Genome-wide identification and characterization of GRAS transcription factors in tomato (*Solanum lycopersicum*)

Yiling Niu [1], **Tingting Zhao** [1], **Xiangyang Xu** [Corresp., 1], **Jingbin Jiang** [1], **He Zhang** [1], **Jingfu Li** [1],

[1] College of Horticulture and Landscape, Northeast Agricultural University, Harbin, China

Corresponding Author: Xiangyang Xu
Email address: nxyyyx@neau.edu.cn

Solanum lycopersicum, belonging to Solanaceae, is one of the commonly used model plants. The GRAS genes are transcriptional regulators, which play a significant role in plant growth and development, and the functions of several GRAS genes have been recognized, such as, axillary shoot meristem formation, radial root patterning, phytohormones (gibberellins) signal transduction, light signaling, and abiotic/biotic stress; however, only a few of these were identified and functionally characterized. In this study, a gene family was analyzed comprehensively with respect to phylogeny, gene structure, chromosomal localization, and expression pattern; the 54 GRAS members were screened from tomato by bioinformatics for the first time. The GRAS genes among tomato, *Arabidopsis*, rice, and grapevine were rebuilt to form a phylogenomic tree, which was divided into ten groups according to the previous classification of *Arabidopsis* and rice. A multiple sequence alignment exhibited the typical GRAS domain and conserved motifs similar to other gene families. Both the segmental and tandem duplications contributed significantly to the expansion and evolution of the GRAS gene family in tomato; the expression patterns across a variety of tissues and biotic conditions revealed potentially different functions of GRAS genes in tomato development and stress responses. Altogether, this study provides valuable information and robust candidate genes for future functional analysis for improving the resistance of tomato growth.

Genome-wide identification and characterization of GRAS transcription factors in tomato (*solanum lycopersicum*)

Yiling Niu[1] Tingting Zhao[1] Xiangyang Xu[1]

[1]College of Horticulture and Landscape, Northeast Agricultural University, Harbin, Heilongjiang, China
[1]College of Horticulture and Landscape, Northeast Agricultural University, Harbin, Heilongjiang, China
[1]College of Horticulture and Landscape, Northeast Agricultural University, Harbin, Heilongjiang, China

Corresponding Author:
Xiangyang Xu
Email address: nxyyyx@neau.edu.cn

ABSTRACT

Solanum lycopersicum, belonging to Solanaceae, is one of the commonly used model plants. The GRAS genes are transcriptional regulators, which play a significant role in plant growth and development, and the functions of several GRAS genes have been recognized, such as, axillary shoot meristem formation, radial root patterning, phytohormones (gibberellins) signal transduction, light signaling, and abiotic/biotic stress; however, only a few of these were identified and functionally characterized. In this study, a gene family was analyzed comprehensively with respect to phylogeny, gene structure, chromosomal localization, and expression pattern; the 54 GRAS members were screened from tomato by bioinformatics for the first time. The GRAS genes among tomato, *Arabidopsis*, rice, and grapevine were rebuilt to form a phylogenomic tree, which was divided into ten groups according to the previous classification of *Arabidopsis* and rice. A multiple sequence alignment exhibited the typical GRAS domain and conserved motifs similar to other gene families. Both the segmental and tandem duplications contributed significantly to the expansion and evolution of the GRAS gene family in tomato; the expression patterns across a variety of tissues and biotic conditions revealed potentially different functions of GRAS genes in tomato development and stress responses. Altogether, this study provides valuable information and robust candidate genes for future functional analysis for improving the resistance of tomato growth.

Keywords:

Introduction

Transcription factors comprise the core of functional genomics. These transcription factors are protein molecules that can either activate or repress the target genes to ensure their specific expression by combining with the genes of 5'-flank cis-element (*Riechmann et al.,2000*). The typical transcription factor consists of DNA-binding domain, transcription regulation domain (active region or suppressor region), oligomerization site, and nuclear localization signal (*Morohashi et al., 2002*). GRAS is a major plant-specific transcription factor gene family among putative transcription factors that are found in a variety of plant species; for example, *Arabidopsis (Arabidopsis thaliana*), rice (*Sativa oryza*), grapevine (*Vitis vinifera*), tobacco (*Nicotiana tabacum*), Chinese cabbage (*Brassica rapa ssp.pekinensis*), and *Prunus mume*. Based on the known first three members, GAI (gibberellic acid insensitive) (*Peng et al., 1997*), RGA (repressor of GA1-3 mutant (*Silverstone et al., 1998*) and SCR (scarecrow) (*Di Laurenzio et al., 1996*), this transcription factor was named GRAS with the characteristic letter from each of the three members. The family members were screened as GRAS gene family as each of them contained the GRAS domain (*Pysh et al. 1999*). The GRAS proteins are usually composed of 400–700 amino acid residues (*Bolle, 2004*). A few GRAS proteins contain two structural domains: one GRAS domain and the other functional domain (*Schumacher et al., 1999*). The typical features of these proteins include a highly conserved C-terminal region and a variable N-terminal region (*Sun et al., 2011*). The conserved C-terminal region harbors five sequence motifs: LHRⅠ (leucine heptad repeat I), LHRII (leucine heptad repeat II), VHIID motif (*Pysh et al.,1999*), SAW motif, and PFYRE motif. The structures of both LHR I and LHR II are two leucine enrichment regions, the VHIID motif is a core structure, which exists in all members of the GRAS gene family and it can combine with LHR I and II to form the complex LHR I-VHIID-LHR II. This structural pattern might be a crucial function for DNA-binding and protein-binding in the interactions of proteins and proteins (*Itoh et al., 2002*). The localization of SAW and PFYRE motifs, for functional specificity, is not yet clearly elucidated; however, the missense mutations in these motifs in RGA and SLR1 proteins exhibit strong mutant phenotypes (*Silverstone et al., 1998*). In addition, the N-terminal region of GRAS proteins in the DELLA subfamily contains the other two motifs: DELLA and VHYNP, and the VHYNP motif is dynamic, implying that the N-terminal domains of GRAS proteins harbor various motifs (*Peng et al., 1997*). Moreover, the GRAS proteins are not only structurally diverse but also exert multiple functions. In recent years, several groups have found that the GRAS proteins are one of the indispensable regulation factors in plant growth development and participate in many biochemical and physiological processes in plants, such as gibberellin signal transduction (*Peng et al., 1997; Silverstone et al., 1998; Ikeda et al., 2001),* axillary shoot meristem formation (*Stuurman et al., 2002*), root radial patterning (*Di Laurenzio et al.,1996; Helariutta et al.,2000*), male gametogenesis (*Morohashi et al., 2003*), phytochrome A signal transduction (*Bolle et al.,2000*), nodulation signal transduction (*Hirsch et al.,2009*), and biotic/abiotic stress (*Huang et al.,2015*). Previously, the GRAS proteins were divided into eight subfamilies, according to

their common feature or member to name those subfamilies; for instance, SHR, SCR, DELLA, LISCL, Ls, HAM, PAT1, and SCL3. The amino acids in each group are mostly homologous, and thus, the GRAS genes of each subfamily might possess similar or related functions (*Tian et al., 2004*).

