A multilayered cross-species analysis of GRAS transcription factors uncovered their functional networks in plant adaptation to the environment

Moyang Liua,b,1, Wenjun Suna,b,1, Chaorui Li a, Guolong Yua, Jiahao Lia, Yudong Wanga, Xu Wang a,⇑

a Joint Center for Single Cell Biology, School of Agriculture and Biology, Shanghai Jiao Tong University, Shanghai 200240, China
b College of Life Science, Sichuan Agricultural University, Ya’an 625014, China

HIGHLIGHTS

• Polyploidization events were the main driving force for GRASs expansion.
• Evolutionary analysis reveals that inter-species GRASs functions may be conserved.
• GRASs helped plants resist stress by regulating flavonoids pathway.
• GRASs regulate flavonoid synthesis by crosstalk with auxin and photosynthetic pathway.
• Polyploidized GRASs play key roles in environment adaption, growth and development.

GRAPHICAL ABSTRACT

Article info

Article history:
Received 31 August 2020
Revised 7 October 2020
Accepted 24 October 2020
Available online 29 October 2020

Keywords:
GRAS
Environmental stress
Polyploidization
Plant population
Systems biology

ABSTRACT

Introduction: Environmental stress is both a major force of natural selection and a prime factor affecting crop qualities and yields. The impact of the GRAS [gibberellic acid-insensitive (GAI), repressor of GA1–3 mutant (RGA), and scarecrow (SCR)] family on plant development and the potential to resist environmental stress needs much emphasis.

Objectives: This study aims to investigate the evolution, expansion, and adaptive mechanisms of GRASs of important representative plants during polyploidization.

Methods: We explored the evolutionary characteristics of GRASs in 15 representative plant species by systematic biological analysis of the genome, transcriptome, metabolite, protein complex map and phenotype.

Results: The GRAS family was systematically identified from 15 representative plant species of scientific and agricultural importance. The detection of gene duplication types of GRASs in all species showed that the widespread expansion of GRASs in these species was mainly contributed by polyploidization events. Evolutionary analysis reveals that most species experience independent genome-wide duplication (WGD) events and that interspecies GRAS functions may be broadly conserved. Polyploidy-related Chenopodium quinoa GRASs (CqGRASs) and Arabidopsis thaliana GRASs (AtGRASs) formed robust networks with flavonoid pathways by crosstalk with auxin and photosynthetic pathways. Furthermore, Arabidopsis thaliana population transcriptomes and the 1000 Plants (OneKP) project confirmed that GRASs are components of flavonoid biosynthesis, which enables plants to adapt to the environment by promoting

https://doi.org/10.1016/j.jare.2020.10.004
2090-1232/© 2021 The Authors. Published by Elsevier B.V. on behalf of Cairo University.
This is an open access article under the CC BY-NC-ND license (http://creativecommons.org/licenses/by-nc-nd/4.0/).
Introduction

Plants live in constantly changing environments that are often adverse to their growth and development [1]. In particular, the frequent occurrence of extreme weather events due to global climate change has made plants more vulnerable to environmental stresses such as salt, abnormal temperatures, and drought stress, greatly increasing the vulnerability of agricultural production. How plants perceive stress signals and survive in adverse environments represents fundamental and key biological problems.

Polyploidy, or whole-genome duplication (WGD), is one of the key forces of ecological and evolutionary processes in plants [2]. Caused by abnormal environments, polyploidization, or WGD, can increase the adaptive plasticity of plants to environments and the genetic variability of plants [3]. Polyploidy is particularly common in plants, with all angiosperms sharing ancestral polyploid events and 24% of existing plant species being recent polyploids [2,4]. During evolution and genome diploidization, many duplicated genes have been lost or highly modified, and the number of duplicated chromosomes has been rearranged and reduced, leaving multiple duplicated genes without obvious cytological evidence of WGD [5]. This genetic redundancy caused by WGD allows new functional evolution and functional division between duplicates [6,7], leading to a reduced probability of lineage extinction [8,9]. There is increasing evidence that polyploids exhibit greater resistance to stress compared to diploids under the same genetic background [10,11]. Consequently, polyploidy may promote plant evolvability and improve the adaptability of the population under certain conditions [12]. After polyploidization, genes involved in signal transduction and transcriptional regulation are often expanded, e.g., transcription factor (TF) families, whose evolution and expansion may be related to WGD events [13] and can enhance plant resistance to environmental stress [14].

