Genome-wide identification and expression of GRAS gene family members in cassava

Zhongying Shan¹,³, Xinglu Luo¹,²*, Meiyuan Wu¹, Limei Wei¹, Zhupeng Fan¹ and Yanmei Zhu¹

Abstract

Background: Cassava is highly tolerant to stressful conditions, especially drought stress conditions; however, the mechanisms underlying this tolerance are poorly understood. The GRAS gene family is a large family of transcription factors that are involved in regulating the growth, development, and stress responses of plants. Currently, GRAS transcription factors have not been systematically studied in cassava, which is the sixth most important crop in the world.

Results: Seventy-seven MeGRAS genes were identified from the cassava genome database. Phylogenetic analysis revealed that the MeGRAS proteins could be divided into 14 subfamilies. The gene structure and motif compositions of the proteins were considerably conserved within the same subfamily. Duplication events, particularly segmental duplication, were identified as the main driving force for GRAS gene expansion in cassava. Global expression analysis revealed that MeGRAS genes exhibited similar or distinct expression profiles within different tissues among different varieties. Moreover, qRT-PCR analysis revealed the expression patterns of MeGRAS genes in response to abiotic stress (drought, salt, cold, and H₂O₂), and the results suggest that these genes may have multiple functions.

Conclusion: This study is the first to provide comprehensive information on GRAS gene family members in cassava. The data will increase our understanding of both the molecular basis and the effects of GRAS genes. In addition, the results will contribute further to identifying the responses to various environmental conditions and provide insights into the potential functions of GRAS genes.

Keywords: Cassava, GRAS genes, Gene expression, Abiotic stress

Background

Cassava (Manihot esculenta Crantz) is the sixth most important cash and food crop in Africa, Asia, Latin America, and the Caribbean. Cassava is cultivated for its starchy roots, which are used for food and many products. Owing to its inherent tolerance to stressful environments and the minimal care required, cassava is often considered a food security source against famine where other food crop species would fail. Under optimal environmental conditions, the energy production of cassava is greater than that of most other major staple food crop species [1].

With the advent of next-generation sequencing technology, a large number of genomic, transcriptomic, proteomic, and metabolomic data have been generated, providing a great opportunity for the development of metabolic engineering [2–10]. High-quality genome sequences of cassava have recently become available, which has greatly increased the understanding of the biological processes and molecular/cellular mechanisms in cassava [2, 3].

Transcription factors play regulatory roles in multiple physiological processes in higher plants. Transcription factors act as switches of gene regulation; promote or inhibit the functional expression of specific genes; and are involved in maintaining the growth, development, and stress responses of plants [11, 12]. The term “GRAS” is derived from the first three transcription factors identified in this family: gibberellic acid insensitive (GAI), repressor of GAI (RGA), and scarecrow (SCR) [13].

* Correspondence: luoxinglu@sina.com
¹Agricultural College, Guangxi University, Nanning 530005, China
²State Key Laboratory for Conservation and Utilization of Subtropical Agro-Bioresources, Nanning 530004, China
Full list of author information is available at the end of the article

© The Author(s). 2020 Open Access This article is distributed under the terms of the Creative Commons Attribution 4.0 International License (http://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. The Creative Commons Public Domain Dedication waiver (http://creativecommons.org/publicdomain/zero/1.0/) applies to the data made available in this article, unless otherwise stated.
Typically, GRAS proteins exhibit some C-terminal homology but diversification in sequence and length at their N-termini [13]. The leucine heptad repeat (LHR) I, VHIID, LHR II, PFYRE and SAW motifs are conserved, which is conducive to the function of proteins within the C-terminal region [13, 14]. The structure of VHIID as well as its two flanking regions (LHRs and LER II) is important for protein-protein interactions. Mutations in the PFYRE and SAW motifs result in distinct phenotypic abnormalities in Arabidopsis [15–17]. The N-termini of the GRAS proteins are diverse; however, the DELLA and TVHYNP motifs are conserved within the DELLA subfamily in the N-terminal region.

GRAS genes have recently been studied in various plant species, i.e., Arabidopsis [18], rice [18], Chinese cabbage [19], Populus [20], Prunus mume [21], tobacco [22], castor bean [23], grapevine [24], Medicago truncatula [25, 26], Malus domestica [28], pepper [29], and tea plant [30]. According to previous studies in Arabidopsis and rice [18], the GRAS family members can be classified into eight subfamilies: the DELLA, HAM, LAS, PAT1, SCR, SHR, SCL3, and LISCL subfamilies. However, the number of subfamilies ranges from 8 to 16 in other plant species. GRAS proteins function in various physiological processes during plant growth and development. Considering the highly different amino acid (aa) sequences between each subfamily, each subfamily may have distinct functions. For example, DELLA members mainly function as repressors of gibberellin (GA) signalling [15, 16, 31–33]. The HAIRY MERISTEM gene from the HAM subfamily controls shoot meristem maintenance by mediating signals from differentiating cells [34]. Three genes (MOCI, LS, and LAS) from the LAS subfamily play important roles in axillary meristem initiation [35–37]. Three Arabidopsis genes (PAT1, SCL5, and SCL21) from the PAT1 subfamily are positive regulators of phytochrome-A signal transduction [38, 39], whereas SCL13 from the same subfamily is involved mainly in phytochrome-B signal transduction [40]. SCR and SHR form an SCR/SHR complex, which plays an essential role in root and shoot radial organization [41–43]. SCL3 acts downstream of the GA/DELLA and SCR/SHR pathways and controls GA homeostasis during root development [44]. Moreover, the LISCL gene from the LISCL subfamily is involved in the microsporogenesis of anthers [45].