Hitherto, the RNA-seq is developed to identify the specific genes of the GRAS family; subsequently, the functions of these genes are investigated. Some GRAS genes have been characterized based on *Arabidopsis* and rice; for example, the DELLA proteins that function as negative regulators in gibberellin signal transduction, GA signaling pathway regulates the plant growth and development by degrading the DELLA proteins (*Zhang et al., 2011; Heo et al., 2011*). Firstly, the signal perception of GID1 proteins receive the GA signal; then the GID1 proteins combine with the DELLA proteins to form the complexes of GA-GID1-DELLA; subsequently, the DELLA proteins specifically bind the F-box protein SLY1, which subordinates the $SCF^{SLY1/GID2/SNE}$ protein complex. Finally, the degradation of the DELLA proteins medaited by the 26S proteasome released the inhibition, and thus, the plants show normal growth (*Day et al.,2004*). The study showed that the DELLA motif in the N-terminal is indispensable for the interaction between the two proteins DELLA and GID1; however, the motifs in the C-terminal are redundant. The SCR and SHR groups and the analysis of SHR/SCR mutant showed short root phenotype (*Di Laurenzio et al., 1996; Helariutta et al., 2000*), providing evidence that both proteins act as positive regulators in the radial organization of the root. Previous studies have shown that the SCR proteins combine with SHR proteins to form a complex, while the SHR proteins were transported to the endodermis of the root (*Sabatini et al., 2003; Wysocka et al., 2000; Helariutta et al., 2000*). Similarly, SCL3 proteins were involved in the elongation and differentiation region of the root and over-ground organs, respectively. In the root, the subsequent elongation is regulated to control the GA signaling pathway, while in the meristematic tissue, the combination of SHR/SCR leads to the organizational maturity within the GA signaling pathway (*Heo et al., 2011*). Some studies indicate that the overexpression of *OsMOC1* gene results in increasing the tiller numbers and decreasing the length; moreover, it can only be detected in the axillary bud (*Li et al., 2003*). In addition to rice, the *Ls* gene in tomato and the *AtLAS/SCL18* gene in *Arabidopsis* are closely linked to the growth of the lateral bud (*Schumacher et al., 1999; Greb et al., 2003*). The *PAT1*, *SCL13*, and *SCL21* genes belong to the *PAT1* branch, which mediates the phytochrome signaling pathways. Genetic evidence suggests that three genes act as positive regulators; the *SCL13* gene participates in phytochrome B transduction in dependently, whereas the*PAT1*and *SCL21* adjust the phytochrome A signaling network by the mutual effect (*Bolle et al., 2000; Torres-Galea et al., 2006; Torres-Galea et al., 2013*). The *NSP1* and *NSP2* genes exist in the downstream of symbiosis signal transduction pathway CCaMK (Ca/calmodulin-dependent protein kinases), which is related to the nodulation. TheNPS1-NPS2 heterodimer is induced by the nodulation factors, following which, the heterodimer can specifically combine the promoter of the early nodulation gene *ENOD2* to promote the expression of related genes, thereby forming the stage of nodulation

(*Hirsch et al., 2009*). The homologous genes of *NSP1/NSP2*are extensively encompassed in several non-leguminous plants. Another analysis about GRAS proteins found that some members in the gene family are regulated by miRNA171; for instance, *AT2G45160*, *AT3G60630*, and *AT4G00150* in *Arabidopsis* (*Schulze et al., 2010*), Pm017821 and Pm023512 in *Prunus mume* (*Wang et al., 2014*), *Solyc01g090950.2.1* and *Solyc08g078800.1.1* in tomato (*Huang et al., 2015*), and four genes in rice (*Llave et al., 2002*) are complementary to miRNA171.

In the past few years, some investigations showed that the GRAS genes respond to different hormones and abiotic stress treatments (*Huang et al., 2015*). In his study, which identified 53 GRAS genes and the phylogenetic tree and the expression patterns in different abiotic stress treatment were investigated, but only 48 tomato GRAS genes were selected for phylogenetic analysis, the abiotic stress analysis focused on salt, cold, hat, osmotic and drought stress, only about 40 tomato GRAS genes were investigated, and the analysis of bioinformatics was poor. This study choosing more tomato GRAS genes for a number of bioinformatic analyses, and analyzed the relationship between pstDC3000 and GRAS genes from tomato for the first time. In this study, a comprehensive and systematic analysis would provide a deep insight on the GRAS family and precede the sequencing studies. Tomato is an adequate good model plant for the study of Solanaceae due to its great economic value. With the whole-genome analyses of tomato, the genomic data can highlight the connection between genes and plants.

Materials & methods

Identification of tomato GRAS genes

we retrieved the tomato GRAS genome sequences, protein sequences and annotation information from SGN (https://solgenomics.net/) (*Consortium, 2012*). The Arabidopsis GRAS genes' family sequences and annotation information was download from TAIR (http://www.arabidopsis.org/) (*Swarbreck et al., 2008*), whereas the rice GRAS transcription factor sequences and annotation information was obtained from RGAP (http://rice.plantbiology.msu.edu/index.shtml) (*Ouyang et al., 2007*). The grape GRAS gene family sequences were download from plantTFDB (http://planttfdb.cbi.pku.edu.cn/). The HMM model of GRAS transcription factor was download from PFAM (http://pfam.xfam.org/) (*Finn et al., 2010*), known as PF03514. The HMM model was used as a query to identify the tomato GRAS genes containing the GRAS domain with a cut-off E-value of $1e^{-5}$ in the HMMER software. Consequently, we identified a total of 54 GRAS genes in tomato, 34 from Arabidopsis, 56 from rice, and 37 from the grapevine. Then, we conducted a quality check using the Simple Modular Architecture Research Tool SMART (http://smart.embl-heidelberg.de/) (*Letunic et al., 2012*) to confirm the presence of GRAS domain on the candidate GRAS genes. These data were used for subsequent analysis.

Phylogenetic analysis for tomato GRAS genes

MEGA 6.0 was used to analyze the phylogenetics of genome-wide GRAS gene family based on the whole set of GRAS protein sequences from tomato (*solanum lycopersicum*), *Arabidopsis*

(*Arabidopsis thaliana*), rice (*Sativa oryza*) and grapevine (*Vitis vinifera*). The *Arabidopsis* is one of the most commonly used plants in Cruciferae for studying the genetic correlations; *Arabidopsis* to *Cruciferae*, grapevine to *Vitaceae*, and rice to Gramineae (*Tamura et al., 2011*). In this building process, several shorter amino acids (*Solyc01g090950.1.1, Solyc04g011630.1.1, Solyc12g049320.1.1*) were excluded, the domain length of these sequences were shorter than half of the typical GRAS domain (350 amino acids), and the low similarity among the tomato GRAS family (*Huang et al., 2015*). According to a previous study, the classification was made on the phylogenetic tree using the Evolview software (http://www.evolgenius.info/).

Structure analysis of tomato GRAS genes

Using the NCBI platform (https://www.ncbi.nlm.nih.gov/), the conserved domains of 54 GRAS genes were visualized. In order to search for the potentially conserved motifs in the complete amino acid sequence of tomato GRAS proteins, the Multiple EM for Motif Elicitation (http://meme-suite.org/) (*Bailey et al., 2006*) was used with default parameters, except that the number of motifs was set to 10. In order to present the characteristic of every subfamily, the 54 GRAS genes were used to build a Maximum Likelihood tree based on the JTT matrix-based model. Other parameters were same to above. The secondary structures of the tomato GRAS genes were generated using the Gene Structure Display Sever (http://gsds.cbi.pku.edu.cn/) (*Guo et al., 2007*).

Evolution analysis of tomato GRAS genes

The sequences were compared using the GRAS genes in tomato, *Arabidopsis*, and rice; the entire protein sequences were used to identify the orthologous and paralogous genes using the software OrthoMCL (http://orthomcl.org/orthomcl/) with an E-value of $1e^{-5}$ and a match cut-off value 50 for against-all BLASTp alignment (*Li et al., 2003*). The MCscanX software (*Tanget al., 2008*) was used to identify the collinear block based on the tomato genomes; if a gene had more than one transcript, only the first transcript in the annotation was used. The underlying mechanism of tandem duplication showed that two genes were physically close to each other with the genes residing within 20kb (*Liu et al., 2014*), segmental duplication resulted from the whole genome duplication accompanied by a comprehensive gene loss (*Tang et al., 2008*), and those large duplication events can be deduced by anchor genes in collinear blocks (*Cannon et al., 2004*). The chromosomal localization and homologous collinear relationship of GRAS genes were visualized using the Circos program (*Krzywinski et al., 2009*).

Expression pattern analysis of GRAS genes

We utilized the Illumina RNA-seq data of tomato download from SGN (https://solgenomics.net/), which were reported previously. To confirm the expression patterns of the GRAS genes, the FPKM was used to represent the expression level of each tomato gene. The data on GRAS genes' expression was retrieved to display the consequences using the HemI program (http://hemi.biocuckoo.org/); the expression data were amplified 100-fold (*Deng et al.,2014*). The transcriptomic data were extracted from twelve tissues of *S. pimpinellifolium* (LA1589), and *S. pimpinellifolium* performances more abundant in genetic variation, which can

better assess the evolution of the GRAS genes. Which comprised of: A: newly developed leaves approximately 5 mm long; B: mature green leaflets; C: flower buds 10 days before anthesis or younger; D: flowers at anthesis (0DPA); E: 10 days post anthesis fruit (10DPA); F: 20 days post anthesis fruit (20 DPA); G: breaker stage ripening fruit(33DPA); H: another set of 10 DPA fruit was collected in a separate greenhouse for comparison, the following tissues were collected from seeds that were germinated and grown for 7 days in a Petri dish under growing lights, I: whole root; J: hypocotyl from below the cotyledons to above the root zone; K: cotyledons; L: vegetative meristems.

