-specific GRAS transcription factors diversely participate in the regulation of multiple biological processes including growth and development, signal cross-talking and biotic/abiotic responses. However, this gene family was not characterized detailed in pepper ( *Capsicum annuum* L.), an economically important vegetable crop. Here, a total of 50 Ca GRAS members were identified in the pepper genome and renamed by their respective chromosomal distribution. Genomic organization revealed that most *CaGRAS* genes (84%) have no intron. A phylogenetic analysis was carried out using *Arabidopsis thaliana* to classify pepper *GARS* genes into at least ten subfamilies. Multiple sequence alignment showed GRAS-typical domains present in those proteins, with the members from the same phylogenetic subfamily exhibiting the similar motif composition. The presence of highly divergent N-terminus may be associated with functional specificity of each CaGRAS protein. Expression of 12 *CaGRAS* genes was not detected in all tissues tested, suggesting that their functions may be lost during evolution. By contrast, the rest 38 *CaGRAS* genes were expressed largely in several organs, showing their important roles in pepper life activities. Moreover, 21 *CaGRAS* genes were differentially expressed under cold, drought, salt and GA treatments, indicating that they play vital roles in response to abiotic stress in pepper. The first comprehensive analysis of GRAS gene family in the pepper genome in this study provide insights into understanding the CRAS-mediated regulation network, benefiting the genetic improvements in pepper and some other relative plants.

# Genome-wide characterization and expression analysis of GRAS gene family in pepper (*Capsicum annuum* L.)

Baoling Liu, Yan Sun, Jinai Xue, Runzhi Li*
Institute of Molecular Agriculture and Bioenergy, Shanxi Agricultural University,
Shanxi 030801, P. R. China

**Abstract**

Plant-specific GRAS transcription factors diversely participate in the regulation of multiple biological processes including growth and development, signal cross-talking and biotic/abiotic responses. However, this gene family was not characterized detailed in pepper (*Capsicum annuum* L.), an economically important vegetable crop. Here, a total of 50 CaGRAS members were identified in the pepper genome and renamed by their respective chromosomal distribution. Genomic organization revealed that most *CaGRAS* genes (84%) have no intron. A phylogenetic analysis was carried out using *Arabidopsis thaliana* to classify pepper *GARS* genes into at least ten subfamilies. Multiple sequence alignment showed GRAS-typical domains present in those proteins, with the members from the same phylogenetic subfamily exhibiting the similar motif composition. The presence of highly divergent N-terminus may be associated with functional specificity of each CaGRAS protein. Expression of 12 *CaGRAS* genes was not detected in all tissues tested, suggesting that their functions may be lost during evolution. By contrast, the rest 38 *CaGRAS* genes were expressed largely in several organs, showing their important roles in pepper life activities. Moreover, 21 *CaGRAS* genes were differentially expressed under cold, drought, salt and GA treatments, indicating that they play vital roles in response to abiotic stress in pepper. The first comprehensive analysis of GRAS gene family in the pepper genome in this study provide insights into understanding the CRAS-mediated regulation network, benefiting the genetic improvements in pepper and some other relative plants.

**Keywords**: *GRAS* genes; Pepper; Phylogeny; Gene expression; Genome-wide

## 1. Introduction

GRAS proteins, a group of plant-specific transcription regulators, are named after the acronyms of three initially identified members: GAI, RGA and SCR. Typically, the length distributions of GRAS proteins range from 400-770 amino acids (*Bolle 2004; Pysh et al. 1999*). Based on considerable sequence alignments, a typical GRAS protein usually contains five consecutive conserved motifs: LHR I, VHIID, LHR II, PFYRE and SAW at its conserved C-terminal region despite of highly variable N-termini (*Pysh et al. 1999; Sun et al. 2011*). VHIID with its flanking two leucine heptad repeats (LHR I and LHR II) are critical for protein-protein interactions. The mutagenesis of PFYRE and SAW motifs displayed distinct phenotype abnormality in *Arabidopsis thaliana*, indicating that they may contribute to the structural integrity of GRAS proteins (*Itoh et al. 2002; Silverstone et al. 1998*). In contrast, except for the members of DELLA subgroup characterized by two conserved N-terminal motifs (DELLA and TVHYNP), GRAS proteins displayed variable N-termini in their length and sequence, Such divergence in N-terminus may determine the various roles of GRAS proteins (*Sun et al. 2011*). Previous studies on *Arabidopsis* and rice (*Oryza sativa* L.) classified the GRAS family into 8 distinct subfamilies, namely DELLA, HAM, LISCL, PAT1, LAS, SCR, SHR and SCL3 (*Tian et al. 2004*). However, the number of distinct subgroups was ranged from 8 to 16 in other plants such as *Prunus mume*, tomato (*Solanum lycopersicum*) and maize (*Zea mays*). So far, GRAS family has been systematically investigated in nearly 30 plant species, including rice, *Arabidopsis*, tomato, poplar, Chinese cabbage, maize, *Medicago truncatula*, lily and pine (*Huang et al. 2015; Lu et al. 2015; Song et al. 2014; Tian et al. 2004*) , with a number of GRAS proteins functionally characterized.

Members of GRAS family perform diverse functions in plant growth, development, and physiological

processes, including axillary meristem formation, root development, gametogenesis, phytochrome and gibberellin acid (GA) signal transduction, and the response to biotic and abiotic stresses. For example, SCR and SHR, two independent sub-families of GRAS proteins, are both found to regulate root and shoot radial organization via SCR/SHR complex (*Cui et al. 2007; Helariutta et al. 2000*). DELLA members, which are distinctly different from other GRAS proteins because of the existence of DELLA and TVHVNP domains, usually act as inhibitors of GA signaling perception (*Sun & Gubler 2004*). Studies on *Arabidopsis* repressor of ga1-3 (AtRGA) and rice SLR1 indicated that DELLA members function within the nucleus, and the loss-of-function mutant of DELLA domain manifested a GA-insensitive dwarf status. SCARECROW-LIKE 3 (SCL3) expressed mainly in endodermis is essential for integrating downstream pathways of SCR/SHR and GA/DELLA, and controlling GA homeostasis during root development (*Zhang et al. 2011*). SCL13 (PAT1 subfamily) in *Arabidopsis* has been reported to participate in phytochrome-B (phyB) signal transduction (*Bolle et al. 2000*), whereas other members of the same subfamily including PAT1, SCL5 and SCL21 mainly function as positive regulators mediating phyA signaling pathway to control plant development (*Torres-Galea et al. 2006*). Another GRAS member MOC1 is a positive regulator of rice tillering, which is directly related to the increase of grain yield.

Pepper (*Capsicum annuum* L.) is an economically important vegetable and has tremendous value for providing food, spice, coloring agent, pharmaceuticals and ornamental products (*Kim et al. 2014; Qin et al. 2014*). In 2013, the total pepper production of the world already reached 34.9 million tons, making it the second largest Solanaceae crop after tomato (*Kim et al. 2014*). The accomplishment of whole pepper genome sequencing project in 2014 provides a platform for us to conduct a genome-wide analysis for an entire gene family and explore the right gene which is critical for pepper growth and development (*Kim et al. 2014; Qin et al. 2014*). By far, transcription factor families, such as WRKY, Dof, SBP-Box and Hsp70 have been characterized in pepper (*Guo et al. 2016; Guo et al. 2015b; Wu et al. 2016*). However, no information is available regarding the GRAS proteins of pepper despite their important roles in plant growth regulation. Here, we firstly describe the entire members of GRAS family in pepper using comparative genome analysis tools and experimental verification. Total of 50 *CaGRAS* genes were identified based on pepper genome sequence. The intron/exon organization and protein structure of each *GRAS* gene were also characterized, together with their phylogenetic relationships and chromosomal locations. Subsequently, we examined the function diversity of *CaGRAS* members by conserved motif analysis, followed by real-time PCR to profile their expression patterns in different tissues and various stress treatments. The present data provide essential information for further studies on molecular functions of *GRAS* genes in regulation of pepper growth and development as well as environmental responses.

## 2. MATERIALS AND METHODS

### 2.1 Identification and annotation of pepper *GRAS* genes.

Whole genome data for pepper cv. CM334 and cv. Zunla-1 were used for this study, and their genome information were downloaded from http://peppergenome.snu.ac.kr/download.php and http://peppersequence.genomics.cn/ respectively (*Kim et al. 2014; Qin et al. 2014*). *Arabidopsis* GRAS protein sequences previously reported were obtained from *Arabidopsis* Information Resource (https://www.Arabidopsis.org/) (*Tian et al. 2004*). The latest Hidden Markov Model (HMM) of GRAS domain (PF03514.11) (http://pfam.sanger.ac.uk/) was used as a BLAST query to search against the entire protein datasets of cv. CM334 and cv. Zunla-1 with an E-value of $1e^{-5}$ using HMMER 3.0 (*Huang et al. 2015*). Meanwhile, all AtGRAS proteins were used as queries to search against the two pepper databases using default parameters. The length of all hits out of the range from 350 to 820 aa was rejected. In order to validate their

putative accuracy, conserved domains essential for GRAS proteins were evaluated by SMART (http://smart.embl-heidelberg.de/) and PFAM database. Finally, all outputs from two independent databases were aligned and those having similar GRAS core domain were deemed as the same gene. After these stringent criterions, sequences with the presence of GRAS domain were retained for further analysis. In our study, we refer to the variety cv. CM334 as the reference for subsequent whole genome-wide analysis.

**2.2 Phylogenetic and evolutionary analysis of *GRAS* genes**

All screened GRAS proteins from *Arabidopsis* and pepper were used for multiple alignments by ClustalW program (*Larkin et al. 2007*). The gene IDs of GRAS members were listed in Table S1. Maximum likelihood method was adopted to generate unrooted phylogenetic tree using MEGA 6.0 based on alignment results. Reliability of phylogenetic tree was estimated with 1,000 bootstrapping replicates (*Tamura et al. 2013*). GRAS members in pepper were further categorized into different subfamilies based on the well-classified *GRAS* genes in *Arabidopsis (Tian et al. 2004)*.

**2.3 Chromosome localization and gene duplication analysis**

Physical positions of *CaGRAS* genes were extracted from pepper genome annotation file, and these genes were plotted onto the chromosomes using Mapchart 2.3 (*Voorrips 2002*). We then renamed every *GRAS* gene according to its ascending chromosomal distribution. Existing tandem duplications (TDs) were characterized as contiguous homologous genes located in a 100-kb single region or separated by less than 5 genes, while the whole blocks of genes copying from one chromosome region to another were defined as segmental duplications (SDs) (*Tang et al. 2008*). The mean Ks (synonymous rate) value between duplicated gene pairs was effective to deduce the selection modes and determines the time (Mya, million years ago) of duplication events. The ratio of non-synonymous (Ka) to synonymous substitution rates between pepper segmental duplicated gene pairs were calculated by PAL2NAL (http://www.bork.embl.de/pal2nal/) (*Suyama et al. 2006*). The approximate time of segmental duplicated events (T) was subsequently calculated using the formula: $T = Ks/(2*\lambda)*10^{-6}$ based on universal clock-like rate of $6.1*10^{-9}$ substitutions per site per year for pepper (*Guo et al. 2015b*).