Until recently, the GRAS gene family in cassava has not been characterized. Thus, in the present study, members of the GRAS gene family were identified from a previous cassava genomic database. A phylogenetic tree was constructed, conversed motif and gene exon/intron structural analyses were performed. MeGRAS genes were mapped onto the cassava chromosomes, and cis-elements of MeGRAS genes and interaction network of MeGRAS proteins were analyzed. In addition, the expression patterns of the MeGRAS genes among different cassava varieties and different tissues were surveyed via available transcriptome data, and the expression patterns in response to various abiotic stresses were investigated via qRT-PCR in cassava. This research is the first to provide evidence concerning the cassava GRAS gene family, which may help elucidate the molecular mechanisms underlying stress responses in cassava.

Results
Identification of the GRAS gene family members in cassava
A total of 77 non-redundant MeGRAS genes were confirmed and used for subsequent analyses (Additional file 1: Table S1). With the exception of MeGRAS4, whose GRAS domain was divided into two parts by a sequence of 15 aa, all of the MeGRAS proteins contained a complete GRAS domain (PF03514). Three of the MeGRAS members (MeGRAS12, 72, and 73) had a DELLA domain (PF12041) and were thus considered DELLA proteins (Additional file 1: Table S2). The protein properties and subcellular localization were analysed, and the results are summarized in Additional file 1: Table S3. The length and molecular mass of the MeGRAS proteins varied greatly, with lengths ranging from 230 to 829 aa and molecular weights (MWs) ranging from 26.08 to 89.79 kDa. The average theoretical isoelectric point (pI) was 5.7, suggesting that most MeGRAS proteins were weakly acidic. With the exception of eight MeGRAS proteins (MeGRAS11, 19, 31, 34, 55, 63, 71, and 74), which were stable, all the MeGRAS proteins were considered unstable. The majority of the MeGRAS proteins contained a high percentage of aliphatic aas, with a predicted aliphatic index ranging from 68.16 to 99.77. Owing to their relatively low average hydropathy (GRAVY) value (< 0), all MeGRASs were predicted to be hydrophilic. It was predicted that 67.53% of the MeGRAS proteins localize to the nucleus. There were no transmembrane helices within the MeGRAS proteins except within GRAS21, which contained one helix that is targeted to the inside of cell membranes. The secondary structure prediction indicated that alpha helices and random coils were dominant in all of the MeGRAS aa sequences, followed by extended strands and beta turns, with average incidences of 44.90, 40.73, 10.11, and 4.26%, respectively (Additional file 1: Table S4).

Phylogenetic analysis of the MeGRAS family
To uncover the evolutionary relationships of the GRAS gene family in cassava and to help in their classification, a total of 211 GRAS proteins, comprising 77 from cassava, 33 from Arabidopsis, 50 from rice, 13 from castor bean, 13 from Populus, 13 from tomato, and 12 from tea plant (Additional file 1: Table S5) [18, 20, 23, 29, 48], were performed to construct an unrooted phylogenetic tree using maximum likelihood (ML) method by
MEGA-X (Fig. 1). Fourteen subfamilies were identified, including 15, 11, 9, 8, 4, 4, 4, 3, 3, 2, 2, 2, and 1 MeGRAS members in the PAT1, LISCL, Pt20, HAM, SHR, SCL3, DELLA, SCR, DTL, Os4, LAS, SCL4/7, Os19, and Os43 subfamilies, respectively (Fig. 1, Fig. 2a).

**Motif composition and gene structure of MeGRAS family members**

To investigate the structural features of cassava GRASs further, the protein conserved motifs and gene intron/exon distributions were analysed. A total of 20 conserved motifs (referred to as motifs 1–20) were identified by MEME (http://meme-suite.org/tools/meme), with more motifs located within the C-terminal region than within the N-terminal region (Fig. 2b); the features of these protein motifs are listed in Additional file 1: Table S6. The motifs from the same subfamily display nearly similar patterns. For example, members of SCL3 had the same motifs. This discovery provides additional evidence to support the close evolutionary relationship of MeGRAS members in the same subfamily. The motifs were matched with the corresponding GRAS domain. Motifs 1 and 2 were in the LHRI domain within the N-terminal region, followed by motifs 3, 4, and 5 in the VHIID domain; motifs 6 and 7 in the LHRII domain; motifs 8, 9, and 10 in the PFYRE domain; and motifs 12 and 13 in the SAW domain within the C-terminal region (Fig. 2b). Gene structural diversity is an important
part of gene family evolution and further supports phylogenetic groupings [46]. In the present study, the number of introns varied from one to three (Fig. 2c). Among the 77 MeGRAS genes, 36 had introns, 34 had only one intron, and 41 had no introns. Generally, MeGRAS genes within the same subfamily in the phylogenetic tree had similar exon-intron structures. The LAS, DELLA, and Os4 subfamilies had no introns, and the SCL4/7, Os43, and PAT1 subfamilies had 1, 2, and 0–3 introns, respectively. The other subfamilies had 0–1 introns.
Chromosomal localization and gene duplication analysis of MeGRAS genes

All of the MeGRAS genes were unevenly distributed on the cassava chromosomes except MeGRAS77 (Fig. 3). There were no MeGRAS members mapped onto Chr16. Chr2 contained the most MeGRAS genes (n = 16; 21.05%), followed by Chr1 (n = 8; 10.53%) and then both Chr3 and Chr15, each of which had seven members. In addition, six MeGRAS genes were distributed on Chr13, and five genes were distributed on Chr8. Four MeGRAS genes were found on both Chr11 and Chr12, three MeGRAS genes were distributed each on Chr5, Chr9, Chr14, and Chr18, and two MeGRAS genes were located on both Chr4 and Chr10. Only one GRAS gene was distributed on Chr6, Chr7, and Chr17. Notably, many MeGRAS genes were concentrated at both ends of the chromosomes.