In the case of difficulty, other public data were retrieved to reveal the relationship between GRAS genes and biotic stress. The transcriptome of leaves of resistant (RG-PtoR) and susceptible (RG-prf3 and RG-prf19) tomato plants treated with pstDC3000 in 4 and 6 h were sequenced.

Results

Genome-wide identification and annotation of the GRAS genes in tomato

In order to identify the GRAS proteins in tomato, we downloaded the raw data sequence of tomato GRAS transcription factors from SGN database. The bioinformatics approach retrieved 54 tomato GRAS genes, 34 *Arabidopsis* GRAS genes, 56 rice GRAS genes, and 37grape GRAS genes. In addition, we obtained basic information in connection with tomato GRAS proteins (Table S1). The length of GRAS proteins ranged from 125–864 aa and the tomato GRAS genes were almost distributed across the 12 chromosomes uniformly; the highest content was on chromosome 1, containing 8 GRAS genes. However, at least 3 GRAS genes were located on other chromosomes. The annotation information represented that the GRAS domain always assembled on the C-terminal, suggesting that the conserved terminal is supported by the stable structure of the GRAS domain; this phenomenon could be used for the subsequent comparison analysis. The smaller HMM E-value provided reliable screening results. Notably, a number of amino acid residues in some GRAS proteins were less than that in the typical GRAS domain, such as *Solyc04g011630.1.1* or *Solyc06g076290.1.1* genes.

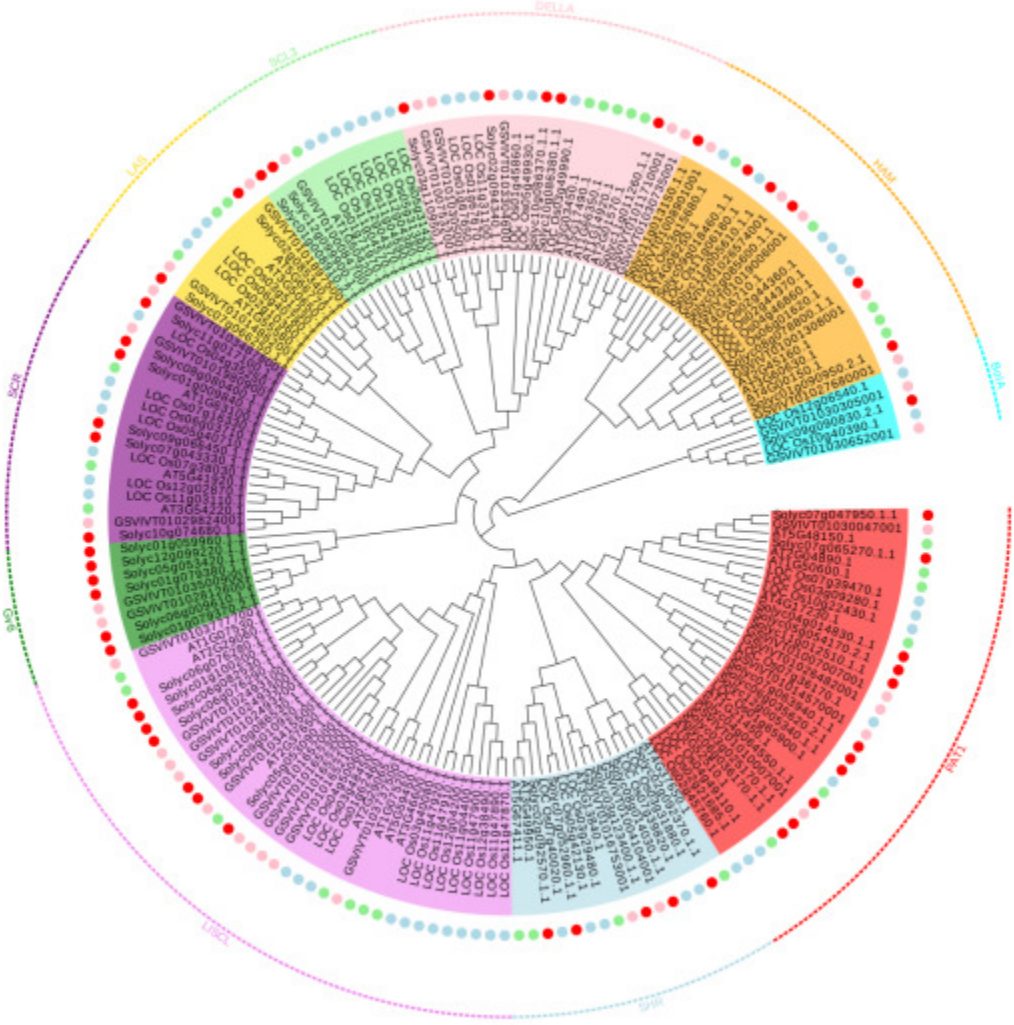
Genome-wide sequence alignment and phylogenetic analysis

The topology tree contains 178 proteins: 56 from rice, 34 from *Arabidopsis*, 37 from grapevine, and 51 from tomato. However, the remaining 3 proteins from tomato were neglected because of the shorter domain sequence. Basing on the topology structure, clade support value and previous classification from rice and *Arabidopsis*, as shown in Fig.1 (*Tian et al., 2004*), all of these proteins were divided into ten groups, named specifically after a common feature or one of the members: PAT1, SHR, LISCL, Gv6, SCR, LAS, SCL3, DELLA, HAM, and BolA, respectively. Of these, eight were named according to a previous report (*Tian et al., 2004*). The remaining two groups were named by the characteristic of the members. The tree showed that the four species GRAS proteins were distributed in the ten groups, randomly, and hence, the members of GRAS transcription factors in these four species are not represented equally. For

example, the subfamily SCL3 contains one or two proteins from *Arabidopsis*, tomato, and grapevine, respectively, and seven members from rice. Strikingly, some clades do not contain the members of GRAS proteins from *Arabidopsis* or rice; for instance, Gv6 and BolA groups. This might be attributed to the loss of the corresponding member following the separation of the last common ancestor if not the problem of assembling or annotation of the *Arabidopsis* or rice genomes. We found that the Gv6 group consisted of 6 tomato proteins and 2 grapevine proteins, indicating that it is a fruit tree species-specific clade.

The genes with orthologs frequently tend to be clustered together, and the subfamily members in a major group share similar gene structure and function. Therefore, we observed that the distribution of specific genes provides information on the role of the other genes in the same clade. For example, within the PAT1 group, *AT5G48150.1* of *Arabidopsis* was previously demonstrated to participate in the process of phytochrome signal transduction. In addition, the *Solyc07g047950.1.1* was highly similar to the *AT5G48150.1* protein, and hence, we inferred that *Solyc07g047950.1.1* protein also played a crucial role in phytochrome signal transduction. PAT1 and LISCL subfamily comprised of more proteins than the other subfamilies as well as that reported in previous studies.

Members in the same sub-branch were marked by the same color and surrounded with an appropriate color line that was commanded by the subfamily name. The red circle corresponds to the tomato GRAS proteins, the green circle corresponds to the *Arabidopsis* GRAS proteins, the blue circle corresponds to the rice GRAS proteins, and the pink circle corresponds to the grapevine GRAS proteins.



