Comparative analysis of GRAS genes in six Cucurbitaceae species provides novel insights into their evolution and function

Qiqi Zhang

Chinese Academy of Agricultural Sciences

Jun He

Southwest University

Yuanchao Xu

Chinese Academy of Agricultural Sciences

Sen Chai

Qingdao Agricultural University

Tianshu Sun

Chinese Academy of Agricultural Sciences

Hongbo Li

Chinese Academy of Agricultural Sciences

Hongjia Xin

Chinese Academy of Agricultural Sciences

Kuipeng Xu

Qingdao Agricultural University

Zhonghua Zhang (✉ zhangzhonghua@qau.edu.cn)

Qingdao Agricultural University

Research Article

Keywords: GRAS, Phylogeny, Synteny, Expression, Cucurbitaceae

Posted Date: February 26th, 2021

DOI: <https://doi.org/10.21203/rs.3.rs-258961/v1>

License: © ⓘ This work is licensed under a Creative Commons Attribution 4.0 International License.

[Read Full License](#)

Abstract

Background

GRAS proteins are important transcription factors that play essential roles in diverse fundamental processes of plant growth and development. However, the features, evolution and potential functions of *GRAS* genes among the Cucurbit crops remain largely unknown.

Results

In this study, we identified 237 *GRAS* gene sequences from six Cucurbitaceae species. The genomes of *Cucumis sativus* L., *Cucumis melo* L., *Benincasa hispida* L., *Citrullus lanatus* L. and *Lagenaria siceraria* L. contain 35 to 37 *GRAS* genes, whereas 55 are present in *Cucurbita moschata* L. which should be derived from a recent whole genome duplication. These *GRAS* genes were divided into sixteen subfamilies on the basis of their phylogenetic relationships. In the SCLB and RAD1 subclade, three specific motifs were found in Cucurbitaceae species, implying the potential Cucurbitaceae-specific functions. Duplication and synteny analysis revealed that segmental and tandem duplication may be the main reasons for the expansion of the *GRAS* family in Cucurbitaceae species, and that whole-genome duplication is the main reason for the greater number of *GRAS* genes in *Cucurbita moschata* L.. Some *GRAS* genes exhibit tissue-specific, subfamily-specific or Cucurbitaceae-specific expression patterns, implying that they may play key roles in specific developmental processes in cucurbit crops.

Conclusions

This study provides insights into the features, evolution and potential functions of *GRAS* genes in Cucurbitaceae species, which may be associated with important agronomic traits.

Backgrounds

Transcription factors (TFs) play important roles in plant development and resistance to abiotic/biotic stresses. The *GRAS* gene family is a large plant-specific TF family named after its first three founding members: GIBBERELLIC-ACID INSENSITIVE (GAI) [1], REPRESSOR OF GAI (RGA) [2] and SCARECROW (SCR) [3]. It has been divided into different subfamilies (8–17) such as DELLA, Hairy Meristem (HAM), Lateral Suppressor (Ls), LISCL, NSP2, PHYTOCHROME-A SIGNAL TRANSDUCTION 1 (PAT1), SCARECROW (SCR), SCARECROW-LIKE 3 (SCL3), and SHORT-ROOT (SHR), etc [4–8]. *GRAS* genes have been reported to be involved many biological processes such as gibberellin signaling [1–3], light signaling [9, 10], meristem initiation [11–13], cell proliferation and elongation [14, 15], fruit development and ripening [16–18]).

Cucurbitaceae species include a number of important crops such as cucumber (*Cucumis sativus* L.), melon (*Cucumis melo* L.), watermelon (*Citrullus lanatus* L.), wax gourd (*Benincasa hispida* L.), bottle gourd (*Lagenaria siceraria* L.) and pumpkin (*Cucurbita moschata* L.), etc. They exhibit numerous specific traits such as melon fruit and diverse floral sex types [19, 20]. However, the characterization, evolution and function of *GRAS* genes has been rarely reported [21, 22]. The release of the whole-genome sequences of several Cucurbitaceae species (*Cucumis sativus* L. [23], *Cucumis melo* L. [24], *Citrullus lanatus* L. [25], *Benincasa hispida* L. [19], *Cucurbita moschata* L. [26] and *Lagenaria siceraria* L. [27]) has facilitated the comprehensive analysis of the entire *GRAS* gene family in these species. Here, we report the identification of 237 *GRAS* genes from these six Cucurbitaceae species and present the evolutionary history, conserved motif distribution, and synteny of these *GRAS* genes. Additionally, we have analyzed the expression patterns of these genes in different tissues (roots, stems, leaves, flowers and fruits) using transcriptome data. This study provides an important resource and novel insights into the evolutionary and functional characterization of *GRAS* genes in Cucurbitaceae crops and angiosperms.

Results

Identification and characterization of *GRAS* genes in six Cucurbitaceae crops

A total of 237 *GRAS* genes were identified in the genomes of six Cucurbitaceae crops. The number of *GRAS* genes was relatively consistent among these species: 37 in *C. sativus*, 36 in *C. melo*, 35 in *B. hispida*, 37 in *C. lanatus* and 37 in *L. siceraria*, with a considerably greater number of 55 in *C. moschata* (Additional file 1: Table S1; Additional file 2: Table S2). Cucurbitaceae *GRAS* genes contained 1–12 exons, with most containing only one exon (160, 67.51%). Among the 237 *GRAS* proteins, the length of the putative encoded proteins varied from 258 (Bhi11P000300) to 1,466 (Lsi06P016090.1) amino acids, the predicted molecular weight (MW) ranged from 29.83 (Bhi11P000300) to 164.43 kDa (Lsi06P016090.1), and the isoelectric point (pI) varied from 4.70 (Lsi01P011330.1, CmoCh01G003940) to 9.67 (Cla004408) with a mean of 5.66, demonstrating that most of the proteins are weakly acidic. These *GRAS* genes (except for *BhiUN508M4*, *BhiUN508P6* and *CmoCh00G001590.1*, which are not anchored onto the chromosomes) were unevenly distributed across the genomes of the six Cucurbitaceae species (Additional file 3: Fig. S1a–f). For example, among the seven chromosomes of *C. sativus*, chromosome 3 is the longest, but it had only 7 *GRAS*s, compared with the 10 *GRAS* genes present on chromosome 6 (Additional file 4: Table S3).