**2.4 Protein property and gene structure analysis**

With the help of MEME (Multiple Expectation maximization for Motif Elicitation, http://meme-suite.org/), conserved motifs of GRAS proteins were searched with the following parameters: 1. maximum number of motif was 12; 2. optimum motif width was set from 6 to 50aa (*Bailey et al. 2009*). These identified motifs were further validated using InterProScan (http://www.ebi.ac.uk/Tools/pfa/iprscan/) (*Mulder & Apweiler 2007*). The properties of GRAS proteins were calculated on ExPASy online server (http://web.expasy.org/), such as molecular weight (MW), isoelectric point (pI), instability index and GRAVY (grand average of hydropathy) value (*Gasteiger et al. 2003*). Based on the relationships of coding sequence and its corresponding genomic DNA sequence, the final exon/intron distribution of each *CaGRAS* gene was illustrated by GSDS 2.0 (Gene Structure Display Server, http://gsds.cbi.pku.edu.cn/) (*Hu et al. 2015*).

**2.5 Prediction of CaGRAS protein-protein interaction network**

To further clarify the relationships between CaGRASs, a protein-protein interaction network was predicted using their interolog members from *Arabidopsis*. First, specific homologous relationships between *Arabidopsis* AtGRASs and pepper CaGRASs were mapped from INPARANOID database (http://inparanoid.sbc.su.se/cgi-bin/gene_search.cgi) (*Remm et al. 2001*). Then, we retrieved the interaction information among AtGRASs from AraNet database (http://www.functionalnet.org/aranet/) and mapped these attributions to CaGRASs to generate corresponding interaction relationships for pepper (*Guo et al. 2015b; Lee et al. 2010*). Finally, these interaction networks among CaGRASs were visualized using Cytoscape version 3.4.0 (*Shannon et al. 2003*).

**2.6 Expression analysis of *CaGRAS* genes in different tissues**

The transcriptome data of leaf, stem, root, pericarp and placenta at mature green, breaker, 5 and 10 days post-breaker, 6, 16 and 25 days post-anthesis (PC-MG, PL-MG, PC-B, PL-B, PC-B5, PC-B10, PL-B5, PL-B10, PC-6DPA, PC-16DPA, PC-25DPA, PL-6DPA, PL-16DPA, PL-25DPA) for pepper cv. CM334 have been previously generated (*Guo et al. 2015a; Kim et al. 2014*). We retrieved the FPKM (fragments per kilobase per million reads) value representing the expression level of each *CaGRAS* gene and displayed the result using BAR Heatmapper Plus.

**2.7 Pepper plant preparation and stress treatments**

Pepper plants were grown on soil in greenhouse with conditions: 14/10 h photoperiod, 25/20 °C day/night temperature and 60 % relative humidity. In this study, pepper seedlings with 6-8 true leaves were randomly divided into five groups, namely control (untreated) and treatment with cold (4±1 ℃), salt (300 mM NaCl), drought (400 mM mannitol) and gibberellin solution (20 μM GA). Leaves were sampled at 3 h after the treatment. For each treatment, leaves from five randomly selected seedlings were bulked to form one sample, and six biological replicate samples were immediately frozen in liquid nitrogen and then stored at -80°C before use.

**2.8 RNA isolation and qRT-PCR analysis**

Total RNA from leaves was extracted using Total RNA kit (BioTeke, Beijing, China) and reversely transcribed into cDNA using M-MLV Reverse Transcriptase (Promega). Real-time quantitative PCR (qRT-PCR) experiment was done using SYBR GREEN I Master Mix (Applied Biosystems) on iCycler iQ™ thermocycle (Bio-Rad). Each reaction volume contained 12.5 μl of SYBR GREEN Mix, 1 μl of each primer, 5 μl of 10 × diluted cDNA, and 5.5 μl of nuclease-free water. The reaction program was set as follows: initial polymerase incubation at 95 °C for 10 min, then 40 cycles of 95 °C for 15 s, 60 °C for 45 s. Melting curve analysis was conducted with heating the PCR product from 60 °C to 95 °C for verifying the specificity of the primers. The relative expression levels of *CaGRAS* genes were calculated based on the comparative Ct method using the $2^{-\triangle\triangle Ct}$ method with the *actin1* as an internal reference gene. Primer pairs (Table S3) were designed by Primer Premier 5.0 and checked by NCBI Primer BLAST.

**Results**

**3.1 Genome-wide identification of *CaGRAS* gene family**

We employed two different approaches to identify GRAS members in pepper genome. Totally, 50 non-redundant *CaGRAS* genes were found from variety cv. CM334, concurrent with the corresponding genes from cv. Zunla-1 (Table 1). Nearly all these proteins were detected to have one representative GRAS domain (PF03514.11), with the exception of two CaGRAS (CA00g84110, CA01g26680) having two and one CaGRAS (CA00g84090) having three such domains. The molecular mass and length of CaGRAS proteins varied greatly, with molecular weights ranging from 48 to 87 KDa and length from 419 to 801aa. The average theoretical pI was 6.1, implying that most CaGRAS proteins were weakly acidic. Only CaGRAS21 was stable for its instability index less than 40 and the rest were considered as unstable. All CaGRASs were predicted to be hydrophilic due to the GRAVY of each protein was less than 0. Most of CaGRAS proteins contained large percentage of aliphatic amino acids, with predicted aliphatic index ranging from 65.74 to 95.76. Interestingly, most of *CaGRAS* genes (84%) were intronless, while seven members had just one intron. Only one *CaGRAS* gene had more than one intron (Fig. 1) (*Chen et al. 2015; Huang et al. 2015; Wu et al. 2015*).

**3.2 Chromosomal localization and gene duplication analysis**

Except for six members (CaGRAS45-50) unmapped to a specific chromosome, 44 of the 50 *CaGRAS* genes were unevenly distributed across 11 out of 12 pepper chromosomes (Chr), with the exclusion of Chr11. This indicates that the *GRAS* genes may have been abundant across the genome of common ancestor. Among

those anchored members, Chr7 occupied the largest number of *GRAS* genes (n=7; 15.22%), followed by Chr1 (n=6; 13.04%) and the other three chromosomes (Chr2, Chr5 and Chr12) each having five *GRAS* genes. Additionally, four *GRAS* genes were located on Chr4 while three genes were detected on Chr3, Chr4 and Chr9, respectively. Only one and two *GRAS* genes were separately found on Chr8 and Chr10. Notably, most of *CaGRAS* genes were gathered at both ends of chromosomes.

Furthermore, we analyzed duplication events of *CaGRA*S gene in pepper genome since gene duplication acts importantly on the occurrence of novel functions and gene family expansion (*Zhang et al. 2016*). As shown in Fig. 2, two tandem duplication regions (CaGRAS4/5 and CaGRAS20/21) were distributed on Chr1 and Chr5, respectively. Three pairs of CaGRAS members (CaGRAS2/44, CaGRAS13/24 and CaGRAS18/40) were confirmed as the products of segmental duplications by Plant Genome Duplication Database analysis (http://chibba.agtec.uga.edu/duplication/) (*Lee et al. 2013*). The mean Ks value was used to estimate the time of duplication events. The ratio of Ka/Ks for segmental duplicated gene pairs ranged from 0.0982 (CaGRAS2/44) to 0.2612 (CaGRAS18/40) with an average of 0.1809, indicating that these genes experienced strong purifying selection pressure during evolution processes. In addition, three pairs of segmental duplicated genes occurred mainly between 53.5 (CaGRAS18/40) and 101.8Mya (CaGRAS13/24). These data implied there were large-scale genome duplication events during this period. Considering that the species differentiation time between tomato and pepper is approximately 19.1 Mya (*Kim et al. 2014*), we estimated that a large-scale intragenomic duplication occurred before the split of pepper and tomato.

**3.3 Phylogenetic analysis, classification and functional characterization of CaGRAS family**

To uncover the evolutionary relationships among CaGRAS proteins and their classifications, we performed a phylogenetic analysis using 82 full-length GRAS proteins (32 from *Arabidopsis* and 50 from pepper). An unrooted phylogenetic tree was constructed (Fig. 3), demonstrating that 50 pepper CaGRAS proteins could be classified into ten distinct subfamilies based on clade support values and classification of *Arabidopsis* GRAS proteins. The 10 subfamilies were termed as DELLA, PAT1, SCL3, SHR, SCR, LISCL, HAM, LAS, DLT and Ca_GRAS, respectively. Of them, subfamily of Ca_GRAS just contains six CaGRAS members. The 10 subfamily classification of GRAS family in pepper is well in agreement with the classification in castor bean (*Xu et al. 2016*). To date, only the function of CaGRAS3 known as CaHAM (NCBI accession: XP_016569270.1), has been clearly described in pepper (*David-Schwartz et al. 2013*), showing CaGRAS3 involved in shoot apical meristem organization.

To explore the potential functions of *GRAS* genes in each subfamily, we conducted a comparative analysis between CaGRAS and AtGRAS members in the same subfamily. DELLA subfamily contained two pepper GRAS (CsGRAS14 and CaGRAS41) and five AtGRAS (GAI, RGA, RGL1, 2 and 3). The complete DELLA and TVHYNPS motifs were all detected in these group members (Fig. 4). Previous studies reported that DELLA proteins mainly regulate GA signal transduction pathway which affect plant growth and development, implying the similar roles of CsGRAS14 and CaGRAS41 (*Zhang et al. 2011*). PAT1 subfamily contained ten CaGRAS and six AtGRAS (PAT1, SCL1, 5, 8, 13, and 21) proteins. In this subfamily, PAT1 and SCL13 from *Arabidopsis* were shown to be involved in phyA and phyB signaling pathway, respectively, suggesting that pepper GRAS homologs might possess the identical functions of PAT1 and SCL13 (*Bolle et al. 2000*). SCL3 subfamily consisted of two CaGRAS (CaGRAS2 and CaGRAS44) and one AtGRAS (SCL3). The members in this subfamily may mediate GA homeostasis through integrating other signals during root growth because AtSCL3 was found to regulates root cell elongation by integrating multiple signals in Arabidopsis (*Zhang et al. 2011*). For subfamily SHR and SCR, AtSHR and AtSCR were detected to function importantly in maintaining stem cell and root meristem. It is reasonable to predict that those pepper GRAS homologs in these two

subfamilies may possess the similar functions (*Di Laurenzio et al. 1996*). For subfamily LISCL consisting of nine CaGRAS and six AtGRAS members, the biological roles of those GRAS members are mostly unknown although a homolog member (LiSCL) from *Lilium longiflorum* was proven to play an important regulatory role during microsporogenesis (*Morohashi et al. 2003*). The first HAM gene member in HAM subfamily was isolated from petunia and proved to promote shoot indeterminacy (*Stuurman et al. 2002*). CaGRAS3 in HAM subfamily was also demonstrated to be involved in SAM (shoot apical meristem) organization and axillary meristem development (*David-Schwartz et al. 2013*). LAS subfamily comprised two members from pepper and three from Arabidopsis. AtLAS proteins in this subfamily mainly function to regulate and promote the initiation of axillary meristems (*Liang et al. 2014*). DLT subfamily, the smallest group, only contained two members (one from pepper, and the other from Arabidopsis). The members of this group have been previously shown to participate in brassinosteroid signal pathway responsible for the plant height (*Tong et al. 2009*). For Ca_GRAS subfamily having six CaGRAS members, no *Arabidopsis* GRAS homologs were grouped into this subfamily, indicating that this subfamily may be pepper-specific.