The GRAS family is named after the three first identified members, including gibberellic acid-insensitive (GAI), repressor of GA1–3 mutant (RGA), and scarecrow (SCR) [15]. As a plant-specific TF family, it plays a pivotal role in multiple signal transduction pathways controlling plant development and stress responses [16,17]. Most GRAs contain a highly conserved C-terminal GRAS domain and a variable N-terminal domain [18]. The GRAS domain is composed of five motifs, including LHRI, VHIID, LHR1, PFYRE, and SAW. In contrast, the N-termini are highly variable and involved mainly in molecular recognition [19,20]. Genome-wide identification of GRAS families has been performed in some species, including Glycine max [21], cassava [22], cucumber [23], barley [24], etc. The GRAS family is further divided into different subfamilies, including DELLA, HAM, LS, LISCI, NSP2, PAT1, SCR, SCL3, and SHR, according to different functions [19]. Although static characterization of GRAS has been carried out within individual plant genomes, the cross-species evolution of the GRAS family and GRAS regulatory network in plant adaptation to environments after polyploidization is still unclear.

A deep understanding of GRAS evolution in different plants after polyploidization is indispensable to improving crop quality and environmental adaptability. In addition to Arabidopsis thaliana, other plant species have been used as models to study more specific traits or characteristics that do not exist in the model plant. For example, rice has been widely studied as a model plant for species important for human food security, such as wheat, maize, and sorghum. In addition, Chenopodium quinoa (quinoa), as a polyploid crop with high nutrition, adapts to many environmental stresses and can thus be used as an ideal model of polyploid plants for GRAS research on stress resistance.

To remedy the gap in our knowledge of GRAS evolution after polyploidization, some representative plants were selected to explore the adaptive evolution of the GRAS regulatory network and to identify important genes in the network related to stress resistance and other agronomic traits. The 15 representative plant species of scientific and agricultural importance include model plants rice and Arabidopsis thaliana, a basal angiosperm Amborella trichopoda, Chenopodium quinoa, Solanum tuberosum, Solanum lycopersicum, etc. Most species in this study experienced independent WGD events, which further confirmed that the widespread expansion of GRASs in these species was associated with polyploidization. Genome synteny and evolutionary rates reveal that the function of GRASs may be broadly conserved among Chenopodium quinoa, Arabidopsis thaliana, Solanum tuberosum and Fagopyrum tataricum. According to a systematic integration of phenotype, genome, transcriptome, metabolome, and protein interaction data from different plant species and populations, GRASs interact with multiple pathways, e.g., auxin signaling and photosynthesis, to regulate flavonoid biosynthesis and plant resistance to various environmental stresses. With these results taken together, this study analyzed the expansion and evolution of GRASs and provides a comprehensive dissection of the GRAS regulatory network that improves plant adaptation to environments after polyploidization.

Materials and methods

Data sources and sequence retrieval

The genome sequences of the studied species Oryza sativa (diploid) [25], Arabidopsis thaliana (diploid) [26], Fagopyrum tataricum (diploid) [27], Beta vulgaris (diploid) [28], Chenopodium quinoa (allopolyploid) [29,30], Chenopodium palidicaule (diploid) [29,31], Chenopodium succedum (diploid) [29], Daucus carota (diploid) [32], Lactuca sativa (diploid) [33], Helianthus annuus (diploid) [34], Olea europaea (diploid) [35], Nicotiana attenuata (diploid) [36], Solanum tuberosum (autotetraploid) [37], Solanum lycopersicum (diploid) [38], Actinidia chinesis (heterozygous diploid) [39], Aquilegia coerulea (diploid) [40] and Amborella trichopoda (diploid) [41] were downloaded from the NCBI.

