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

Yiling Niu [1] , Jingbin Jiang [1] , He Zhang [1] , Xiuling Chen [1] , Jingfu Li [1] , Xiangyang Xu Corresp. [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 often 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, rational 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, including 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] Jingbin Jiang[1] He Zhang[1] Xiuling Chen[1] Jingfu Li[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
[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 often 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, rational 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, including 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: GRAS transcription factors, tomato, genome-wide analysis, characterization

37

38

39

40

Introduction

41

42   The transcription factors are essential in functional genomics. Transcription factors are
43   protein molecules that can either activate or repress the target genes to ensure their specific
44   expression by combining with the genes of 5'-flank cis-element(*Riechmann et al.,2000*). The
45   typical transcription factor consists of DNA-binding domain, transcription regulation
46   domain(activeregion or suppressor region), oligomerization site, and nuclear localization
47   signal(*Morohashi et al., 2002*). The GRAS transcription factor is a major plant-specific gene
48   family among putative transcription factors that are found in a variety of plant species; for
49   example, *Arabidopsis(Arabidopsisthaliana*), rice(*Sativa oryza*), grapevine(*Vitis vinifera*),
50   tobacco (*Nicotiana tabacum*), Chinese cabbage (*Brassica rapa ssp.pekinensis*), and *Prunus*
51   *mume*. Based onthe knownfirst three members, GAI(gibberellicacidinsensitive) (*Peng et al.,*
52   *1997*), RGA(repressor of GA1-3 mutant(*Silverstone et al., 1998*) and SCR(scarecrow) (*Di*
53   *Laurenzio et al., 1996*), this transcription factor was named GRAS with the characteristic letter
54   from each of the three members The family members were screened as GRAS gene family as
55   each of them contained the GRAS domain(*Pysh et al. 1999*). The GRAS proteins are usually
56   composed of 400–700 amino acid residues (*Bolle, 2004*). A few GRAS proteins contain two
57   structural domains: one GRAS domain andthe other functional domain (*Schumacher et al.,*
58   *1999*). The typical features of these proteins include a highly conserved C-terminal region and a
59   variable N-terminal region (*Sun et al., 2011*). The conserved C-terminal region harbors five

60   sequence motifs: LHRI(leucine heptad repeat I), LHRII(leucine heptad repeat II), VHIID motif

61   (*Pysh et al.,1999*), SAW motif, and PFYRE motif. The structures of both LHR I and LHR II are
62   two leucine enrichment regions, the VHIID motif is a core structure, which exists in all members
63   of the GRAS gene family and it can combine with LHR I and II to form the complex LHR I-
64   VHIID-LHR II. This structural pattern might be a crucial function for DNA-binding and protein-
65   binding in the interactions of proteins and proteins (*Itoh et al., 2002*). The functional localization
66   of SAW and PFYRE motifs are not yet clearly elucidated; however, the missense mutations in
67   these motifs in RGA and SLR1 proteins exhibit strong mutant phenotypes (*Silverstone et al.,*
68   *1998*). In addition, the N-terminal region of GRAS proteins in the DELLA subfamily contains
69   the other two conserved motifs: DELLA and VHYNP, indicating that the N-terminal domains of
70   GRAS proteins are variable (*Peng et al., 1997*). Moreover, the GRAS proteins are not only
71   structurally diverse but also exert multiple functions. In recent years, several groups have found
72   that the GRAS proteins are one of the indispensable regulation factors in plant growth
73   development and participate in many biochemical and physiological processes in plants, such as

74 gibberellin signal transduction (*Peng et al., 1997; Silverstone et al., 1998; Ikeda et al., 2001*),
75 axillary shoot meristem formation (*Stuurman et al., 2002*), root radial patterning (*Di Laurenzio et*
76 *al.,1996; Helariutta et al.,2000*), male gametogenesis (*Morohashi et al., 2003*), phytochrome A
77 signal transduction (*Bolle et al.,2000*), nodulation signal transduction   (*Hirsch et al.,2009*), and
78 biotic/abiotic stress (*Huang et al.,2015*). Previously, the GRAS proteins were divided into eight
79 subfamilies, according to their common feature or member to name those subfamilies; for
80 instance, SHR, SCR, DELLA, LISCL, Ls, HAM, PAT1, and SCL3. The amino acids in each
81 group are mostly homologous, and thus, the GRAS genes of each subfamily might possess
82 similar or related functions (*Tian et al., 2004*).

83   Hitherto, the RNA-seq is developed, following the function of GRAS genes, which are
84 deeply excavated. Some GRAS genes have been characterized based on *Arabidopsis* and rice; for
85 example, the DELLA proteins that function as negative regulators in gibberellin signal
86 transduction, GA signaling pathway regulates the plant growth and development by degrading
87 the DELLA proteins (*Zhang et al., 2011; Heo et al., 2011*). Firstly, the signal perception of GID1
88 proteins receive the GA signal; then the GID1 proteins combine with the DELLA proteins to
89 form the complexes of GA-GID1-DELLA; subsequently, the DELLA proteins specifically bind
90 the F-box protein SLY1, which subordinates the SCF$^{SLY1/GID2/SNE}$ protein complex. Finally, the
91 inhibition of DELLA proteins was released immediately by the 26S proteasome, and thus, the
92 plants show normal growth (*Day et al.,2004*). The study showed that the DELLA motif in the N-
93 terminal is indispensable for the interaction between the two proteins DELLA and GID1;
94 however, the motifs in the C-terminal are redundant. The SCR and SHR groups and the analysis
95 of SHR/SCR mutant showed short root phenotype (*Di Laurenzio et al., 1996; Helariutta et al.,*
96 *2000*), providing evidence that both proteins act as positive regulators in the radial organization
97 of the root. Previous studies have shown that the SCR proteins combine with SHR proteins to
98 form a complex, while the SHR proteins were transported to the endodermis of the root (*Sabatini*
99 *et al., 2003; Wysocka et al., 2000; Helariutta et al., 2000*). Similarly, SCL3 proteins were
100 involved in the elongation and differentiation region of the root and over-ground organs,
101 respectively. In the root, the subsequent elongation is regulated to control the GA signaling
102 pathway, while in the meristematic tissue, the combination of SHR/SCR leads to the
103 organizational maturity within the GA signaling pathway (*Heo et al., 2011*). Some studies
104 indicate that the overexpression of *OsMOC1* gene results in increasing the tiller numbers and
105 decreasing the length; moreover, it can only be detected in the axillary bud (*Li et al., 2003*). In
106 addition to rice, the *Ls* gene in tomato and the *AtLAS/SCL18* gene in *Arabidopsis* are closely
107 linked to the growth of the lateral bud (*Schumacher et al., 1999; Greb et al., 2003*). The *PAT1*,
108 *SCL13*, and *SCL21* genes belong to the *PAT1* branch, which mediates the phytochrome signaling
109 pathways. Genetic evidence suggests that these three genes act as positive regulators; the *SCL13*
110 gene participates in phytochrome B transduction in dependently, whereas the*PAT1*and *SCL21*
111 adjust the phytochrome A signaling network by the mutual effect (*Bolle et al., 2000; Torres-*
112 *Galea et al., 2006; Torres-Galea et al., 2013*). The *NSP1* and *NSP2* genes exist in the

