

RESEARCH ARTICLE

# Genome-wide characterization of *GRAS* family genes in *Medicago truncatula* reveals their evolutionary dynamics and functional diversification

Hailing Zhang<sup>1,2,3</sup>, Yingping Cao<sup>3,4</sup>, Chen Shang<sup>2</sup>, Jikai Li<sup>2</sup>, Jianli Wang<sup>2</sup>, Zhenying Wu<sup>3,4</sup>, Lichao Ma<sup>3,4</sup>, Tianxiong Qi<sup>3,4</sup>, Chunxiang Fu<sup>3,4</sup>, Zetao Bai<sup>3,4\*</sup>, Baozhong Hu<sup>1,5\*</sup>

**1** College of Life Science, Northeast Agricultural University, Harbin, China, **2** Pratacultural Sciences Institute of Heilongjiang Academy of Agricultural Sciences, Harbin, Heilongjiang, China, **3** Key Laboratory of Biofuels, Shandong Provincial Key Laboratory of Energy Genetics, Qingdao Institute of Bioenergy and Bioprocess Technology, Chinese Academy of Sciences, Qingdao, China, **4** Qingdao Engineering Research Center of Biomass Resources and Environment, Qingdao Institute of Bioenergy and Bioprocess Technology, Chinese Academy of Sciences, Qingdao, China, **5** Harbin University, Harbin, China

\* [baizt@qibebt.ac.cn](mailto:baizt@qibebt.ac.cn) (ZB); [bzhu@neau.edu.cn](mailto:bzhu@neau.edu.cn) (BH)



**OPEN ACCESS**

**Citation:** Zhang H, Cao Y, Shang C, Li J, Wang J, Wu Z, et al. (2017) Genome-wide characterization of *GRAS* family genes in *Medicago truncatula* reveals their evolutionary dynamics and functional diversification. PLoS ONE 12(9): e0185439. <https://doi.org/10.1371/journal.pone.0185439>

**Editor:** Hong Luo, Clemson University, UNITED STATES

**Received:** June 14, 2017

**Accepted:** September 12, 2017

**Published:** September 25, 2017

**Copyright:** © 2017 Zhang et al. This is an open access article distributed under the terms of the [Creative Commons Attribution License](https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

**Data Availability Statement:** All relevant data are within the paper and its Supporting Information files.

**Funding:** The work was supported by the “100-Talent Program of the Chinese Academy of Sciences” foundation, the National Science Foundation for Fostering Talents in Basic Research of China (J1210069), Postdoc Youth Elite Project in Helongjiang province (LBH-TZ1209) and the National Natural Science Foundation of Shandong province in China (ZR201702170122). The funders

## Abstract

The *GRAS* gene family is a large plant-specific family of transcription factors that are involved in diverse processes during plant development. *Medicago truncatula* is an ideal model plant for genetic research in legumes, and specifically for studying nodulation, which is crucial for nitrogen fixation. In this study, 59 *MtGRAS* genes were identified and classified into eight distinct subgroups based on phylogenetic relationships. Motifs located in the C-termini were conserved across the subgroups, while motifs in the N-termini were subfamily specific. Gene duplication was the main evolutionary force for *MtGRAS* expansion, especially proliferation of the LISCL subgroup. Seventeen duplicated genes showed strong effects of purifying selection and diverse expression patterns, highlighting their functional importance and diversification after duplication. Thirty *MtGRAS* genes, including *NSP1* and *NSP2*, were preferentially expressed in nodules, indicating possible roles in the process of nodulation. A transcriptome study, combined with gene expression analysis under different stress conditions, suggested potential functions of *MtGRAS* genes in various biological pathways and stress responses. Taken together, these comprehensive analyses provide basic information for understanding the potential functions of *GRAS* genes, and will facilitate further discovery of *MtGRAS* gene functions.

## Introduction

Transcription factors play important roles in regulating various plant development and physiological processes. The plant-specific *GRAS* gene family has been studied in nearly 30 plant species from more than 20 genera [1, 2]. Evolutionary analyses have suggested that the *GRAS* gene family possibly originated from bacteria through horizontal gene transfer [3]. The name

have participated in decision to publish and preparation of the manuscript.

**Competing interests:** The authors have declared that no competing interests exist.

GRAS is derived from the first three transcription factors identified in this family: gibberellic-acid insensitive (*GAI*) [4], repressor of *GAI* (*RGA*) [5], and scarecrow (*SCR*) [6]. GRAS proteins typically contain 400–770 amino acids [7]. Their C-terminal regions are highly conserved, and include several ordered motifs, namely leucine heptad repeat I (LHRI, LRI), VHIID, leucine heptad repeat II (LHRII, LRII), PFYRE, and SAW, which are crucial in interactions between GRAS and other proteins [8]. In *Arabidopsis*, mutation of the SAW and PFYRE motifs in *SCR1* and *RGA* proteins resulted in distinct phenotypic variation [5]. However, the length and sequence of N-terminal regions in GRAS proteins are more divergent, indicating their potential function in protein specificity [1].

Many studies have defined categories of GRAS proteins [1, 9–13]. According to a study in *Arabidopsis thaliana* and rice (*Oryza sativa* L.), the GRAS family is divided into 8 sub-branches, including LISCL, PAT1, SCL3, DELLA, SCR, SHR, LAS, and HAM [14]. However, in other studies in poplar (*Populus trichocarpa* L.), bean (*Ricinus communis* L.), and tomato (*Solanum lycopersicum* L.), the number of distinct clades ranged from 8 to 13 [12, 15, 16]. A large number of GRAS genes have been functionally characterized in various species [17–21]. In lily (*Lilium longiflorum* L.), *LiSCL* (LISCL subfamily) participated in microsporogenesis of anthers [22]. In *Arabidopsis*, *SCL14* (LISCL subfamily) interacted with TGA transcription factors to activate a broad-spectrum detoxification network [23]. Three *Arabidopsis* genes, *PAT1*, *SCL5*, and *SCL21* (PAT1 subfamily), are positive regulators in phytochrome-A signal transduction [24, 25], whereas *SCL13* (PAT1 subfamily) is mainly involved in phytochrome-B signal transduction [26]. *GAI*, *RGA*, and *RGL* (DELLA subfamily) proteins mainly function as repressors of gibberellin signaling [19]. *SCR* and *SHR* form a *SCR/SHR* complex, which plays an essential role in root radial patterning [27]. *SCL3* acts downstream of the GA/DELLA and *SCR/SHR* pathways, and mediates cell elongation during root development [28]. The role of *MOC1*, *LS*, and *LAS* (LAS subfamily) in axillary meristem initiation has been validated [29–31]. The *microRNA171* (*miR171*) family is one of the most ancient and well conserved miRNA families which have diverse roles in plant development, such as flowering, meristem identity, and phase transition [32, 33]. Overexpressing *miR171* had pleiotropic phenotypes including plant height, flowering time, leaf architecture, phase transitions and floral meristem determinacy [34–36]. This family is known to target the HAM genes. Three *HAM* homologs in *Arabidopsis* (*SCL6-II*, *SCL6-III*, and *SCL6-IV*) were post-transcriptionally regulated by *miRNA171* and play vital roles in the proliferation of meristematic cells [37–39]. Furthermore, the triple *scl6* mutants and overexpressing *miR171* showed similar pleiotropic phenotypes [34].

*Medicago truncatula* is an annual, diploid legume plant. Because of its small genome size, self-pollination, and a well-established transformation platform, *M. truncatula* is an ideal model for genetic studies of legumes. The genome sequence of *M. truncatula* was released in 2011 and was recently updated, which provides the opportunity for gene family analysis on a genomic level [40]. To absorb nitrogen, leguminous plants have established a symbiotic relationship with nitrogen fixing rhizobial bacteria, forming special lateral organs called nodules. Certain types of transcription factors have been reported to play crucial roles in regulating the nodulation including GRAS, AP2/ERF, and NF-Y [41, 42]. *Medicago truncatula* can serve as a model for the molecular mechanism mediating nodulation, which is very important for understanding nitrogen acquisition and fixation in legumes [43]. Recent studies have suggested that GRAS proteins are involved in initiating nodulation [2]. *NSP1* and *NSP2*, two GRAS proteins in *M. truncatula*, play an essential role in nodule morphogenesis, serving as possible regulators of Nod-factor-inducible gene expression [44]. To explore the functional role of GRAS genes, especially in the process of nodulation, we performed a genome-wide study of the entire GRAS gene family in *M. truncatula*.

Genomic analyses of the *GRAS* gene family have been conducted in various species including *Arabidopsis* [1], rice [1], Chinese cabbage (*Brassica rapa* ssp. *pekinensis*) [9], tomato [12], castor beans [16] and grapevine (*Vitis vinifera* L.) [10], but have not been explored in legumes. In this study, we systematically and comprehensively analyzed *GRAS* genes in *M. truncatula* using comparative genomic strategies and experimental validation. The aims of this study were as follows: (1) identify and classify *GRAS* genes in *M. truncatula*; (2) explore the evolutionary dynamics of *MtGRAS* gene proliferation and uneven distribution; and (3) determine the functional diversity of *MtGRAS* genes by structure conservation analysis and expression profile analysis in different tissues and stress treatments. These findings provide insights into the molecular functions of *MtGRAS* genes, and will be helpful for future functional characterization of *GRAS* genes in legumes.

## Materials and methods

### Identification of *MtGRAS* genes

The current genome sequence and annotation files (Mt4.0v1) of *M. truncatula* were downloaded from Phytozome (<https://phytozome.jgi.doe.gov/pz/portal.html>). The most updated Hidden Markov Model (HMM) for the *GRAS* gene family, PF03514.11, was downloaded from the Pfam database (<http://pfam.xfam.org>). Using PF03514.11 as a query, we conducted a BLAST search against the entire protein dataset of *M. truncatula* with a cut-off E-value of  $1e-10$  using the blastall v2.2.18 package. Subsequently, all hit protein sequences were extracted using custom Perl scripts. Then, the integrity of the *GRAS* domain was evaluated using SMART tools [45], and candidate *MtGRAS* proteins composed of a truncated *GRAS* domain were identified. To obtain the gene structure, a GFF3 annotation file involving precise position information of introns and exons in each *MtGRAS* was retrieved from the genomic dataset, and uploaded to the Gene Structure Display Server (<http://gsds.cbi.pku.edu.cn/>) [46]. Peptide length, molecular weight, and isoelectric point of each *MtGRAS* protein were calculated using the online ExPasy program (<http://www.expasy.org/>) [47].

### Classification and conservation analysis of *MtGRAS* genes

The identified *MtGRAS* proteins were combined with the well-classified *Arabidopsis* and rice *GRAS* proteins and aligned using ClustalW [48]. Then, a phylogenetic tree was constructed in MEGA5 software using the neighbor-joining method with 1000 bootstrap replicates [49]. The phylogenetic tree was visualized using Evolview (<http://www.evolgenius.info/evolview/>) [50]. The *MtGRAS* genes were further categorized into different subgroups according to homology with *GRAS* genes in *Arabidopsis* and rice. The conserved motif analysis of *MtGRAS* was conducted using the motif finding tool, MEME (Multiple EM for Motif Elicitation, v4.10.0) with 20 motif numbers, and the order of motifs in each *MtGRAS* was evaluated by MAST [51]. The targets of *miR171* were predicted *in silico* using the website (<http://plantgrn.noble.org/psRNATarget/>). The *Pv-miR171* genes were identified based on the homology searching stem-loop sequence of *osa-miR171*, which obtained from the website of miRbase (<http://www.mirbase.org/index.shtml>).

### Chromosomal distribution and gene duplication analysis of *MtGRAS* genes

Physical positions of *MtGRAS* genes were retrieved from the GFF3 annotation file using a Perl script, and diagrams of their chromosomal locations and duplication events were drawn using Circos software (<http://circos.ca/>) [52]. Homologous gene pairs were defined as having protein

similarity of more than 70% and coverage greater than 75%. In addition, gene duplication information was also identified based on public data in the Plant Genome Duplication Database (PGDD, <http://chibba.agtec.uga.edu/duplication/>) [53]. If two homologous genes were separated by five or fewer genes, they were identified as tandem duplications (TD), while if two genes were separated by more than five genes or distributed in different chromosomes, they were referred to as segmental duplications (SD). To determine the evolutionary pressure acting on duplicated genes,  $K_a$  (non-synonymous substitution) and  $K_s$  (synonymous substitution) values were calculated using KaKs\_Calculator 2.0 [54].