Phylogenetic relationships, conserved motifs and gene structures

To examine the phylogenetic relationships among the *GRAS* proteins of the Cucurbitaceae species, a phylogenetic tree was constructed using the complete set of 237 Cucurbitaceae *GRAS* proteins, together

with 165 GRAS proteins selected from five other representative angiosperms: *A. thaliana* (*Arabidopsis thaliana* L.) (33), *S. lycopersicum* (*Solanum lycopersicum* L.) (49), *O. sativa* (*Oryza sativa* L.) (24), *V. vinifera* (*Vitis vinifera* L.) (29) and *A. trichopoda* (*Amborella trichopoda* L.) (30). All 237 GRAS proteins were classified into 16 subfamilies (Fig. 1; Additional file 5: Table S4). The PAT1 subfamily has the largest number of *GRAS* genes, followed by DELLA (27), HAM (26), LISCL (21) subfamily. Orthologous genes of the Scarecrow-Like A (SCLA) subfamily and Required for Arbuscule Development 2 (RAD2) subfamily, OG-NSP2-Amb, OG-NSP2-3 orthologous groups were not identified in the Cucurbitaceae species, suggesting these genes may be not needed and have lost during the evolution (Additional file 6: Fig. S2).

The GRAS proteins in Cucurbitaceae species shared a highly conserved C-terminus, which contained five distinct conserved motifs in the following order: LHR I, VHIID, LHR II, PFYRE and SAW (Additional file 7: Fig. S3). We identified an additional 20 conserved motifs among the 237 GRAS proteins (Additional file 8, 9: Fig. S4, 5), named motifs 1–20. Most of the motifs had a similar distribution pattern within the same subfamily (Additional file 8: Fig. S4; Additional file 10: Table S5). Some motifs were present only in specific subfamilies; for example, motif 20 was only present in the LISCL and HAM subfamilies (Additional file 8: Fig. S4b). A comparison of members from different GRAS subfamilies showed that those from most closely related subfamilies contained similar motifs. For example, members of the SHR group contained motifs 9, 16, 8, 2, 1, 13, 5, 14, 7, 6, 12, 3, 10, 19 and 4, but those in the NSP1 group contained motifs 9, 16, 8, 2, 1, 13, 5, 14, 7, 10, 19 and 4 (Additional file 8: Fig. S4b). In addition, we analyzed the gene structures of 237 *GRAS* genes by comparing the patterns of exon-intron architecture among the six Cucurbitaceae species (Additional file 8: Fig. S4c). The majority of *GRAS* genes within the same subgroup or between orthologous groups contained a similar gene structure. For instance, all genes in the OG-SHR-2, OG-SCL32-2, RAM1 and Ls subfamilies contained only one intron and two CDS.

Further, we compared the predicted motifs in the six cucurbits with those in other species including *Arabidopsis*, *S. lycopersicum* (Additional file 11: Fig. S6; Additional file 13: Table S6). Some motifs were specific in Cucurbitaceae species; for example, motif 7 (EFGDFNFPSANQSGFYQQDISKIGDQTNYYQQPNSDCLIFDELLFGNDFTI) in the SCLB clade, (Fig. 2a) and motifs 9 (LDDTTAASRWVISFSDEF RHK) and 10 (MALDGDGGSFFSTDFSTVGKEDEDTVGD) in the RAD1 clade (Fig. 2b). These specific motifs may be responsible for the specific processes or traits in Cucurbitaceae species.

Duplication and syntenic analysis of *GRAS* genes among six Cucurbitaceae species

Gene duplication events were surveyed to examine the expansion of the *GRAS* gene family in the six Cucurbitaceae genomes. Preliminary results showed that four types of duplicated *GRAS* genes (156 dispersed, 7 proximal, 8 tandem and 66 segmental genes) were present in the six Cucurbitaceae species (Additional file 14: Table S7). Homologous gene pairs (2 dispersed, 4 proximal, 8 tandem and 18

segmental genes) with a protein similarity greater than 80% and coverage greater than 80% were analyzed among the duplicated *GRAS* genes (Additional file 14: Table S7). In total, 18 segmental duplicated genes were identified in the *C. moschata* genome and two tandem duplicated genes were located in *C. sativus*, *C. melo*, *C. lanatus* and *L. siceraria*, whereas no tandem and segmental duplications occurred during the evolution of *B. hispida* *GRAS* genes. Additionally, we observed that all duplicated gene pairs possessed Ka/Ks ratios lower than 0.5, suggesting that these genes underwent strong purifying selection during genome evolution (Additional file 15,16: Table S8,S9). Intra-genome synteny analysis for *C. moschata* showed that the duplicated *GRAS* genes occurred equally in subgenomes A and B (Fig. 3).

To infer the evolution of *GRAS* genes, synteny analysis was carried out among the six Cucurbitaceae species (Fig. 4; Additional file 17: Table S10). A total of 225 *GRAS* genes (37 in *C. sativus*, 36 in *C. melo*, 35 in *B. hispida*, 36 in *C. lanatus*, 37 in *L. siceraria* and 44 in *C. moschata*) were located within synteny blocks of the six Cucurbitaceae genomes. Some orthologous gene pairs showed a two-to-one syntenic relationship between *C. moschata* and five other Cucurbitaceae species. We found six gene pairs which exist among all five other species have two counterparts in *C. moschata*, for example, the *Csa4P196810.1*, *MELO3C025282T1*, *Bhi03P001806*, *Clu012151* and *Lsi01G016490.1* are orthologs, and have two collinear gene pairs (*CmoCh03G002750.1* and *CmoCh07G012270.1*) in *C. moschata* (Fig. 4; Additional file 17: Table S10; Additional file 18: Fig. S8). These results show that these *GRAS* genes were conserved during the evolution of all the six cucurbit species, and the amplification in the *C. moschata* genome due to a recent whole genome duplication (WGD) event was well preserved, suggesting their conserved roles and the specific traits in *C. moschata*. In addition, some *GRAS* genes were found to have been lost in *C. moschata*. For example, some *GRAS* genes were existed in *C. sativus*, *C. melo*, *B. hispida*, *C. lanatus* and *L. siceraria*, but were lost in *C. moschata* (the *Csa7P322070.1/MELO3C023684T1/Bhi09P001228/Clu015025/Lsi02G020110.1* collinear gene pair) (Fig. 4; Additional file 17: Table S10; Additional file 18: Fig. S8).

Expression analysis of *GRAS* genes in six Cucurbitaceae species

To investigate the potential function of Cucurbitaceae *GRAS* genes, we profiled their expression in different tissues, including roots, stems, leaves, flowers and fruits. Among the 237 *GRAS* genes, 191 were expressed in all analyzed tissues (Additional file 19: Table S11), and 60 were relatively high expressed (FPKM > 10) (Additional file 19: Table S11). Almost all the *GRAS* genes in the HAM, DELLA, LISCL, SCL4/7 and PAT1 subfamilies were expressed in all five tissues analyzed (Additional file 20, 21: Fig. S9, S10). Transcripts of three *GRAS* genes (*MELO3C008036T1*, *MELO3C008170T1* and *Lsi03G004130.1*) could barely be detected in any of the five tissues (FPKM < 0.001) (Additional file 19: Table S11). Expression of some *GRAS* genes were barely detected in specific tissues (Additional file 19: Table S11; Additional file 21: Fig. S10). For example, *MELO3C020907T1* (SHR subfamily) was not expressed in stems and transcripts of *CmoCh00G001590.1* and *CmoCh17G005960.1* (SHR subfamily) in leaves were nearly absent.