Gene duplication plays an important role in the occurrence of novel functions and gene family expansion; therefore, the duplication events of the MeGRAS gene within the cassava genome were analysed. As shown in Figs. 3 and 4, two groups of tandem duplicated genes (MeGRAS14/15/16 and MeGRAS22/23) were located on Chr2, while two other groups of tandem duplicated genes (MeGRAS29/30/31 and MeGRAS68/69/70/71) were located on Chr3 and Chr15, respectively. In addition, 34 pairs of MeGRAS genes were identified as being segmental duplications (Fig. 3).
Analysis of Cis-elements in the MeGRAS promoters

The cis-elements were scanned in the promoter regions (1.5 kb upstream of the translation start site) of MeGRAS to better understand potential regulatory mechanisms of MeGRAS genes (Fig. 4). These cis-elements could be divided into four groups: 1) light responsive elements; 2) associated with defense and stress, such as drought, low-temperature, wounding and hypoxia; 3) related to plant hormone responses, such as ABA, MeJA, GA, auxin, salicylic acid and...

![Diagram showing predicted cis-elements in the promoter regions of MeGRAS genes. The number at the bottom indicates the nucleotides to the translation initiation codon, ATG.](image-url)
ethylene; and 4) involved in temporal and spatial gene expression, such as meristem, endosperm and seed. The identified motifs showed that MeGRAS may be regulated by various cis-elements within the promoter during growth.

**Interaction network of GRAS proteins in cassava**

To understand the interactions of the MeGRAS proteins further, an interaction network was constructed via STRING software on the basis of the orthologues in *Arabidopsis*. The orthologous proteins with the highest bit score were considered STRING proteins, and only 46 MeGRAS proteins were selected because of the consideration of reliability (Fig. 5). In general, the MeGRAS proteins of the SCL3 subfamily (MeGRAS8 and 18) and LISCL subfamily (MeGRAS14, 38, and 40) had more interaction partners than did members of the other subfamilies. These findings were consistent with the working mechanisms in consideration of the regulation of GA homeostasis by AtSCL3 proteins, which integrate other signalling pathways; however, this relationship needs to be confirmed [47]. These interaction networks may provide significant clues to understanding the functions of unknown proteins.

**Expression analysis of MeGRAS genes in different tissues**

Cumulative evidence has confirmed that GRASs play important roles in plant growth and development. To understand the function of MeGRAS genes in cassava better, the transcript levels of MeGRAS genes in different tissues, i.e., leaves, stems, early-storage roots, middle-storage roots, and late-storage roots, of the cultivated varieties Arg7 and KU50 and the wild subspecies W14 were examined via publicly available transcriptome data [2]. The fragments per kilobase of transcript per million mapped reads (FPKM) values of the MeGRAS genes are listed in Additional file 1: Table S7, and a heat map of the hierarchical clustering was generated to display the expression profiles of the MeGRAS genes (Fig. 6).

The expression of twelve MeGRAS genes (MeGRAS21, 22, 23, 24, 28, 29, 30, 41, 51, 55, 61, and 74) was not detected in any of the analysed tissues, which is possibly due to differences in spatial and temporal expression patterns. With respect to the Arg7 variety, the expression of 63 out of 77 MeGRAS genes was detected in at least one tissue, in which 26, 28, 34, 30, and 26 genes presented high transcript abundance (FPKM > 5) in the leaf, stem, early-storage root, middle-storage root and late-root storage tissues, respectively. With respect to the KU50 variety, the expression of 58 out of 77 MeGRAS genes was detected in at least one tissue, in which 26, 28, 34, 30, and 26 genes presented high transcript abundance (FPKM > 5) in the leaf, stem, early-storage root, middle-storage root and late-root storage tissues, respectively.
genes was detected in at least one tissue, of which 26, 31, 29, and 30 genes presented high transcript abundance (FPKM > 5) in the leaf, early-storage root, middle-storage root and late-storage root tissues, respectively. With respect to the W14 variety, the expression of 61 out of 77 MeGRAS genes was detected in at least one tissue, of which 28, 33, and 31 genes presented high transcript abundance (FPKM > 5) in the leaf, stem, and middle-storage root tissues, respectively. Overall, with the exceptions of MeGRAS25 and MeGRAS64 in the leaf and stem tissues of Arg7, respectively, the expression levels of 15 MeGRAS genes (MeGRAS37, 72, 75, 12, 73, 25, 4, 44, 7, 64, 62, 20, 15, 26, and 3) from five subfamilies (PAT1, DELLA, LISCL, HAM, and SCL4/7) were high (FPKM > 5) in all of the tested tissues in the three varieties, suggesting key roles of these genes in tissue development.

Most MeGRAS genes exhibited similar expression profiles in the same tissues of Arg7, KU50, and W14, demonstrating that most MeGRAS genes play similar roles in tissue development in the three genotypes. However, a number of genes displayed different expression profiles. For example, the MeGRAS33 transcript abundance was high (FPKM > 5) in the middle-storage roots of W14 but low in the middle-storage roots of Arg7 and KU50. In contrast, the MeGRAS42 transcript abundance was high (FPKM > 5) in the middle-storage roots of Arg7 and KU50 but low in the middle-storage roots of W14. This phenomenon was also detected in other tissues. These findings indicate different roles of these genes in tissue development within different genotypes.