113 downstream of symbiosis signal transduction pathway CCaMK (Ca/calmodulin-dependent
114 protein kinases), which is related to the nodulation. TheNPS1-NPS2 heterodimer is induced by
115 the nodulation factors, following which, the heterodimer can specifically combine the promoter
116 of the early nodulation gene *ENOD2* to promote the expression of related genes, thereby forming
117 the stage of nodulation (*Hirsch et al., 2009*). The homologous genes of *NSP1/NSP2*are
118 extensively encompassed in several non-leguminous plants. Another analysis about GRAS
119 proteins found that some members in the gene family are regulated by miRNA171; for instance,
120 *AT2G45160, AT3G60630*, and *AT4G00150* in Arabidopsis (*Schulze et al., 2010*), Pm017821 and
121 Pm023512 in *Prunus mume* (*Wang et al., 2014*), *Solyc01g090950.2.1* and *Solyc08g0700.1.1* in
122 tomato (*Huang et al., 2015*), and four genes in rice (*Llave et al., 2002*) are complementary to
123 miRNA171.
124   In the past few years, some investigations showed that the GRAS genes respond to different
125 hormones and abiotic stress treatments (*Huang et al., 2015*). Tomato is an adequate good model
126 plant for the study of Solanaceae due to its great economic value. With the whole-
127 genome analyses of tomato, the genomic data can highlight the connection between genes and
128 plants. The previous study of tomato GRAS genes focused on several genes' functions. In this
129 study, a comprehensive and systematic analysis would provide a deep insight on the GRAS
130 family and precede the sequencing studies.

131 ## Materials & methods

132 ### Identification of tomato GRAS genes
133   we retrieved the tomato GRAS genome sequences, protein sequences and annotation
134 information from SGN (https://solgenomics.net/) (*Consortium, 2012*). The Arabidopsis GRAS
135 genes' family sequences and annotation information was download from TAIR
136 (http://www.arabidopsis.org/) (*Swarbreck et al., 2008*), whereas, the rice GRAS transcription
137 factor sequences and annotation information was obtained from RGAP
138 (http://rice.plantbiology.msu.edu/index.shtml) (*Ouyang et al., 2007*). The grape GRAS gene
139 family sequences were download from plantTFDB (http://planttfdb.cbi.pku.edu.cn/). The HMM
140 model of GRAS transcription factor was download from PFAM (http://pfam.xfam.org/) (*Finn et*
141 *al., 2010*), known as PF03514. The HMM model was used as a query to identify the tomato
142 GRAS genes containing the GRAS domain with a cut-off E-value of $1e^{-5}$ in the HMMER
143 software. Consequently, we identified a total of 54 GRAS genes in tomato, 34 from Arabidopsis,
144 56 from rice, and 37 from the grapevine. Then, we conducted a quality check using the Simple
145 Modular Architecture Research Tool SMART (http://smart.embl-heidelberg.de/) (*Letunic et al.,*
146 *2012*) to confirm the presence of GRAS domain on the candidate GRAS genes. These data were
147 used for subsequent analysis.

148 ### Phylogenetic analysis for tomato GRAS genes
149   MEGA 6.0 was used to analyze the phylogenetics of genome-wide GRAS gene family
150 based on the whole set of GRAS protein sequences from tomato(*solanum lycopersicum*),

151   *Arabidopsis*(*Arabidopsis thaliana*), rice(*Sativa oryza*) and grapevine(*Vitis vinifera*). The
152   *Arabidopsis* is one of the most commonly used plants in Cruciferae for studying the genetic
153   correlations; *Arabidopsis* to Cruciferae, grapevineto Vitaceae and rice to Gramineae. (*Tamura et*
154   *al., 2011*). In this building process, several shorter amino acids were excluded. According to a
155   previous study, the classification was made on the phylogenetic tree using the Evolview software
156   (http://www.evolgenius.info/).

157   **Structure analysis of tomato GRAS genes**

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

167   **Evolution analysis of tomato GRAS genes**

168   The sequences were compared using the GRAS genes in tomato, *Arabidopsis* and rice; the
169   entire protein sequences were used to identify the orthologous and paralogous genes using the
170   software OrthoMCL (http://orthomcl.org/orthomcl/) with an E-value of 1e-5 and a match cut-off
171   value 50 for against-all BLASTp alignment (*Li et al., 2003*). The MCscanX software (*Tanget al.,*
172   *2008*) was used to identify the collinear block based on the tomato genomes; if a gene had more
173   than one transcript, only the first transcript in the annotation was used. The underlying
174   mechanism of tandem duplication showed that two genes were physically close to each other
175   with the genes residing within 20kb (*Liu et al., 2014*), segmental duplication resulted from the
176   whole genome duplication accompanied by a comprehensive gene loss (*Tang et al., 2008*), and
177   those large duplication events can be deduced by anchor genes in collinear blocks (*Cannon et al.,*
178   *2004*). The chromosomal localization and homologous collinear relationship of GRAS genes
179   were visualized using the Circos program (*Krzywinski et al., 2009*).

180   **Expression pattern analysis of GRAS genes**

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

188   In the case of difficulty, other public data were retrieved to reveal the relationship between
189   GRAS genes and biotic stress. The transcriptome of leaves of resistant (RG-PtoR) and

190 susceptible (RG-prf3 and RG-prf19) tomato plants treated with pstDC3000 in 4 and 6 h was
191 sequenced.