The expression profiles of *GRAS* genes from some subfamilies or orthologous groups varied greatly. For example, most OG-SCR-3 members were hardly expressed in any analyzed tissues, whereas expression level in most OG-SCR-1 and OG-SCR-2 members were high in all tissues (Fig. 5a). In the NSP2 subfamily, OG-NSP2-2 members were barely expressed in any tested tissues, in contrast to the low level of expression of all Cucurbitaceae OG-NSP2-1 genes in leaves (Fig. 5b).

We found 38 *GRAS* genes showing tissue-specific expression (tissue-specificity index, $\tau > 0.85$), including 19 in roots, 1 in stems, 9 in leaves, 2 in flowers and 7 in fruits (Additional file 22: Table S12). For instance, *Cla014025*, *Cla020203* and *Lsi10G005540.1* (SHR subfamily); *Cla014874*, *Bhi09P001022* and *Lsi02G021610.1* (SCR subfamily) (Fig. 5a), *Lsi01G009100.1*, *Lsi10G005540.1* and *Cla014014* (SCL32 subfamily) (Fig. 5c) were highly expressed in roots; most genes in the SCL32 subfamily (*Bhi03P000775*, *MELO3C007619T1*, *Cla007701*, *Csa6P495010.1*, *Csa7P044940.1*) (Fig. 5c) were high expressed in leaves, etc.

Besides tissue-specific genes, genes in some subfamilies showed expression patterns that were specific to Cucurbitaceae species. For example, all Cucurbitaceae genes in the SCLB clade were relatively high expressed in fruits, where other clade members of species outside the Cucurbitaceae were hardly detectable (Fig. 5d); the same was true for members of the RAD1 clade (Fig. 5e).

In most cases, similar expression patterns were shown by duplicated gene pairs. However, a few exceptions were observed (Additional file 23: Fig. S11; Additional file 24: Table S13). For example, for the *Cla_TanDup_1.1/1.2* duplicated gene pair, *Cla_TanDup_1.2* was highly expressed in roots, flowers and fruits, whereas *Cla_TanDup_1.1* weakly expressed in roots, flowers and fruits; for the *Cmo_SegDup_6.1/6.2* duplicated gene pair, *Cmo_SegDup_6.2* was highly expressed in roots and fruits, whereas *Cmo_SegDup_6.1* was hardly detected in these tissues, etc. These differences in expression suggest that duplicated *GRAS* genes may have acquired divergent functions.

Discussion

Recent studies in several plants such as *A. thaliana* [1, 2, 6, 9, 13, 28-33], *O. sativa* [8, 11, 34-38] and *S. lycopersicum* [16, 18, 39-43] showed that GRAS transcription factors play important roles in multiple biological processes. However, limited information is available for GRAS proteins in Cucurbitaceae species. Here, we report the genome-wide identification and characteristics of *GRAS* genes in six Cucurbitaceae species and describe their evolutionary history and expression patterns. This study provides insights into understanding of the functions of *GRAS* genes in Cucurbitaceae plants.

Identification and evolutionary history of *GRAS* genes in six Cucurbitaceae species

The GRAS gene family contains different numbers of members in different species, such as 32–34 in *A. thaliana* [4, 6], 60 in *O. sativa* [4, 8], 53–54 in *S. lycopersicum* [7, 39] and 106 in *P. euphratica* [4]. The

number of *GRAS* genes in the genomes of the Cucurbitaceae species *C. sativus*, *C. melo*, *B. hispida*, *C. lanatus* and *L. siceraria* of the Benincasae group is similar, whereas *C. moschata* contains 55 *GRAS* genes, indicating that the *GRAS* family diverged together with the evolutionary separation of the Cucurbitaceae and Benincasae groups. Collinearity analysis showed that segmental duplication events detected in *C. moschata* occurred between the two *C. moschata* subgenomes (Fig. 3), indicating that a whole-genome duplication event was the main reason for the difference in the number of *GRAS* genes between the Benincasae and Cucurbitaceae group [26, 44, 45]. The number of *GRAS* genes (except for *CmoCh00G001590.1* whose chromosome location is unclear) distributed on the subgenome A of *C. moschata* (32 *GRAS* genes) was greater than that on subgenome B (22 *GRAS* genes) (Additional file 3: Fig. S1f), indicating that the rate of gene loss within this family in subgenome A was lower than that for subgenome B [26].

Our characterization of *GRAS* family members was expanded by the analysis of conserved CDSs and UTRs in Cucurbitaceae genes and motifs in their encoded proteins (Additional file 8: Fig. S4, Additional file 11: Fig. S6; Additional file 12: Fig. S7). Homologous *GRAS* genes that clustered together in a subgroup or clade tended to share similar numbers and positions of the CDS and encoded protein motifs (Additional file 8: Fig. S4). However, some genes within the same subgroup also contained different structures, reflecting evolutionary differences of these *GRAS* gene family members and suggesting that they might have different functions (Additional file 11: Fig. S6).

Gene expression profiles reveal diverse roles for *GRAS* genes in plant growth and development

To identify the potential functions of the Cucurbitaceae *GRAS* genes, we profiled their expression patterns and observed different abundances of *GRAS* gene transcripts in different tissues and organs, implying that the genes might differ in function (Additional file 20, 21: Fig. S9, S10).

The tissue-specific expression of some *GRAS* genes suggests that they play potential roles in the corresponding tissue. For example, *Cla020023* and *Cla014025* from the SHR subfamily were particularly expressed in roots (Additional file 20, 21: Fig. S9, S10; Additional file 22: Table S12), which is consistent with the function of their homologous gene *AtSHR* in root development [15, 29, 46, 47]. Similarly, the root expression of *Bhi09P001022*, *Cla014874*, *Lsi02G021610.1* (Fig. 5a) from the SCR subfamily is consistent with the function of their homologue *AtSCR* in root development [29].

If the tissue-specific expression of genes suggests a potential function in a particular tissue, the species-specific expression of some genes may imply they have a specific role in that particular species. Genes from the SCLB subfamily were expressed in fruits of Cucurbitaceae plants, whereas their homologous genes were not expressed in fruits of the other analyzed species (Fig. 5d). Therefore, these genes might participate in fruit development only in Cucurbitaceae species. Consistent with this, we also identified some motifs in members of the SCLB subfamily that were specific to the Cucurbitaceae (Fig. 2a).