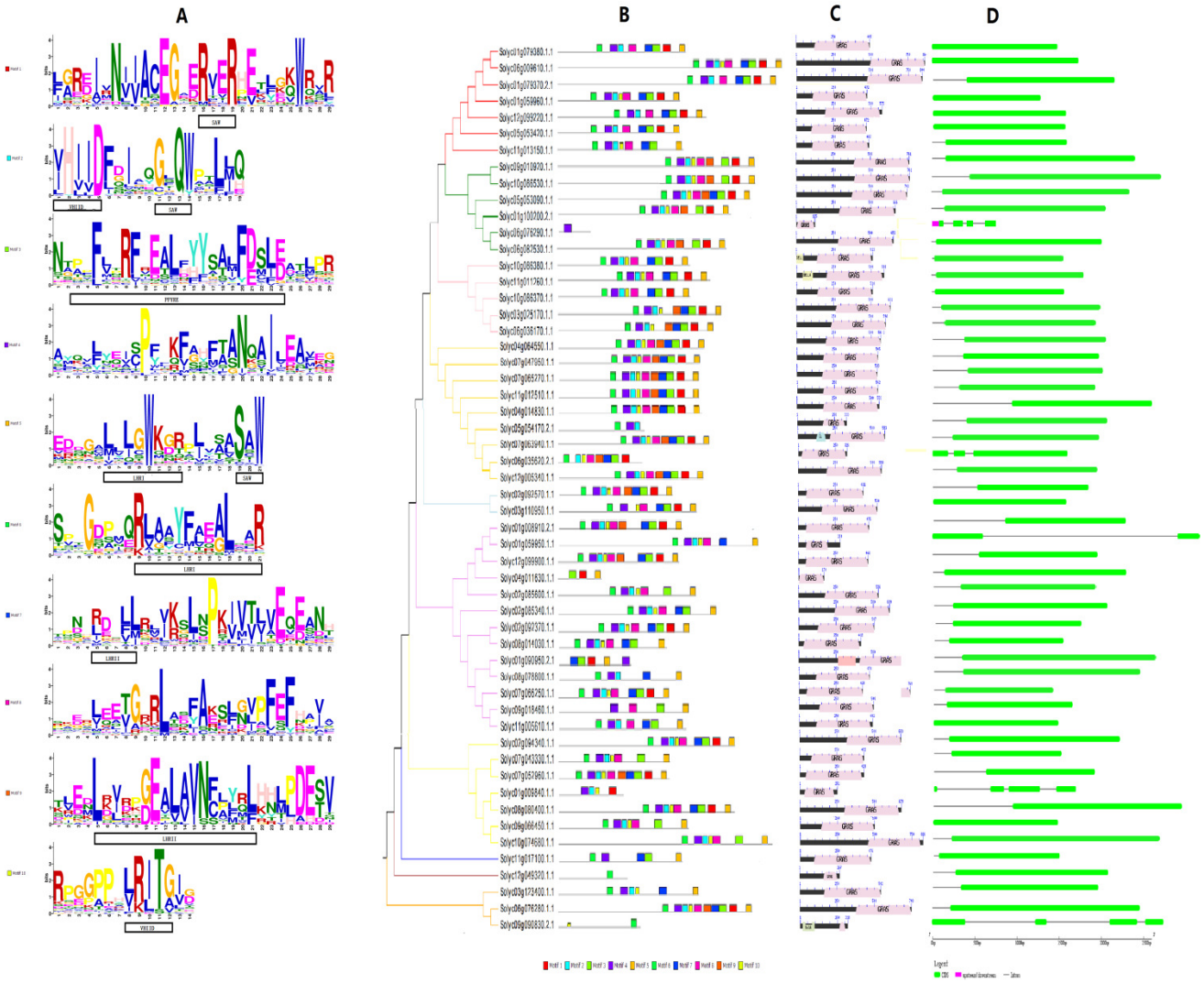
Members in the same sub-branch were marked by the same color and surrounded with an appropriate color line that was commanded by the subfamily name. The red circle corresponds to the tomato GRAS proteins, the green circle corresponds to the *Arabidopsis* GRAS proteins, the blue circle corresponds to the rice GRAS proteins, and the pink circle corresponds to the grapevine GRAS proteins.

Fig.1 Phylogenetic tree of GRAS proteins from tomato, *Arabidopsis*, rice and grapevine, respectively.

The structural analysis of GRAS genes (motif analysis and exon/intron analysis)

To achieve a general overview of the conserved features of the tomato GRAS proteins, we performed a multiple sequence alignment of the 54 tomato GRAS genes using MEME software; the conserved motifs of all proteins are shown in Fig.2-B. Ten is the default number of motifs in MEME analysis. We found that 80% GRAS proteins contained >7 motifs, and 24% GRAS proteins encompassed 10 motifs. Moreover, members of the same clade of phylogenetic tree shared similar motif organization with respect to either gene length or motif number (*Liu et al., 2014*). On the other hand, several proteins only contained one or two motifs, which might be

attributed to the short duration of those and cause the shorter domain. Also, the motifs were more likely to be located in the C-terminal than the N-terminal. The motif logo is shown in Fig.2-A. According to the previous studies on GRAS motifs (*Tian et al., 2004*), the five putative GRAS domains were exhibited in the logo of motifs; we speculated that each LHRI, LHRII, VHIID, PFYRE, and SAW motifs could be divided into different units. This phenomenon demonstrated that maximum tomato GRAS genes are conserved, and the conserved region is localized on the C-terminal. Furthermore, we conducted a domain analysis (Fig.2-C) and exon/intron analysis (Fig.2-D) of tomato GRAS genes in order to explore the diversity and functionality. The visualization of the domains revealed that nearly all the GRAS proteins domains were primarily distributed in the C-terminal, further supporting the theory that the C-terminal is highly conserved. Nevertheless, there were some specific genes, such as the whole *Solyc06g076290.1.1* gene is a GRAS domain, *Solyc10g086380.1.1*, *Solyc11g011260.1.1*, *Solyc07g063940.1.1*, *Solyc01g090950.2.1*, and *Solyc09g090830.2.2* genes contained two domains. In addition, the exon/intron analysis displayed similar distribution characteristics. The *Solyc06g076290.1.1* and *Solyc090830.2.2* genes are comprised of multiple exons; however, the majority of GRAS proteins possessed only one exon and one intron, which might result from intron gain or loss event during evolution.



A: Motif logo, the amino acid composition of each motif; B: The motif distribution in each GRAS gene; C: The location of GRAS domain in genes; D: The length of exons, introns and UTRs.

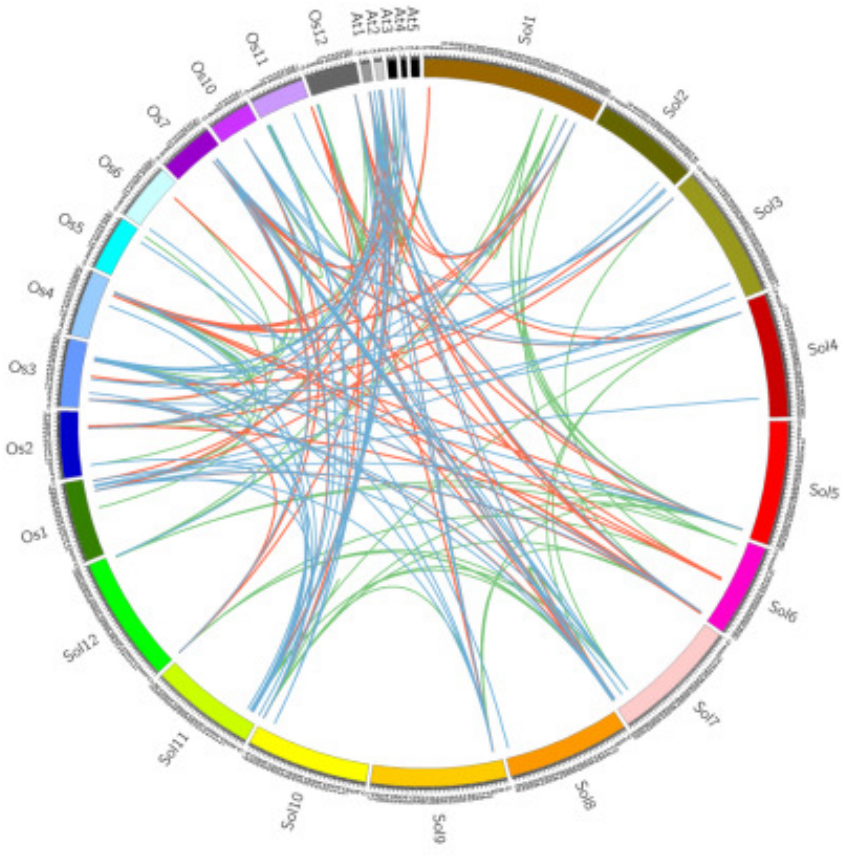
Fig. 2 The structure of GRAS genes in tomato.