### Expression profile analysis of *MtGRAS* genes using RNA-seq

Illumina single read sequencing data for the transcriptome of *M. truncatula* were obtained from the NCBI Short Read Archive (accession numbers SRX099059-SRX099062). This dataset contained six different tissues including root, flower, bud, seedpod, blade, and nodule. We aligned all reads from each tissue to the reference genome of *M. truncatula* (Mt4.0v1) using tophat v2.1.0 [55]. Subsequently, the expression level for each *MtGRAS* was measured using Cufflink v2.1.1 [56], and the FPKM (fragment per kilobase per million mapped reads) representing the gene expression level of each *MtGRAS* was extracted with custom Perl scripts. A heatmap of the *MtGRAS* expression profile was created using Mev v4.8.1 [57].

### Biotic and abiotic stress treatments

For hormone treatment, three-week-old *M. truncatula* (cv. Jemalong A17) seedlings were soaked in liquid MS medium with 30  $\mu$ M gibberellin (GA3). For cold treatment, seedlings were grown in a greenhouse (12/12h photoperiod, 18–24°C) and transferred to a cold chamber maintained at 4 $\pm$ 1°C. For salt treatment, 200 mM NaCl was sprayed on the leaves. Seedlings soaked in liquid MS medium without any treatment were used as control. Whole plants were sampled at 3h and 6h after treatment. For each treatment, six randomly chosen seedlings were pooled together to form a biological replicate. All plant samples were frozen in liquid nitrogen and stored at -80°C until use.

### Expression levels of *MtGRAS* genes under stress treatments

Total RNA was extracted from control and stress-treated samples using Trizol reagent (Invitrogen) based on the manufacturer's instructions. cDNA was synthesized using approximately 2  $\mu$ g of RNA according to the manufacturer's protocol. Real-time quantitative PCR (qRT-PCR) was performed using SYBR Green mix (TaKaRa) on a LightCycler480 Real-Time PCR Detection System (Roche). The fold-change of expression was calculated with *ACTIN* as the internal reference gene. All the primers for qRT-PCR are listed in S1 Table.

## Results

### Identification and structural analysis of *MtGRAS* genes

Using PF03154.11 as a query, we identified 59 *GRAS* genes in *M. truncatula*. Most of these genes contained a complete *GRAS* domain except five (*MtGRAS6*, 23, 25, 26, and 49). To further elucidate the cause and consequences of *MtGRAS* gene expansion, we selected all the genes identified for further analysis, and named them based on their distribution and linear order on the respective chromosomes (Table 1). The peptide length of *MtGRAS* varied greatly ranging from 69 amino acids (*MtGRAS25*) to 1,155 amino acids (*MtGRAS55*). Nearly 52 (88%) *MtGRAS* genes were intronless, which is consistent with most previous studies [1, 12, 58], while three members (*MtGRAS25*, 37, and 58) contained just one intron, and four members



**Table 1. Detailed information for 59 GRAS genes in the *M. truncatula* genome.**

Groups	Gene Symbol	Gene ID	ORF(aa)	Number of Exons	Molecular Weight/Da	Theoretical PI
SHR	<i>MtGRAS12</i>	<i>Medtr2g089100</i>	458	1	51691.23	5.58
	<i>MtGRAS21</i>	<i>Medtr2g099110</i>	452	1	51905.63	5.3
	<i>MtGRAS27</i>	<i>Medtr3g053270</i>	448	1	50729.91	6.22
	<i>MtGRAS40</i>	<i>Medtr4g095500</i>	470	1	52854.85	5.78
	<i>MtGRAS41</i>	<i>Medtr4g097080</i>	504	1	57826.3	5.09
	<i>MtGRAS47</i>	<i>Medtr5g015490</i>	491	1	55787.56	5.34
	<i>MtGRAS48</i>	<i>Medtr5g015950</i>	448	1	50729.91	6.22
	<i>MtGRAS60</i>	<i>Medtr8g020840</i>	554	1	61826.68	5.76
SCR	<i>MtGRAS3</i>	<i>Medtr1g069725</i>	468	1	52996.13	5.81
	<i>MtGRAS4</i>	<i>Medtr1g086970</i>	480	1	54823.4	6.84
	<i>MtGRAS8</i>	<i>Medtr2g034250</i>	587	1	67184.17	5.23
	<i>MtGRAS9</i>	<i>Medtr2g034260</i>	586	1	67296.74	5.17
	<i>MtGRAS10</i>	<i>Medtr2g034280</i>	577	1	65847.85	5.14
	<i>MtGRAS37</i>	<i>Medtr4g076020</i>	438	2	48640.81	5.04
	<i>MtGRAS58</i>	<i>Medtr7g074650</i>	805	2	89030.24	6.1
	<i>MtGRAS59</i>	<i>Medtr7g109580</i>	567	1	65717.01	5.77
SCL3	<i>MtGRAS6</i>	<i>Medtr1g106590</i>	340	5	38454.02	5.63
	<i>MtGRAS23</i>	<i>Medtr3g022580</i>	186	3	21812.25	8.27
	<i>MtGRAS24</i>	<i>Medtr3g022830</i>	438	1	49706.32	6.61
	<i>MtGRAS25</i>	<i>Medtr3g025340</i>	69	2	7939.17	6.7
	<i>MtGRAS26</i>	<i>Medtr3g027430</i>	333	5	38417.96	9.51
	<i>MtGRAS38</i>	<i>Medtr4g076140</i>	472	1	53482.58	6.64
	<i>MtGRAS46</i>	<i>Medtr5g009080</i>	481	1	54037.85	5.52
	<i>MtGRAS2</i>	<i>Medtr1g029420</i>	592	1	65944.87	4.79
PAT1	<i>MtGRAS7</i>	<i>Medtr2g026250</i>	598	1	67703.5	5.39
	<i>MtGRAS11</i>	<i>Medtr2g082090</i>	579	1	64449.5	5.88
	<i>MtGRAS30</i>	<i>Medtr3g089055</i>	570	1	64226.14	4.93
	<i>MtGRAS45</i>	<i>Medtr4g133660</i>	554	1	61196.12	5.6
	<i>MtGRAS51</i>	<i>Medtr5g094450</i>	524	1	59328.15	5.03
	<i>MtGRAS52</i>	<i>Medtr5g097480</i>	544	1	61006.97	5.96
	<i>MtGRAS53</i>	<i>Medtr6g047750</i>	624	1	73400.28	6.28
	<i>MtGRAS55</i>	<i>Medtr7g057230</i>	1155	1	129425.86	8.5
LISCL	<i>MtGRAS13</i>	<i>Medtr2g097310</i>	640	1	73256.09	5.71
	<i>MtGRAS14</i>	<i>Medtr2g097350</i>	642	1	73409.68	5.4
	<i>MtGRAS15</i>	<i>Medtr2g097380</i>	563	8	64537.94	7.66
	<i>MtGRAS16</i>	<i>Medtr2g097390</i>	689	1	78373.19	5.38
	<i>MtGRAS17</i>	<i>Medtr2g097410</i>	743	1	84140.61	5.42
	<i>MtGRAS18</i>	<i>Medtr2g097463</i>	657	1	74536.97	5.16
	<i>MtGRAS19</i>	<i>Medtr2g097467</i>	657	1	74704.13	5.77
	<i>MtGRAS20</i>	<i>Medtr2g097473</i>	656	1	74496.99	5.63
	<i>MtGRAS32</i>	<i>Medtr4g064120</i>	628	1	71758.98	5.86
	<i>MtGRAS33</i>	<i>Medtr4g064150</i>	735	1	83871.84	5.19
	<i>MtGRAS34</i>	<i>Medtr4g064160</i>	686	1	78061.03	5.72
	<i>MtGRAS35</i>	<i>Medtr4g064180</i>	628	1	72032.35	5.86
LAS	<i>MtGRAS5</i>	<i>Medtr1g096030</i>	445	1	52766.11	5.73
	<i>MtGRAS39</i>	<i>Medtr4g077760</i>	515	1	60352.59	5.14

(Continued)

Table 1. (Continued)

Groups	Gene Symbol	Gene ID	ORF(aa)	Number of Exons	Molecular Weight/Da	Theoretical PI
HAM	<i>MtGRAS1</i>	<i>Medtr0092s0100</i>	732	1	81808.23	5.56
	<i>MtGRAS29</i>	<i>Medtr3g072710</i>	508	1	56336.98	5.71
	<i>MtGRAS31</i>	<i>Medtr4g026485</i>	625	1	70120.89	5.44
	<i>MtGRAS49</i>	<i>Medtr5g019750</i>	295	2	33956.58	8.61
	<i>MtGRAS50</i>	<i>Medtr5g058860</i>	506	1	56739.58	4.8
	<i>MtGRAS57</i>	<i>Medtr7g069740</i>	585	1	67022.34	4.72
	<i>MtGRAS61</i>	<i>Medtr8g077940</i>	500	1	60492.44	5.58
	<i>MtGRAS62</i>	<i>Medtr8g093070</i>	507	1	58326.49	4.76
DELLA	<i>MtGRAS28</i>	<i>Medtr3g065980</i>	547	1	60002.63	5.01
	<i>MtGRAS43</i>	<i>Medtr4g104020</i>	521	1	58794.28	6.43
	<i>MtGRAS54</i>	<i>Medtr7g027190</i>	674	2	74869.29	5.65
	<i>MtGRAS63</i>	<i>Medtr8g442410</i>	536	1	60329.3	4.84

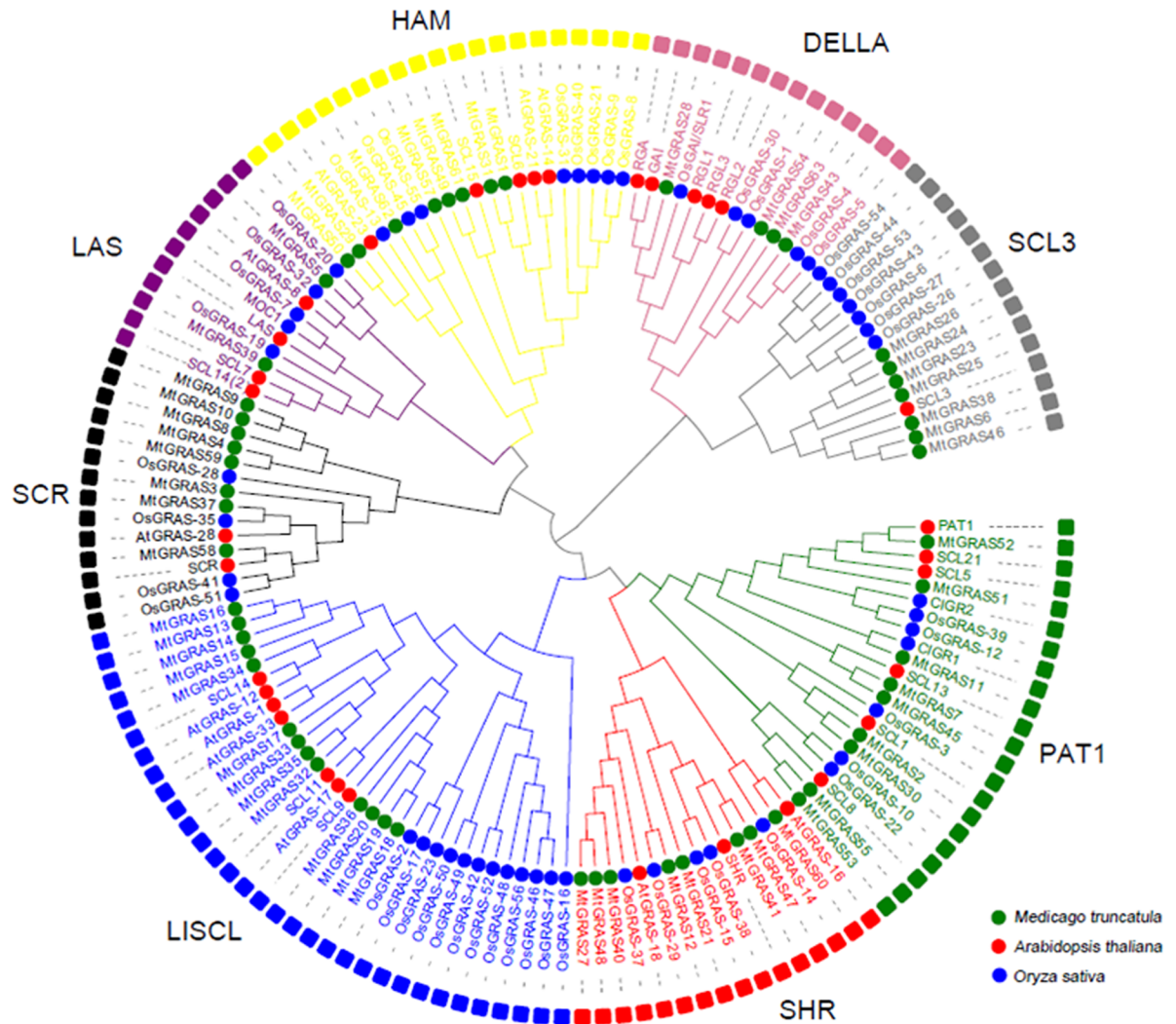
<https://doi.org/10.1371/journal.pone.0185439.t001>

(*MtGRAS6*, 15, 23, and 26) contained more than one intron (Table 1 and S1 Fig). Molecular weights were significantly different among *MtGRAS* genes, with the smallest at 7 kDa (*MtGRAS25*) and the largest at 129 kDa (*MtGRAS55*). The predicted theoretical pI ranged from 4.72 to 9.8, with a mean of 5.83 (Table 1), which indicates that most of them were weakly acidic.