Furthermore, homologous genes may share similar functions. In this discussion, we particularly focus on genes associated with fruit development, from model plants such as *S. lycopersicum*. For instance, *Lsi07G014150.1* and *Csa3P043910.1* in the DLT clade, were more highly expressed during the flower–fruit transition, implying that they potentially function in fruit set and development (Additional file 21: Fig. S10; Additional file 19: Table S11). This pattern is consistent with the expression pattern of their homologous gene *SIGRAS41* (*Solyc08g080400.1.1*), which might play a role in fruit development by modulating brassinosteroid signaling in *S. lycopersicum* [14, 35, 39]. Within the DELLA subfamily, *Lsi05G003650.1* was highly expressed in fruits (Additional file 20, 21: Fig. S9, S10; Additional file 19: Table S11), which is consistent with the reported function of its homologue *SIDELLA* in controlling cell division, expansion, and the auxin-signaling pathway throughout fruit set and growth in *S. lycopersicum* [31, 48]. It has been reported that silencing of *SIGRAS2* (*Solyc07g063940.1.1*, PAT1 subfamily) led to reduced fruit weight during *S. lycopersicum* fruit development through the modulation of ovary growth and cell size [16]. In Cucurbitaceae species, *CmoCh09G009100.1*, *CmoCh01G012140.1*, *MELO3C018144T1* (Additional file 21: Fig. S10; Additional file 19: Table S11), which are homologous to the *SIGRAS2* gene, were highly transcribed in fruits, implying that they may play a role in fruit development.

Conclusions

In this study, 237 *GRAS* genes were identified in the genomes of six Cucurbitaceae species. The *GRAS* family was divided into 16 subgroups on the basis of their phylogenetic relationships and previous classifications. We observed Cucurbitaceae-specific motifs and expression patterns in *GRAS* proteins of the SCLB clade, which may play a specific role in Cucurbitaceae species. Duplication and synteny analysis revealed that segmental and tandem duplication might be the main drivers of the expansion of the *GRAS* family in Cucurbitaceae species, with whole-genome duplication as the main reason for the presence of more *GRAS* genes in *C. moschata* than in the other five analyzed species. The expression patterns of the *GRAS* members among tissues or at different development stages were surveyed using publicly available transcriptome data, and some tissue-specific genes were identified, and their potential functions were discussed, especially for the genes that might be involved in fruit development. In conclusion, this study should facilitate improvements in agronomic traits of Cucurbitaceae crops.

Methods

Identification and characterization of *GRAS* genes in Cucurbitaceae species

We downloaded the genome sequences, protein sequences and annotation information for *C. sativus* [23], *C. melo* [24], *B. hispida* [19], *C. lanatus* [25], *L. siceraria* [27] and *C. moschata* [26]. The Pfam database (<http://pfam.xfam.org/>) [49] was used to retrieve the *GRAS* Hidden Markov Model (HMM) model (PF03514) for HMM searches against annotated protein databases from different genomes with

an E-value cutoff of 0.01 with the program HMMER (version: 3.1) (<http://hmmer.org/>). From the proteins obtained using the raw GRAS HMM, a high-quality protein set (E-value < 1×10^{-20} and manual verification of an intact GRAS domain) was aligned and used to construct a species-specific GRAS HMM using hmmbuild from the HMMER v3 suite. This new species-specific HMM was used, and all proteins with an E-value lower than 0.01 were selected [50]. To identify amino-acid sequences that contained GRAS domains, we used the NCBI Conserved Domain Search (<https://www.ncbi.nlm.nih.gov/Structure/cdd/>) to ensure the accuracy of the identification of these transcription factors. In the same way, we identified *GRAS* genes from the *A. thaliana* and *S. lycopersicum* genomes and compared the results with the number of *GRAS* genes previously reported, to verify the accuracy of the method. Finally, *GRAS* genes were further filtered based on GRAS proteins whose ratio of the GRAS conserved domain length to the GRAS HMM length was less than 60%. The *GRAS* genes were mapped to the six chromosomes of Cucurbitaceae species based on their genome annotation files, and diagrams of the chromosomal locations were plotted with Mapchart software [51]. The chemical features of each GRAS protein, such as isoelectric point (pI) and molecular weight (MW), were extracted using the ExPASy proteomics server (<http://web.expasy.org/protparam/>) [52]. We used OrthoFinder [53] to identify single-copy homologous genes from each of the 11 species and used EasySpeciesTree (<https://github.com/dongwei1220/EasySpeciesTree>) to construct species phylogenetic trees.

Analysis of *GRAS* gene structures and the phylogenetic relationships, and conserved motifs among the GRAS proteins

The 237 candidate Cucurbitaceae GRAS proteins and 165 additional well-described GRAS proteins (encoded by 33 *GRAS* genes in *A. thaliana*, 49 *GRAS* genes in *S. lycopersicum*, 24 *GRAS* genes in *O. sativa*, 29 *GRAS* genes in *V. vinifera* and 30 *GRAS* genes in *A. trichopoda*.) were multiply aligned using MUSCLE software (version 3.8.1551) [54] to further search the evolutionary relationships of the GRAS gene family and a neighbor-joining (NJ) consensus tree was subsequently constructed using MEGA software (version 6.06) [55] under the Jones-Taylor-Thornton (JTT) amino-acid substitution model and the Partial deletion option of 1,000 bootstrap replicates to assess the reliability of internal branches. Finally, we divided the sequences into different subfamilies according to previous reports [5] and modified the phylogenetic trees by iTOL software (<https://itol.embl.de>) [56]. The other trees that contained 318 GRAS proteins from six Cucurbitaceae plants, *A. thaliana* and *S. lycopersicum* were constructed for each subfamily in the same way. Multiple alignments of 237 Cucurbitaceae sequences containing a conserved GRAS domain were performed with DNAMAN software (version 8.0), and the maps were constructed.

To obtain the gene structure, a GFF3 annotation file containing the precise positional information of coding sequences (CDSs) and untranslated regions (UTRs) in *GRAS* genes was retrieved from the genomic dataset. The conserved and shared motifs for protein sequences from each subfamily were annotated by Multiple Expectation for motif Elicitation tool (version: 4.9.1) (MEME, [Page 9/23](http://meme-</p></div><div data-bbox=)

suite.org/) [57] with the following parameters: “minsites 6, maxsites 149, minw 6, maxw 200 and nmotifs 20”. Gene structures and conserved motifs were visualized using TBtools software [58].

Duplication and synteny analysis of *GRAS* genes from six Cucurbitaceae species

Duplication events among Cucurbitaceae plants were identified using the Basic Local Alignment Search Tool (BLAST) (E-value: $\leq 10^{-5}$) and the duplicate_gene_classifier program with the default parameters of the Multiple Collinearity Scan toolkit (MCScan) [59]. The duplicate_gene_classifier program can classify the duplicate genes of a single species into WGD/segmental, tandem, proximal and dispersed duplicates. WGD/segmental duplicates are inferred by the anchor genes in collinear blocks. Tandem duplicates are defined as paralogs that are adjacent to each other on chromosomes, which are suggested to arise from illegitimate chromosomal recombination. Proximal duplicates are paralogs physically located near each other but interrupted by several other genes. Proximal duplicates are inferred to result from localized transposon activities, or ancient tandem arrays interrupted by more recent gene insertions. Dispersed duplicates are paralogs that are neither near each other on chromosomes, nor do they show conserved synteny [60]. Subsequently, non-synonymous (K_a) and synonymous substitution (K_s) between duplicated *GRAS* gene pairs were calculated by the add_ka_and_ks_to_collinearity program of the MCScan toolkit. The duplicated *GRAS* genes in *C. moschata* were visualized as circle plots generated by Circos software (version: 0.69-8) [61]. Syntenic gene pairs among the *C. sativus*, *C. melo*, *B. hispida*, *C. lanatus*, *L. siceraria* and *C. moschata* genomes were identified using MCScan with default parameters. Chromosome-scale synteny blocks plots between six Cucurbitaceae genomes were constructed using the python version of MCScan ([https://github.com/tanghaibao/jcvi/wiki/MCscan-\(Python-version\)](https://github.com/tanghaibao/jcvi/wiki/MCscan-(Python-version))).