Responses of MeGRAS genes to different abiotic treatments

To measure the transcript levels of MeGRAS genes under different abiotic stresses (drought, salt, cold, and H2O2) in different cassava varieties (D346, NZ199, and GR891), fifteen MeGRAS genes from different subfamilies were subjected to qRT-PCR.
Under drought treatment (Fig. 7), the expression of most \textit{MeGRAS} genes was induced in the three cassava varieties. The expression of five \textit{MeGRAS} genes (\textit{MeGRAS}1, 3, 11, 17, and 51) peaked at 24 h and decreased after 3 d in the three cassava varieties. The expression of seven \textit{MeGRAS} genes (\textit{MeGRAS}2, 4, 12, 32, 41, 53, and 63) first increased but then decreased in D346 and tended to be relatively consistent in NZ199 and GR891, but the peaks differed. The expression of these seven \textit{MeGRAS} genes was highly induced in GR891. The expression of \textit{MeGRAS}27 was clearly upregulated in both D346 and NZ199 but downregulated in GR891, whereas that of \textit{MeGRAS}37 clearly increased in both D346 and GR891 but decreased in NZ199. The expression of the \textit{MeGRAS}1 and \textit{MeGRAS}27 was higher in NZ199 than in D346 and GR891 under drought treatment.

Under salt treatment (Fig. 8), the expression of most \textit{MeGRAS} genes first increased but then decreased in the three cassava varieties. The expression of four \textit{MeGRAS} genes (\textit{MeGRAS}2, 11, 22, and 32) peaked at 6 h but decreased after 3 d in the three cassava varieties. The expression of four other \textit{MeGRAS} genes (\textit{MeGRAS}1, 3, 17, and 51) peaked at 6 h in both D346 and NZ199, whereas it rapidly peaked at 2 h in GR891. The expression of three \textit{MeGRAS} genes (\textit{MeGRAS}12, 41, and 63) in NZ199 and three other \textit{MeGRAS} genes (\textit{MeGRAS}27, 37, and 41) in GR891 tended to increase. The expression of five \textit{MeGRAS} genes (\textit{MeGRAS}27, 12, 22, and 63) was highly induced in GR891 under salt treatment, and the expression of two \textit{MeGRAS} genes (\textit{MeGRAS}17 and 51) was also highly induced in NZ199.

Under cold treatment (Fig. 9), the expression of eight \textit{MeGRAS} genes (\textit{MeGRAS}2, 3, 11, 17, 27, 37, 51, and 53) first increased but then decreased in both D346 and NZ199, peaking at 6 h and 2 h, respectively. The expression of five \textit{MeGRAS} genes (\textit{MeGRAS}1, 3, 17, and 53) first decreased but then increased in GR891, reaching the lowest point at 2 h, and the expression of four \textit{MeGRAS} genes (\textit{MeGRAS}22, 41, and 51) was upregulated in GR891. \textit{MeGRAS}12 was the most highly induced gene in GR891, and \textit{MeGRAS}32 was highly induced in NZ199 under cold treatment.

Under H$_2$O$_2$ treatment (Fig. 10), the expression of nine \textit{MeGRAS} genes (\textit{MeGRAS}1, 2, 3, 11, 17, 32, 41, and 63) was upregulated in D346, and that of nine \textit{MeGRAS} genes (\textit{MeGRAS}1, 2, 3, 11, 17, 32, 51, 53, and 63) was upregulated in GR891. The expression of nine \textit{MeGRAS} genes (\textit{MeGRAS}2, 3, 4, 12, 22, 32, 37, 53, and 63) first
increased but then decreased in NZ199; the expression of five of those genes (MeGRAS4, 12, 32, 37, and 63) rapidly peaked at 2 h, and the expression of the other four MeGRAS genes (MeGRAS2, 3, 22, and 53) peaked at 6 h. The expression of seven MeGRAS genes first increased but then decreased in GR891; the expression of two of them (MeGRAS2 and 37) rapidly peaked at 2 h, whereas the expression of three others (MeGRAS4, 22, and 27) peaked at 6 h. The expression of five MeGRAS genes (MeGRAS17, 32, 22, 1, and 17) was highly induced in NZ199, and that of four MeGRAS genes (MeGRAS12, 41, 63, and 53) was highly induced in GR891. Last, the expression of the MeGRAS32 and 51 genes was highly induced in D346 under H$_2$O$_2$ treatment.

**Discussion**

GRAS transcription factors play vital roles in regulating the growth, development, and stress responses of plants. However, the prevalence and functional diversity of the members of the GRAS family in cassava have not been comprehensively investigated. In the present study, the GRAS gene family in cassava was thoroughly analysed. We explored the features of MeGRAS genes, including their phylogenetic classification, gene structure, chromosomal distribution, cis-elements, expression profile, and responses to various stresses. The results enable us to research the evolution of the GRAS family and hypothesize about the potential functions of unknown genes.

In the present study, a total of 77 GRAS genes were identified in cassava. This number is lower than that in *Malus domestica* (127) [28], *Populus* (106) [20], and maize (86) [27] but higher than that in *Arabidopsis* (34) [20], rice (60) [20], Chinese cabbage (48) [19], *Prunus mume* (46) [21], tobacco (53) [22], castor (48) [23], grapevine (52) [24], *Medicago truncatula* (68) [25, 26], pepper (50) [29], and tea plant (52) [30]. This variation in GRAS gene numbers might be related to gene duplication events or genome size [24]. Four groups of tandem duplicated MeGRAS genes and 34 pairs of segmental duplicated MeGRAS genes were detected in this study. It appears that segmental duplication contributes more to cassava GRAS expansion than does tandem duplication. The MeGRAS genes were located on nearly all of the chromosomes except Chr16 and were unevenly distributed, with the “hot regions” on Chr2 (16 members). Consistent with that which occurs in other species such as *Medicago truncatula* [25], tomato [48], *Arabidopsis*, rice and *Populus* [20], most GRAS genes in cassava lack introns (53.2%) or have only a
single intron (44.2%). The high proportion of intronless genes in the GRAS gene family suggests the close evolutionary relationship of GRAS members. Intronless genes have also been found in other large gene families, such as the DEAD-box RNA helicase [49] and F-box transcription factor families [50]. Although intronless genes are archetypical in prokaryotic genomes, one study [51] showed that plant GRAS genes originated from the prokaryotic genomes mainly by horizontal gene transfer as well as by duplication events throughout their evolution. This phenomenon may explain the formation of a substantial number of intronless GRAS genes.