Based on phylogenetic analysis, the orthologous relationships of *GRAS* genes from *Arabidopsis* and pepper were classified into three categories, namely pepper-specific group, the one-to-one group and n-to-n group. In one-to-one group, each pepper gene corresponded to one *Arabidopsis* gene. This group contained 12 gene pairs and these genes in pepper were thought to have well-conserved functions with *Arabidopsis* orthologs *(Chen et al. 2015)*. Pepper-specific group only have Ca_GRAS subfamily (six CaGRAS members). The remaining 32 CaGRAS members were defined as n-to-n group, which means n *CaGRAS* genes in pepper correspond to n *AtGRAS* genes in the same sub-branch.

To investigate the common features of pepper GRAS proteins in more detail, we used MEME suite to identify their conserved motifs and sequence logos. Total of 11 conserved motifs (named Motif 1-11) (Fig. 4) were identified, with more motifs locating at C-terminus than at N-terminus, Moreover, the motifs from the same subfamily nearly hold the similar patterns (Fig. 4). We then matched up the motifs with corresponding GRAS domains. It was found that Motif 10 and 4 is in LHRI domain at N-terminus, followed by Motif 7 and 1in VHIID domain, Motif 6 and 8 in LHRII domain, Motif 9, 3 and 11in PFYRE domain, and Motif 2 and 5 in SAW domain at C-terminus (Fig. 4). Among the ten subfamilies, the CaGRAS members from PAT1 and LISCL subfamilies all contained the 11 conserved motifs identified.

Calculation of Ka/Ks ratio could help us to understand which selection pressure exists in pepper during gene evolutionary process (*Chen et al. 2015*). Based on the results of phylogenetic analysis, we selected 12 pairs of *GRAS* orthologous genes from one-to-one group for calculating their Ka and Ks substitution rates (Table 2). The value of Ka/Ks between orthologous gene pairs suggested that most of amino acid substitutions have been cleared up by stabilizing selection, resulting in less numbers of amino acid substitutions and l slow evolution space. (*Kondrashov et al. 2002*).

**3.4 Prediction of CaGRAS protein-protein interaction network**

Due to unavailable reference for pepper interactome data, we predicted the protein-protein interaction relationships of CaGRAS members based on the interologs from *Arabidopsis (Guo et al. 2015b)*. We only obtained the interaction information for 19 CaGRAS proteins, and generated a complex interaction network using these proteins (Fig. 5). In general, the members from SCL3 subfamily (CaGRAS2 and CaGRAS44) owned more interaction partners than others. These were consistent with their working mechanisms considering the facts that AtSCL3 protein could regulate GA homeostasis by integrating other signal pathway (*Zhang et al. 2011*). CaGRAS33, a member of LAS subfamily directly interacted with nine CaGRAS members, while CaGRAS7 from the same subfamily only had three interaction partners. Surprisingly, no interaction partner was

detected for the CaGRAS proteins from DELLA and DLT subfamily.

### 3.5 Expression analysis of *CaGRAS* genes in various tissues and fruit development stages

The online available expression data of 38 *CaGRAS* genes in 17 pepper samples were investigated, including five tissues (leaf, stem, root, pericarp and placenta) and seven developmental stages of pericarp and placenta (mature green, breaker, 5 and 10 days post-breaker, 6, 16, 25 days post-anthesis) (Fig. 6). The RPKM value for each of those *CaGRAS* genes was listed in Table S2. The transcripts for the other 12 *CaGRAS* genes were not detected in any tissues (RPKM < 0.001), which may be the result of pseudogenes. Generally, 25 *CaGRAS* genes were detected to express in all tissues, with only five members (*CaGRAS8*, *CaGRAS16*, *CaGRAS29*, *CaGRAS38* and *CaGRAS48*) showing high expression levels (PPKM > 10). A number of *CaGRAS* genes exhibited a certain degree of tissue specificity. For example, *CaGRAS18* and *CaGRAS27* expressed only in pericarp, *CaGRAS35* and *CaGRAS43* were highly expressed in leaf while the transcripts of *CaGRAS30* and *CaGRAS34* largely accumulated in stem rather than in other tissues. Tissue-specific expression showed that these genes may highly participate in the corresponding tissue development. *CaGRAS28* homologous with AtPAT1 showed high expression level in leaves, which is in line with AtPAT1 function as a positive regulator in phyA signal pathway (*Bolle et al. 2000*). Several *CaGRAS* genes exhibited constitutive expression levels at most stages of pericarp development. For example, *CaGRAS7* and *CaGRAS42* displayed a relatively higher expression at green fruit stage (PC_6DPA and PC_16DPA), and then decreased gradually towards fruit ripening. This expression pattern implied that CaGRAS7 and CaGRAS42 may function importantly in the early fruit development. In addition, the similar expression patterns were often detected for gene pairs from duplication event, but not for all such genes. For instance, in the CaGRAS18/40 duplicated region, *CaGRAS40* was highly expressed, whereas the other showed the opposite expression pattern. These differences implied that duplicated *GRAS* gene pairs may have diverged evolutionary outcomes.

### 3.6 Response of *CaGRAS* genes to different stress treatments

In order to elucidate the functions of *CaGRAS* genes responsive to GA stimuli, qRT-PCR was performed to examine the expression of such genes in seedling leaves after treatment with GA. The results (Fig. 7) showed that GA treatment resulted in the expression changes of 14 *CaGRAS* genes, with high upregulation for three genes (*CaGRAS17, 28, and 37*) and significant downregulation for five genes (*CaGRAS8, 14, 19, 38, and 41*). To broaden our knowledge regarding how these genes are affected by GA, we conducted a comprehensive analysis on *cis*-elements in the promoter regions of such 14 *CaGRAS* genes. The 12 *CaGRAS* genes were detected to contain at least one *GARE* (GA responsive element) in their promoter sequences, again confirming the function of these genes in mediating GA signal pathway in pepper.

We further examined the expression levels of *CaGRAS* genes under abiotic stresses, including salt, drought and cold treatment. Compared to the control group, the expression of 12 *CaGRAS* genes were highly affected by these treatments, indicating that the 12 *CaGRAS* genes may involve in pepper responses to these abiotic stresses. The downregulated expression was detected for 6, 2 and 4 *CaGRAS* genes, respectively, under cold, drought and salt stresses. The upregulated genes exhibited a group-specific expression. For example, the expression of *CaGRAS* genes from DELLA subfamily was significantly induced under cold stress. The genes in SCL3 subfamily was highly upregulated under drought stress, and the genes in PAT1 subfamily were highly induced by GA and other four stress treatments. Therefore, it is possible that different CaGRAS members function in different stress responses.

## Discussion

With the rapid development of bioinformatics, information stored in genome sequences is increasingly to become the targets to explore the mechanisms about plant growth and development. Recent studies in a number

of higher plants by comparative genomics tools shown that GRAS transcription factors play significant roles in regulating plant development and physiological processes (*Huang et al. 2015; Lee et al. 2008; Wu et al. 2015; Xu et al. 2016*). However, limited knowledge was available for the function of *GRAS* genes in pepper. Hence, we conducted a systematic analysis on this important transcription factor family in pepper, including genome-wide identification of CaGRAS members, chromosomal localization, intron-exon structure, physical-chemical features, phylogenetic analysis and expression profiles in various pepper tissues as well as their responses to different stresses.

A total of 50 *CaGRAS* genes were obtained from 34,903 protein-coding genes in pepper genome. The number of *CaGRAS* genes is actually more than that in *Arabidopsis* (32), *P. mume* (45), castor bean (46) and cabbage (48), respectively (*Huang et al. 2015; Lee et al. 2008; Lu et al. 2015; Xu et al. 2016*), but less than those in tomato (53), rice (60) and *Populus* (106) (*Huang et al. 2015; Tian et al. 2004*). It is known that gene duplication events might be the major driving forces to the expansion of *GRAS* genes (*Huang et al. 2015*). Analysis of pepper and tomato genomes indicated that no recent genome duplication event occurred since species differentiation. Thus, we could rule out the influence of recent genome duplication events on the *GRAS* number. Taken tandem duplication as an example, 15 GRAS members originated from tandem duplicated regions were identified in tomato and only 4 members was detected to derive from such tandem-duplicated regions in pepper. However, pepper genome size (3.48 Gb) was about fourfold larger than tomato genome (900 Mb). It is likely that expansion mechanisms of *GRAS* genes are different among lineages.

The 50 CaGRAS proteins could be classified into ten subfamilies according to their conserved domains and sequence homology in *Arabidopsis (Tian et al. 2004)*. Although the conserved motifs were identical among all CaGRAS proteins, a number of differences in chemical-physical characteristics were also detected for CaGRAS members. These differences may due to the amino acid discrepancies in the non-conserved regions of CaGRAS members, implying that different CaGRAS proteins may act different functions in their own microenvironments (*Huang et al. 2015*).Notably, we found that pepper had a specific GRAS subfamily Ca_GRAS, which only contained six *CaGRA*S genes. In agreement of this, the species-specific GRAS subfamily also widely existed in other plant species, such as Os4 subfamily of OsGRAS being rice-specific, and Pt20 subfamily of PtGRAS being *Populus*-specific. These species-specific *GRAS* genes may be lost from some other plants or become highly specialized during evolution.

Another important finding is that most *CaGRAS* genes (84%) contain just one exon. The high percentage of such intronless *GRAS* genes is detected as 67.6%, 54.7%, 82.2% and 83.3% in *Arabidopsis*, *Populus*, *P. mume* and Chinese cabbage (*Lee et al. 2008; Lu et al. 2015; Song et al. 2014; Tian et al. 2004*), respectively, implying the close evolutionary relationship of GRAS proteins among these plant species. Besides *GRAS* genes, intronless genes were also enriched among some other gene families, such as *SAUR* genes, F-box gene families and DEAD box helicases (*Aubourg et al. 1999; Jain et al. 2007; Jain et al. 2006*). Given the fact that intronless genes are archetypical in prokaryotic genomes, the recent work by Zhang et al. (*Zhang et al. 2012*) showed that the origin plant *GRAS* genes is come from the prokaryotic genomes by horizontal gene transfer, followed by duplication events in evolutionary history. This may explain the formation of substantial intronless *GRAS* genes in pepper genome.