The expression of *GRAS* genes in different tissues

We analyzed the expression profiles of *GRAS* genes in different tissues, including roots, stems, leaves, flowers and fruits, using public RNA-seq data [26, 62-68]. RNA-seq data from five tissues were mapped to the six Cucurbitaceae plant genomes using Hisat2 (version: 2.1.0) [69] with default settings for parameters and were assembled using stringtie (version: 1.3.4) [70]. The bam files of uniquely mapped reads were used as inputs for the stringtie software, and fragments per kilobase million (FPKM) values were calculated to measure the gene expression levels. Heat maps for *GRAS* genes were plotted by pheatmap in the R package using the FPKM values as $\log_2(\text{FPKM value} + 1)$. To examine the specificity of the *GRAS* gene expression patterns, we calculated a tissue-specificity index, τ , which is a quantitative, graded scalar measure of the specificity of an expression profile. The τ values interpolate the entire range between 0 for housekeeping genes and 1 for genes that are strictly specifically expressed in a single tissue. In this study, $0.85 \leq \tau \leq 1$ was regarded as a tissue-specific gene, whereas $0 \leq \tau \leq 0.15$ identified a housekeeping gene. All needed RNA sequencing data in this study from NCBI

(<https://www.ncbi.nlm.nih.gov/>) Sequence Read Archive (SRA) database with the accession number are shown in Additional file 25: Table S14.

Abbreviations

TFs: Transcription factors

GAI: GIBBERELLIC-ACID INSENSITIVE

RGA: REPRESSOR OF GAI

SCR: SCARECROW

HAM: Hairy Meristem

Ls: LATERAL SUPPRESSOR

PAT1: PHYTOCHROME-A SIGNAL TRANSDUCTION 1

SCR: SCARECROW

SCL3: SCARECROW-LIKE 3

SHR: SHORT-ROOT

SCLA: Scarecrow-Like A

RAD2: Required for Arbuscule Development 2

LHRI: leucine heptad repeat I

LHRII: leucine heptad repeat II

Chr: Chromosome

kb: Kilobase

aa: Amino acid

MW: Molecular weight

pI: Isoelectric point

NJ: Neighbor-joining

OG: Orthologous Group

CDS: Coding sequence

UTRs: Untranslated regions

WGD: Whole genome duplication

HMM: Hidden Markov Model

BLAST: Basic Local Alignment Search Tool

Ka: Non-synonymous substitution

Ks: Synonymous substitution

FPKM: Fragments per kilobase million

SRA: Sequence Read Archive

Declarations

Availability of data and materials

All needed genome sequences and genome annotation files of Cucurbitaceae species were obtained from the Cucurbit Genome Database (*C. sativus*: <http://cucurbitgenomics.org/organism/2>; *C. melo*: <http://cucurbitgenomics.org/organism/18>; *B. hispida*: <http://cucurbitgenomics.org/organism/22>; *C. lanatus*: <http://cucurbitgenomics.org/organism/1>; *L. siceraria*: <http://cucurbitgenomics.org/organism/13>; *C. moschata*: <http://cucurbitgenomics.org/organism/9>). The genome assemblies and annotation of *A. thaliana* and *S. lycopersicum* are available at TAIR database (<https://www.arabidopsis.org/>, TAIR10) and Sol Genomics Network (<https://solgenomics.net>, ITAG3.2). All needed RNA sequencing data in this study from NCBI (<https://www.ncbi.nlm.nih.gov/>) Sequence Read Archive (SRA) database with the accession number are shown in Additional file 25: Table S14.

Acknowledgments

Not applicable

Funding

This work was supported by National Natural Science Foundation of China (31772304 to Z.Z.), China National Key Research and Development Program for Crop Breeding (2016YFD0100307 to Z.Z.), and National high-level talent special support Program in China (Z.Z.). This work was also supported by the “Taishan Scholar” Foundation of the People’s Government of Shandong Province, the Chinese Academy

of Agricultural Sciences (ASTIP-CAAS), the Shenzhen municipal (The Peacock Plan KQTD2016113010482651).

Ethics declarations

Ethics approval and consent to participate

Not applicable

Consent for publication

Not applicable

Competing interests

The authors declare that they have no conflict of interest.

Author information

Affiliations

Key Laboratory of Biology and Genetic Improvement of Horticultural Crops of the Ministry of Agriculture, Sino-Dutch Joint Laboratory of Horticultural Genomics, Institute of Vegetables and Flowers, Chinese Academy of Agricultural Sciences, Beijing 100081, China.

Qiqi Zhang, Yuanchao Xu, Tianshu Sun, Hongbo Li, Hongjia Xin & Zhonghua Zhang

Engineering Laboratory of Genetic Improvement of Horticultural Crops of Shandong Province, College of Horticulture, Qingdao Agricultural University, Qingdao 266109, China.

Sen Chai, Kuipeng Xu & Zhonghua Zhang

Citrus Research Institute, Southwest University, Chongqing 400715, China

Jun He

Guangdong Laboratory of Modern Agriculture, Genome Analysis Laboratory of the Ministry of Agriculture, Agricultural Genomics Institute at Shenzhen, Chinese Academy of Agricultural Sciences, Shenzhen 518120, China.

Contributions

Z.Z. and Q.Z. conceived and designed the research. Q.Z. performed the data analysis and wrote the manuscript. J.H., Z.Z., S.C., T.S., H.L. and K.X. revised the manuscript. Y.X., H.L. and H.X. provided bioinformatics technical assistance to Q.Z. All the authors read and approved the final manuscript.