### Phylogenetic categories and conserved motif analysis of *MtGRAS* genes

To fully classify the *MtGRAS* gene family, 59 *MtGRAS* genes were analyzed with 32 and 53 *GRAS* genes in *Arabidopsis* and rice, respectively, to construct an unrooted phylogenetic tree using the neighbor-joining (NJ) method in MEGA5. Eight subfamilies were identified based on clade support values, the topology of the phylogenetic tree, and the previous classification of *GRAS* families in *Arabidopsis* and rice. We identified 9, 13, 8, 8, 7, 8, 4, and 2 *MtGRAS* genes in the PAT1, LISCL, SHR, SCR, SCL3, HAM, DELLA, and LAS sub-branches, respectively (Fig 1).

To explore the potential biological functions of genes in each sub-branch, a detailed comparative analysis was performed (Fig 1). The PAT1 subfamily included nine *MtGRAS* genes including one member (*MtGRAS52*) and two members (*MtGRAS7* and *MtGRAS45*) that have high sequence similarity to *AtPAT1* and *AtSCL13*; *MtGRAS11* was also in the PAT1 subfamily and was the closest homolog of CIGR1 in rice. The LISCL subfamily consists of 13 *MtGRAS* members. Four *MtGRAS* genes (*MtGRAS36*, 20, 19, and 18) share high homology with *AtSCL9*. In addition, another five members (*MtGRAS34*, 15, 14, 13, and 16) had high similarity with *AtSCL14*, demonstrating that they may function in stress-related processes [59, 60]. The SCR and SHR subfamilies are crucial for stem cell maintenance that occurs during root and shoot development [27]. In our study, two homologous genes (*MtGRAS47* and *MtGRAS41*) of SHR were identified, and one gene (*MtGRAS58*) shared high similarity with SCR. *SCL3* regulates root cell elongation by integrating multiple signals in *Arabidopsis* [28]. Seven *MtGRAS* genes (*MtGRAS26*, 24, 23, 25, 38, 6 and 46) belonged to the SCL3 sub-branch, implying a similar function in root development. *MtGRAS29*, which belonged to the HAM subgroup, has been suggested to participate in nodule morphogenesis [2]. *MtGRAS50*, the closest paralog of *MtGRAS29*, may function in the same pathway as well. In the DELLA subfamily, four members (*MtGRAS54*, 63, 43, and 28) had the highest similarity to *RGA* and *GAI*. Two *MtGRAS*



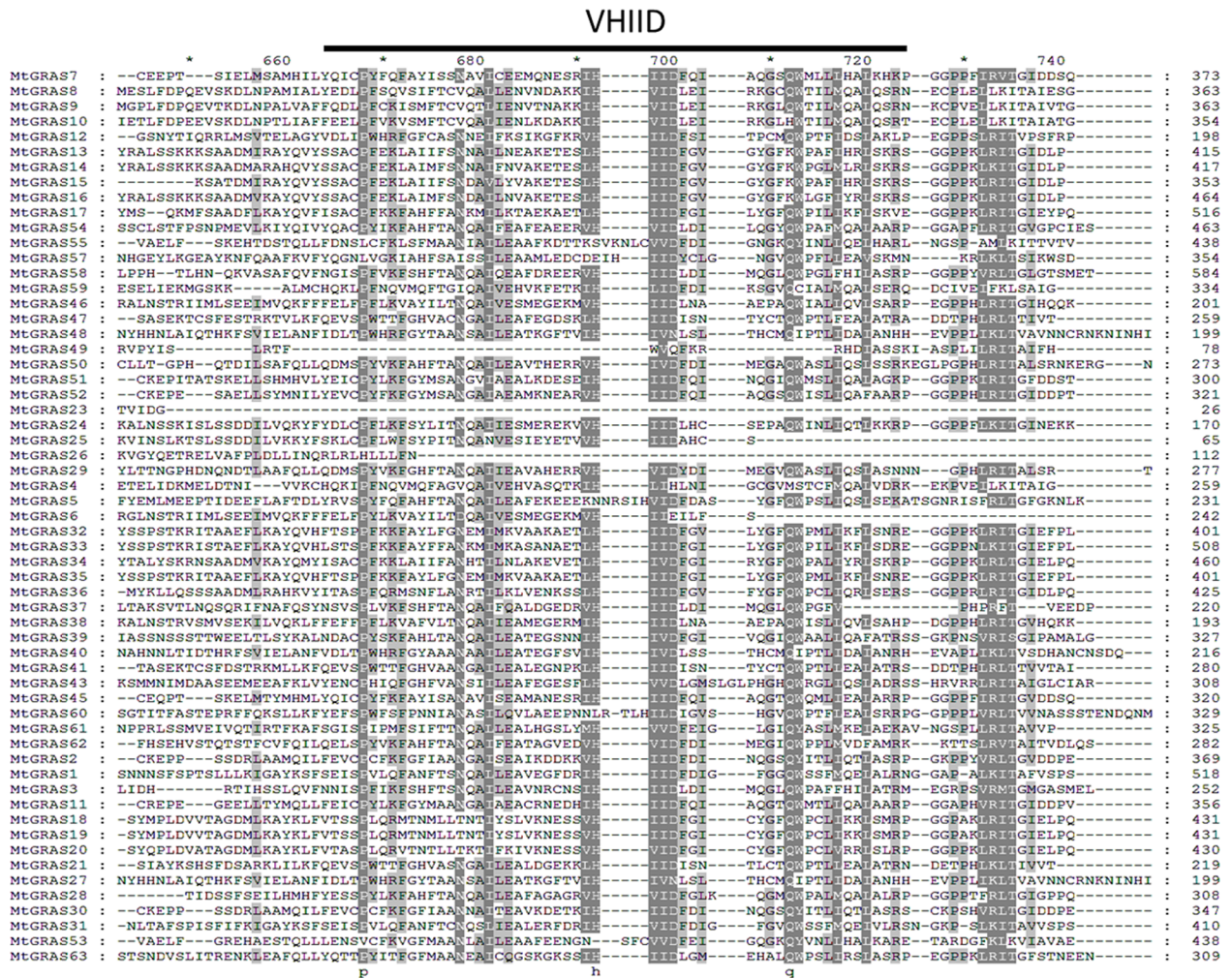
**Fig 1. Phylogenetic tree constructed using GRAS proteins from *Medicago truncatula*, *Arabidopsis thaliana*, and *Oryza sativa*.** The phylogenetic tree was constructed using MEGA5. Members in the same clade are marked by the same color.

<https://doi.org/10.1371/journal.pone.0185439.g001>

members (*MtGRAS5* and 39) were identified as part of the LAS subgroup, which has several genes that have been found to regulate meristem formation [29–31].

Using multiple sequence alignment, the characteristic conserved domains located in the C-termini were identified including VHIID, LHRI, LHRII, PFYRE, and SAW (Fig 2 and S2–S5 Figs). We further explored conserved motifs in *MtGRAS* using MEME tools [51]. In total, 20 conserved motifs were found, and most of them had a similar distribution within the same subgroup (Fig 3). The logo of these motifs is listed in S6 Fig. The motifs located in the GRAS domain, including LHRI (motif6, motif9), VHIID (motif5, motif1), LHRII (motif13, motif7, motif10), PFYRE (motif4, motif11), and SAW (motif2, motif14, motif3), were shared across almost all *MtGRAS* members. In addition, motif8 was situated between LHRII and PFYRE and conserved among most *MtGRAS* subfamilies, suggesting its functional importance. Other motifs located outside the GRAS domain regions showed subgroup specific patterns. Motif16 was located between LHR1 and VHIID, and was specific to the PAT1 and LISCL subgroups,





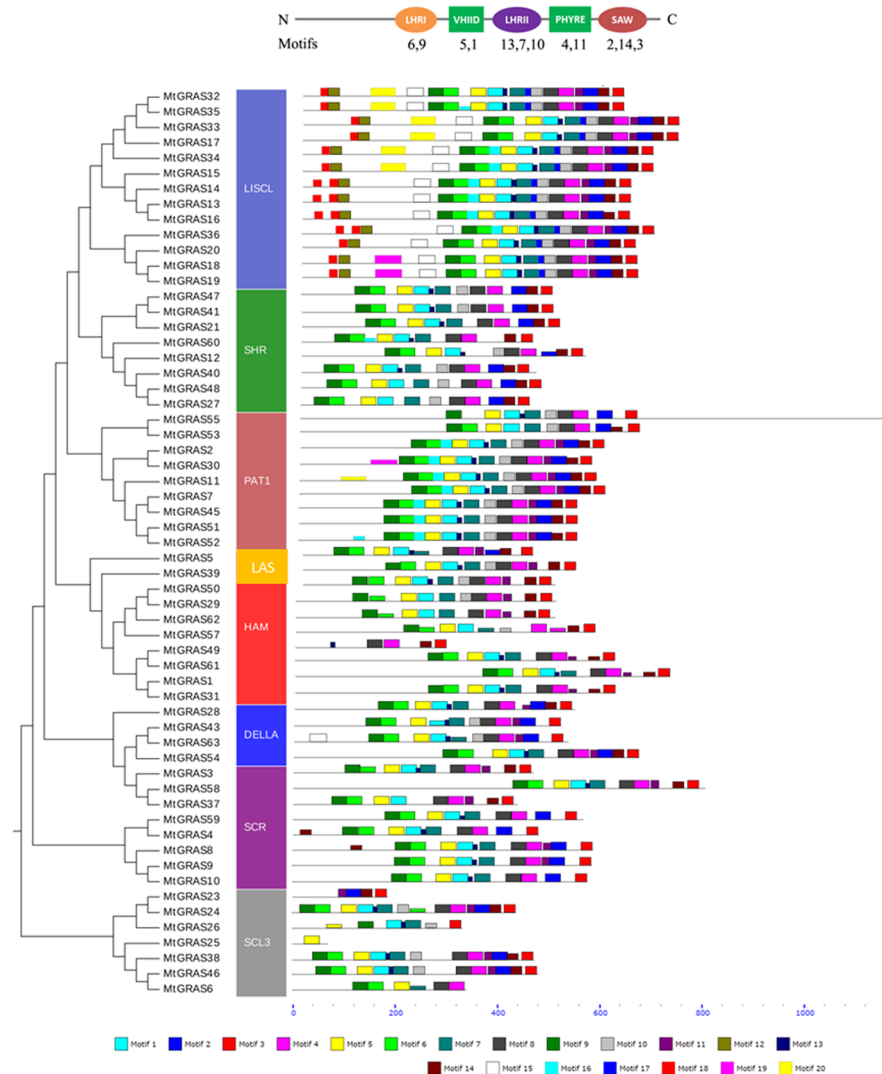
**Fig 2. Multiple sequence alignment of 59 *MtGRAS* genes by ClustalW.** The most conserved motif of VHIID is underlined with a black solid line.

<https://doi.org/10.1371/journal.pone.0185439.g002>

while motif17 was nested within LHR1 and was LISCL specific. Motif12 and motif18 were located in the N-termini and were also only present in the LISCL subgroup (Fig 3).