Generally, an intrinsically disordered region (IDR) in an intrinsically disordered protein (IDP) allows protein to recognize and interact with various partners, which are crucial for molecular function. Bioinformatics analysis showed that GRAS protein is a kind of IDP (*Sun et al. 2013*). One of a typical IDR in GRAS protein is its highly variable N-terminus, which possess short interaction-prone segments and molecular recognition features responsible for recognizing and binding the specific partner of GRAS proteins. Here, pepper GRAS

proteins were found to contain a highly variable N-terminal region, which is consistent with the notion that N-terminus of GRAS proteins were intrinsically disordered, contributing the functional divergence of CaGRAS proteins.

For functional characterization of those *CaGRAS* genes, an effective method is to identify highly homologous genes between *Arabidopsis* and pepper (*Chen et al. 2015*). Another approach is to profile expression patterns of *CaGRAS* genes, particularly for those members in pepper-specific *GRAS* subfamily without function information deduced from Arabidopsis. Our data showed that *CaGRAS4* may be a pseudogene because of no expression level detected in any tissues. CaGRAS5 might be involved in pericarp and placenta development, showing a relatively high abundance during all consecutive stages. On the whole, the expression profiles of *CaGRAS* genes varied greatly not only among different tissues, but members from the same clade. Likely, such a great expression variation was also observed for *GRAS* genes in *Populus* and *P.mume* (*Huang et al. 2015*). These results indicated that *GRAS* genes may have experienced neo-functionalization or sub-functionalization in many higher plants. The RPKM values of twelve *CaGRAS* genes from seven subfamilies (DELLA, PAT1, SHR, SCR, LISCL, LAS and Ca_GRAS) were not detected in any tissues, suggesting these genes may lose their functions during evolution. By contrast, higher expression levels of *GRAS* genes in several organs signified their important roles. For example, *CaGRAS29* from SHR subfamily was highly transcribed in root tissue, which is in agreement with the function of its homologous *AtSHR* responsible for root development (*Cui et al. 2007*). *CaGRAS41* from DELLA subfamily expressed in all tissues played critical roles in controlling a variety of signal hubs, whereas no expression of *CaGRAS2* from the same group was detected in any tissues. It seems that functional diversification is occurred for the two *CaGRAS* genes from the same subfamily. Overall, the current expression data obtained for *CaGRAS* genes in different tissues lay a foundation for further functional analysis of pepper GRAS members.

In general, hormones could regulate plant growth and development via the modulation of the related gene expression. GA is found to play important roles in many aspects of plant development such as organ elongation, germination and flowering time. It has been reported that expression of *GRAS* genes in tomato showed dose-dependent response to GA (*Huang et al. 2017*). Our results demonstrated that the majority of *CaGRAS* genes detected here displayed dramatic changes after GA treatment. The promoters of these *CaGRAS* genes contained at least one GA response element (*GARE*), implying that a set of CaGRAS proteins could regulate plant adaptability to adversity through a complex regulatory network. Additionally, previous studies revealed that *GRAS* genes could affect plant responses to abiotic stresses. For example, BnLAS and PeSCL7, GRAS members from *Brassica napus* and poplar, were identified as the good targets for engineering to increase plant drought and salt tolerance (*Ma et al. 2010; Yang et al. 2011*). Combined analysis of all qPCR results revealed that several pepper *GRAS* genes were associated with the three stress responses (cold, salt, and drought), showing the cross- talking of *GRAS* genes in regulation of plant responses against various adversity. Notably, we found that *CaGRAS* members belonging to PAT1 group exhibit the similar expression patterns when stressed by GA and other abiotic treatments. Consistently, *OsGRAS* genes from rice PAT1 group were also reported to be involved in GA and stress responses. All these indicate that some *GRAS* genes may specifically coordinate plant responses to multiple stresses.

**Conclusions**

In this study, 50 GRAS members were characterized from pepper genome, and classified into ten subfamilies based on phylogenetic analyses. Duplication event was identified as the main driving force to *GRAS* gene expansion in pepper. Interaction network and expression profiles among *GRAS* genes were examined, illustrating important roles of CaGRAS proteins in regulating GA and abiotic stress responses. Taken

together, the present study is the first comprehensive characterization of *GARS* genes in pepper. All these data provide the foundation to elucidate the GRAS-mediated molecular mechanism underlying plant growth and development as well as stress biology, showing that GRAS members could be selected as the targets for genetic improvement of stress tolerance in pepper and other related plants.

**ACKNOWLEDGEMENTS**

This work was supported by National Natural Science Foundation of China (Grant No. 30971806, 31201266, and 31401430), the State Ministry of Agriculture "948" Project (2014-Z39), Shanxi Province Key Projects of Coal-based Science and Technology (FT-2014-01), Research Project Supported by Shanxi Scholarship Council of China (2015-064), and the Key Project of the Key Research and Development Program of Shanxi Province, China (Grant No. 201603D312005).

**Authors' Contributions**

Baoling Liu: Design, formulation of bioinformatics process with writing manuscript. Yan Sun and Jinai Xue: Collection of data. Runzhi Li: Designed the study and corrected the article.

**REFERENCES**

**Aubourg S, Kreis M, and Lecharny A. 1999.** The DEAD box RNA helicase family in Arabidopsis thaliana. *Nucleic Acids Res* 27:628-636.

**Bailey TL, Boden M, Buske FA, Frith M, Grant CE, Clementi L, Ren J, Li WW, and Noble WS. 2009.** MEME SUITE: tools for motif discovery and searching. *Nucleic Acids Res* 37:W202-208.

**Bolle C. 2004.** The role of GRAS proteins in plant signal transduction and development. *Planta* 218:683-692.

**Bolle C, Koncz C, and Chua NH. 2000.** PAT1, a new member of the GRAS family, is involved in phytochrome A signal transduction. *Genes Dev* 14:1269-1278.

**Chen YQ, Tai SS, Wang DW, Ding AM, Sun TT, Wang WF, and Sun YH. 2015.** Homology-based analysis of the GRAS gene family in tobacco. *Genet Mol Res* 14:15188-15200.

**Cui H, Levesque MP, Vernoux T, Jung JW, Paquette AJ, Gallagher KL, Wang JY, Blilou I, Scheres B, and Benfey PN. 2007.** An evolutionarily conserved mechanism delimiting SHR movement defines a single layer of endodermis in plants. *Science* 316:421-425.

**David-Schwartz R, Borovsky Y, Zemach H, and Paran I. 2013.** CaHAM is autoregulated and regulates CaSTM expression and is required for shoot apical meristem organization in pepper. *Plant Sci* 203-204:8-16.

**Di Laurenzio L, Wysocka-Diller J, Malamy JE, Pysh L, Helariutta Y, Freshour G, Hahn MG, Feldmann KA, and Benfey PN. 1996.** The SCARECROW gene regulates an asymmetric cell division that is essential for generating the radial organization of the Arabidopsis root. *Cell* 86:423-433.

**Gasteiger E, Gattiker A, Hoogland C, Ivanyi I, Appel RD, and Bairoch A. 2003.** ExPASy: The proteomics server for in-depth protein knowledge and analysis. *Nucleic Acids Res* 31:3784-3788.

**Guo M, Liu JH, Lu JP, Zhai YF, Wang H, Gong ZH, Wang SB, and Lu MH. 2015a.** Genome-wide analysis of the CaHsp20 gene family in pepper: comprehensive sequence and expression profile analysis under heat stress. *Front Plant Sci* 6:806.

**Guo M, Liu JH, Ma X, Zhai YF, Gong ZH, and Lu MH. 2016.** Genome-wide analysis of the Hsp70 family genes in pepper (Capsicum annuum L.) and functional identification of CaHsp70-2 involvement in heat stress. *Plant Sci* 252:246-256.

**Guo M, Lu JP, Zhai YF, Chai WG, Gong ZH, and Lu MH. 2015b.** Genome-wide analysis, expression profile of heat shock factor gene family (CaHsfs) and characterisation of CaHsfA2 in pepper (Capsicum annuum L.). *BMC Plant Biol* 15:151.

**Helariutta Y, Fukaki H, Wysocka-Diller J, Nakajima K, Jung J, Sena G, Hauser MT, and Benfey PN**. **2000.** The SHORT-ROOT gene controls radial patterning of the Arabidopsis root through radial signaling. *Cell* 101:555-567.

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

**Huang W, Peng S, Xian Z, Lin D, Hu G, Yang L, Ren M, and Li Z**. **2017.** Overexpression of a tomato miR171 target gene SlGRAS24 impacts multiple agronomical traits via regulating gibberellin and auxin homeostasis. *Plant Biotechnol J* 15:472-488.

**Huang W, Xian Z, Kang X, Tang N, and Li Z**. **2015.** Genome-wide identification, phylogeny and expression analysis of GRAS gene family in tomato. *BMC Plant Biol* 15:209.

**Itoh H, Ueguchi-Tanaka M, Sato Y, Ashikari M, and Matsuoka M**. **2002.** The gibberellin signaling pathway is regulated by the appearance and disappearance of SLENDER RICE1 in nuclei. *Plant Cell* 14:57-70.

**Jain M, Nijhawan A, Arora R, Agarwal P, Ray S, Sharma P, Kapoor S, Tyagi AK, and Khurana JP**. **2007.** F-box proteins in rice. Genome-wide analysis, classification, temporal and spatial gene expression during panicle and seed development, and regulation by light and abiotic stress. *Plant Physiol* 143:1467-1483.

**Jain M, Tyagi AK, and Khurana JP**. **2006.** Genome-wide analysis, evolutionary expansion, and expression of early auxin-responsive SAUR gene family in rice (Oryza sativa). *Genomics* 88:360-371.

**Kim S, Park M, Yeom SI, Kim YM, Lee JM, Lee HA, Seo E, Choi J, Cheong K, Kim KT et al.** . **2014.** Genome sequence of the hot pepper provides insights into the evolution of pungency in Capsicum species. *Nat Genet* 46:270-278.

**Kondrashov AS, Sunyaev S, and Kondrashov FA**. **2002.** Dobzhansky-Muller incompatibilities in protein evolution. *Proc Natl Acad Sci U S A* 99:14878-14883.

**Larkin MA, Blackshields G, Brown NP, Chenna R, McGettigan PA, McWilliam H, Valentin F, Wallace IM, Wilm A, Lopez R et al.** . **2007.** Clustal W and Clustal X version 2.0. *Bioinformatics* 23:2947-2948.

**Lee I, Ambaru B, Thakkar P, Marcotte EM, and Rhee SY**. **2010.** Rational association of genes with traits using a genome-scale gene network for Arabidopsis thaliana. *Nat Biotechnol* 28:149-156.