Results

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

194 In order to identify the GRAS proteins in tomato, we downloaded the raw data sequence of
195 tomato GRAS transcription factors from SGN database. The bioinformatics approach retrieved
196 54tomato GRAS genes, 34 *Arabidopsis* GRAS genes, 56 rice GRAS genes, and 37grape GRAS
197 genes. In addition, we obtained basic information in connection with tomato GRAS proteins
198 (Table S1). The length of GRAS proteins ranged from 125–864aa and the tomato GRAS genes
199 were almost distributed across the 12 chromosomes uniformly; the highest content was on
200 chromosome 1, containing 8 GRAS genes. However, at least 3 GRAS genes were located on
201 other chromosomes. The annotation information represented that the GRAS domain always
202 assembled on the C-terminal, suggesting that the conserved terminal is supported by the
203 stable structure of the GRAS domain; this phenomenon could be used for the subsequent
204 comparison analysis. The smaller HMM E-value provided reliable screening results. Notably, a
205 number of amino acid residues in some GRAS proteins were less than that in the typical GRAS
206 domain, such as *Solyc04g011630.1.1* or *Solyc06g076290.1.1*genes, which might be related to the
207 GRAS domain including conservation-specific regions.

Genome-wide sequence alignment and phylogenetic analysis

209 The topology tree contains 178 proteins: 56 from rice, 34 from *Arabidopsis*, 37 from
210 grapevine, and 51 from tomato. However, the remaining 3 proteins from tomato were neglected
211 because of the shorter sequence. Basing on the topology structure, clade support value and
212 previous classification from rice and *Arabidopsis*, as shown in Fig.1 (*Tian et al., 2004*), all of
213 these proteins were divided into ten groups, named specifically after a common feature or one of
214 the members: PAT1, SHR, LISCL, Gv6, SCR, LAS, SCL3, DELLA, HAM and BolA,
215 respectively. Of these, eight were named according to a previous report (*Tian et al., 2004*). The
216 remaining two groups were named by the characteristic of the members. The tree showed that the
217 four species GRAS proteins were distributed in the ten groups, randomly, and hence, the
218 members of GRAS transcription factors in these four species are not represented equally. For
219 example, the subfamily SCL3 contains one or two proteins from *Arabidopsis*, tomato, and
220 grapevine, respectively, and seven members from rice. Strikingly, some clades do not contain the
221 members of GRAS proteins from *Arabidopsis* or rice; for instance, Gv6 and BolA groups. This
222 might be attributed to the loss of the corresponding member following the separation of the last
223 common ancestor. We found that the Gv6 group consisted of 6 tomato proteins and 2 grapevine
224 proteins, indicating that it is a fruit tree species-specific clade.

225 The genes with orthologs frequently tend to be clustered together, and the subfamily
226 members in a major group share similar gene structure and function. Therefore, we observed that
227 the distribution of specific genes provide information on the role of the other genes in the same

228 clade. For example, within the PAT1 group, *AT5G48150.1* of Arabidopsis was previously
229 demonstrated to participate in the process of phytochrome signal transduction. In addition, the
230 *Solyc07g047950.1.1* was highly similar to the *AT5G48150.1* protein, and hence, we inferred that
231 *Solyc07g047950.1.1* protein also played a crucial role in phytochrome signal transduction. PAT1
232 and LISCL subfamily comprised of more proteins than the other subfamilies as well as that
233 reported in previous studies.
234 Members in the same sub-branch were marked by the same color and surrounded with an appropriate
235 color line that was commanded by the subfamily name. The red circle corresponds to the tomato GRAS
236 proteins, the green circle corresponds to the *Arabidopsis* GRAS proteins, the blue circle corresponds to
237 the rice GRAS proteins, and the pink circle corresponds to the grapevine GRAS proteins.

238
239 Members in the same sub-branch were marked by the same color and surrounded with an appropriate
240 color line that was commanded by the subfamily name. The red circle corresponds to the tomato GRAS
241 proteins, the green circle corresponds to the *Arabidopsis* GRAS proteins, the blue circle corresponds to
242 the rice GRAS proteins, and the pink circle corresponds to the grapevine GRAS proteins.