### Chromosomal distributions and duplication analysis of *MtGRAS* genes

Fifty-nine *MtGRAS* genes were mapped to the chromosomes of *M. truncatula*; *MtGRAS1* was excluded because it was positioned on a scaffold (Table 1). The distribution of *MtGRAS* genes among the chromosomes was uneven. Chr2 and chr4 are the “hot regions”, and contained 15 (25.9%) and 13 (22.4%) *MtGRAS* genes, respectively; chr6 is the “cold region”, and contained only one (1.7%) *MtGRAS* gene. Moreover, 5, 8, 7, 5, and 4 *MtGRAS* genes were found on chr1, chr3, chr5, chr7, and chr8, respectively (Fig 4). Based on these distributions, we explored duplication events of *MtGRAS* genes. Seventeen duplicated *MtGRAS* gene pairs were identified. Most duplication events occurred in chr2 (n = 7), chr3 (n = 6), and chr4 (n = 5); one duplication event occurred in chr1, chr5, and chr8; and no duplication events were identified in chr6 and chr7 (Fig 4). We also verified the types of duplication. The results suggested that



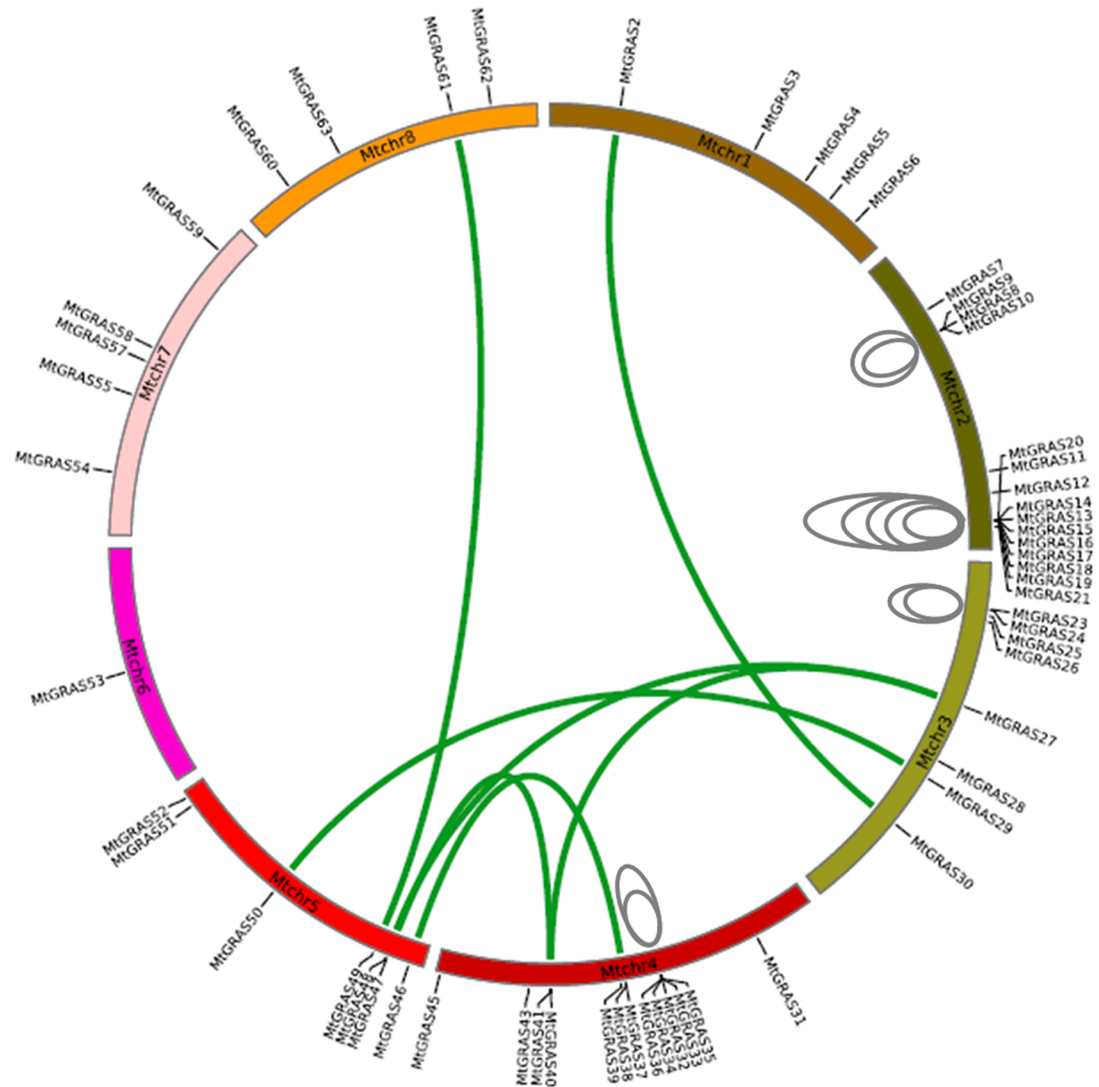
**Fig 3. The distribution of conserved motifs in MtGRAS proteins.** Neighbor-joining tree of MtGRAS proteins is shown on the left. MtGRAS proteins are categorized into eight distinct clusters including LISCL, SHR, PAT1, LAS, HAM, DELLA, SCR, and SCL3, and are represented by different colored vertical boxes. The horizontal colored boxes indicate conserved motifs within each protein. Conserved domains and corresponding motifs are shown at the top. A scale of protein length is shown at the bottom.

<https://doi.org/10.1371/journal.pone.0185439.g003>

11 duplicated gene pairs arose from tandem duplications, while six pairs were segmental duplications (Table 2).

To understand the evolutionary process of gene duplications, we evaluated the positions of duplicated genes. Seven duplicated *MtGRAS* genes (*MtGRAS9*, *14*, *15*, *13*, *19*, *24*, *2*, and *49*) were positioned near the telomeres of each chromosome, and three genes (*MtGRAS24*, *29*, and *38*) were located around the centromeres, implying that the highly repetitive nature of these regions may lead to gene duplication (Fig 4). To reveal the evolutionary dynamics of duplicated *MtGRAS* genes, we calculated nonsynonymous substitution rates ( $K_a$ ) and synonymous substitution rates ( $K_s$ ) between duplicated genes. All of the duplicated gene pairs have  $K_a/K_s$  values  $<1$ , suggesting that purifying selection acted on them (Table 2).





**Fig 4. Chromosomal distribution and expansion analysis of *MtGRAS* genes in *M. truncatula*.** Green lines show paralogous genes duplicated by segmental duplication (SD), and grey lines indicate tandem duplication (TD).

<https://doi.org/10.1371/journal.pone.0185439.g004>

### Expression pattern analysis of *MtGRAS* genes in different tissues

*GRAS* transcription factors have crucial roles in various biological pathways. In this study, we analyzed the expression profiles in different tissues including root, blade, flower, bud, and nodule using publicly available Illumina RNA-seq data. Based on the transcriptome analysis, 53,777 expressed genes were identified, including: 52,236 in blade; 51,236 in bud; 54,032 in flower; 55,041 in nodule; and 52,437 in root.

In our data, expression of most *MtGRAS* genes was identified in at least in one tissue. Expression of nine *MtGRAS* genes (*MtGRAS4*, 10, 23, 25, 26, 27, 48, 49 and 54) was not detected in our transcriptome data, which may be the result of spatial and temporal expression patterns or uncharacterized pseudogenes. The FPKM values of each *MtGRAS* are shown in [S2 Table](#) and the expression profiles were clustered across six tissues ([Fig 5](#)). Among the 50 expressed *MtGRAS* genes, 38 were highly expressed (FPKM >1), while the other 12 had low

**Table 2. Detailed information of duplicated *MtGRAS* genes.**

Groups	Gene symbol	Homolog	Protein identity	Types of duplication	Ka <sup>c</sup>	Ks <sup>d</sup>	Ka/Ks
PAT1	<i>MtGRAS2</i>	<i>MtGRAS30</i>	76.21%	SD <sup>a</sup>	0.149	0.819	0.181
LISCL	<i>MtGRAS14</i>	<i>MtGRAS16</i>	84.12%	TD <sup>b</sup>	0.081	0.092	0.882
LISCL	<i>MtGRAS15</i>	<i>MtGRAS16</i>	76.21%	TD	0.177	0.292	0.607
LISCL	<i>MtGRAS13</i>	<i>MtGRAS16</i>	85.85%	TD	0.070	0.099	0.708
LISCL	<i>MtGRAS19</i>	<i>MtGRAS18</i>	71.40%	TD	0.106	0.261	0.406
LISCL	<i>MtGRAS19</i>	<i>MtGRAS20</i>	91.62%	TD	0.040	0.085	0.467
LISCL	<i>MtGRAS32</i>	<i>MtGRAS35</i>	97.77%	TD	0.010	0.047	0.218
LISCL	<i>MtGRAS32</i>	<i>MtGRAS33</i>	78.15%	TD	0.121	0.404	0.300
SHR	<i>MtGRAS27</i>	<i>MtGRAS40</i>	69.18%	SD	0.206	1.115	0.185
SHR	<i>MtGRAS41</i>	<i>MtGRAS47</i>	74.01%	SD	0.169	1.048	0.161
HAM	<i>MtGRAS49</i>	<i>MtGRAS61</i>	60.54%	SD	0.315	1.407	0.224
HAM	<i>MtGRAS29</i>	<i>MtGRAS50</i>	66.67%	SD	0.248	1.543	0.161
SCL3	<i>MtGRAS24</i>	<i>MtGRAS23</i>	78.38%	TD	0.122	0.333	0.367
SCL3	<i>MtGRAS24</i>	<i>MtGRAS25</i>	68.12%	TD	0.221	0.364	0.607
SCL3	<i>MtGRAS38</i>	<i>MtGRAS46</i>	77.66%	SD	0.134	0.968	0.139
SCR	<i>MtGRAS9</i>	<i>MtGRAS8</i>	83.89%	TD	0.082	0.267	0.307
SCR	<i>MtGRAS9</i>	<i>MtGRAS10</i>	82.24%	TD	0.093	0.284	0.326

<sup>a</sup> Segmental duplication.

<sup>b</sup> tandem duplication.

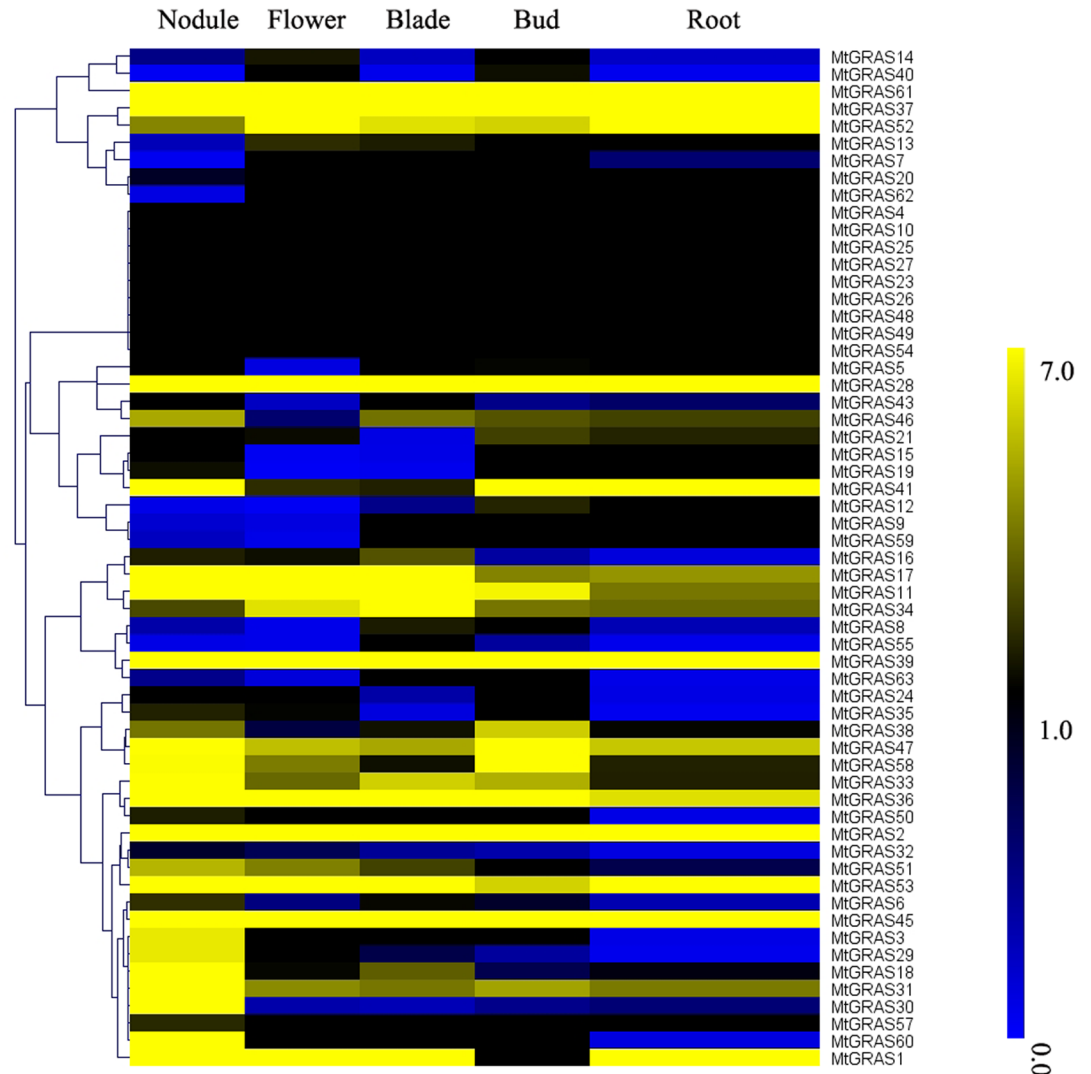
<sup>c</sup> Non-synonymous substitution rate.

<sup>d</sup> Synonymous substitution rate.