References

1. Peng J, Carol P, Richards DE, King KE, Cowling RJ, Murphy GP, Harberd NP: **The *Arabidopsis* *GAI* gene defines a signaling pathway that negatively regulates gibberellin responses.** *Genes & development* 1997, **11**(23):3194-3205.
2. Silverstone AL, Ciampaglio CN, Sun T-p: **The *Arabidopsis* *RGA* Gene Encodes a Transcriptional Regulator Repressing the Gibberellin Signal Transduction Pathway.** *The Plant Cell* 1998, **10**(2):155-169.
3. Laura Di Laurenzio, Joanna Wysocka-Diller, Jocelyn E. Malamy, Leonard Pysh, Yrjo Helariutta, Glenn Freshour, Michael G. Hahn, Kenneth A. Feldmann, Benfey PN: **The *SCARECROW* Gene Regulates an Asymmetric Cell Division That Is Essential for Generating the Radial Organization of the *Arabidopsis* Root.** *Cell* 1996.
4. Liu X, Widmer A: **Genome-wide Comparative Analysis of the GRAS Gene Family in *Populus*, *Arabidopsis* and Rice.** *Plant Molecular Biology Reporter* 2014, **32**(6):1129-1145.
5. Cenci A, Rouard M: **Evolutionary Analyses of GRAS Transcription Factors in Angiosperms.** *Front Plant Sci* 2017, **8**:273.
6. Lee MH, Kim B, Song SK, Heo JO, Yu NI, Lee SA, Kim M, Kim DG, Sohn SO, Lim CE *et al*: **Large-scale analysis of the GRAS gene family in *Arabidopsis thaliana*.** *Plant Mol Biol* 2008, **67**(6):659-670.
7. Niu Y, Zhao T, Xu X, Li J: **Genome-wide identification and characterization of GRAS transcription factors in tomato (*Solanum lycopersicum*).** *PeerJ* 2017, **5**:e3955.
8. 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):519-532.
9. Cordelia Bolle CK, Nam-Hai Chua: **PAT1, a new member of the GRAS family, is involved in phytochrome A signal transduction.** *GENES & DEVELOPMENT* 2000.
10. Torres-Galea P, Huang L-F, Chua N-H, Bolle C: **The GRAS protein SCL13 is a positive regulator of phytochrome-dependent red light signaling, but can also modulate phytochrome A responses.** *Molecular Genetics and Genomics* 2006, **276**(1):13-30.
11. Xueyong Li, Qian Qian, Zhiming Fu, Yonghong Wang, Guosheng Xiong, Dali Zeng, Xiaoqun Wang, Xinfang Liu, Sheng Teng, Fujimoto Hiroshi *et al*: **Control of tillering in rice.** *Nature* 2003.

12. SCHUMACHER. K, SCHMITT. T, ROSSBERG. M, SCHMITZ. G, THERES. AK: **The *Lateral suppressor (Ls)* gene of tomato encodes a new member of the VHLID protein family.** *PNAS* 1999.
13. Greb T, Clarenz O, Schafer E, Muller D, Herrero R, Schmitz G, Theres K: **Molecular analysis of the *LATERAL SUPPRESSOR* gene in *Arabidopsis* reveals a conserved control mechanism for axillary meristem formation.** *Genes Dev* 2003, **17**(9):1175-1187.
14. Sun L, Li X, Fu Y, Zhu Z, Tan L, Liu F, Sun X, Sun X, Sun C: ***GS6*, a member of the GRAS gene family, negatively regulates grain size in rice.** *J Integr Plant Biol* 2013, **55**(10):938-949.
15. Stuurman J, Jaggi F, Kuhlemeier C: **Shoot meristem maintenance is controlled by a GRAS-gene mediated signal from differentiating cells.** *Genes Dev* 2002, **16**(17):2213-2218.
16. Li M, Wang X, Li C, Li H, Zhang J, Ye Z: **Silencing *GRAS2* reduces fruit weight in tomato.** *J Integr Plant Biol* 2018, **60**(6):498-513.
17. Zhang L, Zhu M, Ren L, Li A, Chen G, Hu Z: **The *SIFSR* gene controls fruit shelf-life in tomato.** *J Exp Bot* 2018, **69**(12):2897-2909.
18. Liu Y, Shi Y, Su D, Lu W, Li Z: ***SIGRAS4* accelerates fruit ripening by regulating ethylene biosynthesis genes and *SIMADS1* in tomato.** *Hortic Res* 2021, **8**(1):3.
19. Xie D, Xu Y, Wang J, Liu W, Zhou Q, Luo S, Huang W, He X, Li Q, Peng Q *et al*: **The wax gourd genomes offer insights into the genetic diversity and ancestral cucurbit karyotype.** *Nat Commun* 2019, **10**(1):5158.
20. Zhang H, Li S, Yang L, Cai G, Chen H, Gao D, Lin T, Cui Q, Wang D, Li Z *et al*. (2020) **Gain-of-function of the 1-aminocyclopropane-1-carboxylate synthase gene *ACS1G* induces female flower development in cucumber gynoecy.** *The Plant Cell* 2020, koaa018.
21. Haywood V, Yu TS, Huang NC, Lucas WJ: **Phloem long-distance trafficking of GIBBERELLIC ACID-INSENSITIVE RNA regulates leaf development.** *Plant J* 2005, **42**(1):49-68.
22. Lu X, Liu W, Xiang C, Li X, Wang Q, Wang T, Liu Z, Zhang J, Gao L, Zhang W: **Genome-Wide Characterization of GRAS Family and Their Potential Roles in Cold Tolerance of Cucumber (*Cucumis sativus* L.).** *Int J Mol Sci* 2020, **21**(11).
23. Huang S, Li R, Zhang Z, Li L, Gu X, Fan W, Lucas WJ, Wang X, Xie B, Ni P *et al*: **The genome of the cucumber, *Cucumis sativus* L.** *Nat Genet* 2009, **41**(12):1275-1281.
24. Garcia-Mas J, Benjak A, Sanseverino W, Bourgeois M, Mir G, Gonzalez VM, Henaff E, Camara F, Cozzuto L, Lowy E *et al*: **The genome of melon (*Cucumis melo* L.).** *Proceedings of the National Academy of Sciences* 2012, **109**(29):11872-11877.
25. Guo S, Zhang J, Sun H, Salse J, Lucas WJ, Zhang H, Zheng Y, Mao L, Ren Y, Wang Z *et al*: **The draft genome of watermelon (*Citrullus lanatus*) and resequencing of 20 diverse accessions.** *Nat Genet* 2013, **45**(1):51-58.
26. Sun H, Wu S, Zhang G, Jiao C, Guo S, Ren Y, Zhang J, Zhang H, Gong G, Jia Z *et al*: **Karyotype Stability and Unbiased Fractionation in the Paleo-Allotetraploid Cucurbita Genomes.** *Mol Plant* 2017, **10**(10):1293-1306.