The 77 MeGRAS proteins were divided into 14 subfamilies on the basis of their sequence homology and classification from Arabidopsis and rice [20]. It is noteworthy that some GRAS proteins considered to be species-specific in previous publications have homologs in cassava. For example, nine cassava MeGRAS genes (MeGRAS22, 23, 27, 28, 29, 30, 31, 34, and 66), together with PtGRAS20 (belong to “Pt20” subfamily) and RcGRAS27 (29,889.m003282; belong to “Rc_GRAS” subfamily) which was previously regarded as species-specific subfamily [20, 23]. They were also clustered with tomato SI5GRAS22, which was clustered with pepper-specific “Ca_GRAS” subfamily [29]. To summarize, all GRAS genes from “Pt20”, “Rc_GRAS” and “Ca_GRAS” subfamilies came from a same subfamily. Three (MeGRAS57, 63, and 65), two (MeGRAS21 and 76) and one (MeGRAS51) cassava GRAS genes, were clustered into “Os4”, “Os19” and “Os43” subfamily, respectively, which were previously reported as rice-specific protein subfamilies [20]. These four subfamilies did not include any Arabidopsis genes, implying lineage-specific gene loss in Arabidopsis. During evolution, other plant species may have lost these species-specific GRAS genes. Another possibility is that they became very specialized during the course of evolution. Analysis of the conserved motifs of the cassava proteins further corroborates the categorization of the MeGRAS family. Conserved motifs have been found within the GRAS domain regions and may have important functions. Although the conserved motifs were similar for all of these MeGRAS proteins, there were some differences in physicochemical features, which were also found among MeGRAS members. These differences may be due to the aa discrepancies within the regions of non-conserved

![Figure 9](image.png)
MeGRAS members, suggesting that MeGRAS proteins function differently in their microenvironments [48].

Expression profiles of various tissues of different varieties with far affinity (wild subspecies and cultivated varieties) have been analysed extensively with respect to the functional characterization of MeGRAS genes. The RPKM data revealed no expression of 12 MeGRAS genes among five subfamilies (HAM, SHR, Os19, Os43, and Pt20) in any tissue but high expression of 15 MeGRAS genes (FPKM > 5) from 5 subfamilies (PAT1, DELLA, LISCL, HAM, and SCL4/7) in every tested tissue of these 3 varieties. There are differences in the expression patterns among these tissues, as demonstrated previously in Chinese cabbage [19], Prunus mume [21], and pepper [29]. Important roles have been inferred from high expression levels of MeGRAS genes. For example, MeGRAS 73, 72, and 12 of the DELLA subfamily are highly important with respect to the control of various signalling hubs. Moreover, 7 MeGRAS genes (MeGRAS 75, 64, 44, 37, 26, 25 and 20) from the PAT1 subfamily are expressed in a wide variety of tissues; these genes are probably involved in various developmental processes via phytochrome signalling regulatory systems, as is the case in Arabidopsis [13, 40]. Together, these results indicate that GRAS genes might have undergone sub-functionalization or neo-functionalization.

Harmful environmental conditions can cause substantial damage to the growth and development of cassava. GRAS genes may play key roles in plant responses against abiotic stresses [23, 25, 30, 48]. A GRAS gene in poplar, PeSCL7, is considered beneficial for engineering salt- and drought-tolerant trees [52], and overexpression of the Brassica napus BnLAS gene in Arabidopsis increases drought tolerance [53]. DELLA proteins are related to the response to many abiotic stresses, i.e., nitric oxide, cold, and phosphate starvation [54–56]. In the present study, expression analysis revealed that the expression of most GRAS genes in cassava is affected by various stress treatments (drought, salt, cold, and oxidative stress), suggesting that MeGRAS genes play crucial roles in the response to abiotic stress. Some expression trend differences occurred among the varieties D346, NZ199, and GR891. For example, under H2O2 treatment, the expression of MeGRAS4 first increased but then decreased in NZ199 and GR891, but it peaked at 2 h in NZ199 and at 6 h in GR891; on the other hand, the expression of MeGRAS4 increased in D346 but did not peak within 24 h. These different trends of MeGRAS
gene expression may be related to different responses to abiotic stress in the three varieties.

This comprehensive study provides a basis for further investigation of GRAS genes in cassava and could also have potential value for the genetic improvement of cassava as well as other related species.

Methods
Identification of GRAS genes in cassava
The latest versions of the genome annotations of cassava (*Manihot esculenta* v6.1) were downloaded from the Phytozone v12 database (https://phytozone.jgi.doe.gov/). Annotated protein databases were scanned using HMMER 3.0 (http://hmmer.org/) with the Hidden Markov model (HMM) of the GRAS domain (PFD03514), which was downloaded from Pfam (http://pfam.xfam.org/). On the basis of the proteins acquired through the GRAS HMM, a high-quality protein set was aligned (E-value \(<1\times10^{-20}\)) and used to construct a cassava-specific GRAS HMM using hmmbuild in HMMER 3.0. This new cassava-specific HMM was used to select all of the proteins with an E-value lower than \(1\times10^{-5}\). In addition, all of the OsGRAS and AtGRAS proteins were used as queries to explore the cassava database via the default parameters. With the application of Pfam database and the Conserved Domain Database (CDD, https://www.ncbi.nlm.nih.gov/Structure/cdd/wrpsb.cgi), only those sequences having a full-length GRAS domain were selected as MeGRAS proteins and used for the subsequent analyses.