<https://doi.org/10.1371/journal.pone.0185439.t002>

expression (FPKM <1). Interestingly, the FPKM values of six genes (*MtGRAS1*, 28, 30, 45, 53, and 61) were higher than 40, indicating their important role in the development of *M. truncatula*. We also compared expression levels among different tissues. Interestingly, 30 *MtGRAS* genes had the highest expression in nodules, and 5, 7, 4, and 4 *MtGRAS* genes had the highest expression in blade, bud, flower, and root, respectively (S2 Table). The high proportion of *MtGRAS* genes expressed in nodules indicates that additional *MtGRAS* genes besides *NSP1* and *NSP2* participate in the process of nodulation. Furthermore, three *Arabidopsis* GRAS genes (*SCL6*, 22, and 27) in the HAM subfamily are post-transcriptionally regulated by *miR171*. Here, the two closest homologs of *SCL6*, *MtGRAS1* and *MtGRAS31*, were identified as having a putative binding site for *miR171* (Fig 6A). Two *Pv-miR171* genes were detected in the *Medicago* genome. The expression pattern of *Pv-miR171* genes were negatively correlated with their targets. Both *MtGRAS1* and 31 exhibited highest expression in nodules, while *Pv-miR171* genes (*Pv-miR171-1* and *Pv-miR171-2*) showed the lowest expression in this tissue. In buds, *MtGRAS1* and 31 accumulated the least transcript, but the transcript of *Pv-miR171* genes, especially *Pv-miR171-1*, was the highest among different tissues (Fig 6B).

Since duplicated genes can exhibit significant variation in gene expression, we next explored the expression divergence of 17 sets of *MtGRAS* duplicated genes. Detailed expression information of these genes is listed in S3 Table. Eleven duplicated gene pairs shared similar expression patterns with different expression levels. For example, both *MtGRAS47* and its duplicated gene *MtGRAS41*, had higher expression in bud and nodule, but showed lower expression in blade and flower (Fig 7A). This pattern was also observed in gene pairs *MtGRAS2/30*, *19/18*, *19/20*, *29/50*, *9/8*, *9/10*, *14/16*, *15/16*, *13/16* and *32/35* (S7 Fig). After the duplications, four genes were not expressed in our transcriptome data, including *MtGRAS49*



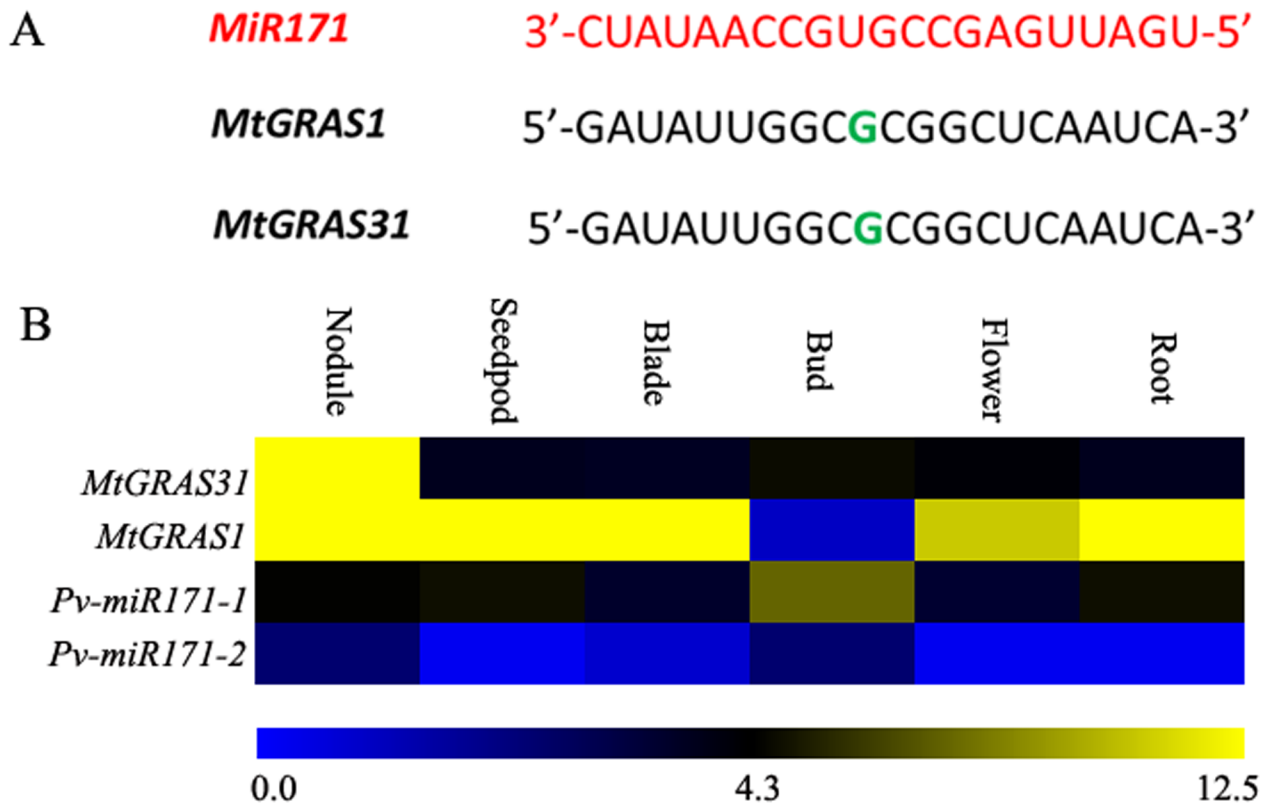
**Fig 5. Expression profile cluster analysis of MtGRAS genes.**

<https://doi.org/10.1371/journal.pone.0185439.g005>

in the gene pair 49/61, 27 in 27/40, 23 in 24/23, and 25 in 24/25 (Fig 7B and S7 Fig). For example, *MtGRAS61* was expressed in six tissues, and had the highest expression in flower, but the transcript level of the duplicated gene *MtGRAS49* was not detected (Fig 7B). Intriguingly, two duplicate gene pairs, *MtGRAS32/33* and *38/46* –showed different expression profiles (Fig 7C and 7D). In the gene pair, *MtGRAS32/33*, *MtGRAS32* had higher expression in nodule and blade, whereas *MtGRAS33* had higher expression in nodule and flower. *MtGRAS38* was highly expressed in bud, whereas the transcripts of its homologous gene *MtGRAS46* were enriched in nodules. In all, the expression differences of duplicated genes presented here implies that genes may have various evolutionary outcomes after duplication.

### Responses of MtGRAS genes to different stress treatments

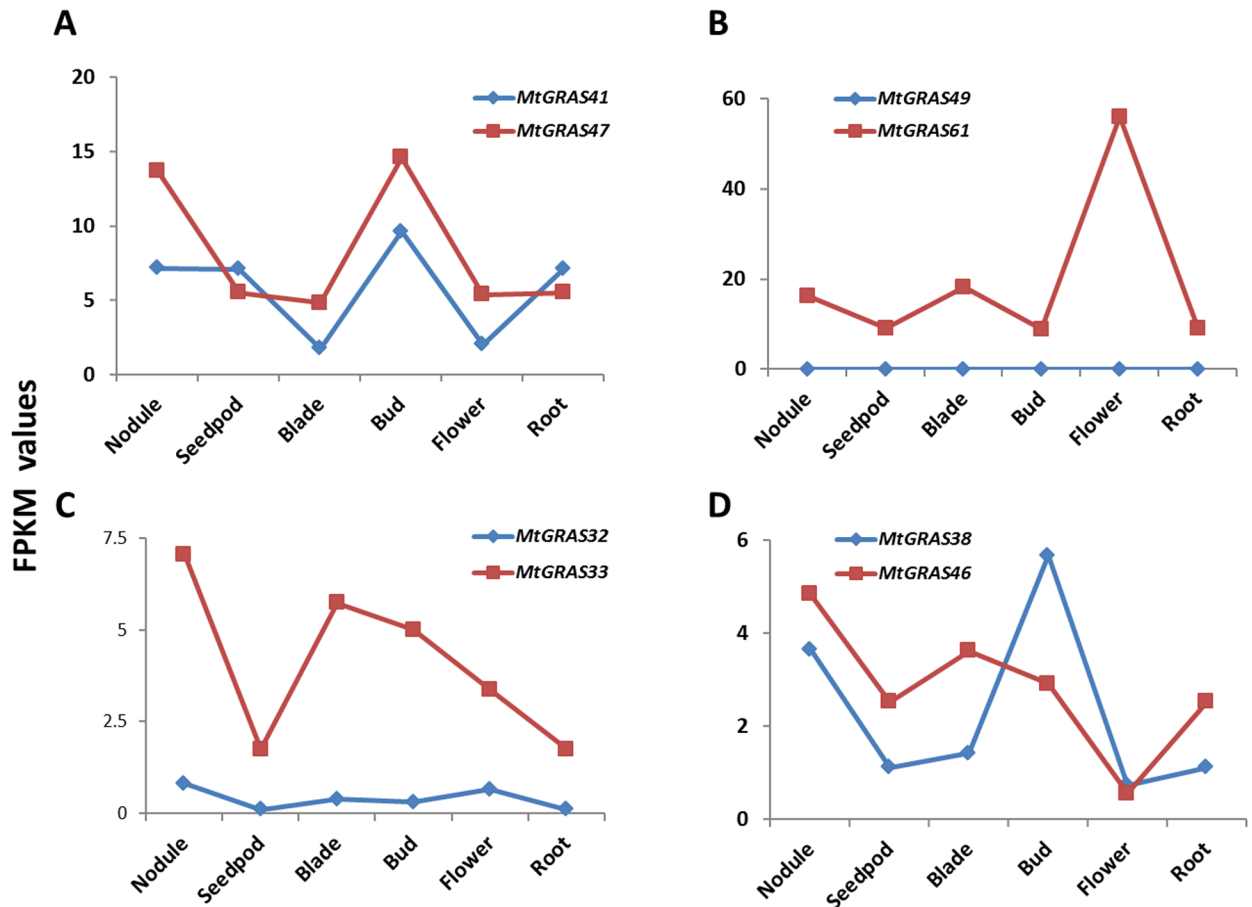
We further examined changes in transcript abundance in response to stress treatments including GA3, salt, and cold using qPCR. Twelve *MtGRAS* genes were used to assess responses to



**Fig 6. Putative *Pv-miR171* genes and their target *MtGRAS* genes.** (A) *miRNA171* cleavage sites in *MtGRAS* genes. (B) The expression patterns of *Pv-miR171* genes and their target *MtGRAS* genes in different tissues. *Pv-miR171-1*: Medtr4g111710; *Pv-miR171-2*: Medtr1g099290.

<https://doi.org/10.1371/journal.pone.0185439.g006>

treatments, including: *MtGRAS32* and *35* (LISCL subfamily); *60* and *47* (SHR subfamily); *45* and *51* (PAT1 subfamily); *50* and *61* (HAM subfamily); *37* (SCR subfamily); *39* (LAS subfamily); and *38* and *46* (SCL3 subfamily). All of these genes exhibited differential expression in response to at least one stress treatment (Fig 8). After GA3 treatment, the expression levels of most *MtGRAS* genes (10) were downregulated. Five genes (*MtGRAS35*, *51*, *39*, *38*, and *46*) had decreased expression at 3h and recovered, to some extent, at 6h. The rest of the genes (*MtGRAS47*, *45*, *50*, *61*, and *37*) had decreased transcripts after GA3 treatment, and reached the lowest expression level at 6h. Only two genes, *MtGRAS32* and *MtGRAS60*, were positively upregulated and reached the highest expression level at 6h. Under salt treatment, the transcripts of *MtGRAS51* and *37* were not changed compared to the control, which indicated that they might not participate in the response to salt stress during the development of *M. truncatula*. In the remaining 10 *MtGRAS* genes, three genes (*MtGRAS45*, *46*, and *32*) were upregulated, and reached the highest expression level at 6h; the other seven genes were clearly downregulated, including four genes (*MtGRAS60*, *50*, *39*, and *38*) that had the lowest expression at 6h, and three genes (*MtGRAS35*, *47*, and *61*) with the lowest expression at 3h. In the cold stress treatment, the transcripts of most *MtGRAS* genes were upregulated, except in two genes (*MtGRAS32* and *35*). The expression levels of six genes (*MtGRAS60*, *47*, *50*, *61*, *39*, and *38*) decreased at 3h but increased at 6h, while three genes (*MtGRAS51*, *MtGRAS37*, and *MtGRAS46*) increased in linear order and reached the highest expression level at 6h. Finally, the expression of *MtGRAS45* was highest at 3h but decreased at 6h (Fig 8).



**Fig 7. Comparative analysis of expression profiles of duplicated *MtGRAS* genes.** The X-axis represents different tissues of *M. truncatula*. The Y-axis shows the expression values (FPKM) obtained using RNA-seq data.