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

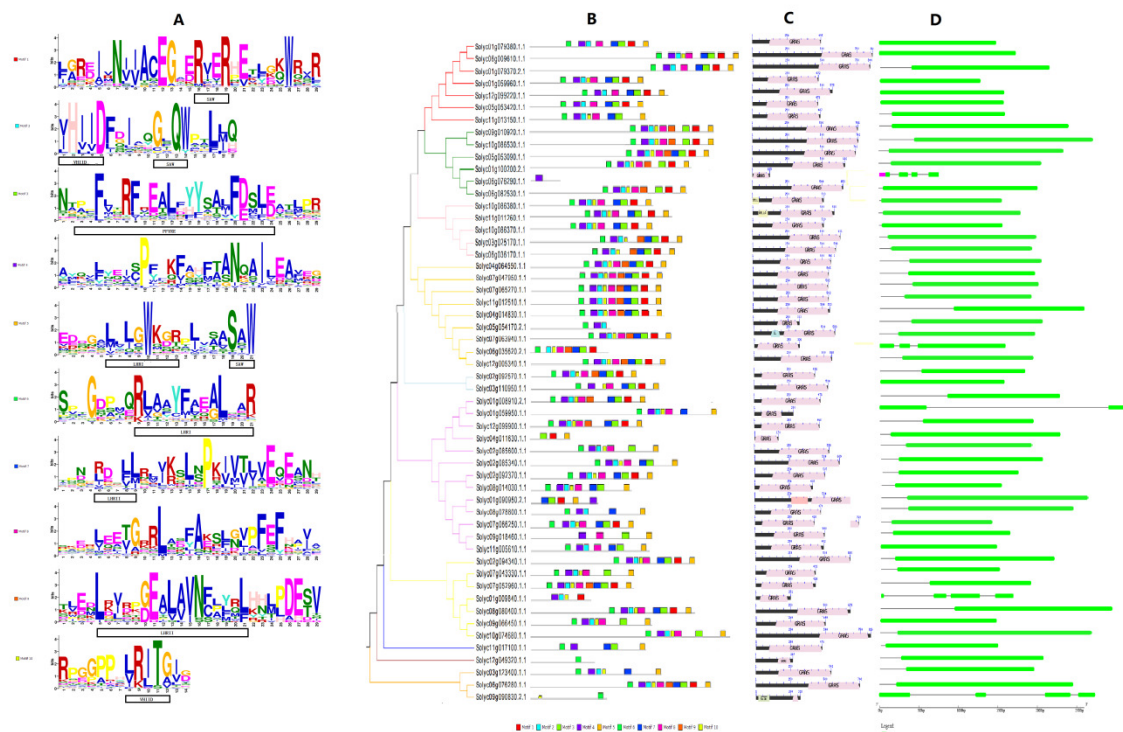
244

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

246   To achieve a general overview of the conserved features of the tomato GRAS proteins, we
247 performed a multiple sequence alignment of the 54 tomato GRAS genes using MEME software;
248 the conserved motifs of all proteins are shown in Fig.2-B. Ten is the default number of motifs in
249 MEME analysis. We found that 80% GRAS proteins contained >7 motifs, and 24% GRAS
250 proteins encompassed 10 motifs. Moreover, members of the same clade of phylogenetic tree
251 shared similar motif organization with respect to either gene length or motif number (*Liu et al.,*
252 *2014*). On the other hand, several proteins only contained one or two motifs, which might be
253 attributed to the short duration of those. Also, the motifs were more likely to be located in the C-
254 terminal than the N-terminal. The motif logo is shown in Fig.2-A. According to the previous
255 studies on GRAS motifs (*Tian et al., 2004*), the five putative GRAS domains were exhibited in
256 the logo of motifs; we speculated that each LHRI, LHRII, VHIID, PFYRE, and SAW motifs
257 could be divided into different units. This phenomenon demonstrated that maximum tomato
258 GRAS genes are conserved, and the conserved region is localized on the C-terminal.
259 Furthermore, we conducted a domain analysis (Fig.2-C) and exon/intron analysis(Fig.2-D) of
260 tomato GRAS genes in order to explore the diversity and functionality. The visualization of the
261 domains revealed that nearly all the GRAS proteins domains were primarily distributed in the C-
262 terminal, further supporting the theory that the C-terminal is highly conserved. Nevertheless,
263 there were some specific genes, such as the whole *Solyc06g076290.1.1* gene is a GRAS domain,
264 *Solyc10g086380.1.1*, *Solyc11g011260.1.1*, *Solyc07g063940.1.1*, *Solyc01g090950.2.1* and
265 *Solyc090830.2.2* genes contained two domains. In addition, the exon/intron analysis displayed
266 similar distribution characteristics. The *Solyc06g076290.1.1* and *Solyc090830.2.2* genes are
267 comprised of multiple exons; however, the majority of GRAS proteins possessed only one exon
268 and one intron, which might result from intron gain or loss event during evolution.

269

270   A: Motif logo, the amino acid composition of each motif; B: The motif distribution in each GRAS gene;

271   C: The location of GRAS domain in genes; D: The length of exons, introns and UTRs.

272   **Fig. 2** The structure of GRAS genes in tomato.

273

274   **Comparison and expansion analysis of tomato GRAS genes**

275   To evaluate the evolutionary relationships among tomato GRAS genes, we made a

276   comparative analysis to identify the co-orthologous, orthologous and paralogous GRAS genes

277   among tomato, *Arabidopsis*, and rice using the OrthoMCL software, as shown in Fig.3. We

278   identified 22 co-orthologous and 46 orthologous gene pairs between tomato and *Arabidopsis*, 45

279   co-orthologous and 48 orthologous gene pairs between tomato and rice and 19 co-orthologous

280   and 29 orthologous gene pairs between *Arabidopsis* and rice. Moreover, 38 tomato GRAS genes

281   (65%) have paralogous genes, which is higher than that in rice (63%) and *Arabidopsis* (19–56%).

282   The orthologous genes commonly share a similar structure and biological function. The number

283   of orthologous gene pairs between tomato and rice was more than that between tomato and

284   *Arabidopsis*, suggesting that rice is more close to tomato.