ProtParam (http://web.expasy.org/protparam/) was used for the prediction of the physical and chemical features of MeGRAS proteins. To verify the subcellular localization of the identified MeGRAS proteins, WoLF PSORT was used for predicting the transmembrane helices in the proteins.

Phylogenetic analysis of GRAS genes
This research investigated the GRAS proteins of cassava, *Arabidopsis*, rice, castor bean, *Populus*, tomato, and tea plant. *Arabidopsis* and rice are the most commonly used model plant species for researching genetic correlations. One GRAS gene was selected from each subfamily of castor bean, *Populus*, tomato, and tea plant for a better classification of MeGRAS genes. Additional file 1: Table S5 lists the gene IDs of the GRAS members. An unrooted phylogenetic tree was constructed via the default parameters, except for the maximum number of motifs, which was set to 20. The gene structure of the cassava GRASs was determined via the Gene Structure Display Server (GSDS) 2.0 (http://gsds.cbi.pku.edu.cn/) program.

Protein conserved motifs and gene structure analysis
The MEME program was used to identify the conserved motifs. The search also involved the default parameters, except for the maximum number of motifs, which was set to 20. The gene structure of the cassava GRASs was determined via the Gene Structure Display Server (GSDS) 2.0 (http://gsds.cbi.pku.edu.cn/) program.

Chromosomal mapping and gene duplication analysis
Every GRAS gene was matched with the chromosomes of cassava on the basis of the genome annotations of cassava. MapGene2Chromosome (http://mg2c.iask.in/mg2c_v2.0/) was used to draft the map. Gene duplication was explored for MeGRAS genes according to the method described in maize [27]; this method involved 1) the alignment of the entire covered protein length, which is \(>80\%\) of the longest gene, 2) \(>80\%\) identity of the aligned region, and 3) the counting of only one duplication event for tightly linked genes. If there were 5 or fewer than 5 genes separated by two homologous genes, they were labelled as tandem duplications. However, when there were more than 5 genes separating these two genes or there were distributions across various chromosomes, they were referred to as segmental duplications.

Promoter cis-elements analysis
The upstream of 1.5 Kb were used for cis-elements in the promoters of the candidate MeGRAS genes. PlantCARE software (http://bioinformatics.psb.ugent.be/webtools/plantcare/html/) was used for searching regulatory elements.

Prediction of the MeGRAS protein-protein interaction network
To illustrate the correlations between MeGRASs further, interologues of *Arabidopsis* were used to predict the protein-protein interaction network. STRING software was used to construct the functional interaction networks of proteins, with the confidence parameter set at 0.15 [57].

Expression analysis of MeGRAS genes in different tissues
Transcriptome data available online were used to explore the expression profiles of MeGRAS genes in various tissues in various cassava varieties [2]. (Additional file 1: Table S8 lists the accession numbers). Tissues from the leaves, stems, late-storage roots, middle-storage roots, and early-storage roots of a wild subspecies (W14) and two cultivated varieties (KU50 and Arg7) were sampled to investigate the expression profiles of MeGRAS genes. RPKM values were subsequently calculated to evaluate the gene expression.
Cassava plant preparation and stress treatments
All of the studied plants were obtained from a glasshouse at Guangxi University (Nanning, China) between April and July 2018. Cassava stem segments were transplanted into individual pots. The plants were then watered regularly. Three-month-old plants of three cassava varieties (D346, NZ199, and GR891) whose resistance to stress was different were subjected to different abiotic stress treatments, which included 20% polyethylene glycol (PEG) 6000 for 2, 6, and 24 h and 3 d; 300 mM NaCl for 2, 6, and 24 h; cold (4 °C) for 2, 6, and 24 h; and 10% H2O2 for 2, 6, and 24 h. Every sample comprised three independent biological replications.

RNA isolation and qRT-PCR expression analysis
An RNA extraction kit (Huayueyang, China) was used to extract the mRNA from the leaves after each treatment. cDNA was used for the reverse transcription of 1 μg of total mRNA via a cDNA synthesis kit (Takara, Japan). Primer 5.0 was used to design the primers used (Additional file 1: Table S9). For normalization, the MeActin gene was used, serving as an endogenous control. The reaction mechanism of PCR contained 0.5 μl of primers, 1 μl of template cDNA and 5 μl of 2X ChamQ SYBR qPCR Master Mix (Vazyme, China). Afterwards, ddH2O was added to reach a final volume of 10 μL. The protocol was as follows: 95 °C for 30 s followed by 40 cycles of 95 °C for 10 s, 55 °C for 10 s, and 72 °C for 20 s. Each reaction was performed three times, and the 2−ΔΔCT method [58] was used to calculate the relative gene expression levels.