**Lee MH, Kim B, Song SK, Heo JO, Yu NI, Lee SA, Kim M, Kim DG, Sohn SO, Lim CE et al.** . **2008.** Large-scale analysis of the GRAS gene family in Arabidopsis thaliana. *Plant Mol Biol* 67:659-670.

**Lee TH, Tang H, Wang X, and Paterson AH**. **2013.** PGDD: a database of gene and genome duplication in plants. *Nucleic Acids Res* 41:D1152-1158.

**Liang WH, Shang F, Lin QT, Lou C, and Zhang J**. **2014.** Tillering and panicle branching genes in rice. *Gene* 537:1-5.

**Lu J, Wang T, Xu Z, Sun L, and Zhang Q**. **2015.** Genome-wide analysis of the GRAS gene family in Prunus mume. *Mol Genet Genomics* 290:303-317.

**Ma HS, Liang D, Shuai P, Xia XL, and Yin WL**. **2010.** The salt- and drought-inducible poplar GRAS protein SCL7 confers salt and drought tolerance in Arabidopsis thaliana. *J Exp Bot* 61:4011-4019.

**Morohashi K, Minami M, Takase H, Hotta Y, and Hiratsuka K**. **2003.** Isolation and characterization of a novel GRAS gene that regulates meiosis-associated gene expression. *J Biol Chem* 278:20865-20873.

**Mulder N, and Apweiler R**. **2007.** InterPro and InterProScan: tools for protein sequence classification and comparison. *Methods Mol Biol* 396:59-70.

**Pysh LD, Wysocka-Diller JW, Camilleri C, Bouchez D, and Benfey PN**. **1999.** The GRAS gene family in Arabidopsis: sequence characterization and basic expression analysis of the SCARECROW-LIKE genes. *Plant J* 18:111-119.

**Qin C, Yu C, Shen Y, Fang X, Chen L, Min J, Cheng J, Zhao S, Xu M, Luo Y et al.** . **2014.** Whole-genome sequencing of cultivated and wild peppers provides insights into Capsicum domestication and specialization. *Proc Natl Acad Sci U S A* 111:5135-5140.

**Remm M, Storm CE, and Sonnhammer EL**. **2001.** Automatic clustering of orthologs and in-paralogs from pairwise species comparisons. *J Mol Biol* 314:1041-1052.

**Shannon P, Markiel A, Ozier O, Baliga NS, Wang JT, Ramage D, Amin N, Schwikowski B, and Ideker T**. **2003.** Cytoscape: a software environment for integrated models of biomolecular interaction networks. *Genome Res* 13:2498-2504.

**Silverstone AL, Ciampaglio CN, and Sun T**. **1998.** The Arabidopsis RGA gene encodes a transcriptional regulator repressing the gibberellin signal transduction pathway. *Plant Cell* 10:155-169.

**Song XM, Liu TK, Duan WK, Ma QH, Ren J, Wang Z, Li Y, and Hou XL**. **2014.** Genome-wide analysis of the GRAS gene family in Chinese cabbage (Brassica rapa ssp. pekinensis). *Genomics* 103:135-146.

**Stuurman J, Jaggi F, and Kuhlemeier C**. **2002.** Shoot meristem maintenance is controlled by a GRAS-gene mediated signal from differentiating cells. *Genes Dev* 16:2213-2218.

**Sun TP, and Gubler F**. **2004.** Molecular mechanism of gibberellin signaling in plants. *Annu Rev Plant Biol* 55:197-223.

**Sun X, Rikkerink EH, Jones WT, and Uversky VN**. **2013.** Multifarious roles of intrinsic disorder in proteins illustrate its broad impact on plant biology. *Plant Cell* 25:38-55.

**Sun X, Xue B, Jones WT, Rikkerink E, Dunker AK, and Uversky VN**. **2011.** A functionally required unfoldome from the plant kingdom: intrinsically disordered N-terminal domains of GRAS proteins are involved in molecular recognition during plant development. *Plant Mol Biol* 77:205-223.

**Suyama M, Torrents D, and Bork P**. **2006.** PAL2NAL: robust conversion of protein sequence alignments into the corresponding codon alignments. *Nucleic Acids Res* 34:W609-612.

**Tamura K, Stecher G, Peterson D, Filipski A, and Kumar S**. **2013.** MEGA6: Molecular Evolutionary Genetics Analysis version 6.0. *Mol Biol Evol* 30:2725-2729.

**Tang H, Bowers JE, Wang X, Ming R, Alam M, and Paterson AH**. **2008.** Synteny and collinearity in plant genomes. *Science* 320:486-488.

**Tian C, Wan P, Sun S, Li J, and Chen M**. **2004.** Genome-wide analysis of the GRAS gene family in rice and Arabidopsis. *Plant Mol Biol* 54:519-532.

**Tong H, Jin Y, Liu W, Li F, Fang J, Yin Y, Qian Q, Zhu L, and Chu C**. **2009.** DWARF AND LOW-TILLERING, a new member of the GRAS family, plays positive roles in brassinosteroid signaling in rice. *Plant J* 58:803-816.

**Torres-Galea P, Huang LF, Chua NH, and Bolle C**. **2006.** The GRAS protein SCL13 is a positive regulator of phytochrome-dependent red light signaling, but can also modulate phytochrome A responses. *Mol Genet Genomics* 276:13-30.

**Voorrips RE**. **2002.** MapChart: software for the graphical presentation of linkage maps and QTLs. *J Hered* 93:77-78.

**Wu Z, Cheng J, Cui J, Xu X, Liang G, Luo X, Chen X, Tang X, Hu K, and Qin C**. **2016.** Genome-Wide Identification and Expression Profile of Dof Transcription Factor Gene Family in Pepper (Capsicum annuum L.). *Front Plant Sci* 7:574.

**Wu ZY, Wu PZ, Chen YP, Li MR, Wu GJ, and Jiang HW**. **2015.** Genome-wide analysis of the GRAS gene family in physic nut (Jatropha curcas L.). *Genet Mol Res* 14:19211-19224.

**Xu W, Chen Z, Ahmed N, Han B, Cui Q, and Liu A**. **2016.** Genome-Wide Identification, Evolutionary Analysis,

and Stress Responses of the GRAS Gene Family in Castor Beans. *Int J Mol Sci* 17.

**Yang M, Yang Q, Fu T, and Zhou Y**. **2011.** Overexpression of the Brassica napus BnLAS gene in Arabidopsis affects plant development and increases drought tolerance. *Plant Cell Rep* 30:373-388.

**Zhang D, Iyer LM, and Aravind L**. **2012.** Bacterial GRAS domain proteins throw new light on gibberellic acid response mechanisms. *Bioinformatics* 28:2407-2411.

**Zhang XH, Liu HQ, Guo QW, Zheng CF, Li CS, Xiang XM, Zhao DF, Liu J, Luo J, Zhao DK et al.** . **2016.** Genome-wide identification, phylogenetic relationships, and expression analysis of the carotenoid cleavage oxygenase gene family in pepper. *Genet Mol Res* 15.

**Zhang ZL, Ogawa M, Fleet CM, Zentella R, Hu J, Heo JO, Lim J, Kamiya Y, Yamaguchi S, and Sun TP**. **2011.** Scarecrow-like 3 promotes gibberellin signaling by antagonizing master growth repressor DELLA in Arabidopsis. *Proc Natl Acad Sci U S A* 108:2160-2165.