27. Wu S, Shamimuzzaman M, Sun H, Salse J, Sui X, Wilder A, Wu Z, Levi A, Xu Y, Ling KS *et al*: **The bottle gourd genome provides insights into Cucurbitaceae evolution and facilitates mapping of a Papaya ring-spot virus resistance locus.** *Plant J* 2017, **92**(5):963-975.
28. 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.** *The Plant Cell* 2008, **20**(11):3122-3135.
29. Della Rovere F, Fattorini L, D'Angeli S, Veloccia A, Del Duca S, Cai G, Falasca G, Altamura MM: **Arabidopsis SHR and SCR transcription factors and AUX1 auxin influx carrier control the switch between adventitious rooting and xylogenesis in planta and in in vitro cultured thin cell layers.** *Ann Bot* 2015, **115**(4):617-628.
30. Goldy C, Pedroza-Garcia JA, Breakfield N, Cools T, Vena R, Benfey PN, De Veylder L, Palatnik J, Rodriguez RE: **The Arabidopsis GRAS-type SCL28 transcription factor controls the mitotic cell cycle and division plane orientation.** *Proc Natl Acad Sci U S A* 2021, **118**(6).
31. Tyler L, Thomas SG, Hu J, Dill A, Alonso JM, Ecker JR, Sun TP: **Della proteins and gibberellin-regulated seed germination and floral development in Arabidopsis.** *Plant Physiol* 2004, **135**(2):1008-1019.
32. Heo JO, Chang KS, Kim IA, Lee MH, Lee SA, Song SK, Lee MM, Lim J: **Funneling of gibberellin signaling by the GRAS transcription regulator scarecrow-like 3 in the Arabidopsis root.** *Proc Natl Acad Sci U S A* 2011, **108**(5):2166-2171.
33. Liu W, Kohlen W, Lillo A, Op den Camp R, Ivanov S, Hartog M, Limpens E, Jamil M, Smaczniak C, Kaufmann K *et al*: **Strigolactone biosynthesis in Medicago truncatula and rice requires the symbiotic GRAS-type transcription factors NSP1 and NSP2.** *Plant Cell* 2011, **23**(10):3853-3865.
34. Tian C, Wan P, Sun S, Li J, Chen M: **Genome-wide analysis of the GRAS gene family in rice and Arabidopsis.** *Plant molecular biology* 2004, **54**(4):519-532.
35. Tong H, Jin Y, Liu W, Li F, Fang J, Yin Y, Qian Q, Zhu L, Chu C: **DWARF AND LOW-TILLERING, a new member of the GRAS family, plays positive roles in brassinosteroid signaling in rice.** *The Plant Journal* 2009, **58**(5):803-816.
36. Li W, Wu J, Weng 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):1383-1396.
37. Tanabe S, Onodera H, Hara N, Ishii-Minami N, Day B, Fujisawa Y, Hagio T, Toki S, Shibuya N, Nishizawa Y *et al*: **The elicitor-responsive gene for a GRAS family protein, CIGR2, suppresses cell death in rice inoculated with rice blast fungus via activation of a heat shock transcription factor, OsHsf23.** *Biosci Biotechnol Biochem* 2016, **80**(1):145-151.
38. Xu K, Chen S, Li T, Ma X, Liang X, Ding X, Liu H, Luo L: **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**:141.

39. Huang W, Xian Z, Kang X, Tang N, Li Z: **Genome-wide identification, phylogeny and expression analysis of GRAS gene family in tomato.** *BMC Plant Biol* 2015, **15**:209.
40. Liu Y, Huang W, Xian Z, Hu N, Lin D, Ren H, Chen J, Su D, Li Z: **Overexpression of *SIGRAS40* in Tomato Enhances Tolerance to Abiotic Stresses and Influences Auxin and Gibberellin Signaling.** *Front Plant Sci* 2017, **8**:1659.
41. Huang W, Peng S, Xian Z, Lin D, Hu G, Yang L, Ren M, Li Z: **Overexpression of a tomato miR171 target gene *SIGRAS24* impacts multiple agronomical traits via regulating gibberellin and auxin homeostasis.** *Plant Biotechnol J* 2017, **15**(4):472-488.
42. Zhou S, Hu Z, Li F, Yu X, Naeem M, Zhang Y, Chen G: **Manipulation of plant architecture and flowering time by down-regulation of the GRAS transcription factor *SIGRAS26* in *Solanum lycopersicum*.** *Plant Science* 2018, **271**:81-93.
43. Liu Y, Shi Y, Zhu N, Zhong S, Bouzayen M, Li Z: **SIGRAS4 mediates a novel regulatory pathway promoting chilling tolerance in tomato.** *Plant Biotechnol J* 2020, **18**(7):1620-1633.
44. Cristina Esteras P, Antonio J Monforte, José Blanca, Nelly Vicente-Dólera, Cristina Roig, Fernando Nuez and Belén Picó: **High-throughput SNP genotyping in Cucurbita pepo for map construction and quantitative trait loci mapping.pdf.** *BMC Genomics* 2012.
45. Shalmani A, Jing XQ, Shi Y, Muhammad I, Zhou MR, Wei XY, Chen QQ, Li WQ, Liu WT, Chen KM: **Characterization of B-BOX gene family and their expression profiles under hormonal, abiotic and metal stresses in Poaceae plants.** *BMC Genomics* 2019, **20**(1):27.
46. Koizumi K, Hayashi T, Gallagher KL: **SCARECROW reinforces SHORT-ROOT signaling and inhibits periclinal cell divisions in the ground tissue by maintaining SHR at high levels in the endodermis.** *Plant Signal Behav* 2012, **7**(12):1573-1577.
47. Clark NM, Hinde E, Winter CM, Fisher AP, Crosti G, Blilou I, Gratton E, Benfey PN, Sozzani R: **Tracking transcription factor mobility and interaction in Arabidopsis roots with fluorescence correlation spectroscopy.** *Elife* 2016, **5**.
48. Ikeda A, Ueguchi-Tanaka M, Sonoda Y, Kitano H, Koshioka M, Futsuhara Y, Matsuoka M, Yamaguchi J: **Slender Rice, a Constitutive Gibberellin Response Mutant, Is Caused by a Null Mutation of the SLR1 Gene, an Ortholog of the Height-Regulating Gene *GAI/RGA/RHT/D8*.** *The Plant Cell* 2001, **13**(5).
49. Finn RD, Coggill P, Eberhardt RY, Eddy SR, Mistry J, Mitchell AL, Potter SC, Punta M, Qureshi M, Sangrador-Vegas A *et al*: **The Pfam protein families database: towards a more sustainable future.** *Nucleic Acids Res* 2016, **44**(D1):D279-285.
50. Lozano R, Hamblin MT, Prochnik S, Jannink JL: **Identification and distribution of the NBS-LRR gene family in the Cassava genome.** *BMC Genomics* 2015, **16**:360.
51. Voorrips RE: **MapChart: Software for the Graphical Presentation of Linkage Maps and QTLs.** *The Journal of Heredity* 2002, **93**(1).
52. Artimo P, Jonnalagedda M, Arnold K, Baratin D, Csardi G, De Castro E, Duvaud S, Flegel V, Fortier A, Gasteiger E: **ExPASy: SIB bioinformatics resource portal.** *Nucleic acids research* 2012, **40**(W1):W597-W603.