The public transcriptome data from studies of A. thaliana responses to different stimuli, including nutrients [42], mitochondrial stresses [42], chloroplast stresses [42], hormones [auxin (IAA) [43], gibberellin (GA3) [43], abscisic acid (ABA) [43], ethylene (ACC) [43], mescaline (MEJ) [43], brassinosteroid (BL) [43], zeatin (ZT) [43], and salicylic acid (SA) [44]), abiotic stresses (ozone [45], high light [46], H2O2 [44], heat [43], salt [43], UV [43], oxidative [43], and osmotic [43]), and biotic stresses (Botrytis cinerea [43], elicitor Fig22 [43], Erysiphe orontii [43], Phytophthora
infectans [43], Blumeria patens [47], elicitor EF-Tu [48] and E. cichoracearum [49]) were obtained. The transcriptome data for C. quinoa includes data from the roots, stems, leaves, flowers, and seeds (GEO: GSE139174). We also obtained proteomics datasets for A. thaliana, O. sativa and C. quinoa [50], as well as the DAP-seq dataset for A. thaliana [51]. In addition, transcriptome datasets from 144 natural A. thaliana accessions (GEO: GSE43858), the A. thaliana 1001 genomes project (GEO: GSE80744) and the 1000 Plants (OneKP) project [52] were obtained.

### Genome-wide identification of GRAs

The Hidden Markov Model (HMM) profile of the GRAS domain (PF03514) was downloaded from the Pfam database [53]. The HMM query in the HMMER v3.1 program was used to retrieve all GRAs from the 15 genomes with an e-value \( \leq 1e^{-5} \) [54] (Table S1, S2). The existence of GRAS domains in all genes identified was validated using HMMER [54], SMART [55], Pfam [56] and InterPro [57].

#### Identification of GRAS orthogroups in multiple species

The orthogroups of GRAs proteins identified from the 15 species were inferred with OrthoFinder 2 [58], with one copy of the most recent common ancestor of each of these species. Multiple sequence alignment of the GRAS proteins in each orthogroup was performed using MUSCLE 3.8.31 [59]. Then, the OrthoFinder 2 results were visualized as a species tree and gene trees with Mega 7 [60]. The orthogroups were divided into LISCL, PAT, HAM, SCL3, SHR, SCR, LAS, DLT, SCL4/7, DELLA, and unclassified subfamilies according to the recent common ancestor of each of these species. Multiple alignments of the duplicate gene pairs in the identified syntenic blocks in each genome or between different species were analyzed by BLASTP. The syntenic maps of 15 species were submitted to the Pfam and CDD databases to find other known domains/motifs beyond the GRAS domain. In addition, MEME v4.9.0 [61] was used to identify new conserved motifs not found in public databases (Table S3).

#### Distribution of GRASs on chromosomes and syntenic and comparative genomics analyses

Chromosome localization information of CqGRASs, AtGRASs, FtGRASs and StGRASs was obtained using gff and genome sequence files and visualized through Cirro [62]. The syntenic maps of A. thaliana, F. tataricum, C. quinoa, and S. tuberosum were constructed using Dual Synteny Plotter software to identify homologous GRASs within or between these species [63] (Table S4, S6). All proteins or GRAs between A. thaliana, C. quinoa, F. tataricum, and S. tuberosum or within each species were analyzed by BLASTP. The syntenic regions within each species and between species were identified by MCscan based on the above BLASTP results [64]. The protein sequences of homologous gene pairs in the identified syntenic region were aligned with MUSCLE and then converted to CDS alignments. The nonsynonymous amino acid substitution rates (Ka) and synonymous amino acid substitution rates (Ks) were calculated using the Ka/Ks calculator [65].

#### Identification of collinearity and specific duplication events

Gene collinearity analysis was performed using the default parameters in BLAST and MCScanX according to previous studies [66,67]. Different types of duplicated genes were speculated by the duplicate_gene_classifier program in MCScanX, and we inferred collinear genes to identify GRAs related to polyploidization. The potential anchoring points (E-value < 1e–5; top five matches) between each possible pair of chromosomes in multiple genomes were found using BLASTP. The loose E-value threshold accommodates the highly differentiated evolution of duplicated genes resulting from polyploidization millions of years ago. Protein sequences were searched against the given genome or genomes of other species. The homologous blocks in each genome or between different genomes were determined with ColinearScan (maximal gap \( \leq 50 \) genes; P-value < 0.05) [68]. The maximum gap between neighboring genes along a chromosome showing collinearity with genes along the corresponding chromosome sequence was set as 50 intervening genes [69].

#### In silico development-related phenotypic analysis

Phenotype data were retrieved from TAIR [70] and RARGE II [71]. The TAIR library contains gain- or loss-of-function phenotypes of GRAS mutants in A. thaliana. In the RARGEII database, the GRASs annotated with increased length, increased size, low saturation, and decreased height were highlighted.