Conclusions
In conclusion, 77 GRAS gene family members from the cassava genome were characterized and classified into 14 subfamilies on the basis of phylogenetic relationships. The gene structure and motif compositions of the proteins were considerably conserved within the same subgroup. Duplication events, particularly segmental duplication, were identified as the main driving force for GRAS gene expansion in cassava. Global expression analysis revealed that the expression profiles of the MeGRAS genes were similar or distinct within different tissues among different varieties. The expression patterns of MeGRAS genes in response to abiotic stress suggested that these genes possibly have multiple functions. Overall, our study is the first comprehensive characterization of GRAS genes in cassava. These data provide a foundation for elucidating the GRAS-mediated molecular mechanism underlying plant growth and development as well as stress biology. This study could serve as a reference for future functional investigations and molecular breeding of cassava.
4. An F, Chen T, Stéphanie DM, Li K, Li QX, Carvalho LJ, Tomlins K, Li J, Gu B, Chen S. Domestication syndrome is investigated by proteomic analysis between cultivated cassava (Manihot esculenta Crantz) and its wild relatives. PLoS One. 2016;11(13):e0152154.
5. Vanderschuren H, Crasemann W. Large-scale proteomics of the cassava storage root and identification of a target gene to reduce postharvest deterioration. Plant Cell. 2014;26(6):1913.
6. Mitprasat M, Roytrakul S, Jiumsup S, Boonsup O, Yokthongwattana K. Leaf proteomic analysis in cassava (Manihot esculenta, Crantz) during plant development, from planting of stem cutting to storage root formation. Planta. 2011;233(6):1209–21.
7. Fu L, Ding Z, Han B, Hu W, Li Y, Zhang J. Physiological investigation and transcriptome analysis of polyethylene glycol (PEG)-induced dehydration stress in cassava. Int J Mol Sci. 2016;17(3):283.
8. Anjanappa RB. Molecular and transcriptomic characterization of natural resistance to cassava brown streak viruses in cassava (Manihot esculenta, Crantz). 2015.
9. Fraser PD. Assessing diversity in cassava through the application of metabolomics. 2016.
10. Uarrotia VG, Marschitz M. Metabolic, enzymatic, and histochemical analyses of cassava roots during postharvest physiological deterioration. BMC Research Notes. 2015;8(1):1–15.
11. Yamaguchi-Shinozaki K, Shinozaki K. Transcriptional regulatory networks in Arabidopsis and rice. Arabidopsis and Rice. Plant Mol Biol Rep. 2014;32(6):1129–45.
12. Liu X, Widmer A. Genome-wide comparative analysis of the GRAS gene family in Arabidopsis and Rice. Plant Mol Biol Rep. 2014;32(6):1129–45.
13. Pflieger C, Koncz C, Chua NH. PAT1, a new member of the GRAS family, is involved in phytochrome signal transduction. Genes Dev. 2000;14(10):1269–78.
14. Torresgalea P, Hirtreiter B, Bolle C. Two GRAS proteins, SCARECROW-LIKE21 and PHOTOCYCLE A SIGNAL TRANSDUCTION1, function cooperatively in phytochrome a signal transduction. Plant Physiol. 2013;161(1):291–304.
15. Koizumi K, Gallagher KL. Identification of SHRUBBY, a SHORT-ROOT and OSHIBBY, a SHORT-ROOT AND OSHIBBY-like gene, from tomato, and their cloning in Arabidopsis. Genet Mol Biol. 2012;35(3):573–9.
16. Sun Y, Xue J, Li X, Li R, Hua M, Han M. The tomato GRAS family gene AtGRAS-35 reveals their evolutionary dynamics and functional diversification. PLoS One. 2017;12(9):e0185439.
17. Song L, Liu L, Cui H, Hu Q, Guo C. Genome-wide identification and expression analysis of the GRAS family genes in Medicago truncatula. Acta Physiol Plant. 2017;39(4):93.
18. Guo Y, Wu H, Li X, Li Q, Zhao X, Duan X, An L, Lv W, An H. Identification and expression of GRAS family genes in maize (Zea mays L.). Plos One. 2017;12(9):e0185418.
19. Fan S, Zhang D, Gao C, Zhao M, Wu H, Li Y, Shen Y, Han M. Identification, classification, and expression analysis of GRAS gene family in Malus domestica. Front Physiol. 2017;8.
20. Liu B, Sun Y, Xue J, Xiu J, Li R. Genome-wide characterization and expression analysis of GRAS gene family in pepper (Capsicum annuum L.). PeerJ, 2018;6.
21. Wang YX, Liu ZW, Wu ZL, Hu H, Wang WL, Cui X, Zhuang J. Genome-wide identification and expression analysis of GRAS family genes transcription in tea plant (Camellia sinensis). Sci Rep-UK. 2018. 8(1).
22. Peng J, Carol P, Richards DE, King KE, Cowling RJ, Murphy GP, Harberd NP. The Arabidopsis GA1 gene defines a signaling pathway that negatively regulates gibberellin responses. Genes Dev. 1997;11(23):3194.
23. Peng J, Richards DE, Hantley NM, Murphy GP, Devos KM, Flintham JE, Beales J, Fish LJ, Worland AJ, Pelica F. ‘Green revolution’ genes encode mutant gibberellin response modulators. Nature. 1999;400(6741):256.
24. Li W, Wu J, Wang S, Zhang Y, Zhang D, Shi C. Identification and characterization of dwarf 62, a loss-of-function mutation in DLT/OsGRAS-32 affecting gibberellin metabolism in rice. Planta. 2010;232(6):1383–96.
25. Stuurman J, Jiggli F, Kuhlmeier C. Shoot meristem maintenance is controlled by a GRAS-gene mediated signal from differentiating cells. Genes Dev. 2002;16(17):2213–8.
26. Schumacher K, Schmitz T, Rossberg M, Schmitz G, Theres K. The Lateral suppressor (Ls) gene of tomato encodes a new member of the VHID protein family. P Natl Acad Sci U S A. 1999;96(1):290–5.
27. Trethewey SM, Ochsendorf FE, Nilsson PM, Tamaroff L. The lateralsuppressor gene is involved in lateral suppression of morphogenesis in Arabidopsis. Plant Physiol. 2000;123(3):963–73.
28. Patel V,shih J, Hulsey T, Carroll C, Suran N, Walsh J, Schemm P. The lateralsuppressor gene is involved in lateral suppression of morphogenesis in Arabidopsis. Plant Physiol. 2000;123(3):963–73.
29. Helariutta Y, Fukaki H, Yamauchi T, McCann A, Wang Q, Wang Q, Wang L, Wang L, Wang L, Wang L. The SHORT-ROOT gene is required for root patterning during root development. Plant J. 2001;26(2):261–73.
30. Wang Q, Wang Q, Wang Q, Wang Q, Wang L, Wang L, Wang L, Wang L, Wang L. The SHORT-ROOT gene is required for root patterning during root development. Plant J. 2001;26(2):261–73.
31. He L, Zou J, Wang Q, Wang Q, Wang Q, Wang L, Wang L, Wang L, Wang L. The SHORT-ROOT gene is required for root patterning during root development. Plant J. 2001;26(2):261–73.
32. Pullmann D, Li-Fang H, Nam-Hai C, Cordelia B. The GRAS protein SCL13 is a positive regulator of phytochrome-dependent red light signaling, but can also modulate phytochrome a responses. Mol Gen Genomics. 2006;270(3):133–50.
33. Li W, Wu J, Wang S, Zhang Y, Zhang D, Shi C. Identification and characterization of dwarf 62, a loss-of-function mutation in DLT/OsGRAS-32 affecting gibberellin metabolism in rice. Planta. 2010;232(6):1383–96.
34. Stuurman J, Jiggli F, Kuhlmeier C. Shoot meristem maintenance is controlled by a GRAS-gene mediated signal from differentiating cells. Genes Dev. 2002;16(17):2213–8.
35. Schumacher K, Schmitz T, Rossberg M, Schmitz G, Theres K. The Lateral suppressor (Ls) gene of tomato encodes a new member of the VHID protein family. P Natl Acad Sci U S A. 1999;96(1):290–5.
36. Trethewey SM, Ochsendorf FE, Nilsson PM, Tamaroff L. The lateralsuppressor gene is involved in lateral suppression of morphogenesis in Arabidopsis. Plant Physiol. 2000;123(3):963–73.
37. Patel V,shih J, Hulsey T, Carroll C, Suran N, Walsh J, Schemm P. The lateralsuppressor gene is involved in lateral suppression of morphogenesis in Arabidopsis. Plant Physiol. 2000;123(3):963–73.
38. Stuurman J, Jiggli F, Kuhlmeier C. Shoot meristem maintenance is controlled by a GRAS-gene mediated signal from differentiating cells. Genes Dev. 2002;16(17):2213–8.
39. Schumacher K, Schmitz T, Rossberg M, Schmitz G, Theres K. The Lateral suppressor (Ls) gene of tomato encodes a new member of the VHID protein family. P Natl Acad Sci U S A. 1999;96(1):290–5.
40. Trethewey SM, Ochsendorf FE, Nilsson PM, Tamaroff L. The lateralsuppressor gene is involved in lateral suppression of morphogenesis in Arabidopsis. Plant Physiol. 2000;123(3):963–73.
41. Koizumi K, Gallagher KL. Identification of SHRUBBY, a SHORT-ROOT and SCARECROW interacting protein that controls root growth and radial patterning. Dev. 2013;140(6):1292–300.
42. Pedersen AL, Ciampaglio CN, Sun T. The Arabidopsis LATERAL SUPPRESSOR gene family. Front Plant Sci. 2016;7:253–8.
43. Di LL, Wysockadiller J, Malamy JE, Pysh L, Helariutta Y, Freshour G, Hahn MG, Feldmann KA, Benfey PN. The SCARECROW gene regulates an asymmetric cell division that is essential for generating the radial organization of the Arabidopsis root. Cell. 1999;96(1):290–5.
44. Stuurman J, Jiggli F, Kuhlmeier C. Shoot meristem maintenance is controlled by a GRAS-gene mediated signal from differentiating cells. Genes Dev. 2002;16(17):2213–8.
45. Schumacher K, Schmitz T, Rossberg M, Schmitz G, Theres K. The Lateral suppressor (Ls) gene of tomato encodes a new member of the VHID protein family. P Natl Acad Sci U S A. 1999;96(1):290–5.
51. Zhang D, Iyer LM, Aravind L. Bacterial GRAS domain proteins throw new light on gibberellic acid response mechanisms. Bioinformatics. 2012;28(19):2407–11.