Comparison and expansion analysis of tomato GRAS genes

To evaluate the evolutionary relationships among tomato GRAS genes, we made a comparative analysis to identify the co-orthologous, orthologous and paralogous GRAS genes among tomato, *Arabidopsis*, and rice using the OrthoMCL software, as shown in Fig.3. We identified 22 co-orthologous and 46 orthologous gene pairs between tomato and *Arabidopsis*, 45 co-orthologous and 48 orthologous gene pairs between tomato and rice and 19 co-orthologous and 29 orthologous gene pairs between *Arabidopsis* and rice. Moreover, 38 tomato GRAS genes (65%) have paralogous genes, which was higher than that in rice (63%) and *Arabidopsis* (19–56%). The orthologous genes commonly share a similar structure and biological function. The

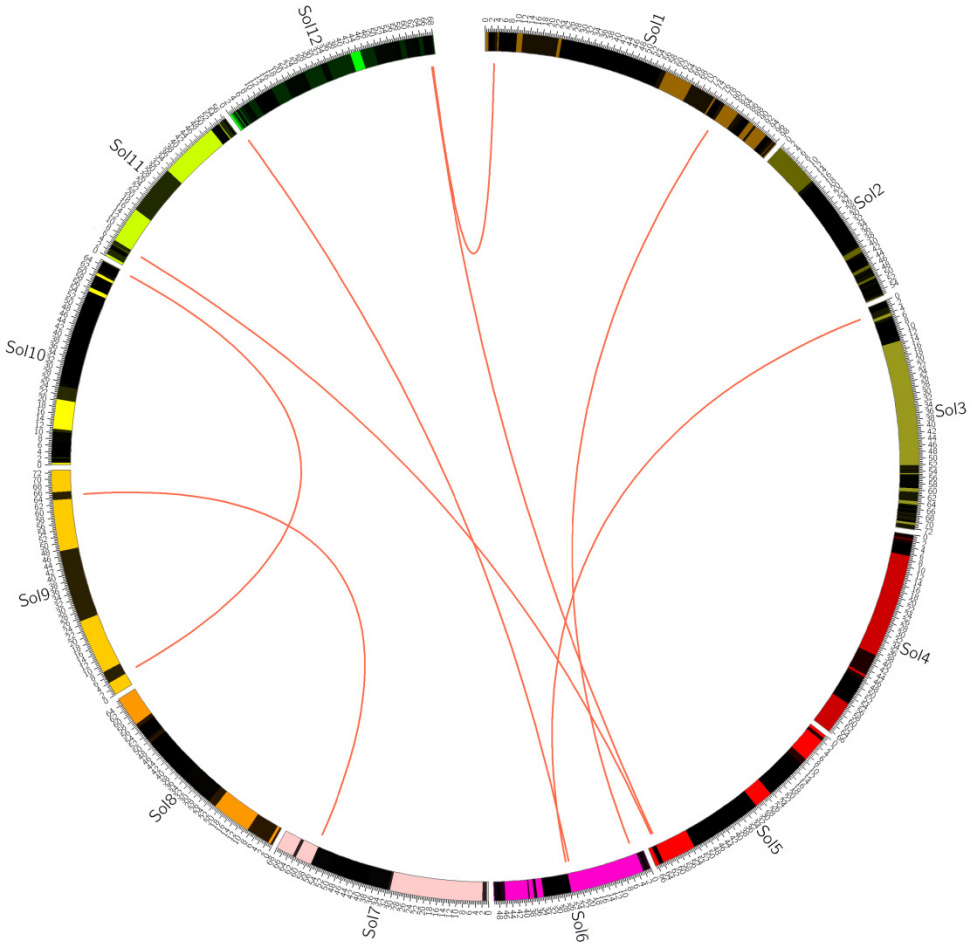
number of orthologous gene pairs between tomato and rice was more than that between tomato and *Arabidopsis*, which indicated that tomato is similar to rice.

Moreover, the duplication events were discovered in the evolution of tomato GRAS transcription factors (*Cannon et al., 2004*). Tandem and segmental duplication were vital for the expansion of the GRAS family. To reveal the relationship between gene duplication and amplification, the syntenic regions were analyzed by MCscanX software. As shown in Fig.4, finally, we obtained five tandem duplication gene pairs (18.5%), which the details of duplication events were referred in Table S2, and the result suggested that the origination of GRAS genes applied to the tandem duplication events. We also assessed the contribution of segmental duplications; 18 genes (33.3%) with duplications were harbored in collinear blocks, indicating their robust participation in the expansion of GRAS genes. The same observations were also made in other protein families (Wang et al., 2016).



Orange, blue and green lines indicate co-orthologous, orthologous and paralogous respectively. The words beginning with A, O and S represent the chromosomes of *Arabidopsis*, rice and tomato, respectively.

Fig. 3 Chromosomal localization of GRAS homologous genes in tomato.



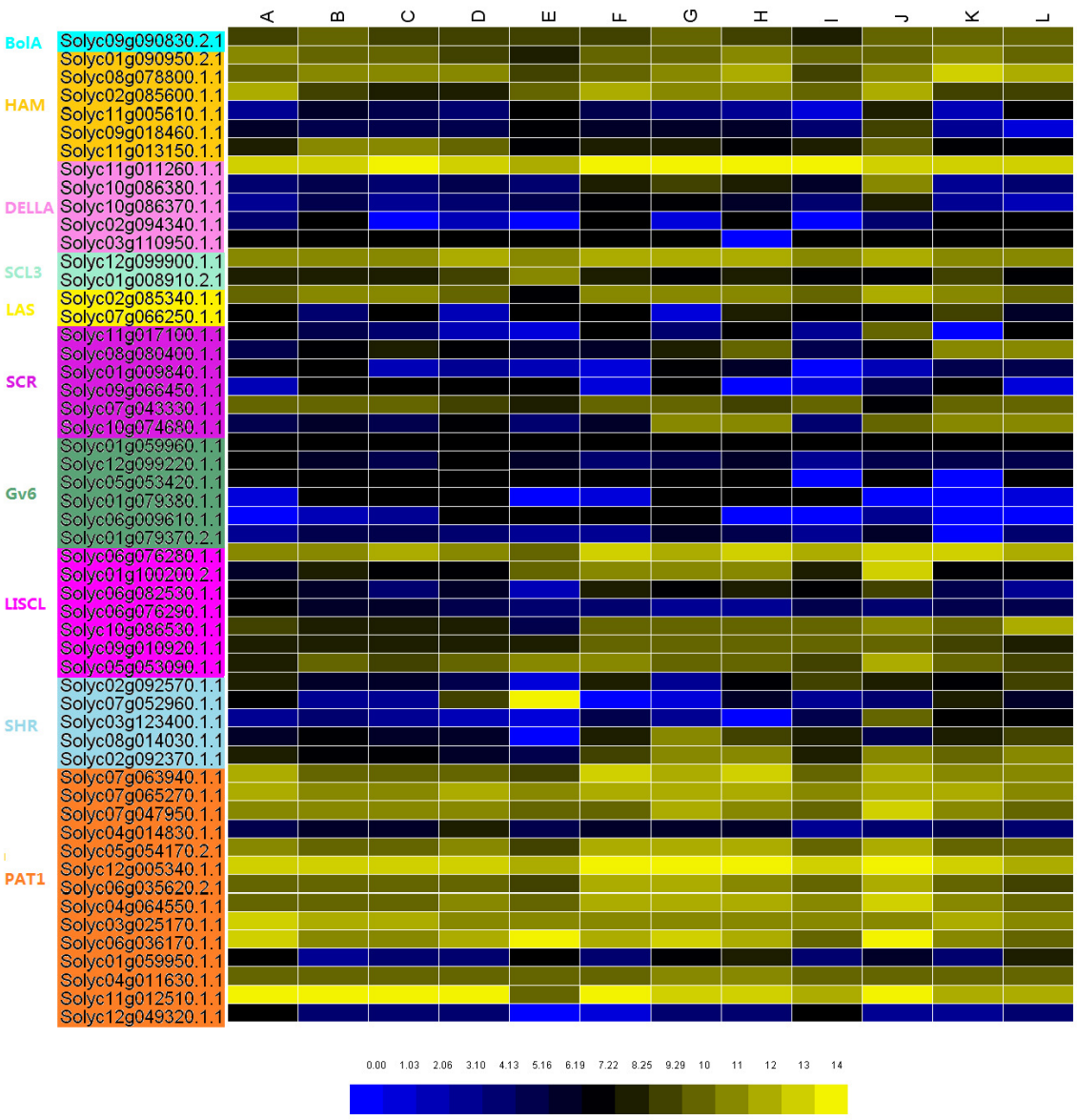
The red lines represent the segmentally duplicated genes, and the black bands represent the collinear block.