#### Microarray and RNA-seq data analysis

Raw data from the A. thaliana Affymetrix ATH1 array were processed into normalized gene expression values for meta-analysis using RMAExpress v1.1.0. The RNA-seq data were derived from C. quinoa leaves, flowers, roots, fruits, and stems. Raw data were processed by FastQC v0.10.1, and the FPKM value of each gene was calculated according to its read count. The hierarchical clustering heatmaps of gene expression were plotted by Tree View software [72].

#### Gene ontology (GO) annotation

GO analysis was performed using Blast2GO gene ontology analysis tools (http://www.blast2go.com) to determine the molecular functions of differentially expressed genes [73]. GO analysis of target genes bound to AtGRASs was performed in the same way.

#### Network visualization and correlation network construction

The interaction networks of GRAs and flavonoid biosynthesis pathway genes, auxin signaling, photosynthesis, and flavonoids under all different stresses were constructed by using the R package imsbinfer in GitHub. The interaction networks in 144 natural A. thaliana accessions, the A. thaliana 1001 Genomes Project and the 1000 Plants (OneKP) project were constructed in the same way. The associations were calculated using Spearman’s rank correlation measure, and the red and blue edges represent positive and negative correlations.

#### Expressions of differentially expressed CqGRASs in two kinds of C. quinoa fruits

The pale-yellow and dark-yellow C. quinoa used in this experiment were grown in a greenhouse at 25°C and 16 h light/8h dark. The expression patterns of differentially expressed CqGRASs in two kinds of C. quinoa mature fruits were determined by qRT-PCR (Table S18). The online software primer 3 (http://frodo.wi.mit.edu/) was used to design the primers (Table S18). The gene Elongation Factor 1 alpha (EF1α) gene was used as an internal reference gene, and SYBR Premix Ex Taq II (TaKaRa) was used to perform qRT-PCR [74]. The 2–ΔΔCT method was used to calculate gene expressions [75]. Three technical replicates were performed for each biological replicate, and a total of three biological replicates was set for this experiment.
Fig. 1. Phylogeny, diversity and motif compositions of GRAS proteins in 15 species. A. The orthogroups of GRAS proteins identified from the 15 species (Oryza sativa, Arabidopsis thaliana, Fagopyrum tataricum, Beta vulgaris, Chenopodium quinoa, Daucus carota, Lactuca sativa, Helianthus annuus, Olea europaea, Nicotiana attenuata, Solanum tuberosum, Solanum lycopersicum, Actinidia chinensis, Aquilegia coerulea and Amborella trichopoda) were inferred with OrthoFinder 2. Then, the OrthoFinder 2 results were visualized as a species tree with Mega 7. The total number of genes, the numbers of genes in the orthogroups, and the number of unassigned genes are provided. Rows represent the species, and columns represent the orthogroups. The orthogroups were named in turn as different subfamilies, including LISCL, PAT, HAM, SCL3, SHR, SCR, LAS, DLT, SCL4/7, DELLA and unclassified subfamilies, according to the A. thaliana GRAS genes in each orthogroup. B. The phylogenetic relationship of GRAS genes of the LISCL subfamily (orthogroup 0) in different species was visualized with Mega 7 (Outer figure). In addition, MEME v4.9.0 was used to identify conserved motifs of GRASs in 15 species (inner figure). The motifs (numbered 1–10) are displayed in differently colored boxes. The sequence information for each motif is provided in Table S3. C. The phylogenetic relationship of GRAS genes of the PAT subfamily (orthogroup 1) in different species was visualized with Mega 7 (Outer figure). In addition, MEME v4.9.0 was used to identify conserved motifs of GRASs in 15 species (inner figure). The motifs (numbered 1–10) are displayed in differently colored boxes. The sequence information for each motif is provided in Table S3. D. The phylogenetic relationship of GRAS genes of the PAT subfamily (orthogroup 2) in different species was visualized with Mega 7 (The Outer figure). In addition, MEME v4.9.0 was used to identify conserved motifs of GRASs in 15 species (inner figure). The motifs (numbered 1–10) are displayed in differently colored boxes. The sequence information for each motif is provided in Table S3. E. The phylogenetic relationship of GRAS genes of the HAM subfamily (orthogroup 3) in different species was visualized with Mega 7 (Outer figure). In addition, MEME v4.9.0 was used to identify conserved motifs of GRASs in 15 species (inner figure). The motifs (numbered 1–10) are displayed in differently colored boxes. The sequence information for each motif is provided in Table S3.
Statistics

All data were analyzed using the Origin Pro 2020b statistics program, and the means were compared by the least significant difference test at the 0.05 level of significance.