52. 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.

53. Yang M, Yang Q, Fu T, Zhou Y. Overexpression of the Brassica napus BnLAS gene in Arabidopsis affects plant development and increases drought tolerance. Plant Cell Rep. 2011;30(3):373–88.

54. Tao Y, Bai S, Li M, Zhang Y. DELLA contribute to tolerance to nitric oxide stress in Arabidopsis seedlings. Chinese Bulletin of Botany. 2011;46(5):481–8.

55. Achard P, Gong F, Cheminant S, Alloua M, Hedden P, Genschik P. The cold-inducible CBF1 factor-dependent signaling pathway modulates the accumulation of the growth-repressing DELLA proteins via its effect on gibberellin metabolism. Plant Cell. 2008;20(8):2117–29.

56. Jiang C, Gao X, Liao L, Harberd NP, Fu X. Phosphate starvation root architecture and anthocyanin accumulation responses are modulated by the gibberellin-DELLA signaling pathway in Arabidopsis. Plant Physiol. 2007;145(4):1460–70.

57. Franceschini A, Szklarczyk D, Frankild S, Kuhn M, Simonovic M, Roth A, Lin J, Minguez P, Bork P, Mering CV, et al. STRING v9.1: protein-protein interaction networks, with increased coverage and integration. Nucleic Acids Res. 2013;41:D808–15.

58. Livak KJ, Schmittgen TD. Analysis of relative gene expression data using real-time quantitative PCR and the 2^{-△△CT} method. Methods. 2001;25(4):402–8.

Publisher’s Note
Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.