Fig. 4 Chromosomal localization of GRAS duplicated genes in tomato.

Expression pattern analysis of tomato GRAS genes in different tissues

In this study, we analyzed different expression levels of GRAS proteins in different tissues regarding the published RNA-seq data. The heat map was constructed to show the expression profile and the genes listed according to their branch (Fig.5). We found that only the *Solyc01g059960.1.1* gene was not detected in the RNA-seq data that might be due to temporal of no expression (*Song et al., 2014*). The other tomato GRAS genes were obtained in at least one tissue. As a whole, we found that the same group of GRAS genes shared a similar expression pattern; for instance, nearly all the PAT1 subfamily members showed a higher expression level than that of other groups, and the Gv6 subfamily showed a low expression level in all tissues. The GRAS genes on some branches exhibited a tissue-specific expression pattern. For example, the *Solyc11g005610.1.1*, *Solyc03g123400.1.1*, and *Solyc09g018460.1.1* genes belonged to the HAM subfamily that shows a low expression in those tissues except the hypocotyls, thereby

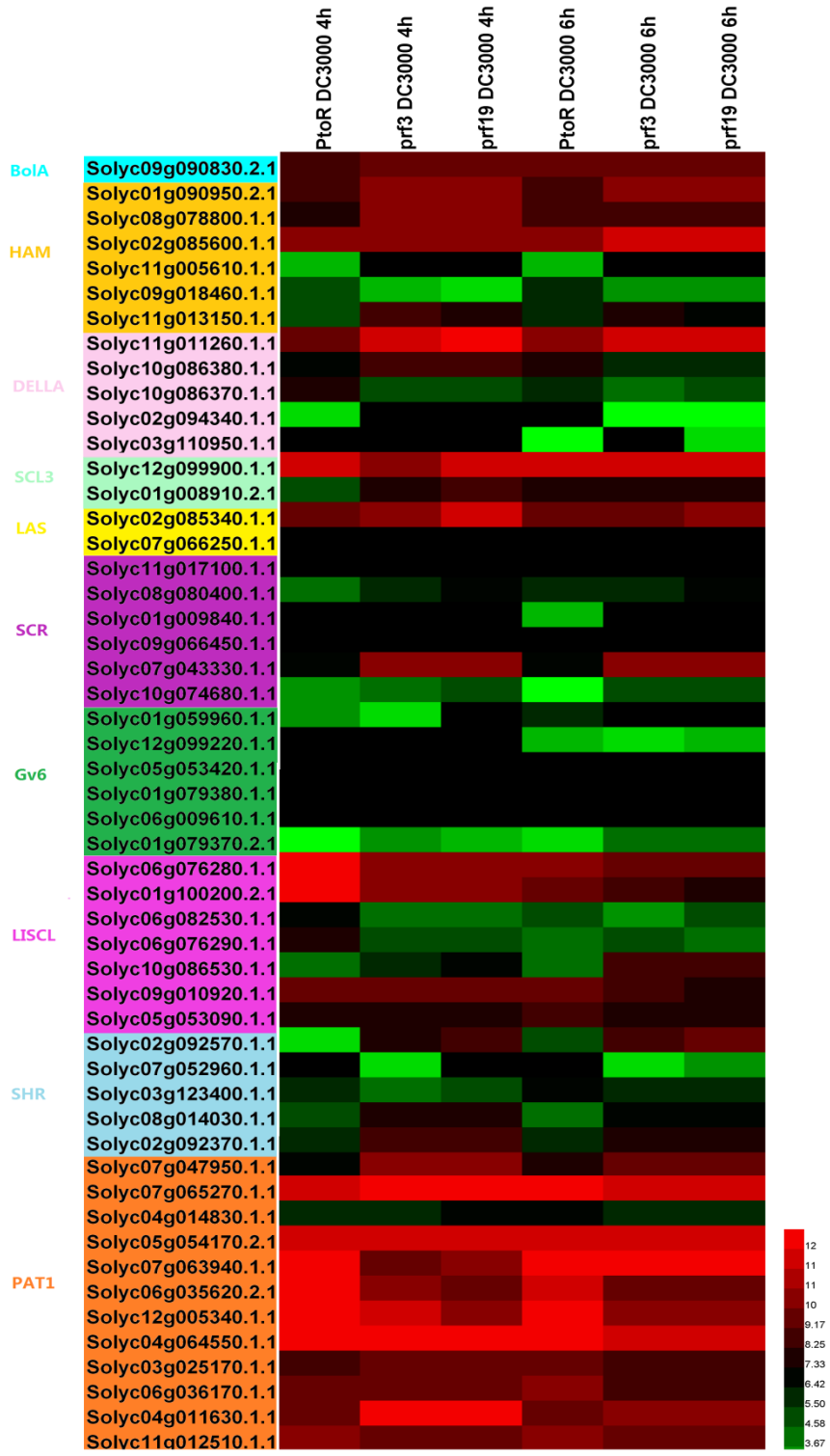
suggesting that the three genes contributed considerably to the development of the hypocotyl. The *Solyc04g014830.1.1* gene only expresses in the anthesis stage, and thus, would be closely related to the flower opening. *Solyc11g011260.1.1*, *Solyc12g005340.1.1*, and *Solyc11g012510.1.1* genes are always expressed at a high level in all the tissues, indicating their crucial role in the whole growth process of tomato. We also determined that some genes present a time-specific expression, such as, the *Solyc01g100200.2.1, Solyc01g008910.2.1*, *and Solyc07g052960.1.1* show high expression level in the stage of fruit ripening; the *Solyc10g074680.1.1* gene was expressed in the later stage. During the mature period of leaf and flower, the expression of the GRAS genes remains stable, and only two or three genes show a difference, such as *Solyc*11g013150.1.1 gene merely didn't express in newly developed leaves approximately, the *Solyc01g*.008910.2.1 only expressed in flowers and *Solyc02g092570.1.1* only expressed in newly developed leaves.



A: newly developed leaves approximately 5 mm long; B: mature green leaflets; C: flower buds 10 days before anthesis or younger; D: flowers at anthesis; E: 10 days post anthesis (DPA) fruit; F: 20 DPA fruit; G: breaker stage ripening fruit; H: another set of 10 DPA fruit was collected in a separate greenhouse for comparison, the following tissues were collected from seeds that were germinated and grown for 7 days in a Petri dish under growing lights, I: whole root; J: hypocotyl from below the cotyledons to above the root zone; K: cotyledons; L: vegetative meristems. and yellow indicating higher expression levels and blue indicating lower expression levels.

Fig. 5 The expression profile of the tomato GRAS genes in different tissues

352 According to the public database, using the visualized tool HemI to present the relationship
353 between pstDC3000 and GRAS genes from tomato (Fig.6), we deduced that the number of
354 higher expression genes in susceptible tomato (54%) were more than that in resistant tomato
355 (41%). The group of PAT1 continued to show the highest expression level, and the members of
356 the PAT1 subfamily might regulate some critical physiological processes in tomato growth that
357 are related to resistance. During the same infection time, the two varieties of susceptible plant
358 shared the same tendency of the expression of GRAS genes. As the infection time continued, the
359 SISCL group genes' expression was on a downward trend in the resistance tomato, such as
360 *Solyc06g076280.1.1*, *Solyc01g100200.2.2*, and *Solyc06g076290.1.1* genes; interestingly, the
361 *Solyc05g053090.1.1* gene was on an increasing trend. In the susceptible tomato, the
362 *Solyc02g085600.1.1* and *Solyc07g063940.1.1* genes' expressions were increasing, whereas, the
363 expression of the other members of the group HAM, LISCL and PAT1 were declining. In 4 h,
364 the expression of *Solyc01g100200.2.1*, *Solyc07g063940.1.1*, and *Solyc12g005340.1.1* genes were
365 different between of the resistance tomato and the susceptible tomato. In 6 h, the
366 *Solyc12g005340.1.1* gene expression still differed largely, and the results suggest that the
367 *Solyc12g005340.1.1* gene regulated the plant disease resistance.