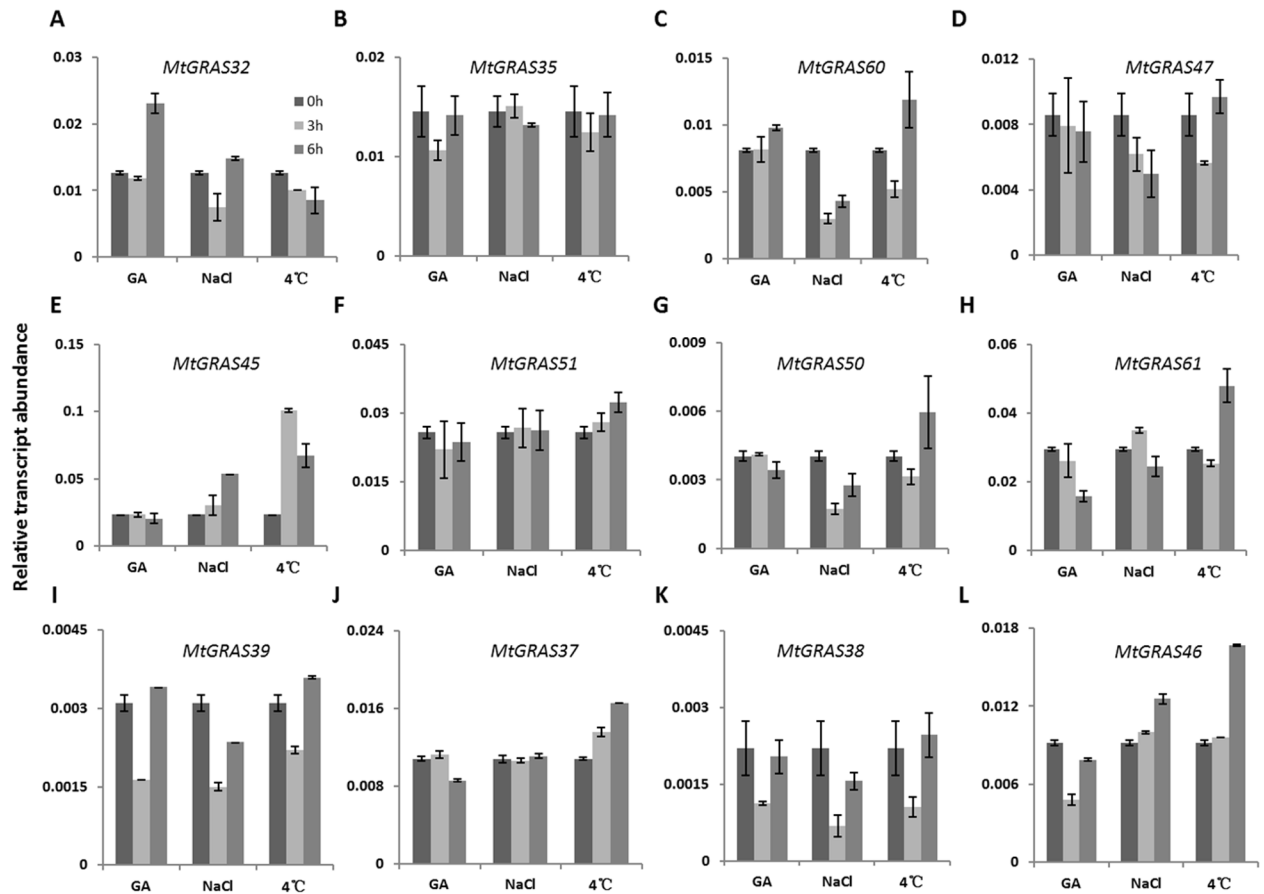
<https://doi.org/10.1371/journal.pone.0185439.g007>

## Discussion

*GRAS* transcription factors play essential roles in regulating plant growth and development. However, the prevalence and functional diversity of the *GRAS* family in *M. truncatula* have not been thoroughly investigated. In this study, we performed a comprehensive analysis of the *GRAS* gene family in *M. truncatula*. The features of *MtGRAS* genes, including their chromosomal distribution, phylogenetic classification, expression profiles, and responses to various stresses were explored. Results of these analyses allowed us to study the evolution of the *GRAS* family and draw hypotheses about the potential functions of unidentified genes.

Our results demonstrated that duplication was the evolutionary force behind *GRAS* gene family expansion. First, the number of *GRAS* was different among species. In this study, we identified 59 *GRAS* genes in *M. truncatula*, which is lower than the number in *Populus* (106) [15], but higher than in other species such as *Arabidopsis* (33) [1], rice (57) [1], Chinese cabbage (46) [9], tomato (53) [12] and grapevine (52) [10]. Gene duplication might cause these differences in numbers of *GRAS* family members. Second, *MtGRAS* genes were unevenly distributed among the chromosomes, with the “hot regions” on chr2 (15 members) and chr4 (13 members), and the “cold region” on chr6 (1 member). Interestingly, more duplication events were found in the hot regions (7/chr2, 4/chr4) (Fig 4). We further related the





**Fig 8. Relative expression levels of *MtGRAS* genes in different abiotic stress treatments.** Three-week-old *M. truncatula* (cv. Jemolong A17) seedlings were subjected to various stress treatments including GA3 gibberellin (30  $\mu$ M), cold (4°C), and salt (200 mM). The expression level was measured at 0h, 3h and 6h after treatment. Error bars show the standard error of three replicates.

<https://doi.org/10.1371/journal.pone.0185439.g008>

duplication events to chromosome positions, and found that duplicated *MtGRAS* gene pairs tend to be located in regions with low levels of conservation (10 members), such as peritelomeres and pericentromeres; tandem and segmental duplication were enriched in these regions because of the composition of repetitive elements (Fig 4). Furthermore, among the 17 duplicated *MtGRAS* gene pairs, seven belonged to the LISCL subfamily and clustered in chr2 and chr4, suggesting that gene duplication might cause proliferation of the LISCL subfamily. Previous studies have found that duplication is common in the *GRAS* gene family. For example, 2/34, 15/53, 17/60, and 40/106 *GRAS* genes were identified as duplicated genes in *Arabidopsis* [1], tomato [12], rice [1], and *Populus* [15], respectively, which further validates the contribution of duplication to expansion of the *GRAS* gene family. Third, nearly 88% of *MtGRAS* genes were intronless, consistent with other species such as tomato (77.4%) [12], *Prunus mume* (82.2%) [58], and *Arabidopsis* (67.6%) [1]. Intronless genes have also been discovered in other large gene families, such as DEAD box RNA helicase [61] and F-box transcription factors [62]. Rapid duplication after horizontal gene transfer from bacteria is the main contributor to the high proportion of intronless genes.

To our knowledge, transcription factors belonging to the same taxonomic clade exhibit recent evolutionary origins and specific conserved motifs associated with functional specification. Because of this, a comparison of homologous genes in the *MtGRAS* family, including

protein sequences and expression profiles, would be an effective method to predict the function of uncharacterized genes. In this study, 50 *MtGRAS* genes were expressed in at least one tissue according to our transcription analysis, and the expression patterns varied across a variety of tissues, as previously reported in *Populus*, *Prunus mume*, tomato, and grapevine. Noticeably, more than half (30/59) of the *MtGRAS* genes had the highest expression in the nodule, and 5, 7, 4, and 4 members were preferentially expressed in blade, bud, flower, and root, respectively. This result suggested that the functions of *MtGRAS* have dramatically diverged. Genes belonging to the LAS subfamily have been found to participate in regulating axillary meristem development. For example, the mutation of *MOC1* resulted in the phenotype of no tillers except for a main culm in rice [29]. In an *Arabidopsis* knockout, the homolog of *MOC1*, named *AtLAS*, led to an inability to form lateral shoots during vegetative development [31]. In *M. truncatula*, two *MtGRAS* genes (*MtGRAS5* and *MtGRAS39*) belonged to the LAS subgroup. Interestingly, both of them were expressed highest in bud, which indicated that they might also play a vital role in axillary meristem formation (Fig 5 and S2 Table). The *DLT* gene in rice and its orthologs, *AtSCL28* in *Arabidopsis* and *VviGRAS8a* in *Vitis vinifera*, modulate the expression of a brassinosteroid-related gene [59, 63]. In tomato, the ortholog of *DLT* was validated to be involved in the flower-fruit transition by mediating brassinosteroid signaling [12]. In our study, *MtGRAS12* was the homolog of *DLT*, and was preferentially highly expressed in bud, but had low expression in flower (Fig 5 and S2 Table). The result suggested that *MtGRAS12* might function in response to brassinosteroid signaling during bud development. DELLA genes participate in various developmental processes including flower development, stem elongation, and seed germination [28]. In addition, the DELLA proteins also participate in hormone signaling pathways, such as the gibberellin, jasmonate, auxin, brassinosteroid, and ethylene pathways [64]. In our data, the closest homolog of RGA and GAI, *MtGRAS28*, was highly expressed in different tissues (FPKM>40) including root, seedpod, and blade, supporting a role in diverse developmental processes (Fig 5 and S2 Table).

Generally, the evolutionary fate of duplicated genes includes nonfunctionalization, neofunctionalization, or subfunctionalization [65, 66]. We further evaluated the evolutionary dynamics and consequences of duplicated *MtGRAS* genes. All of the duplicated genes were under purifying selection ( $K_a/K_s < 1$ ), implying that these genes were still strongly controlled after duplication (Table 2). We next examined the divergence of expression in 17 sets of duplicated *MtGRAS* gene pairs. Eleven duplicated genes showed similar expression patterns to the original gene, but with different expression levels; four duplicated genes were not expressed in our transcription dataset (Fig 7 and S3 Fig). Furthermore, the duplicated gene pairs *MtGRAS38/46* and *MtGRAS32/33* exhibited different expression patterns, suggesting that novel functions might evolve after duplication (Fig 7). Further efforts need to be made to elucidate the functional diversity of duplicated genes.

Previous studies have demonstrated that the GRAS protein could interact with ERN to regulate gene expression during rhizobial infection [67]. Two *MtGRAS* genes, *NSP1* and *NSP2*, belonging to SHR and HAM, respectively, were associated with enhancing Nod factor elicitation [2]. In our study, more than half (30/50) of the *MtGRAS* genes had the highest expression in nodules, which implied that *MtGRAS* genes other than *NSP1* and 2 might participate in the process of nodulation. In *Arabidopsis*, three GRAS members in the HAM subfamily were post-transcriptionally regulated by *miR171* (*AtSCL6*, 22, and 27). Interestingly, in the present study, the two closest homologs of *AtSCL6* (*MtGRAS1* and 31) were found to have a putative binding site for *miR171*. Both of these genes were highly expressed in nodules, especially the *MtGRAS1* gene, which had an FPKM value higher than 40 (Fig 6). These results indicated that *mi172* might also be involved in nodule development of legumes [37–39].

Numerous studies have found that transcription factors in the *GRAS* family could be influenced by various biotic and abiotic stresses. Gibberellin, auxin, brassinosteroid, abscisic acid, ethylene, and salicylic acid also play important roles in a diverse array of developmental processes including germination, flowering time, and stem elongation [64]. GAI, RGA, and RGL in the DELLA subfamily were repressors of gibberellin signaling [4–6]. Loss of function in the *Arabidopsis* mutants *scr* and *shr*, resulted in hypersensitization to abscisic acid [6]. *BnSCL1* in *Brassica napus* showed differential dose responses to auxin in shoots and roots [68]. Recently, a study in tomato demonstrated that the expression level of *GRAS* genes could be modulated by signaling of multiple phytohormones, including gibberellin, auxin, brassinosteroid, ethylene, and salicylic acid [12]. Additionally, several studies have revealed the participation of *GRAS* genes in response to abiotic stresses, such as cold, drought, salt, and heat. In *Arabidopsis*, over-expression of a poplar *GRAS* gene, *PtSCL7*, enhanced tolerance to salt and drought stress [69]. *SCL14* in *Arabidopsis* was involved in the activation of a broad-spectrum detoxification network, and its ortholog in rice, *OsGRAS23*, was involved in regulating the drought stress response [23, 59, 60]. The gene *BoGRAS* was significantly upregulated during heat stress in *Brassica oleracea* [70]. In our study, 12 *MtGRAS* genes from different subgroups were randomly selected to explore their responses to biotic (GA3) and abiotic (NaCl and 4°C) stresses. We found that nearly all *MtGRAS* genes could be affected by different stress treatments (Fig 8). Most *MtGRAS* genes (10/12) were downregulated after treatment with GA3, while only two genes were upregulated, implying that most *MtGRAS* genes had negative roles in response to this hormone. Under salt treatment, the expression levels of seven genes decreased, and the expression levels of three genes increased, suggesting that *MtGRAS* genes modulate the signaling of response to salt through complicated networks. The majority of genes (10/12) increased their expression level under the 4°C treatment, indicating they might positively regulate the response gene in the cold condition. In addition, most *MtGRAS* genes could be influenced by both hormone and abiotic stress treatments, indicating the coordinated response of these two environmental determinants.