53. Emms DM, Kelly S: **OrthoFinder: phylogenetic orthology inference for comparative genomics.** *Genome Biol* 2019, **20**(1):238.
54. Edgar RC: **MUSCLE: multiple sequence alignment with high accuracy and high throughput.** *Nucleic acids research* 2004, **32**(5):1792-1797.
55. Tamura K, Stecher G, Peterson D, Filipinski A, Kumar S: **MEGA6: molecular evolutionary genetics analysis version 6.0.** *Molecular biology and evolution* 2013, **30**(12):2725-2729.
56. Letunic I, Bork P: **Interactive Tree Of Life (iTOL) v4: recent updates and new developments.** *Nucleic Acids Res* 2019.
57. Bailey TL, Boden M, Buske FA, Frith M, Grant CE, Clementi L, Ren J, Li WW, Noble WS: **MEME SUITE: tools for motif discovery and searching.** *Nucleic Acids Res* 2009, **37**(Web Server issue):W202-208.
58. Chen C, Xia R, Chen H, He Y: **TBtools, a Toolkit for Biologists integrating various HTS-data handling tools with a user-friendly interface.** *bioRxiv* 2018:289660.
59. Wang Y, Tang H, Debarry JD, Tan X, Li J, Wang X, Lee TH, Jin H, Marler B, Guo H *et al*: **MCSanX: a toolkit for detection and evolutionary analysis of gene syteny and collinearity.** *Nucleic Acids Res* 2012, **40**(7):e49.
60. Freeling M: **Bias in plant gene content following different sorts of duplication: tandem, whole-genome, segmental, or by transposition.** *Annu Rev Plant Biol* 2009, **60**:433-453.
61. Krzywinski MI, Schein JE, Birol I, Connors J, Gascoyne R, Horsman D, Jones SJ, Marra MA: **Circos: An information aesthetic for comparative genomics.** *Genome Research* 2009.
62. Rhee SJ, Seo M, Jang YJ, Cho S, Lee GP: **Transcriptome profiling of differentially expressed genes in floral buds and flowers of male sterile and fertile lines in watermelon.** *BMC genomics* 2015, **16**:914.
63. Gan X, Stegle O, Behr J, Steffen JG, Drewe P, Hildebrand KL, Lyngsoe R, Schultheiss SJ, Osborne EJ, Sreedharan VT *et al*: **Multiple reference genomes and transcriptomes for *Arabidopsis thaliana*.** *Nature* 2011, **477**(7365):419-423.
64. Jiang B, Xie D, Liu W, Peng Q, He X: **De novo assembly and characterization of the transcriptome, and development of SSR markers in wax gourd (*Benicasa hispida*).** *PLoS One* 2013, **8**(8):e71054.
65. Rhee SJ, Kwon T, Seo M, Jang YJ, Sim TY, Cho S, Han SW, Lee GP: **De novo-based transcriptome profiling of male-sterile and fertile watermelon lines.** *PLoS One* 2017, **12**(11):e0187147.
66. Li Z, Zhang Z, Yan P, Huang S, Fei Z, Lin K: **RNA-Seq improves annotation of protein-coding genes in the cucumber genome.** *BMC Genomics* 2011, **12**:540.
67. Tomato Genome C: **The tomato genome sequence provides insights into fleshy fruit evolution.** *Nature* 2012, **485**(7400):635-641.
68. Yano R, Nonaka S, Ezura H: **Melonet-DB, a Grand RNA-Seq Gene Expression Atlas in Melon (*Cucumis melo* L.).** *Plant Cell Physiol* 2018, **59**(1):e4.
69. Kim D, Langmead B, Salzberg SL: **HISAT: a fast spliced aligner with low memory requirements.** *Nat Methods* 2015, **12**(4):357-360.

70. Pertea M, Pertea GM, Antonescu CM, Chang TC, Mendell JT, Salzberg SL: **StringTie enables improved reconstruction of a transcriptome from RNA-seq reads.** *Nat Biotechnol* 2015, **33**(3):290-295.

Figures

Fig. 1

Figure 1

Phylogenetic tree constructed using 402 GRAS protein sequences from different species. Sequences were extracted for: *Amborella trichopoda*, *Vitis vinifera*, *Oryza sativa*, *Arabidopsis thaliana*, *Solanum lycopersicum* and six Cucurbitaceae species (*Cucumis sativus*, *Cucumis melo*, *Benincasa hispida*, *Citrullus lanatus*, *Lagenaria siceraria* and *Cucurbita moschata*). The branch colors differentiate subfamilies that contain more than one orthologous. Orthologous not grouped into subfamilies are indicated by black branches.

Fig. 2

Figure 2

Cucurbitaceae-specific motifs of GRAS genes in the SCLB and RAD1 subfamilies. The neighbor-joining tree of the SCLB (a) and RAD1 (b) subfamily of GRAS proteins is shown on the left. The horizontal colored boxes indicate conserved motifs within each protein. The motif 7, motif 9 and motif 10 are the Cucurbitaceae-specific motifs in SCLB and RAD1 subfamily, respectively. The motif logo for the subfamily is shown on the right.

Fig. 3

Figure 3

Synteny and duplication analysis of Cucurbita moschata GRAS genes. Chromosomes of Cucurbita moschata are shown in different colors in circular form. The chromosomes are assigned to subgenome A (the outer ring) and subgenome B (the inner ring). The approximate positions of the segmental duplicated GRAS genes are marked with short red lines on the circle. The gray lines in the background represent the

regions of the collinear blocks in *Cucurbita moschata*. Colored curves denote the segmental duplicated genes in *Cucurbita moschata*.

Fig. 4

Figure 4

Synteny analysis of GRAS genes among six Cucurbitaceae species. The collinear GRAS genes are connected by black lines. Gray ribbons represent pairwise collinear blocks connecting the GRAS genes between the genomes of six Cucurbitaceae species; the red lines represent the collinear GRAS genes that have maintained a ratio of 2:1 in *Cucurbita moschata* and the Benincasae group (*Cucumis sativus*, *Cucumis melo*, *Benincasa hispida*, *Citrullus lanatus* and *Lagenaria siceraria*); the magenta lines represent the collinear GRAS genes that have maintained a 1:1 ratio in *Cucurbita moschata* and the Benincasae group and the blue curves represent collinear GRAS genes that have maintained a ratio of 0:1 in *Cucurbita moschata* and the Benincasae group.

Fig. 5

Figure 5

Expression profiles of GRAS genes in *A. thaliana*, *S. lycopersicum* and six Cucurbitaceae species. In the heat maps, columns represent different tissues: roots, stems, leaves, flowers and fruits, and the rows represent genes. The color changes from white to black to represent relative gene expression ($\log_2(\text{FPKM} + 1)$); gray represents missing values. The Neighbor-joining tree of GRAS proteins is shown on the left. a, SCR; b, NSP2; c, SCL32; d, SCLB; e, RAD1 subfamilies.

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- [Supplementaryfile.pdf](#)
- [Supplementaryfile.xls](#)