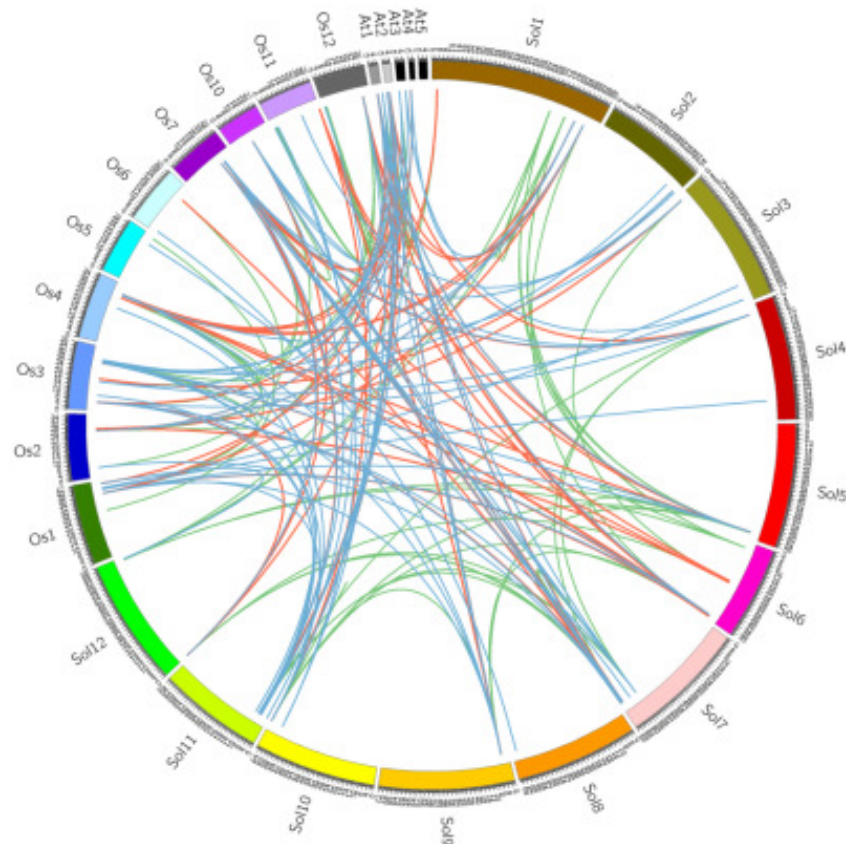
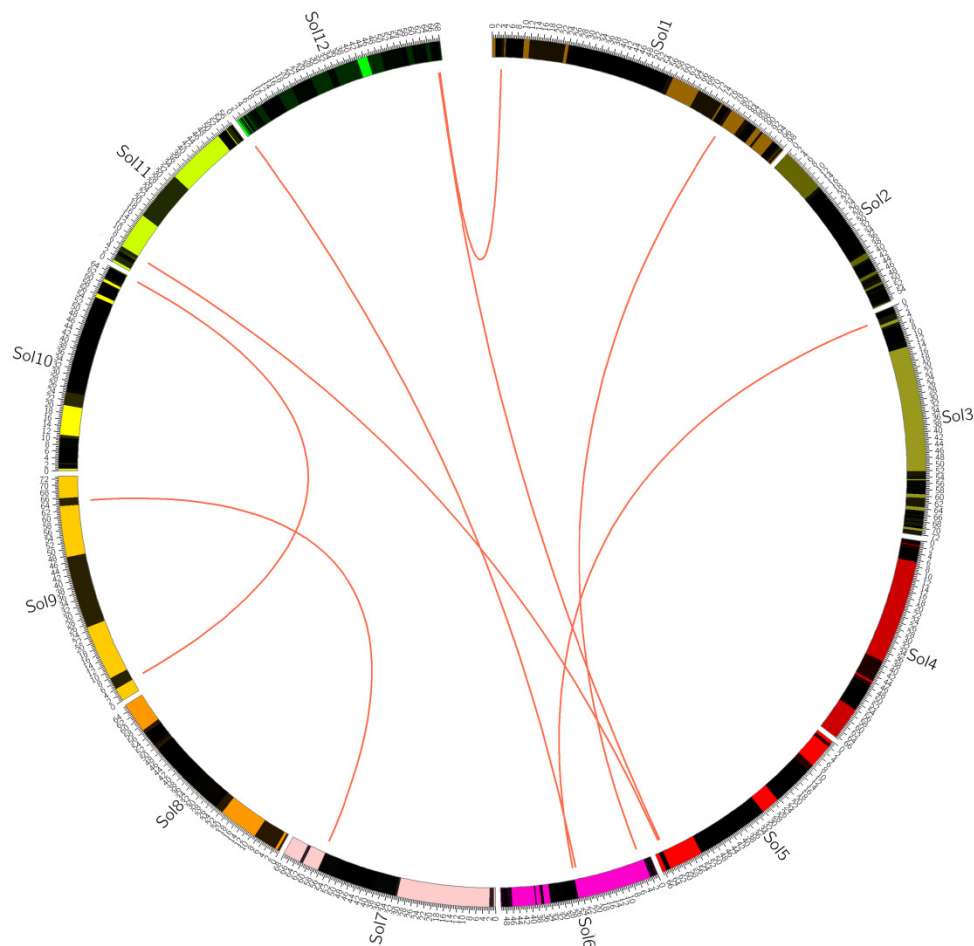
285   Moreover, the duplication events were discovered in the evolution of tomato GRAS

286   transcription factors (*Cannon et al., 2004*). Tandem duplication and segmental duplication were

287   vital to the expansion of the GRAS family. To reveal the relationship between gene duplication

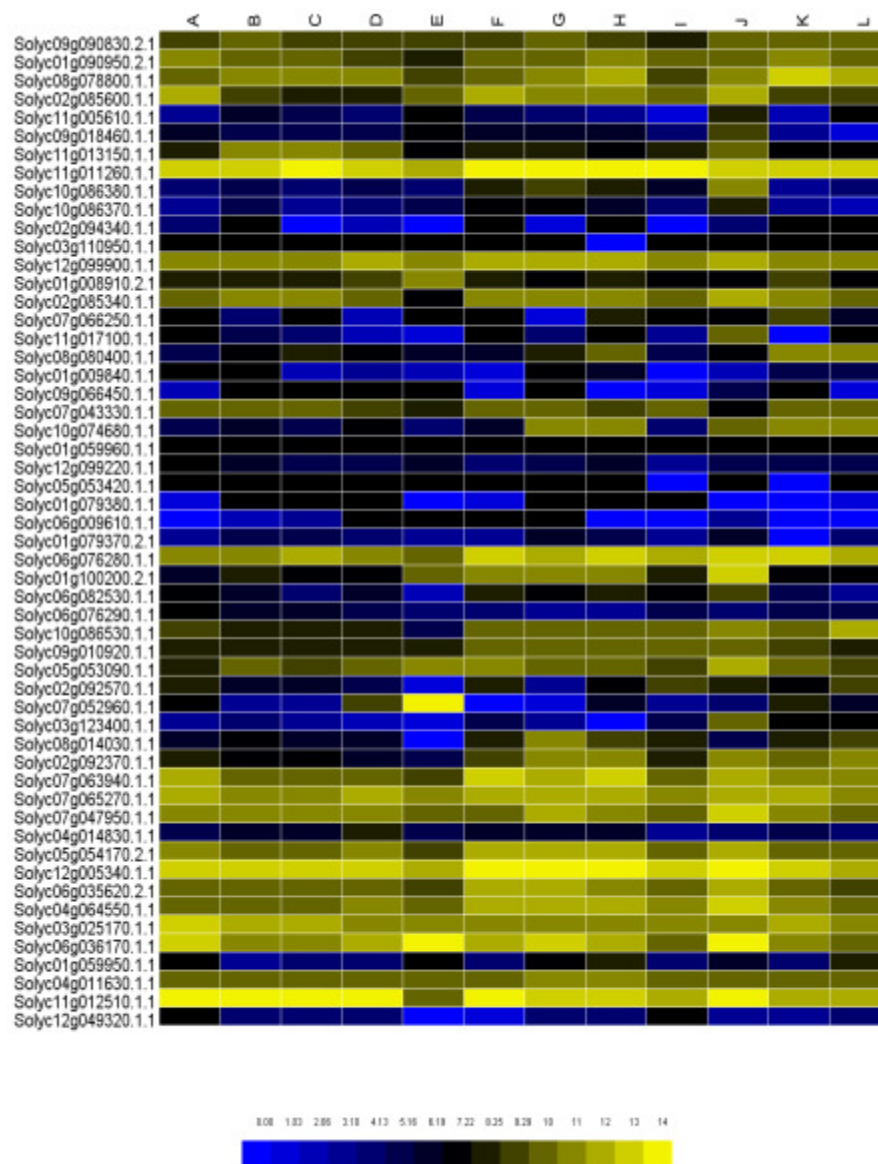
288   and amplification, the syntenic regions were analyzed by MCscanX software. As shown in Fig.3,