X

## Table 1(on next page)

Accession members and characteristics of 50 *CaGRAS* genes in pepper

**Table 1 Accession members and characteristics of 50 *CaGRAS* genes in pepper**

| ID | Name | Chr | Position (Mb) | Group | Length (aa) | MW (KDa) | pI | Aliphatic index | Instability index | GRVY | Corresponding ID in Zunla-1 |
|---|---|---|---|---|---|---|---|---|---|---|---|
| CA01g03260 | CaGRAS1 | Chr1 | 5.207267 | DLT | 679 | 75.1857 | 6.2779 | 83.61 | 62.53 | -0.394837758 | Capana01g000318 |
| CA01g12960 | CaGRAS2 | Chr1 | 58.83842 | SCL3 | 473 | 53.2709 | 6.4116 | 90.93 | 51.59 | -0.255932203 | Capana00g001336 |
| CA01g13150 | CaGRAS3 | Chr1 | 59.375434 | HAM | 694 | 76.1067 | 5.7836 | 86.13 | 60.78 | -0.143001443 | Capana01g000561 |
| CA01g23320 | CaGRAS4 | Chr1 | 178.490414 | Ca_GRAS | 590 | 67.293 | 4.7921 | 85.7 | 43.57 | -0.294057725 | Capana01g002881 |
| CA01g23330 | CaGRAS5 | Chr1 | 178.494383 | Ca_GRAS | 565 | 63.5108 | 6.2048 | 84.72 | 41.97 | -0.291489362 | Capana01g002882 |
| CA01g31850 | CaGRAS6 | Chr1 | 259.625208 | SHR | 419 | 47.7891 | 5.269 | 81.17 | 52.41 | -0.468421053 | Capana01g003866 |
| CA02g22690 | CaGRAS7 | Chr2 | 157.892927 | LAS | 588 | 64.9189 | 5.3529 | 81.6 | 53.65 | -0.291111111 | Capana02g002687 |
| CA02g22940 | CaGRAS8 | Chr2 | 158.229659 | HAM | 549 | 60.6968 | 5.9537 | 85.64 | 42.76 | -0.116058394 | Capana02g002660 |
| CA02g25090 | CaGRAS9 | Chr2 | 161.98608 | SHR | 452 | 50.9813 | 5.7711 | 91.24 | 43.69 | -0.155432373 | Capana02g002989 |
| CA02g25280 | CaGRAS10 | Chr2 | 162.410539 | SHR | 529 | 60.0349 | 6.23 | 67.61 | 50.98 | -0.56875 | Capana02g002967 |
| CA02g29990 | CaGRAS11 | Chr2 | 169.217032 | LISCL | 471 | 53.7333 | 5.0358 | 90.32 | 48.45 | -0.280851064 | Capana02g003543 |
| CA03g07840 | CaGRAS12 | Chr3 | 23.813921 | SHR | 563 | 63.2008 | 6.2047 | 80.34 | 49.19 | -0.497330961 | Capana03g000095 |
| CA03g18670 | CaGRAS13 | Chr3 | 207.249468 | PAT1 | 642 | 70.7949 | 7.1571 | 82.15 | 57.55 | -0.417316693 | Capana03g002179 |
| CA03g37140 | CaGRAS14 | Chr3 | 257.856059 | DELLA | 551 | 60.8579 | 5.0655 | 80.16 | 44.9 | -0.251636364 | Capana03g000088 |
| CA04g11230 | CaGRAS15 | Chr4 | 145.000017 | SCR | 475 | 54.4263 | 6.7004 | 92.57 | 41.19 | -0.284177215 | Capana04g001618 |
| CA04g11770 | CaGRAS16 | Chr4 | 164.396824 | HAM | 508 | 56.4831 | 5.8539 | 80.45 | 40.71 | -0.281620553 | Capana04g002119 |
| CA04g12860 | CaGRAS17 | Chr4 | 178.976341 | PAT1 | 564 | 62.9078 | 4.8145 | 76.41 | 50.61 | -0.387921847 | Capana04g001479 |
| CA05g01670 | CaGRAS18 | Chr5 | 2.926557 | PAT1 | 536 | 59.4974 | 6.0317 | 81.87 | 52.52 | -0.234018692 | Capana05g000176 |
| CA05g03110 | CaGRAS19 | Chr5 | 7.634233 | LISCL | 743 | 83.3188 | 6.0418 | 76.77 | 49.39 | -0.473719677 | Capana05g000332 |
| CA05g12700 | CaGRAS20 | Chr5 | 182.973215 | SHR | 567 | 64.1302 | 6.5097 | 79.12 | 42.35 | -0.510600707 | Capana05g001029 |
| CA05g12710 | CaGRAS21 | Chr5 | 182.978914 | SHR | 594 | 66.9248 | 6.6408 | 78.31 | 37.9 | -0.52250423 | Capana05g001029 |
| CA05g17900 | CaGRAS22 | Chr5 | 227.253914 | PAT1 | 541 | 60.2005 | 5.8782 | 76.2 | 47.11 | -0.382777778 | Capana05g000176 |
| CA06g00220 | CaGRAS23 | Chr6 | 0.170158 | Ca_GRAS | 562 | 63.6674 | 5.6117 | 88.45 | 46.66 | -0.256327986 | Capana00g003286 |
| CA06g07510 | CaGRAS24 | Chr6 | 105.87898 | PAT1 | 608 | 66.6011 | 5.5616 | 85.3 | 47.88 | -0.308731466 | Capana00g005111 |
| CA06g24700 | CaGRAS25 | Chr6 | 231.128198 | LISCL | 751 | 85.0266 | 5.1087 | 74.24 | 48.17 | -0.504266667 | Capana06g000410 |
| CA06g25920 | CaGRAS26 | Chr6 | 233.231592 | LISCL | 681 | 76.867 | 6.433 | 65.74 | 42.65 | -0.593088235 | Capana06g000274 |
| CA07g08560 | CaGRAS27 | Chr7 | 124.080011 | SCR | 440 | 48.6504 | 5.1957 | 95.76 | 60.13 | -0.070615034 | Capana07g001083 |
| CA07g10940 | CaGRAS28 | Chr7 | 179.179162 | PAT1 | 542 | 60.4749 | 6.7096 | 82.61 | 52.9 | -0.239001848 | Capana07g001257 |
| CA07g12530 | CaGRAS29 | Chr7 | 198.623614 | SHR | 432 | 48.7514 | 5.4183 | 89.58 | 43.54 | -0.224361949 | Capana07g001537 |
| CA07g14700 | CaGRAS30 | Chr7 | 213.163858 | HAM | 518 | 59.123 | 5.069 | 79.42 | 51.53 | -0.371760155 | Capana07g001856 |
| CA07g18600 | CaGRAS31 | Chr7 | 226.087613 | PAT1 | 583 | 64.5701 | 6.791 | 83.83 | 51.41 | -0.300343643 | Capana07g002280 |
| CA07g20170 | CaGRAS32 | Chr7 | 228.957285 | PAT1 | 548 | 61.0729 | 5.4493 | 81.3 | 51.34 | -0.27714808 | Capana07g002351 |
| CA07g21550 | CaGRAS33 | Chr7 | 231.511151 | LAS | 449 | 50.6593 | 7.9034 | 93.01 | 49.39 | -0.234821429 | Capana07g002493 |
| CA08g12450 | CaGRAS34 | Chr8 | 132.579365 | LISCL | 688 | 77.8244 | 5.7692 | 74.12 | 54.14 | -0.51516035 | Capana08g001582 |
| CA09g05170 | CaGRAS35 | Chr9 | 30.647439 | LISCL | 763 | 86.3099 | 5.6912 | 78.36 | 41.38 | -0.498031496 | Capana09g001814 |
| CA09g05400 | CaGRAS36 | Chr9 | 34.265329 | LISCL | 748 | 84.6521 | 5.219 | 79.53 | 44.93 | -0.493038822 | Capana09g001799 |
| CA09g13460 | CaGRAS37 | Chr9 | 230.053566 | PAT1 | 574 | 64.3627 | 5.8404 | 85.62 | 42.13 | -0.305061082 | Capana09g000709 |
| CA10g04850 | CaGRAS38 | Chr10 | 30.048136 | LISCL | 769 | 86.7748 | 6.5228 | 79.14 | 44.92 | -0.481510417 | Capana00g002382 |
| CA10g10180 | CaGRAS39 | Chr10 | 157.782881 | SCR | 801 | 87.3531 | 6.1051 | 82.91 | 53.78 | -0.34425 | Capana10g001031 |
| CA12g00700 | CaGRAS40 | Chr12 | 2.029177 | PAT1 | 533 | 59.7566 | 5.6196 | 80.85 | 48.1 | -0.32537594 | Capana12g002864 |
| CA12g02780 | CaGRAS41 | Chr12 | 5.978997 | DELLA | 597 | 64.8974 | 4.7443 | 78.91 | 45.53 | -0.32147651 | Capana05g000798 |
| CA12g08480 | CaGRAS42 | Chr12 | 43.593325 | HAM | 488 | 55.5876 | 5.1055 | 95.24 | 43.35 | -0.131827515 | Capana12g002007 |
| CA12g21180 | CaGRAS43 | Chr12 | 232.987478 | Ca_GRAS | 518 | 58.874 | 5.312 | 91.8 | 43.96 | -0.27040619 | Capana12g000175 |
| CA12g21800 | CaGRAS44 | Chr12 | 233.939464 | SCL3 | 471 | 52.5634 | 6.5729 | 93.57 | 52.8 | -0.142978723 | Capana12g000112 |
| CA00g42950 | CaGRAS45 | Scaffold1070 | 0.654105 | SCR | 519 | 57.5813 | 6.7479 | 83.44 | 42.65 | -0.286293436 | Capana00g002482 |
| CA00g63410 | CaGRAS46 | Scaffold1392 | 0.061625 | LISCL | 787 | 87.7766 | 5.7328 | 83.99 | 45.41 | -0.33129771 | Capana05g002208 |
| CA00g66790 | CaGRAS47 | Scaffold1455 | 0.370366 | SHR | 567 | 64.1572 | 6.4197 | 79.12 | 41.5 | -0.514664311 | Capana00g002290 |
| CA00g67630 | CaGRAS48 | Scaffold1469 | 0.384617 | HAM | 757 | 82.3972 | 5.9771 | 84.13 | 54.44 | -0.218915344 | Capana01g003567 |
| CA00g84090 | CaGRAS49 | Scaffold1805 | 0.019698 | Ca_GRAS | 495 | 57.4293 | 5.5379 | 85.08 | 43.09 | -0.243902439 | Capana00g000912 |
| CA00g84110 | CaGRAS50 | Scaffold1805 | 0.106047 | Ca_GRAS | 445 | 51.0615 | 5.3951 | 88.51 | 45.76 | -0.22454955 | Capana00g000912 |

## Table 2(on next page)

Calculation of Ka and Ks of 12 pairs of GRAS orthologs between pepper and *Arabidopsis*. (Ka indicates nonsynonymous substitution rate, and Ks indicates synonymous substitution rate.)

Ka indicates nonsynonymous substitution rate, and Ks indicates synonymous substitution rate.

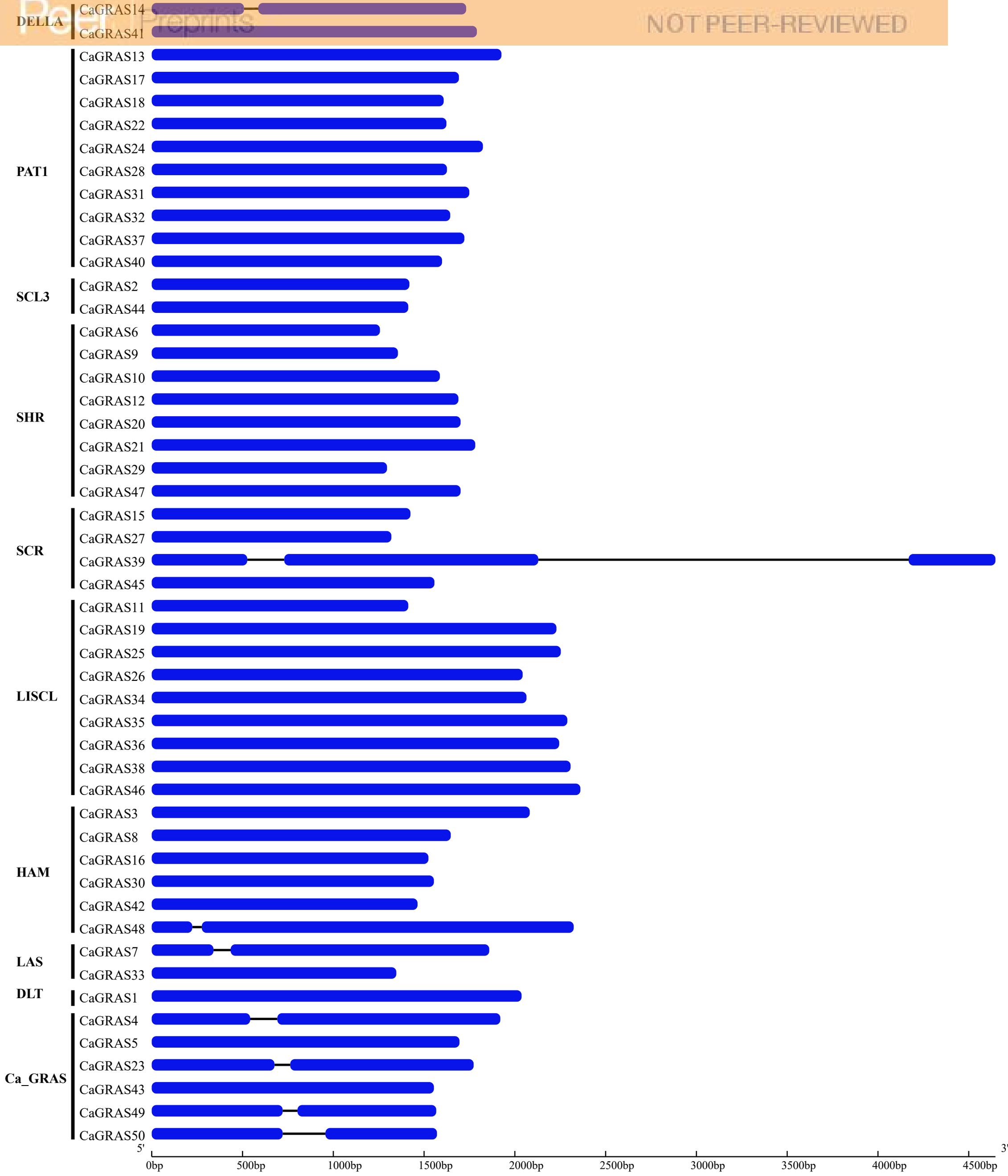
**Table 2 Calculation of Ka and Ks of 12 pairs of GRAS orthologs between pepper and *Arabidopsis*, (Ka indicates nonsynonymous substitution rate, and Ks indicates synonymous substitution rate.)**