Results

Identification, phylogenetic analysis, motif evolution and expansion of GRASs from representative plants

In total, 658 GRASs were identified from 15 representative plant genomes, with copy numbers ranging from 19 to 77 (Fig. 1A, Table S1), and were divided into 22 orthogroups based on their shared primitive ancestral copy genes. The orthogroups were renamed as different subfamilies based on the AtGRASs in each orthogroup (Table S2). Orthogroup 0 was renamed the LISCL subfamily, orthogroups 1, 2, 10 and 15 were renamed the PAT subfamily, orthogroups 3 and 11 were renamed the HAM subfamily, orthogroup 4 was renamed the SCL3 subfamily, orthogroups 5 and 6 were renamed the SHR subfamily, orthogroup 7 was renamed the SCR subfamily, orthogroup 8 was renamed the LAS subfamily, orthogroup 9 was renamed the DLT subfamily, orthogroups 12–14 were renamed the SCL4/7 subfamily, orthogroup 16 was renamed the DELLA subfamily, and orthogroups 17–22 were renamed the unclassified subfamily.
Polyploidization contributed to the expansion of GRASs in these species

The distribution and collinearity of GRASs in different species were analyzed (Fig. 2A-D), and GRASs were not evenly distributed on chromosomes in all species, with the number of segmental duplicated GRASs in C. quinoa being the most abundant (Fig. 2B, Table S4). It is speculated that the differences in the number of GRASs among species may be due to gene duplication. The Ks values of homologous pairs in these syntenic regions, as well as the mean Ks values of individual syntenic blocks, indicated that WGD events had occurred in the evolutionary history of these species (Fig. 2E, Table S4). Nevertheless, C. quinoa experienced an independent WGD event after differentiation from A. thaliana, F. tataricum and S. tuberosum, which may be an important driving force for the production of homologous GRAS in the C. quinoa genome (Fig. 2E).

Genome polyploidy provides rich possibilities for new variants, which probably produce polyploidy-related genes in evolutionary processes that face environmental stress. To further identify GRAS associated with ancestral polyploidization events (WGD or segmental), the Multiple Collinearity Scan toolkit (MCScanX) can be used to detect gene duplication types (Fig. 2F-J). Here, genes related to polyploidy were inferred according to the collinearity persistence of the region around the selected genes. GRASs were divided into four groups, including dispersed, proximal, tandem, and WGD or segmental, and CqGRASs were extremely overrepresented in polyploidy-related groups (Fig. 2H, Table S5). Similarly, the GRAS types in A. thaliana, F. tataricum and S. tuberosum were inferred (Fig. 2G, I, J), and it was found that most of the GRASs in A. thaliana were polyploidy-related (16. 48.48%) (Fig. 2G, Table S5), while 56.06% of the GRASs in F. tataricum were polyploidy-related (Fig. 2I, Table S5). The relationship between GRASs accumulation and polyploidization in the other 12 plants was also explored (Fig. 2K, Table S5). Except in Actinidia chinensis, S. tuberosum, Aquilegia coerulea, Lactuca sativa and O. sativa, polyploidization was the predominant mechanism of GRASs formation, accounting for 46.6%-94.73% of these homologues (Fig. 2K, Table S5). These phenomena indicated that whole-genome or segmental duplication plays an indispensable role in GRAS expansion, and the widespread expansion of GRAS in plants may be related to polyploidization. Further analysis of the correlation network of different types of GRASs expression in multiple species revealed that the expression of polyploidy-related GRASs was closely related (Fig. 2L-O, Table S4).