289   finally, five tandem duplication gene pairs (18.5%) were obtained, the details of duplication

290   events were referred in Table S2, suggesting that the origination of GRAS genes applied to the
291   tandem duplication events. We also assessed the contribution of segmental duplications; 18
292   genes (33.3%) were harbored in collinear blocks with their duplications, indicating that their
293   robust participation in the process of GRAS genes' expansion. The same observations were also
294   made in other protein families.

295
296   Orange, blue and green lines indicate co-orthologous, orthologous and paralogous respectively. The words
297   beginning with A, O and S represent the chromosomes of *Arabidopsis*, rice and tomato, respectively.
298   **Fig. 3** Chromosomal localization of GRAS homologous genes in tomato.

299
300   The red lines represent the segmentally duplicated genes, and the black bands represent the collinear
301   block.
302   **Fig. 4** Chromosomal localization of GRAS duplicated genes in tomato.

303

304   **Expression pattern analysis of tomato GRAS genes in different tissues**
305   In this study, we analyzed different expression levels of GRAS proteins in different tissues
306   regarding the published RNA-seq data. The heat map was constructed to show the expression
307   profile (Fig.5). We found that only one gene *Solyc01g059960.1.1* was not detected in the RNA-
308   seq data that might be due to temporal of no expression (*Song et al., 2014*). The other tomato
309   GRAS genes were obtained in at least one tissue. As a whole, we found that the same group of
310   GRAS genes shared a similar expression pattern; for instance, nearly all the PAT1 subfamily
311   members showed a higher expression level than the other groups, and the Gv6 subfamily showed
312   a low expression level in all tissues. The GRAS genes of some branches exhibited a tissue-
313   specific expression pattern. For example, the *Solyc11g005610.1.1*, *Solyc03g123400.1.1*, and
314   *Solyc09g018460.1.1* genes belong to the HAM subfamily that is lowly expressed in those tissues
315   except the hypocotyls, thereby suggesting that the three genes had contributed considerably to

316 the development of the hypocotyl. The *Solyc04g014830.1.1* gene only expresses in the anthesis
317 stage; it would be closely related to the flower opening. *Solyc11g011260.1.1,*
318 *Solyc12g005340.1.1* and *Solyc11g012510.1.1* genes are always expressed at a high level in all
319 the tissues, indicating their crucial role in the whole growth process of tomato. We also
320 determined that some genes present a time-specific expression, such as, in the stage of fruit
321 ripening; several genes are expressed in the later stage. During the mature period of leaf and
322 flower, the expression of the GRAS genes remains stable, and only two or three genes show a
323 difference.

324
325 A: newly developed leaves approximately 5 mm long; B: mature green leaflets; C: flower buds
326 10 days before anthesis or younger; D: flowers at anthesis; E: 10 days post anthesis (DPA) fruit;

327 F: 20 DPA fruit; G: breaker stage ripening fruit; H: another set of 10 DPA fruit was collected in
328 a separate greenhouse for comparison, the following tissues were collected from seeds that were
329 germinated and grown for 7 days in a Petri dish under growing lights, I: whole root; J: hypocotyl
330 from below the cotyledons to above the root zone; K: cotyledons; L: vegetative meristems. and
331 the the yellow indicating higher expression levels and blue indicating lower expression levels.

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

333   According to the public database, using the visualized tool HemI to present the relationship
334 between pstDC3000 and GRAS genes from tomato (Fig.6), we deduced that the number of
335 higher expression genes in susceptible tomato (54%) were more than that in resistant tomato
336 (41%). The group of PAT1 continued to show the highest expression level, and the members of
337 the PAT1 subfamily might regulate some critical physiological processes in tomato growth that
338 are related to resistance. During the same infection time, the two varieties of susceptible plant
339 shared the same tendency of GRAS genes' expression. As the infection time continued, the
340 SISCL group genes' expression was on a downward trend in the resistance tomato, such as
341 *Solyc06g076280.1.1*, *Solyc01g100200.2.2* and *Solyc06g076290.1.1* genes; interestingly, the
342 *Solyc05g053090.1.1* gene was on an increasing trend. In the susceptible tomato, the
343 *Solyc02g085600.1.1* and *Solyc07g063940.1.1* genes' expressions were increasing, whereas, that
344 of the other members of the group HAM, LISCL and PAT1 were declining. In 4 h, the
345 *Solyc01g100200.2.1*, *Solyc07g063940.1.1* and *Solyc12g005340.1.1* genes' expression differed
346 significantly in resistance tomato from that of the susceptible tomato. In 6 h, the
347 *Solyc12g005340.1.1* gene still differed largely, and the results suggest that the
348 *Solyc12g005340.1.1* gene regulated the plant disease resistance.