## Conclusions

In this study, we performed a genome-wide analysis of the *GRAS* gene family in *M. truncatula* based on publicly available genome data. Fifty-nine *MtGRAS* genes were identified and categorized into eight subfamilies by phylogenetic analysis. Conserved motif analysis combined with expression profile measurement in different tissues and environmental treatments demonstrated the functional conservation and diversity of *MtGRAS* genes. The evolutionary dynamics of *MtGRAS* family members was further inferred by analyzing the cause and consequence of duplicated *MtGRAS* gene pairs. We foresee that these results will be of great value for further functional characterization of the *MtGRAS* gene family and for genetic improvements in agronomic traits or stress tolerance in legumes.

## Supporting information

**S1 Fig. Gene structure of the *MtGRAS* genes.** Pink boxes represent exons, blue arrows represent UTRs, and black lines show introns. The lengths of the exons, introns and UTRs were drawn to scale.

(TIF)

**S2 Fig. The conserved domain of LHRI was underlined by multiple sequence alignment of 59 *MtGRAS* genes.**

(TIF)

**S3 Fig. The conserved domain of LHRII was underlined by multiple sequence alignment of 59 MtGRAS genes.**

(TIF)

**S4 Fig. The conserved domain of PFYRE was underlined by multiple sequence alignment of 59 MtGRAS genes.**

(TIF)

**S5 Fig. The conserved domain of SAW was underlined by multiple sequence alignment of 59 MtGRAS genes.**

(TIF)

**S6 Fig. Conserved motifs detected in MtGRAS proteins using MEME software.** The logo represents conserved amino acids sequences in different motifs, and the heights of letters in the logo represent the frequency of amino acid at specific positions.

(TIF)

**S7 Fig. The comparative analysis of expression profiles of duplicated MtGRAS genes.** X-axis represents different tissues of *M. truncatula*. Y-axis shows the expression values (RPKM) obtained using RNA-seq data.

(TIF)

**S1 Table. The primers used in RT-PCR experiments.**

(DOCX)

**S2 Table. Expression levels of MtGRAS genes measured by transcriptome analysis.**

(DOCX)

**S3 Table. Expression pattern of duplicated MtGRAS genes.**

(DOCX)

## Acknowledgments

We thank Ms. Amy Flanagan for proofreading the manuscript.

## Author Contributions

**Conceptualization:** Chunxiang Fu.

**Data curation:** Jianli Wang, Zhenying Wu.

**Formal analysis:** Yingping Cao, Jikai Li, Tianxiong Qi.

**Funding acquisition:** Chen Shang, Jikai Li, Chunxiang Fu, Zetao Bai.

**Methodology:** Jianli Wang, Lichao Ma.

**Project administration:** Chunxiang Fu, Baozhong Hu.

**Resources:** Zhenying Wu.

**Software:** Jianli Wang, Lichao Ma, Tianxiong Qi.

**Supervision:** Zhenying Wu, Baozhong Hu.

**Validation:** Hailing Zhang.

**Visualization:** Chen Shang, Lichao Ma.

**Writing – original draft:** Zetao Bai.

Writing – review & editing: Chunxiang Fu, Zetao Bai.

## References

1. Tian C, Wan P, Sun S, Li J, Chen M. Genome-wide analysis of the GRAS gene family in rice and *Arabidopsis*. *Plant Mol Biol*. 2004; 54 (4):519–32. <https://doi.org/10.1023/B:PLAN.0000038256.89809.57> PMID: 15316287.
2. Hirsch S, Kim J, Munoz A, Heckmann AB, Downie JA, Oldroyd GED. GRAS Proteins Form a DNA Binding Complex to Induce Gene Expression during Nodulation Signaling in *Medicago truncatula*. *Plant Cell*. 2009; 21 (2):545–57. <https://doi.org/10.1105/tpc.108.064501> PMID: 19252081
3. Zhang D, Iyer LM, Aravind L. Bacterial GRAS domain proteins throw new light on gibberellic acid response mechanisms. *Bioinformatics*. 2012; 28 (19):2407–11. <https://doi.org/10.1093/bioinformatics/bts464> PMID: 22829623.
4. Peng JR, Carol P, Richards DE, King KE, Cowling RJ, Murphy GP, et al. The *Arabidopsis* GAI gene defines a signaling pathway that negatively regulates gibberellin responses. *Gene Dev*. 1997; 11 (23):3194–205. <https://doi.org/10.1101/gad.11.23.3194>. PMID: 9389651
5. Silverstone AL, Ciampaglio CN, Sun T. The *Arabidopsis* RGA gene encodes a transcriptional regulator repressing the gibberellin signal transduction pathway. *Plant Cell*. 1998; 10 (2):155–69. PMID: 9490740.
6. DiLaurenzio L, WysockaDiller J, Malamy JE, Pysh L, Helariutta Y, Freshour G, et al. The SCARECROW gene regulates an asymmetric cell division that is essential for generating the radial organization of the *Arabidopsis* root. *Cell*. 1996; 86 (3):423–33. [https://doi.org/10.1016/S0092-8674\(00\)80115-4](https://doi.org/10.1016/S0092-8674(00)80115-4). PMID: 8756724
7. Bolle C. The role of GRAS proteins in plant signal transduction and development. *Planta*. 2004; 218 (5):683–92. <https://doi.org/10.1007/s00425-004-1203-z> PMID: 14760535.
8. Pysh LD, Wysocka-Diller JW, Camilleri C, Bouchez D, Benfey PN. The GRAS gene family in *Arabidopsis*: sequence characterization and basic expression analysis of the SCARECROW-LIKE genes. *Plant J*. 1999; 18 (1):111–9. PMID: 10341448.
9. Song XM, Liu TK, Duan WK, Ma QH, Ren J, Wang Z, et al. Genome-wide analysis of the GRAS gene family in Chinese cabbage (*Brassica rapa* ssp. *pekinensis*). *Genomics*. 2014; 103 (1):135–46. <https://doi.org/10.1016/j.ygeno.2013.12.004> PMID: 24365788.
10. Grimplet J, Agudelo-Romero P, Teixeira RT, Martinez-Zapater JM, Fortes AM. Structural and Functional Analysis of the GRAS Gene Family in Grapevine Indicates a Role of GRAS Proteins in the Control of Development and Stress Responses. *Front Plant Sci*. 2016; 7. <https://doi.org/10.3389/fpls.2016.00353>.
11. Zhang L, Li Q, Chen JF, Chen WS. Computational identification and systematic classification of novel GRAS genes in *Isatis indigotica*. *Chin J Nat Medicines*. 2016; 14 (3):161–76.
12. Huang W, Xian ZQ, Kang X, Tang N, Li ZG. Genome-wide identification, phylogeny and expression analysis of GRAS gene family in tomato. *Bmc Plant Biol*. 2015; 15. <https://doi.org/10.1186/s12870-015-0590-6>.
13. Wang Y, Shi SL, Zhou Y, Zhou Y, Yang J, Tang XQ. Genome-wide identification and characterization of GRAS transcription factors in sacred lotus (*Nelumbo nucifera*). *Peerj*. 2016; 4. <https://doi.org/10.7717/peerj.2388>.
14. Hirsch S, Oldroyd GE. GRAS-domain transcription factors that regulate plant development. *Plant Signal Behav*. 2009; 4 (8):698–700. <https://doi.org/10.4161/psb.4.8.9176> PMID: 19820314.
15. Liu XY, Widmer A. Genome-wide Comparative Analysis of the GRAS Gene Family in *Populus*, *Arabidopsis* and Rice. *Plant Mol Biol Rep*. 2014; 32 (6):1129–45. <https://doi.org/10.1007/s11105-014-0721-5>.
16. Xu W, Chen ZX, Ahmed N, Han B, Cui QH, Liu AZ. Genome-Wide Identification, Evolutionary Analysis, and Stress Responses of the GRAS Gene Family in Castor Beans. *Int J Mol Sci*. 2016; 17 (7). <https://doi.org/10.3390/ijms17071004>.
17. Lee H, Kim B, Song SK, Heo JO, Yu NI, Lee SA, et al. Large-scale analysis of the GRAS gene family in *Arabidopsis thaliana*. *Plant Molecular Biology*. 2008; 67 (6):659–70. <https://doi.org/10.1007/s11103-008-9345-1> PMID: 18500650
18. Day RB, Tanabe S, Koshioka M, Mitsui T, Itoh H, Ueguchi-Tanaka M, et al. Two rice GRAS family genes responsive to N-acetylchitoooligosaccharide elicitor are induced by phytoactive gibberellins: evidence for cross-talk between elicitor and gibberellin signaling in rice cells. *Plant Molecular Biology*. 2004; 54 (2):261–72. <https://doi.org/10.1023/B:PLAN.0000028792.72343.ee> PMID: 15159627



19. Stuurman J, Jaggi F, Kuhlemeier C. Shoot meristem maintenance is controlled by a GRAS-gene mediated signal from differentiating cells. *Gene Dev.* 2002; 16 (17):2213–8. <https://doi.org/10.1101/gad.230702> PMID: 12208843
20. Fu XD, Richards DE, Ait-Ali T, Hynes LW, Ougham H, Peng JR, et al. Gibberellin-mediated proteasome-dependent degradation of the barley DELLA protein SLN1 repressor. *Plant Cell.* 2002; 14 (12): 3191–200. <https://doi.org/10.1105/tpc.006197> PMID: 12468736
21. Wen CK, Chang C. Arabidopsis *RGL1* encodes a negative regulator of gibberellin responses. *Plant Cell.* 2002; 14 (1):87–100. <https://doi.org/10.1105/tpc.010325> PMID: 11826301
22. Morohashi K, Minami M, Takase H, Hotta Y, Hiratsuka K. Isolation and characterization of a novel GRAS gene that regulates meiosis-associated gene expression. *J Biol Chem.* 2003; 278 (23):20865–73. <https://doi.org/10.1074/jbc.M301712200> PMID: 12657631
23. Fode B, Siemsen T, Thurow C, Weigel R, Gatz C. The Arabidopsis GRAS Protein SCL14 Interacts with Class II TGA Transcription Factors and Is Essential for the Activation of Stress-Inducible Promoters. *Plant Cell.* 2008; 20 (11):3122–35. <https://doi.org/10.1105/tpc.108.058974> PMID: 18984675
24. Torres-Galea P, Hirtreiter B, Bolle C. Two GRAS Proteins, SCARECROW-LIKE21 and PHYTOCHROME A SIGNAL TRANSDUCTION1, Function Cooperatively in Phytochrome A Signal Transduction. *Plant Physiol.* 2013; 161 (1):291–304. <https://doi.org/10.1104/pp.112.206607> PMID: 23109688
25. Bolle C, Koncz C, Chua NH. PAT1, a new member of the GRAS family, is involved in phytochrome A signal transduction. *Gene Dev.* 2000; 14 (10):1269–78. PMID: 10817761
26. Torres-Galea P, Huang LF, Chua NH, Bolle C. The GRAS protein SCL13 is a positive regulator of phytochrome-dependent red light signaling, but can also modulate phytochrome A responses. *Mol Genet Genomics.* 2006; 276 (1):13–30. <https://doi.org/10.1007/s00438-006-0123-y> PMID: 16680434
27. Cui H. An evolutionarily conserved mechanism delimiting SHR movement defines a single layer of endodermis in plants (vol 316, pg 421, 2007). *Science.* 2007; 318 (5858):1866–.
28. Heo JO, Chang KS, Kim IA, Lee MH, Lee SA, Song SK, et al. Funneling of gibberellin signaling by the GRAS transcription regulator SCARECROW-LIKE 3 in the Arabidopsis root. *P Natl Acad Sci USA.* 2011; 108 (5):2166–71. <https://doi.org/10.1073/pnas.1012215108>.
29. Li XY, Qian Q, Fu ZM, Wang YH, Xiong GS, Zeng DL, et al. Control of tillering in rice. *Nature.* 2003; 422 (6932):618–21. <https://doi.org/10.1038/nature01518> PMID: 12687001
30. Schumacher K, Schmitt T, Rossberg M, Schmitz C, Theres K. The *Lateral suppressor (Ls)* gene of tomato encodes a new member of the VHIID protein family. *P Natl Acad Sci USA.* 1999; 96 (1):290–5. <https://doi.org/10.1073/pnas.96.1.290>.
31. Greb T, Clarenz O, Schafer E, Muller D, Herrero R, Schmitz G, et al. Molecular analysis of the LATERAL SUPPRESSOR gene in Arabidopsis reveals a conserved control mechanism for axillary meristem formation. *Gene Dev.* 2003; 17 (9):1175–87. <https://doi.org/10.1101/gad.260703> PMID: 12730136
32. Fattash I, Voss B, Reski R, Hess WR, Frank W. Evidence for the rapid expansion of microRNA-mediated regulation in early land plant evolution. *Bmc Plant Biol.* 2007; 7:13. <https://doi.org/10.1186/1471-2229-7-13> PMID: 17359535.
33. Wu G. Plant microRNAs and development. *J Genet Genomics.* 2013; 40 (5):217–30. <https://doi.org/10.1016/j.jgg.2013.04.002> PMID: 23706297.
34. Wang L, Mai YX, Zhang YC, Luo Q, Yang HQ. MicroRNA171c-targeted *SCL6-II*, *SCL6-III*, and *SCL6-IV* genes regulate shoot branching in Arabidopsis. *Mol Plant.* 2010; 3 (5):794–806. <https://doi.org/10.1093/mp/sss042> PMID: 20720155.
35. Curaba J, Talbot M, Li Z, Helliwell C. Over-expression of microRNA171 affects phase transitions and floral meristem determinacy in barley. *Bmc Plant Biol.* 2013; 13:6. <https://doi.org/10.1186/1471-2229-13-6> PMID: 23294862.
36. Fan T, Li XM, Yang W, Xia KF, Jie OY, Zhang MY. Rice osa-miR171c Mediates Phase Change from Vegetative to Reproductive Development and Shoot Apical Meristem Maintenance by Repressing Four OsHAM Transcription Factors. *Plos One.* 2015; 10 (5). <https://doi.org/10.1371/journal.pone.0125833>.
37. Schulze S, Schafer BN, Parizotto EA, Voinnet O, Theres K. LOST MERISTEMS genes regulate cell differentiation of central zone descendants in Arabidopsis shoot meristems. *Plant Journal.* 2010; 64 (4): 668–78. <https://doi.org/10.1111/j.1365-313X.2010.04359.x> PMID: 21070418
38. Engstrom EM, Andersen CM, Gumulak-Smith J, Hu J, Orlova E, Sozzani R, et al. Arabidopsis Homologs of the Petunia *HAIRY MERISTEM* Gene Are Required for Maintenance of Shoot and Root Indeterminacy. *Plant Physiol.* 2011; 155 (2):735–50. <https://doi.org/10.1104/pp.110.168757> PMID: 21173022
39. Ma ZX, Hu XP, Cai WJ, Huang WH, Zhou X, Luo Q, et al. Arabidopsis *miR171*-Targeted Scarecrow-Like Proteins Bind to GT cis-Elements and Mediate Gibberellin-Regulated Chlorophyll Biosynthesis under Light Conditions. *Plos Genet.* 2014; 10 (8). <https://doi.org/10.1371/journal.pgen.1004519>.