Genomic synteny conservation of GRASs in different species and phenotypic analysis of A6GRASs maps to GRASs of other species that can potentially regulate agronomic traits

It is worth investigating whether the function of polyploidy-related GRASs in multiple species is similar. Thus, the genome sequences of C. quinoa, A. thaliana, S. tuberosum and F. tataricum were compared and identified in multiple syntenic regions, among which C. quinoa and F. tataricum had the most orthologous pairs (Fig. 3A, Table S6). A correlation network analysis was performed on the expression levels of orthologous GRASs of C. quinoa, A. thaliana, S. tuberosum and F. tataricum (Fig. 3B, Table S6). Strikingly, the expression of the orthologous GRAS pairs of these species is tightly correlated (Fig. 3B, Table S6), and this correlation is in sharp contrast to the absence of such a correlation in a random gene set (P < 0.05) (Fig. 3C, Table S6). C. quinoa is a hybrid of the ancestral A-genome diploid C. pallicidae and ancestral B-genome diploid C. suecicum species [76]. By comparing the genome sequences of C. quinoa with those of C. pallicidae or C. suecicum, it was found that the number of orthologous GRAS pairs between them was the same (Fig. 3I, Table S6). Moreover, Ks/Ks analysis of gene homology genes in these species showed that the GRAS family was conserved and retained in the process of species evolution.
These results indicate that the functions of GRASs may be broadly conserved among *C. quinoa*, *A. thaliana*, *S. tuberosum*, and *F. tataricum*.

To further explore the potential functions of GRASs, publicly available phenotypic data were queried from TAIR and RARGE II to generate a list of 7 AtGRASs and the occurrence of mutations that would result in development-related phenotypes (Fig. 3E, F). Six of the AtGRAS mutations affected the seed phenotype, while the AT3G49950 mutant decreased the whole plant height (Fig. 3E). It was found that most of the AtGRASs were classified as WGD or seg-
mental (Fig. 3G). To further excavate important GRASs, these 7 AtGRASs were mapped to CqGRASs, FgGRASs and StGRASs with a view toward identifying syntenic genes that might perform similar functions (Fig. 3H). AtGRASs and CqGRASs formed 5 pairs of one-to-
many synteny pairs, AtGRASs and FgGRASs formed 5 pairs of syn-
teny pairs, among which were 3 one-to-one synteny pairs, while AtGRASs and StGRASs formed 4 pairs of synteny pairs, with 2 one-
to-one synteny pairs (Fig. 3H, Table S7). These results mapped some potential development-related candidate genes by identifying GRASs of other species with synteny to AtGRAS mutants, but the specific functions of these genes need to be confirmed by addi-
tional studies.

CF-MS to detect protein interactions and DNA affinity purification sequencing (DAP-seq) mapping genome-wide TF binding sites to jointly determine GRAS functions and phenotypes

Determining protein-protein interactions is another key method by which to analyze gene and protein functions. The interactions between GRASs and other proteins identified by co-
fractionation MS (CF-MS) [50] in A. thaliana, O. sativa and C. quinoa revealed 831 conserved protein complexes (Fig. 4A-B, Table S8). Gene ontology (GO) analysis of these conserved proteins interact-
ing with GRASs showed that they are mainly concentrated in the catalytic activity pathways that play an important role in organ-
isms (Fig. 4C-D). By further exploring the interaction network of GRAS proteins in A. thaliana (Fig. 4E), it was found that AT2G37650, which affects seed saturation, forms a close network with other proteins (Fig. 3C, Fig. 4F, J). GO analysis of the proteins interacting with AT2G37650 indicated that they were mainly enriched in cellular and metabolic processes and in catalytic activ-
ity and binding functions, indicating that they may jointly partici-
pate in the regulation of seed saturation (Fig. 4C, Table S10). This interactions were analyzed using transcriptome data from 144 A. thaliana accessions (Fig. 4H), which revealed that AT2G37650 formed a robust network with its directly interacting genes (Fig. 4I, Table S9). Further exploration of their expressions in differ-
ent tissues showed that the expressions of 16 genes in fruit were similar to that of AT2G37650, and these genes could jointly regulate seed saturation (Fig. 4K). The GO analysis of interacting genes with similar expressions of AT2G37650 indicated that they were mainly enriched in binding terms, which further illustrates their participa-
tion in biological processes through protein interaction (Fig. 4L, M).