The red indicating higher expression level and green indicating lower expression levels

Fig. 6 The expression profile of the tomato GRAS genes in the leaves of resistant (RG-PtoR) and susceptible (RG-prf3 and RG-prf19) tomato plants treated with Pst DC3000.

Discussion

Recently, the structural and functional genomics of GRAS transcription factors in higher plant model species have shown that a significant number of members were involved in the plant growth and development (*Heo et al., 2011*), including Signal transduction, stress response, meristem formation, cell maintenance and multiplication. The relevant studies have been elaborately conducted in *Arabidopsis*, which serves as a reference. On the other hand, the GRAS gene characteristics in tomato remain unclear. Thus, we identified 54 GRAS genes from tomato using the bioinformatics methods. Subsequently, the classification and annotation information was obtained, and the contrast of full length of the GRAS proteins from tomato showed remarkable differences. The distribution of GRAS genes in tomato is consistent with that of *Arabidopsis*. The next phylogenetic analysis might provide additional functional constituents among the four species. The previous studies demonstrated that the GRAS transcription factors are involved in plant development and stress response; it is acknowledged that the higher of the sequence similarities, the functions were more similar in different species (*Chen et al., 2007*). Additionally, the GRAS proteins have similar functions within the same clade. The structure and domain analysis proved that the topology tree was reliable, and the parallel structural features were clustered to the same subgroup. Every GRAS gene was composed of one or more conserved motifs. A comparative genomics analysis revealed abundant homologous genes in tomato, *Arabidopsis*, and rice, and the segmental duplication commonly promoted the expansion of GRAS proteins (*Cannon et al., 2004*). The expression of GRAS genes in different tissues indicated that the three genes *Solyc11g011260*, *Solyc11g012510*, and *Solyc12g005340* showed a high expression level among those organs, which implicated their vital importance in plant development. Moreover, the members of the same clade shared similar expression profiles (*Wang et al., 2016*). Some of the GRAS genes responded to the biotic stress of pstDC3000; the PAT1 subfamily showed the highest expression level. A similar result was observed in other species for the expression of GRAS genes (*Lu et al., 2015*).

Taken together, the GRAS transcription factor is essential for breeding and cultivation. A total of 54 GRAS members in tomato were identified with respect to gene structures, motifs and domains; the GRAS proteins showed highly conservative characteristics. A comparative analysis suggested that the functional diversity might be sourced from the large-scale genome duplication. The results of expression of the GRAS genes demonstrated that GRAS transcription factors participate in regulating plant development and responding to biotic/abiotic stress. The present study provided useful information for the functional research in the future.

Acknowledgments

References

Riechmann JL, Heard J, Martin G, Reuber L, Jiang C, Keddie J,Adam L, Pineda O, Ratcliffe OJ, Samaha RR, Creelman R, Pilgrim M, Broun P, Zhang JZ, Ghandehari

D, Sherman BK, Yu G. 2000. Arabidopsis transcription factors: genome-wide comparative analysis among eukaryotes. *Science* 290:2105–2110.

Morohashi K, Minami M, Takase H, Hotta Y, Hiratsuka K. 2003. Isolation and characterization of a novel GRAS gene that regulates meiosis-associated gene expression. Journal of Biological Chemistry 278:20865–20873 DOI 10.1074/jbc.M301712200.

Peng J, Carol P, Richards DE, King KE, Cowling RJ, Murphy GP, Harberd NP. 1997. The *Arabidopsis GAI* gene defines a signaling pathway that negatively regulates gibberellin responses. *Gene & Development* 11:3194–3205.

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

Di Laurenzio L, Wysocka-Diller J, Malamy JE, Pysh L, Helariutta Y, Freshour G, Hahn MG, Feldmann KA, 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.

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

Bolle C. 2004. The role of GRAS proteins in plant signal transduction and development. *Planta* 218:683–692 DOI 10.1007/s00425-004-1203-z.

Schumacher K, Schmitt T, Rossberg M, Schmitz G, Theres K. 1999. The Lateral suppressor (Ls) gene of tomato encodes a new member of the VHIID protein family. *Proceedings of the National Academy of Science of the United States of America* 96:290–295.

Sun X, Xue B, Jones WT, Rikkerink E, Dunker AK, 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 Molecular Biology* 77:205–223 DOI 10.1007/s11103-011-9803-z.

Itoh H, Ueguchi-Tanaka M, Sato Y, Ashikari M, Matsuoka M. 2002. The gibberellin signaling pathway is regulated by the appearance and disappearance of SLENDER RICE1 in nuclei. *Plant Cell* 14:57–70 DOI 10.1105/tpc.010319.

Ikeda A, Ueguchi-Tanaka M, Sonoda Y, Kitano H, Koshioka M, Futsuhara Y, Matsuoka M, Yamaguchi J. 2001. slender rice, a constitutive gibberellin response mutant, is caused by a null mutation of the SLR1 gene, an ortholog of the height-regulating gene GAI/RGA/RHT/D8. *Plant Cell* 13:999–1010 DOI 10.1105/tpc.13.5.999.

Stuurman J, Jaggi F, Kuhlemeier C. 2002. Shoot meristem maintenance is controlled by a GRAS-gene mediated signal from differentiating cells. *Genes & Development* 16:2213–2218 DOI 10.1101/gad.230702.

Helariutta Y, Fukaki H, Wysocka-Diller J, Nakajima K, Jung J, Sena G, Hauser MT, Benfey PN. 2000. The SHORT-ROOT gene controls radial patterning of the Arabidopsis root through radial signaling. *Cell* 101:555–567 DOI 10.1016/S0092-8674(00)80865-X.

Bolle C, Koncz C, Chua N-H . 2000. PAT1, a new member of the GRAS family, is involved in phytochrome A signal transduction. *Genes & Development* 14:1269–1278 DOI 10.1101/gad.14.10.1269Genes&Dev.2000 .

Hirsch S, Kim J, Munoz A, Heckmann AB, Downie JA, Oldroyd GED. 2009. GRAS proteins form a DNA binding complex to induce gene expression during nodulation signaling in Medicago truncatula. *Plant Cell* 21:545–557 DOI 10.1105/tpc.108.064501.

Huang W, Xian Z, Kang X, Tang N, Li Z. 2015. Genome-wide identification, phylogeny and expression analysis of GRAS gene family in tomato. *BMC Plant Biology* 15:209 DOI 10.1186/s12870-015-0590-6.

Tian C, Wan P, Sun S, Li J, Chen M. 2004. Genomic-wide analysis of the GRAS gene family in rice and Arabidopsis. Plant Molecular Biology 54:519-532 DOI 10.1023/B:PLAN.000003256.89809.57

Zhang ZL, Ogawa M, Fleet CM, Zentella R, Hu JH, Heo JO, Lim J, Kamiya Y, Yamaguchi S, Sun TP. 2011. SCARECROW-LIKE 3 promotes gibberellin signaling by antagonizing master growth repressor DELLA in Arabidopsis. *Proceedings of the National Academy of Science of the United States of America* 108:2160–2165 DOI 10.1073/pnas.1012232108.

Day RB, Tanabe S, Koshioka M, Mitsui T, Itoh H, Ueguchi-Tanaka M, Matsuoka M, Kaku H, Shibuya N, Minami E. 2004. Two rice GRAS family genes responsive to N-acetylchitooligosaccharide elicitor are induced by phytoactive gibberellins: evidence for cross-talk between elicitor and gibberellin signaling in rice cells. *Plant Molecular Biology* 54:261–272 DOI 10.1023/B:PLAN.0000028792.72343.ee.