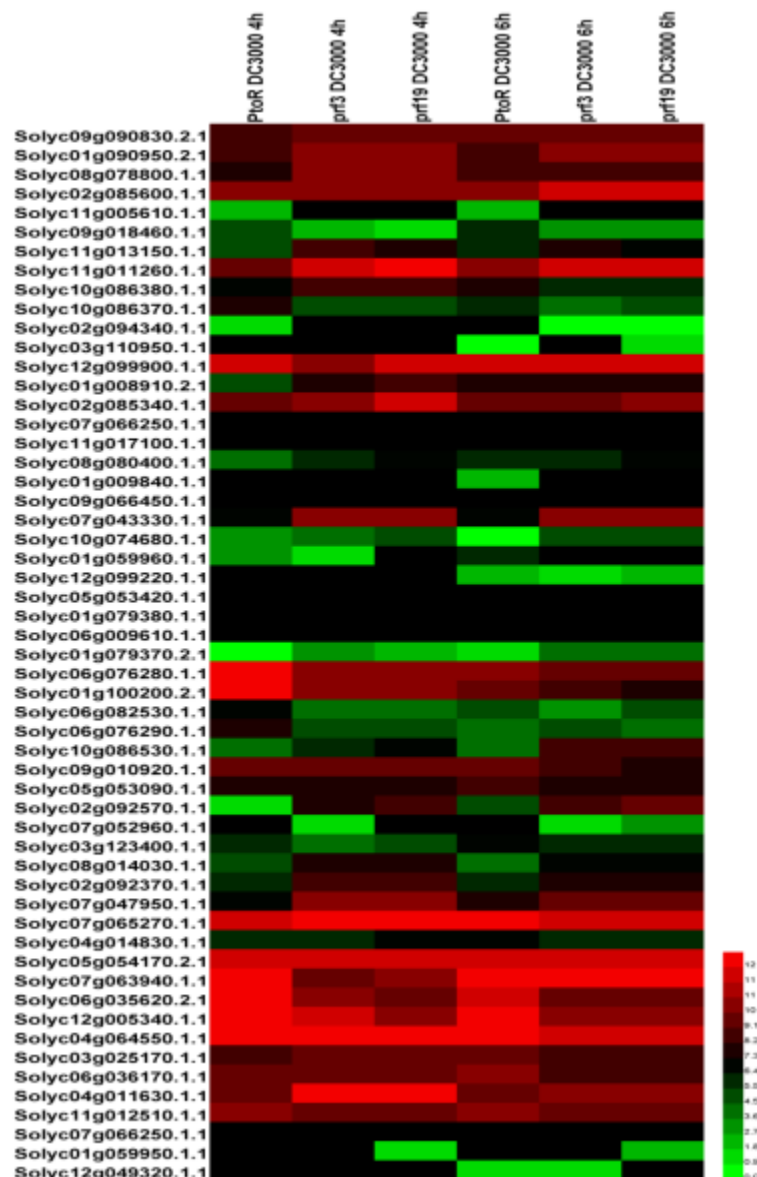
349

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

361 using the bioinformatics methods. Subsequently, the classification and annotation information
362 was obtained, and the contrast of full length of the GRAS proteins from tomato showed
363 remarkable differences. The distribution of GRAS genes in tomato is consistent with that of
364 *Arabidopsis*. The next phylogenetic analysis might provide additional functional constituents
365 among the four species. The previous studies demonstrated that the GRAS transcription factors
366 are involved in plant development and stress response; it is acknowledged that higher the
367 sequence similarities similar the functions in different species (*Chen et al., 2007*). Additionally,
368 the GRAS proteins have similar functions within the same clade. The structure and domain
369 analysis proved that the topology tree was reliable, and the parallel structural features were
370 clustered to the same subgroup. Every GRAS gene was   composed of one or more conserved
371 motifs. A comparative genomics analysis revealed abundant homologous genes in tomato,
372 *Arabidopsis*, and rice, and the segmental duplication commonly promotes the expansion of
373 GRAS proteins (*Cannon et al., 2004*). The expression of GRAS genes in different tissues
374 indicates that the three genes *Solyc11g011260*, *Solyc11g012510* and *Solyc12g005340* showed a
375 high expression level among those organs, implicating their vital importance in plant
376 development. Moreover, the members of the same clade share similar expression profiles (*Wang
377 et al., 2016*). Some of the GRAS genes respond to the biotic stress of pstDC3000; the PAT1
378 subfamily showed the highest expression level. A similar result was observed in other species for
379 the expression of GRAS genes (*Lu et al., 2015*).

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

Acknowledgments

388   This work was support by the National Natural Science Foundation of China (Grant
389 No.31272171) and Province in Heilongjiang outstanding youth science fund (No. JC201204)

References

391 **Riechmann JL, Heard J, Martin G, Reuber L, Jiang C, Keddie J,Adam L, Pineda O,**
392   **Ratcliffe OJ, Samaha RR, Creelman R, Pilgrim M, Broun P, Zhang JZ, Ghandehari**
393   **D, Sherman BK, Yu G. 2000.** Arabidopsis transcription factors: genome-wide comparative
394   analysis among eukaryotes. *Science* 290:2105–2110.
395 **Morohashi K, Minami M, Takase H, Hotta Y, Hiratsuka K. 2003.** Isolation and
396   characterization of a novel GRAS gene that regulates meiosis-associated gene expression.
397   Journal of Biological Chemistry 278:20865–20873

398   DOI 10.1074/jbc.M301712200.

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

402 **Silverstone AL, Ciampaglio CN, Sun T. 1998.** The Arabidopsis RGA gene encodes a
403   transcriptional regulator repressing the gibberellin signal transduction pathway. *Plant Cell*
404   10:155–169.

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

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

412 **Bolle C. 2004.** The role of GRAS proteins in plant signal transduction and development. *Planta* 218:683–
413   692 DOI 10.1007/s00425-004-1203-z.

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

417 **Sun X, Xue B, Jones WT, Rikkerink E, Dunker AK, Uversky VN . 2011.** A functionally
418   required unfoldome from the plant kingdom: intrinsically disordered N-terminal domains of
419   GRAS proteins are involved in molecular recognition during plant development. *Plant
420   Molecular Biology* 77:205–223
421   DOI 10.1007/s11103-011-9803-z.

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

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

430 **Stuurman J, Jaggi F, Kuhlemeier C. 2002.** Shoot meristem maintenance is controlled by a
431   GRAS-gene mediated signal from differentiating cells. *Genes & Development* 16:2213–
432   2218 DOI 10.1101/gad.230702.

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

437 **Bolle C, Koncz C, Chua N-H . 2000.** PAT1, a new member of the GRAS family, is involved in
438   phytochrome A signal transduction. *Genes & Development* 14:1269–1278 DOI
439   10.1101/gad.14.10.1269Genes&Dev.2000 .

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

444 **Huang W, Xian Z, Kang X, Tang N, Li Z. 2015.** Genome-wide identification, phylogeny and
445   expression analysis of GRAS gene family in tomato. *BMC Plant Biology* 15:209 DOI
446   10.1186/s12870-015-0590-6.