Interestingly, the interaction networks of GRAS proteins in C. quinoa were analyzed (Fig. 4N), indicating that AUR62009556 and AUR62039362, which are homologous to AT2G37650, also form a close network with other proteins (Fig. 4O). Furthermore, proteins interacting with AUR62009556 and AUR62039362 were also mainly enriched in metabolic processes and binding and catalytic activity functions (Fig. 4P, Table S11). These results suggest that AUR62009556 and AUR62039362 may regulate seed saturation by participating in metabolic pathways, and the two genes may have functional redundancy. The interaction networks in two kinds of C. quinoa fruits were analyzed, which indicated that only AUR62039362 formed a robust network with other genes, suggest-
ing that AUR62039362 might regulate seed phenotype with other genes (Fig. 4Q, Table S12). The interaction gene sets of AUR62039362 in pale-yellow and dark-yellow C. quinoa fruits were also analyzed, and 31 conserved genes interacted with them (Fig. 4R, Table S11). GO analysis of these conserved genes revealed mainly the enrichment of catalytic activity categories, while the category differences in specific interaction sets in two kinds of C. quinoa fruits also illustrated the importance of AUR62039362 and interaction genes in regulating seed saturation (Fig. 4S-U).

Polyplody-related CqGRASs regulate flavonoid synthesis through auxin and photosynthesis pathway crosstalk

To further verify the functions of the above potential regulators, the transcripts of 44 CqGRASs in RNA-seq data from two kinds of C. quinoa were analyzed. All CqGRASs were hierarchically clustered based on expression in at least one tissue (Fig. 5A). Furthermore, 12 differentially expressed genes (DEGs) in two kinds of C. quinoa fruits were identified, including AUR62039362, which potentially regulates seed saturation (Fig. 5B). The qRT-PCR results are basi-
cally consistent with those of transcriptome analysis, which fur-
ther confirmed the differential regulatory role of these genes in C. quinoa fruits (Fig. S2). The GO enrichment of these DEGs indi-
cated that they have functional differences in protein binding (Fig. 5C). To further explore the function of DEGs, the expressions of differentially expressed proteins bound to AUR62039362 in pale-yellow and dark-yellow C. quinoa fruits were determined (Fig. 5D). Meanwhile, the differential accumulation of anthocyanin, flavonoid and isoflavonol in the pale-yellow and dark-yellow C. quinoa fruits was found (Fig. 5E, F, Table S13). It is speculated that differentially expressed CqGRASs may interact with other genes to regulate the differential accumulation of metabolites. Furthermore, the directed correlation network of differentially expressed CqGRASs with their binding DEGs and differential metabolites was evaluated in C. quinoa fruits (Fig. 5G). Strikingly, the expres-
sions of the CqGRASs are tightly correlated with their binding genes and metabolite network in C. quinoa (P < 0.05) (Fig. 5G, Table S14). Meanwhile, AUR62039362 and its interacting genes and flavonoids, such as rutin, formed a robust directed regulatory network