Heo JO, Chang KS, Kim IA, Lee MH, Lee SA, Song SK, Lee MM, Lim L. 2011. Funneling of gibberellin signaling by the GRAS transcription regulator SCARECROW-LIKE3 in the Arabidopsis root. *Proceedings of the National Academy of Science of the United States of America* 108:2166–2171 DOI 10.1073/pnas.1012215108.

Li X, Qian Q, Fu Z, Wang Y, Xiong G, Zeng D, Wang X, Liu X, Teng S, Hiroshi F, Yuan M, Luo D, Han B, Li J. 2003. Control of tillering in rice. *Nature* 422:618–621 DOI 10.1038/nature01518.

Schumacher K, Schmitt T, Rossberg M, Schmitz G, Theres K. 1999. The Lateral suppressor . Ls. gene of tomato encodes a new member of the VHIID protein family. *Proceedings of the*

National Academy of Science of the United States of America 96:290–5 DOI 10.103/nature01518.

Greb T, Clarenz O, Schafer E, Muller D, Herrero R, Schmitz G, Theres K. 2003. Molecular analysis of the LATERAL SUPPRESSOR gene in Arabidopsis reveals a conserved control mechanism for axillary meristem formation. *Gene & Development* 17:1175–1187 DOI 10.1101/gad.260703.

Torres-Galea P, Huang LF, Chua NH, Bolle C. 2006. The GRAS protein SCL13 is a positive regulator of phytochrome-dependent red light signaling, but can also modulate phytochrome A responses. *Molecular Genet and Genomics* 276:13–30 DOI 10.1007/s00438-006-0123-y.

Torres-Galea P, Hirtreiter B, Bolle C. 2013. Two GRAS proteins, SCARECROW-LIKE21 and PHYTOCHROME A SIGNAL TRANSDUCTION1, function cooperatively in phytochrome A signal transduction. *Plant Physiol.* 161:291–304 DOI 10.1104/pp.112.206607.

Schulze S, Schafer BN, Parizotto EA, Voinnet O, Theres K. 2010. LOST MERISTEMS genes regulate cell differentiation of central zone descendants in Arabidopsis shoot meristems. *Plant Journal* 64:668–678 DOI 10.1111/j.1365-313X.2010.

Wang T, Pan H, Wang J, Yang W, Cheng T, Zhang Q. 2014. Identifcation and profling of novel and conserved microRNAs during the flower opening process in Prunus mume via deep sequencing. *Molecular Genet and Genomics* 289:169–183 DOI 10.1007/s00438-013-0800-6.

Llave C, Xie Z, Kasschau KD, Carrington JC. 2002. Cleavage of Scarecrow-like mRNA targets directed by a class of Arabidopsis miRNA, *Science* 297 2053–2056 DOI 10.1126/science.1076311.

Consortium TG. 2012. The tomato genome sequence provides insights into fleshy fruit evolution. *Nature* 485:635–641 DOI 10.1038/nature11119.

Swarbreck D, Wilks C, Lamesch P, Berardini TZ, Garcia-Hernandez M, Foerster H, Li D, Mever T, Muller R, Ploetz L, Radenbaugh A, Singh S, Swing V, Tissier C, Zhang P, Huala E. 2008. The *Arabidopsis* information resource (TAIR):gene structure and function annotation. *Nucleic Acids Research* 36:D1009–D1014 DOI 10.1093/nar/gkm965.

Ouyang S, Zhu W, Hamilton J, Lin H, Campbell M, Childs K, ThibaudNissen F, Malek RL, Lee Y, Zheng L . 2007. The TIGR rice genome annotation resource: improvements and new features. *Nucleic Acids Research 35*:D883–D887 DOI 10.1093/nar/gkl976.

Finn RD, Mistry J, Tate J, Mistry J, Coggill PC, Sammut SJ, Hotz HR, Ceric G, Forslund K, Eddy SR, Sonnhammer EL, Bateman A. 2010. The Pfam protein families database. *Nucleic Acids Research* 38:D211–D222 DOI 10.1093/nar/gkm960.

Letunic I, Doerks T, Bork P . 2012. SMART 7: recent updates to the protein domain annotation resource. *Nucleic Acids Research* 40:D302–D305 DOI 10.1093/nar/gkr931.

Tamura K, Peterson D, Peterson N, Stecher G, Nei M, Kumar S. 2011. MEGA5: molecular evolutionary genetics analysis using maximum likelihood, evolutionary distance, and maximum parsimony methods. *Molecular Biology and Evolution* 28:2731–2739 DOI 10.1093/molbev/msr121.

Bailey TL, Williams N, Misleh C, Li WW. 2006. MEME: discovering and analyzing DNA and protein sequence motifs. *Nucleic Acids Research* 34:W369–W373 DOI 10.1093/nar/gkl198.

Guo A, Zhu Q, Chen X, Luo J. 2007. GSDS: a gene structure display server. Yi Chuan 29:1023-1026.

Li L, Stoeckert CJ Jr, Roos DS. 2003. OrthoMCL: identification of ortholog groups for eukaryotic genomes. *Genome Research* 13:2178–2189 DOI 10.1101/gr.1224503.

Tang H, Bowers JE, Wang X, Ming R, Alam M, Paterson AH. 2008. Synteny and collinearity in plant genomes. *Science* 320:486–488 DOI 10.1126/science.1153917.

Liu C, Wang X, Xu Y, Deng X, Xu Q. 2014. Genome-wide analysis of the R2R3-MYB transcription factor gene family in sweet orange (Citrus sinensis). *Molecular Biology Reports* 41:6769–6785 DOI 10.1007/s11033-014-3563-1.

Cannon SB, Mitra A, Baumgarten A, Young ND, May G . 2004. The roles of segmental and tandem gene duplication in the evolution of large gene families in Arabidopsis thaliana. *BMC Plant Biology* 4:10 DOI **1**0.1186/1471-2229-4-10.

Wang Y, Shi S, Zhou Y, Zhou Y, Yang J, Tang X. 2016. Genome-wide identification and characterization of GRAS transcription factors in sacred lotus (Nelumbo nucifera). *PeerJ* 4:e2388 DOI 10.7717/peerj.2388.

Krzywinski M, Schein J, Birol I, Connors J, Gascoyne R, Horsman D, Jones SJ, Marra MA. 2009. Circos: an information aesthetic for comparative genomics, *Genome Research* 19:1639–1645 DOI 10.1101/gr.092759.109.

Deng WK, Wang YB, Liu ZX, Cheng H, Xue Y. 2014. HemI: a toolkit for illustrating heatmaps. *PLoS ONE* 9:e111988 DOI 10.137/journal.pone.0111988.

Song XM, Liu TK, Duan WK, Ma QH, Ren J, Wang Z, Li Y, Hou XL. 2014. Genome-wide analysis of the GRAS gene family in Chinese cabbage (Brassica rapa ssp. Pekinensis). Genomics 103:135–146 DOI 10.1016/j.ygeno.2013.12.004.

Liu X, Alex W. 2014. Genome-wide Comparative Analysis of the GRAS Gene Family in Populus,Arabidopsis and Rice. *Plant Molecular Biology Reports* 32:1129 DOI 10.1007/s11105-014-0721-5.

Lee MH, Kim B, Song SK, Heo JO, Yu NI, Lee SA, Kim M, Kim DG, Sohn DO, Lim CE, Chang KS, Lee MM, Lim J. 2008. Large-scale analysis of the GRAS gene family in Arabidopsis thaliana. *Plant Molecular Biology* 67:659–670 DOI 10.1007/s11103-008-9345-1.

Tang K, Dong C, Liu J. 2016. Genome-wide analysis and expression profiling of the phospholipase D gene family in Gossypium arboretum. *Life Sciences* 59:130-141 DOI 10.1007/s11427-015-4916-2.

Chen F, Mackey AJ, Vermunt JK, Roos DS.2007. Assessing performance of orthology detection strategies applied to eukaryotic genomes. *PLoS ONE* 2:e383 DOI 10.1371/journal.pone.0000383.

Lu J, Wang T, Xu Z, Sun L, Zhang Q. 2015. Genome-wide analysis of the GRAS gene family in Prunus mume. *Molecular Genet and Genomics* 290:3.3-317 DOI 10.1007/s00438-014-0918-1.