40. Young ND, Debelle F, Oldroyd GED, Geurts R, Cannon SB, Udvardi MK, et al. The *Medicago* genome provides insight into the evolution of rhizobial symbioses. *Nature*. 2011; 480 (7378):520–4. <https://doi.org/10.1038/nature10625> PMID: 22089132
41. Laloum T, Baudin M, Frances L, Lepage A, Billault-Penneteau B, Cerri MR, et al. Two CCAAT-box-binding transcription factors redundantly regulate early steps of the legume-rhizobia endosymbiosis. *Plant Journal*. 2014; 79 (5):757–68. <https://doi.org/10.1111/tpj.12587> PMID: 24930743
42. Baudin M, Laloum T, Lepage A, Ripodas C, Ariel F, Frances L, et al. A Phylogenetically Conserved Group of Nuclear Factor-Y Transcription Factors Interact to Control Nodulation in Legumes. *Plant Physiol*. 2015; 169 (4):2761–73. <https://doi.org/10.1104/pp.15.01144> PMID: 26432878
43. Miwa H, Sun J, Oldroyd GED, Downie JA. Analysis of nod-factor-induced calcium signaling in root hairs of symbiotically defective mutants of *Lotus japonicus*. *Mol Plant Microbe In*. 2006; 19 (8):914–23. <https://doi.org/10.1094/Mpmi-19-0914>.
44. Kalo P, Gleason C, Edwards A, Marsh J, Mitra RM, Hirsch S, et al. Nodulation signaling in legumes requires *NSP2*, a member of the GRAS family of transcriptional regulators. *Science*. 2005; 308 (5729):1786–9. <https://doi.org/10.1126/science.1110951> PMID: 15961668.
45. Letunic I, Doerks T, Bork P. SMART: recent updates, new developments and status in 2015. *Nucleic Acids Res*. 2015; 43 (Database issue):D257–60. <https://doi.org/10.1093/nar/gku949> PMID: 25300481.
46. Hu B, Jin J, Guo AY, Zhang H, Luo J, Gao G. GSDS 2.0: an upgraded gene feature visualization server. *Bioinformatics*. 2015; 31 (8):1296–7. <https://doi.org/10.1093/bioinformatics/btu817> PMID: 25504850.
47. Artimo P, Jonnalagedda M, Arnold K, Baratin D, Csardi G, de Castro E, et al. ExPASy: SIB bioinformatics resource portal. *Nucleic Acids Res*. 2012; 40 (Web Server issue):W597–603. <https://doi.org/10.1093/nar/gks400> PMID: 22661580.
48. McWilliam H, Li W, Uludag M, Squizzato S, Park YM, Buso N, et al. Analysis Tool Web Services from the EMBL-EBI. *Nucleic Acids Res*. 2013; 41 (Web Server issue):W597–600. <https://doi.org/10.1093/nar/gkt376> PMID: 23671338.
49. Hall BG. Building phylogenetic trees from molecular data with MEGA. *Mol Biol Evol*. 2013; 30 (5): 1229–35. <https://doi.org/10.1093/molbev/mst012> PMID: 23486614.
50. He Z, Zhang H, Gao S, Lercher MJ, Chen WH, Hu S. Evolview v2: an online visualization and management tool for customized and annotated phylogenetic trees. *Nucleic Acids Res*. 2016; 44 (W1): W236–41. <https://doi.org/10.1093/nar/gkw370> PMID: 27131786.
51. Bailey TL, Boden M, Buske FA, Frith M, Grant CE, Clementi L, et al. MEME SUITE: tools for motif discovery and searching. *Nucleic Acids Research*. 2009; 37:W202–W8. <https://doi.org/10.1093/nar/gkp335> PMID: 19458158
52. Krzywinski M, Schein J, Birol I, Connors J, Gascoyne R, Horsman D, et al. Circos: An information aesthetic for comparative genomics. *Genome Res*. 2009; 19 (9):1639–45. <https://doi.org/10.1101/gr.092759.109> PMID: 19541911
53. Lee TH, Tang H, Wang X, Paterson AH. PGDD: a database of gene and genome duplication in plants. *Nucleic Acids Res*. 2013; 41 (Database issue):D1152–8. <https://doi.org/10.1093/nar/gks1104> PMID: 23180799.
54. Zhang Z, Li J, Zhao XQ, Wang J, Wong GK, Yu J. KaKs\_Calculator: calculating Ka and Ks through model selection and model averaging. *Genomics Proteomics Bioinformatics*. 2006; 4 (4):259–63. [https://doi.org/10.1016/S1672-0229\(07\)60007-2](https://doi.org/10.1016/S1672-0229(07)60007-2) PMID: 17531802.
55. Trapnell C, Pachter L, Salzberg SL. TopHat: discovering splice junctions with RNA-Seq. *Bioinformatics*. 2009; 25 (9):1105–11. <https://doi.org/10.1093/bioinformatics/btp120> PMID: 19289445.
56. Trapnell C, Roberts A, Goff L, Pertea G, Kim D, Kelley DR, et al. Differential gene and transcript expression analysis of RNA-seq experiments with TopHat and Cufflinks. *Nat Protoc*. 2012; 7 (3):562–78. <https://doi.org/10.1038/nprot.2012.016> PMID: 22383036.
57. Saeed AI, Sharov V, White J, Li J, Liang W, Bhagabati N, et al. TM4: a free, open-source system for microarray data management and analysis. *Biotechniques*. 2003; 34 (2):374–8. PMID: 12613259.
58. Lu J, Wang T, Xu Z, Sun L, Zhang Q. Genome-wide analysis of the *GRAS* gene family in *Prunus mume*. *Mol Genet Genomics*. 2015; 290 (1):303–17. <https://doi.org/10.1007/s00438-014-0918-1> PMID: 25245166.
59. Tong H, Jin Y, Liu W, Li F, Fang J, Yin Y, et al. *DWARF AND LOW-TILLERING*, a new member of the *GRAS* family, plays positive roles in brassinosteroid signaling in rice. *Plant J*. 2009; 58 (5):803–16. <https://doi.org/10.1111/j.1365-313X.2009.03825.x> PMID: 19220793.
60. Xu K, Chen SJ, Li TF, Ma XS, Liang XH, Ding XF, et al. *OsGRAS23*, a rice *GRAS* transcription factor gene, is involved in drought stress response through regulating expression of stress-responsive genes. *Bmc Plant Biol*. 2015; 15. <https://doi.org/10.1186/s12870-015-0532-3>.

61. Aubourg S, Kreis M, Lecharny A. The DEAD box RNA helicase family in *Arabidopsis thaliana*. *Nucleic Acids Res.* 1999; 27 (2):628–36. PMID: [9862990](https://pubmed.ncbi.nlm.nih.gov/9862990/).
62. Jain M, Nijhawan A, Arora R, Agarwal P, Ray S, Sharma P, et al. F-box proteins in rice. Genome-wide analysis, classification, temporal and spatial gene expression during panicle and seed development, and regulation by light and abiotic stress. *Plant Physiol.* 2007; 143 (4):1467–83. <https://doi.org/10.1104/pp.106.091900> PMID: [17293439](https://pubmed.ncbi.nlm.nih.gov/17293439/).
63. Fortes AM, Teixeira RT, Agudelo-Romero P. Complex Interplay of Hormonal Signals during Grape Berry Ripening. *Molecules.* 2015; 20 (5):9326–43. <https://doi.org/10.3390/molecules20059326> PMID: [26007186](https://pubmed.ncbi.nlm.nih.gov/26007186/).
64. Wild M, Daviere JM, Cheminant S, Regnault T, Baumberger N, Heintz D, et al. The Arabidopsis *DELLA RGA-LIKE3* Is a Direct Target of *MYC2* and Modulates Jasmonate Signaling Responses. *Plant Cell.* 2012; 24 (8):3307–19. <https://doi.org/10.1105/tpc.112.101428> PMID: [22892320](https://pubmed.ncbi.nlm.nih.gov/22892320/)
65. Prince VE, Pickett FB. Splitting pairs: The diverging fates of duplicated genes. *Nat Rev Genet.* 2002; 3 (11):827–37. <https://doi.org/10.1038/nrg928> PMID: [12415313](https://pubmed.ncbi.nlm.nih.gov/12415313/)
66. Sui Y, Li B, Shi JF, Chen MS. Genomic, regulatory and epigenetic mechanisms underlying duplicated gene evolution in the natural allotetraploid *Oryza minuta*. *Bmc Genomics.* 2014; 15. <https://doi.org/10.1186/1471-2164-15-11>.
67. Vernie T, Moreau S, de Billy F, Plet J, Combier JP, Rogers C, et al. *EFD* Is an *ERF* Transcription Factor Involved in the Control of Nodule Number and Differentiation in *Medicago truncatula*. *Plant Cell.* 2008; 20 (10):2696–713. <https://doi.org/10.1105/tpc.108.059857> PMID: [18978033](https://pubmed.ncbi.nlm.nih.gov/18978033/)
68. Gao MJ, Parkin I, Lydiate D, Hannoufa A. An auxin-responsive SCARECROW-like transcriptional activator interacts with histone deacetylase. *Plant Mol Biol.* 2004; 55 (3):417–31. <https://doi.org/10.1007/s11103-004-0892-9> PMID: [15604690](https://pubmed.ncbi.nlm.nih.gov/15604690/).
69. Ma HS, Liang D, Shuai P, Xia XL, Yin WL. The salt- and drought-inducible poplar GRAS protein SCL7 confers salt and drought tolerance in *Arabidopsis thaliana*. *J Exp Bot.* 2010; 61 (14):4011–9. <https://doi.org/10.1093/jxb/erq217> PMID: [20616154](https://pubmed.ncbi.nlm.nih.gov/20616154/).
70. Park HJ, Jung WY, Lee SS, Song JH, Kwon SY, Kim H, et al. Use of Heat Stress Responsive Gene Expression Levels for Early Selection of Heat Tolerant Cabbage (*Brassica oleracea* L.). *Int J Mol Sci.* 2013; 14 (6):11871–94. <https://doi.org/10.3390/ijms140611871> PMID: [23736694](https://pubmed.ncbi.nlm.nih.gov/23736694/)