| Gene pairs | Ka | Ks | Ka/Ks |
|---|---|---|---|
| *CaGRAS17 vs SCL1* | 0.3519 | 3.2457 | 0.184 |
| *CaGRAS28 vs PAT1* | 0.2246 | 3.6917 | 0.0608 |
| *CaGRAS2 vs SCL3* | 0.2305 | 2.6265 | 0.0878 |
| *CaGRAS10 vs SHR* | 0.3046 | 4.2501 | 0.0717 |
| *CaGRAS9 vs SCL32* | 0.2975 | 3.1564 | 0.0943 |
| *CaGRAS27 vs SCL23* | 0.2668 | 3.0907 | 0.0863 |
| *CaGRAS39 vs SCR* | 0.316 | 4.9112 | 0.0643 |
| *CaGRAS16 vs SCL26* | 0.4648 | 7.8487 | 0.0592 |
| *CaGRAS8 vs SCL25* | 0.4185 | 4.1404 | 0.1011 |
| *CaGRAS48 vs SCL6* | 0.3707 | 5.8567 | 0.0633 |
| *CaGRAS33 vs LAS* | 0.4388 | 5.3261 | 0.0824 |
| *CaGRAS1 vs SCL28* | 0.3704 | 3.4156 | 0.1085 |

## Figure 1(on next page)

Exon-intron structure of *CaGRAS* genes. Blue box indicates exon, and black line indicates intron

Blue box indicates exon, and black line indicates intron.

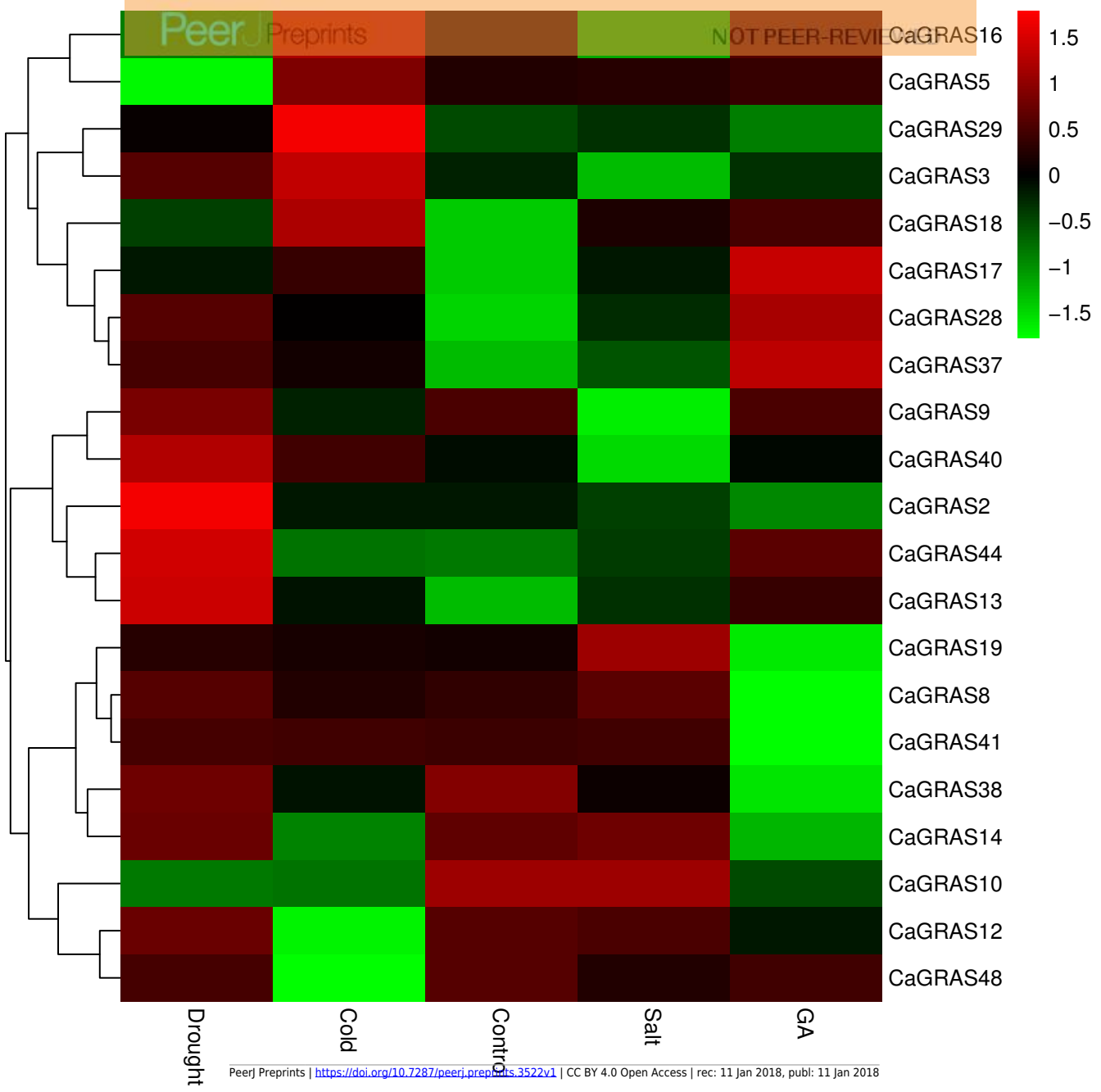



PeerJ Preprints | https://doi.org/10.7287/peerj.preprints.3522v1 | CC BY 4.0 Open Access | rec: 11 Jan 2018, publ: 11 Jan 2018

## Figure 2(on next page)

Differential expression analyses of 21 *GRAS* genes under GA, drought, salt and cold treatment in pepper seedlings. The color scale represents log2 expression values

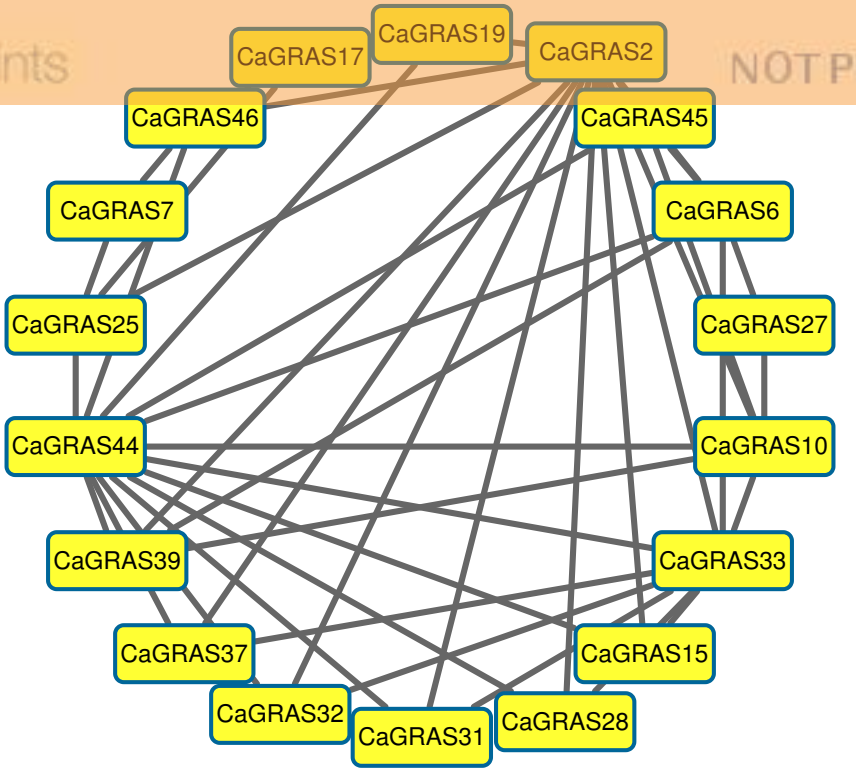
The color scale represents log2 expression values.




PeerJ Preprints | https://doi.org/10.7287/peerj.preprints.3522v1 | CC BY 4.0 Open Access | rec: 11 Jan 2018, publ: 11 Jan 2018

## Figure 3(on next page)

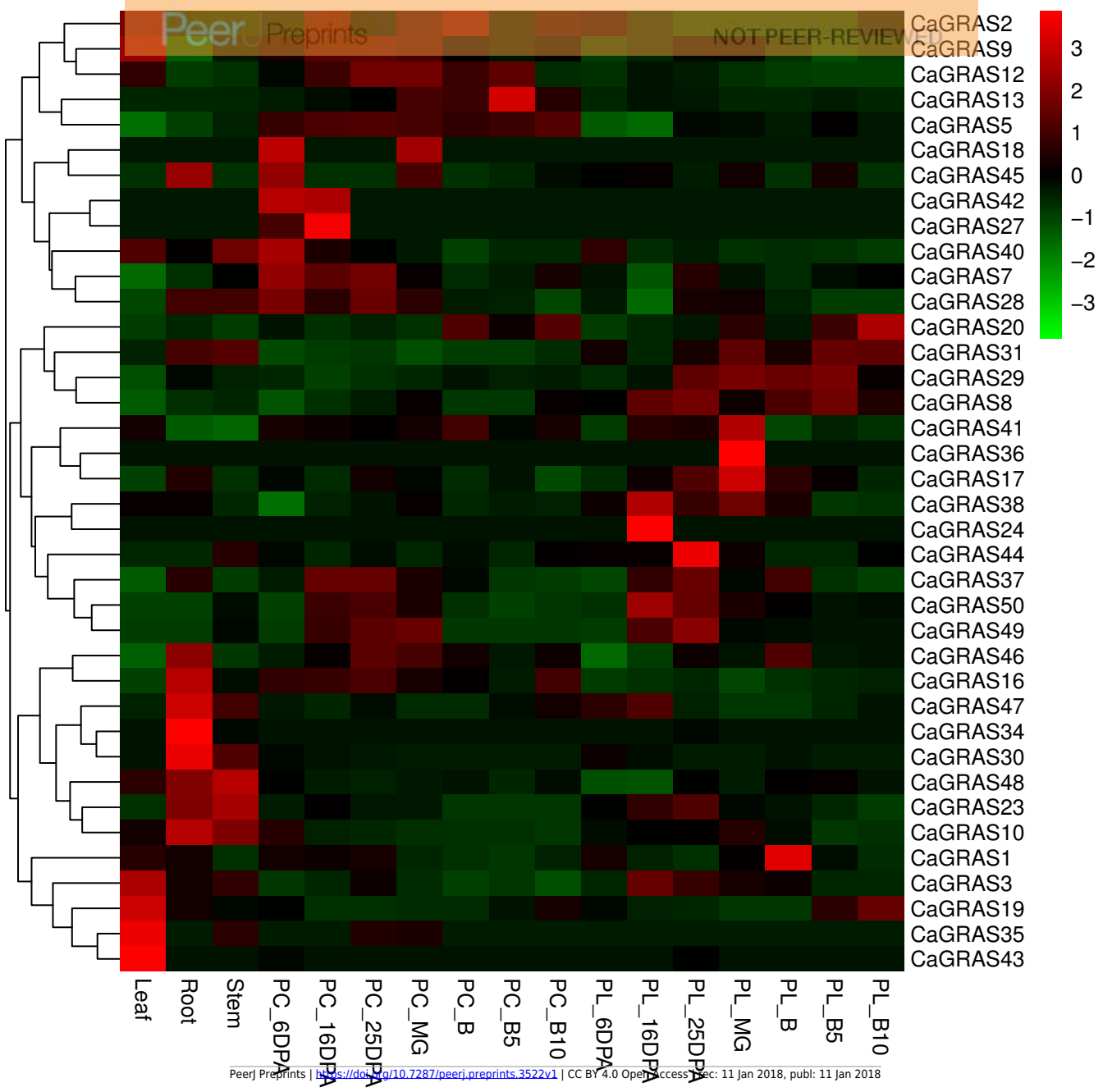
The interaction network of CaGRAS proteins in pepper according to interologs from *Arabidopsis*