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

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

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

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

464 **Li X, Qian Q, Fu Z, Wang Y, Xiong G, Zeng D, Wang X, Liu X, Teng S, Hiroshi F, Yuan M,**
465   **Luo D, Han B, Li J. 2003.** Control of tillering in rice. *Nature* 422:618–621 DOI
466   10.1038/nature01518.

467 **Schumacher K, Schmitt T, Rossberg M, Schmitz G, Theres K. 1999.** The Lateral suppressor .
468   Ls. gene of tomato encodes a new member of the VHIID protein family. *Proceedings of the*
469   *National Academy of Science of the United States of America* 96:290–5 DOI
470   10.103/nature01518.

471 **Greb T, Clarenz O, Schafer E, Muller D, Herrero R, Schmitz G, Theres K. 2003.** Molecular
472   analysis of the LATERAL SUPPRESSOR gene in Arabidopsis reveals a conserved control
473   mechanism for axillary meristem formation. *Gene & Development* 17:1175–1187 DOI
474   10.1101/gad.260703.

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

478   **Torres-Galea P, Hirtreiter B, Bolle C. 2013.** Two GRAS proteins, SCARECROW-LIKE21
479   and PHYTOCHROME A SIGNAL TRANSDUCTION1, function cooperatively in
480   phytochrome A signal transduction. *Plant Physiol.* 161:291–304 DOI
481   10.1104/pp.112.206607.

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

486   **Wang T, Pan H, Wang J, Yang W, Cheng T, Zhang Q. 2014.** Identifcation and profling of
487   novel and conserved microRNAs during the flower opening process in Prunus mume via
488   deep sequencing. *Molecular Genet and Genomics* 289:169–183 DOI 10.1007/s00438-013-
489   0800-6.

490   **Llave C, Xie Z, Kasschau KD, Carrington JC. 2002.** Cleavage of Scarecrow-like mRNA
491   targets directed by a class of Arabidopsis miRNA, *Science* 297 2053–2056 DOI
492   10.1126/science.1076311.

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

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

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

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

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

510   **Tamura K, Peterson D, Peterson N, Stecher G, Nei M, Kumar S. 2011.** MEGA5: molecular
511   evolutionary genetics analysis using maximum likelihood, evolutionary distance, and
512   maximum parsimony methods. *Molecular Biology and Evolution* 28:2731–2739 DOI
513   10.1093/molbev/msr121.

514 **Bailey TL, Williams N, Misleh C, Li WW. 2006.** MEME: discovering and analyzing DNA and
515   protein sequence motifs. *Nucleic Acids Research* 34:W369–W373 DOI 10.1093/nar/gkl198.
516 **Guo A, Zhu Q, Chen X, Luo J. 2007.** GSDS: a gene structure display server. Yi Chuan
517   29:1023-1026.
518 **Li L, Stoeckert CJ Jr, Roos DS. 2003.** OrthoMCL: identification of ortholog groups for
519   eukaryotic genomes. *Genome Research* 13:2178–2189
520   DOI 10.1101/gr.1224503.
521 **Tang H, Bowers JE, Wang X, Ming R, Alam M, Paterson AH. 2008.** Synteny and collinearity
522   in plant genomes. *Science* 320:486–488
523   DOI 10.1126/science.1153917.
524 **Liu C, Wang X, Xu Y, Deng X, Xu Q. 2014.** Genome-wide analysis of the R2R3-MYB
525   transcription factor gene family in sweet orange (Citrus sinensis). *Molecular Biology*
526   *Reports* 41:6769–6785 DOI 10.1007/s11033-014-3563-1.
527 **Cannon SB, Mitra A, Baumgarten A, Young ND, May G . 2004.** The roles of segmental and
528   tandem gene duplication in the evolution of large gene families in Arabidopsis thaliana.
529   *BMC Plant Biology* 4:10 DOI **1**0.1186/1471-2229-4-10.
530 **Krzywinski M, Schein J, Birol I, Connors J, Gascoyne R, Horsman D, Jones SJ, Marra MA.**
531   **2009.** Circos: an information aesthetic for comparative genomics, *Genome Research*
532   19:1639–1645 DOI 10.1101/gr.092759.109.
533 **Deng WK, Wang YB, Liu ZX, Cheng H, Xue Y. 2014.** HemI: a toolkit for illustrating
534   heatmaps. *PLoS ONE* 9:e111988 DOI 10.137/journal.pone.0111988.
535 **Song XM, Liu TK, Duan WK, Ma QH, Ren J, Wang Z, Li Y, Hou XL. 2014.** Genome-wide
536   analysis of the GRAS gene family in Chinese cabbage (Brassica rapa ssp. Pekinensis).
537   Genomics 103:135–146
538   DOI 10.1016/j.ygeno.2013.12.004.
539 **Liu X, Alex W. 2014.** Genome-wide Comparative Analysis of the GRAS Gene Family in
540   Populus,Arabidopsis and Rice. *Plant Molecular Biology Reports* 32:1129 DOI
541   10.1007/s11105-014-0721-5.
542 **Lee MH, Kim B, Song SK, Heo JO, Yu NI, Lee SA, Kim M, Kim DG, Sohn DO, Lim CE,**
543   **Chang KS, Lee MM, Lim J. 2008.** Large-scale analysis of the GRAS gene family in
544   Arabidopsis thaliana. *Plant Molecular Biology* 67:659–670
545   DOI 10.1007/s11103-008-9345-1.
546 **Tang K, Dong C, Liu J. 2016.** Genome-wide analysis and expression profiling of the
547   phospholipase D gene family in Gossypium arboretum. *Life Sciences* 59:130-141 DOI
548   10.1007/s11427-015-4916-2.
549 **Chen F, Mackey AJ, Vermunt JK, Roos DS.2007.**Assessing performance of orthology
550   detection strategies applied to eukaryotic genomes. *PLoS ONE* 2:e383 DOI
551   10.1371/journal.pone.0000383.

552   **Wang Y, Shi S, Zhou Y, Zhou Y, Yang J, Tang X. 2016.** Genome-wide identification and
553   characterization of GRAS transcription factors in sacred lotus (Nelumbo nucifera). *PeerJ*
554   4:e2388 DOI 10.7717/peerj.2388.

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

558