## Figure 4(on next page)

Heatmap and hierarchical clustering of *CaGRAS* genes in leaf, stem, root, and mature green (MG), breaker (B), 5 and 10 days post-breaker (B5, B10), 6, 16, 25 days post-anthesis (6DPA, 16DPA, 25DPA) of pericarp (PC) and placenta (PL).

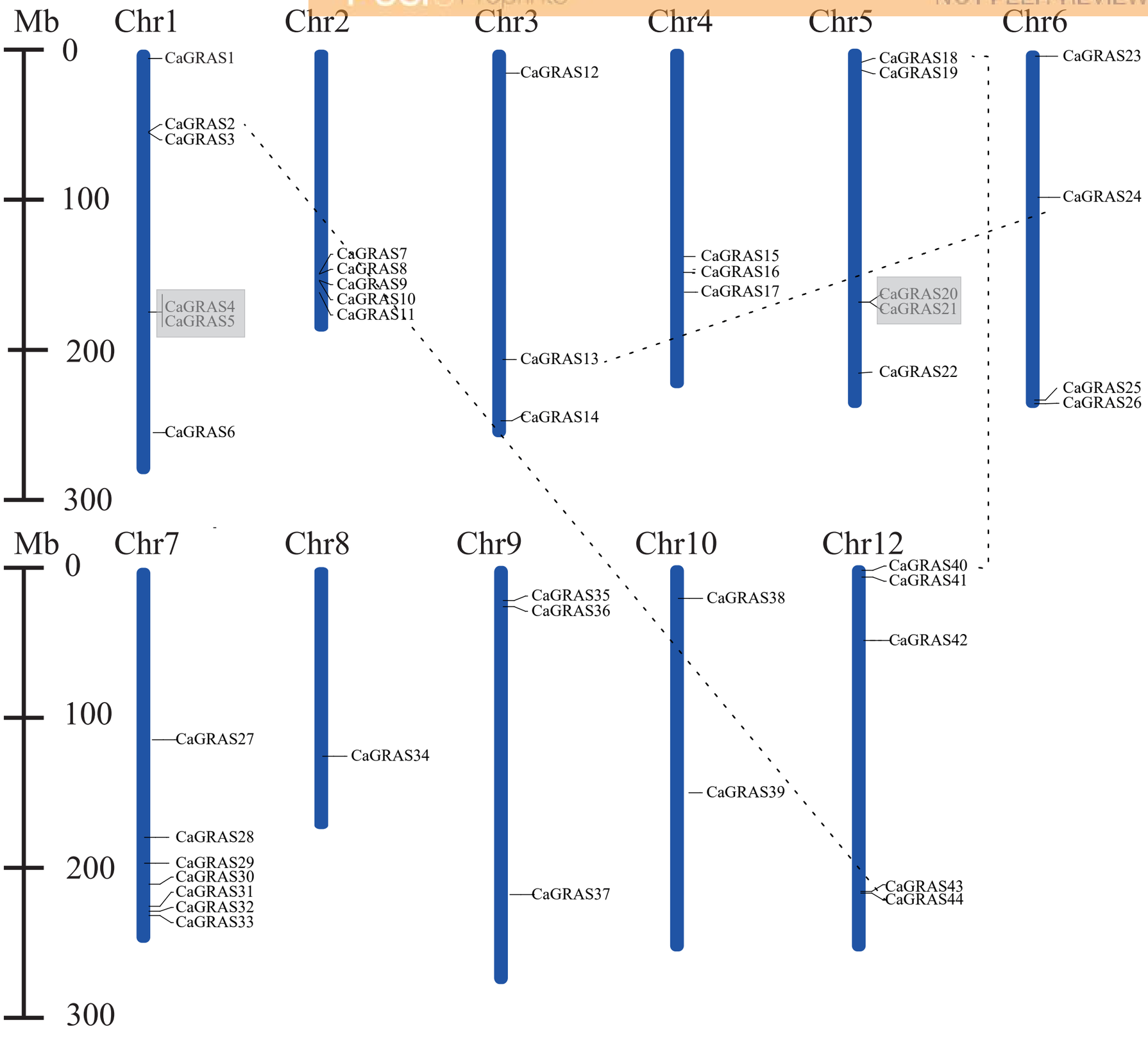
The expression values were calculated by RPKM measure and then were log2 transformed before generating heat maps.




PeerJ Preprints | https://doi.org/10.7287/peerj.preprints.3522v1 | CC BY 4.0 Open Access rec: 11 Jan 2018, publ: 11 Jan 2018

## Figure 5(on next page)

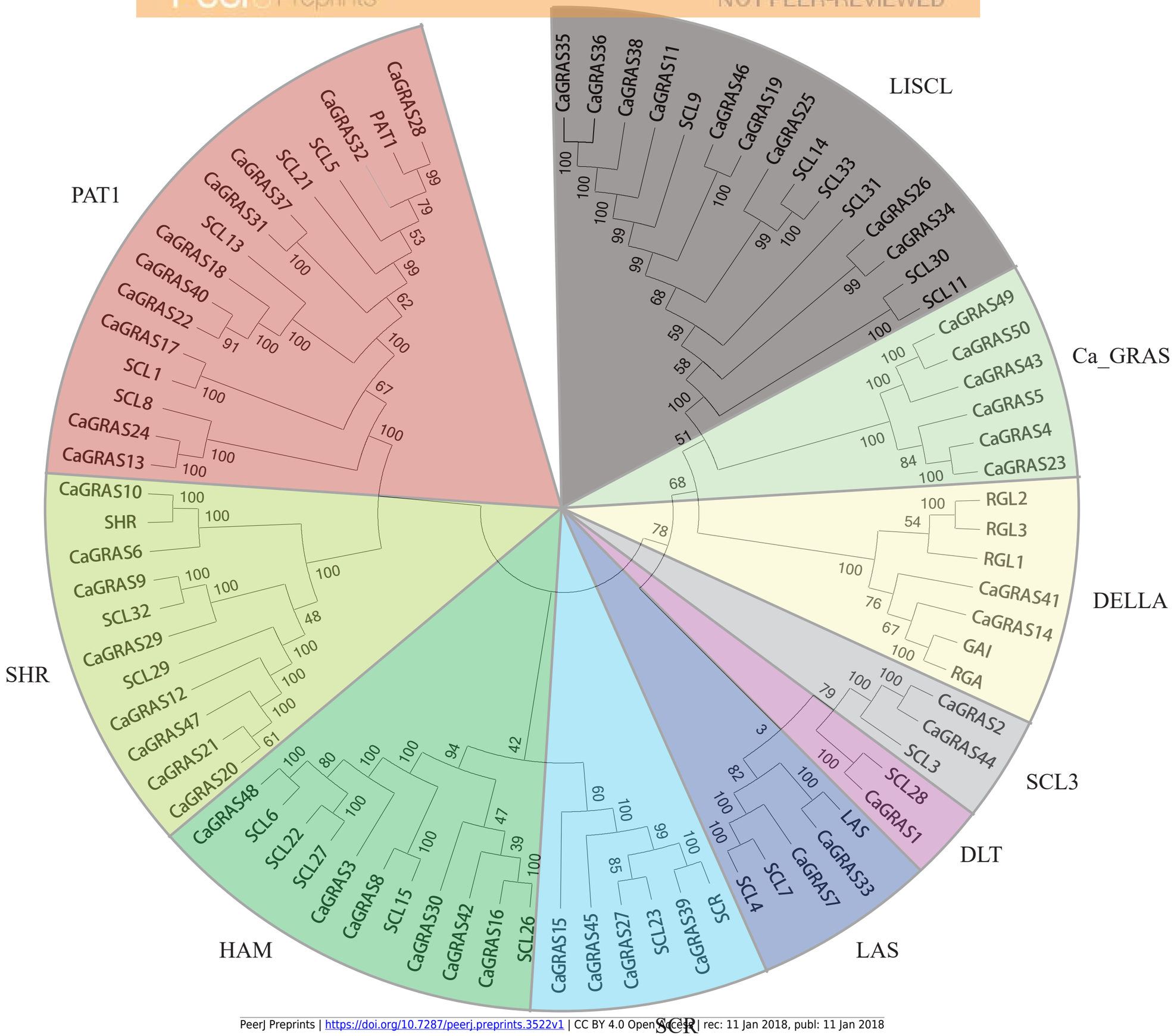
Positions of *CaGRAS* genes on pepper chromosomes. Grey shading indicates tandem duplicated regions. Genes in segmental duplicated repeats are linked by black dashed line.



## Figure 6(on next page)

Phylogenetic analyses of GRAS proteins from pepper and *Arabidopsis*. The phylogenetic tree was constructed using Neighbor-Joining (NJ) method by MEGA6.0. Ten subfamilies were indicated by ten different colors, respectively

The phylogenetic tree was constructed using Neighbor-Joining (NJ) method by MEGA6.0. Ten subfamilies were indicated by ten different colors, respectively.



## Figure 7 (on next page)

Distribution of conserved motifs in CaGRAS proteins. The phylogenetic tree is shown on the left of figure, different motifs and domain features are indicated by different colors numbered 1 – 11 at the top of figure

The phylogenetic tree is shown on the left of figure, different motifs and domain features are indicated by different colors numbered 1 – 11 at the top of figure.