Fig. 3. Comparative syntenic maps and correlation network of GRASs in these representative species and phenotypic analysis of AtGRASs maps on to GRASs of other species that can potentially regulate agronomic traits A. The synteny relationship of orthologous GRASs in Chenopodium quinoa and Arabidopsis thaliana, Chenopodium quinoa and Fagopyrum tataricum, Chenopodium quinoa and Solanum tuberosum, Fagopyrum tataricum and Arabidopsis thaliana, Solanum tuberosum and Arabidopsis thaliana, Solanum tuberosum and Fagopyrum tataricum. Gray lines in the background indicate synteny blocks within plant genomes, while red lines highlight syntenic GRAS gene pairs. B. The correlation network of the expression levels of syntenic GRAS gene pairs of Arabidopsis thaliana, Fagopyrum tataricum, Chenopodium quinoa, and Solanum tuberosum. Diamonds represent GRAS, blue diamonds represent SGRAS, green diamonds represent AGRAS, dark-yellow diamonds represent CqGRASs and pale-yellow diamonds represent FGRASs. The diamonds size represent the degree of genes in the correlation network, that is, the larger the circle, the more genes associated with the gene C. The correlation network of the expression levels of GRAS and random genes of non-syntenic GRASs. Circles represent random genes that are not GRAS. Blue circles represent random genes of Solanum tuberosum, green circles represent random genes of Arabidopsis thaliana, dark-yellow circles represent random genes of Chenopodium quinoa and pale-yellow circles represent random genes of Fagopyrum tataricum. The circles size represent the degree of genes in the correlation network, that is, the larger the circle, the more genes associated with the gene D. The Ka/Ks ratios represent the evolution rate of GRAS genes. Different colored lines represent the Ka/Ks ratios of homologous GRAS gene pairs within or between different species, The Ka/Ks ratios less than 1 indicates that the genes have undergone purifying selection E. 7 Arabidopsis GRAS gene mutations will lead to different phenotypic changes. Clustering of 7 Arabidopsis according to the type of plant phenotype change after mutation. Orange squares indicate corresponding phenotypic changes after the mutation. F. The proportion of GRAS mutants to all GRAS genes in Arabidopsis thaliana. G. The distribution of gene duplication types of AtGRASs and CqGRASs. Gene types of GRAS mutant genes in Arabidopsis thaliana. Multiple Collinearity Scan toolkit (MCScanX) was used to detect duplication types to identify GRAS associated with ancestral polyploidization events (WGD or segmental). The GRAS related to polyploidy were inferred according to the collinearity persistence of the region around the selected genes. GRASs were divided into four groups, including dispersed, proximal, tandem, and WGD or segmental. H. The collinearity between AtGRASs with altered phenotypes after mutation and GRAS genes of Arabidopsis thaliana, Fagopyrum tataricum, Chenopodium quinoa, and Solanum tuberosum. The orange line represents the collinear gene pairs formed by AtGRASs and CqGRASs. The purple line represents the collinear gene pair formed by AtGRASs and SGRASs. The yellow lines represent collinearity gene pairs formed by AtGRASs and FGRASs. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)
Fig. 4. Co-fractionation MS detects protein interactions and DNA affinity purification sequencing (DAP-seq) mapping genome-wide TF binding sites. A. Detection of GRAS protein interacting orthogroups in Arabidopsis, rice and quinoa by Co-fractionation MS. Orange represents Arabidopsis, green represents quinoa, and purple represents rice. The red bubbles represent the orthogroups of proteins interacting with GRASS. B. Venn diagram of the orthogroups of proteins interacting with GRASS in rice, Arabidopsis, and quinoa. The intersection represents a conserved orthogroup of proteins interacting with GRAS proteins in Arabidopsis, rice and quinoa. C. Molecular function analysis of conserved orthogroup of proteins interacting with GRAS proteins in Arabidopsis, rice and quinoa. The color of the graph represents the enrichment degree of interacting proteins with GRAS in GO term. The deeper the color is, the more significant the enrichment is. D. GO enrichment analysis of conserved orthogroup of proteins interacting with GRAS proteins in Arabidopsis, rice and quinoa. Bubble size represents the amount of proteins that interact with GRAS proteins in GO term, and the larger the bubble, the more proteins that interact with GRAS proteins. E. Interaction network between ATGRAS proteins and its interacting proteins in Arabidopsis thaliana. F. At2g37650 mutation will affect seed saturation. The red dotted circle represents an interaction network between AT2G37650 and other proteins. G. The GO analysis of the proteins interacting with AT2G37650. H. This interaction of Arabidopsis ENOG411DSHQ orthogroup was analyzed using transcriptome data from 144 Arabidopsis accessions. I. The red dotted circle represents a robust interaction network between AT2G37650 and its directly interacting genes analyzed using transcriptome data from 144 Arabidopsis accessions. J. Phenotypic changes induced by AT2G37650 mutation. K. Expression levels of genes interacting with AT2G37650 in roots, stems, flowers, leaves and fruits. The closer it is to red, the higher the level of gene expression. L. The genes with similar expression as AT2G37650 in fruits were regarded as concerned genes. This figure shows the GO analysis of genes interacting with concerned genes. Bubble size represents the amount of proteins that interact with concerned genes in GO term, and the larger the bubble, the more proteins that interact with concerned genes. M. The genes with similar expression as AT2G37650 in fruits were regarded as concerned genes. This figure shows the molecular function analysis of genes interacting with concerned genes. The color of the graph represents the enrichment degree of interacting proteins with concerned genes in GO term. The deeper the color is, the more significant the enrichment is. N. Interaction network between CpGRAS proteins and its interacting proteins in quinoa. O. The red dotted circle represents an interaction network between the CpGRAS proteins homologous to AT2G37650 and other proteins. P. GO annotation analysis of proteins interacting with AUR6200956 and AUR62039362. Q. Correlation analysis between AUR62039362 and interaction gene sets in pale-yellow and dark-yellow C. quinoa fruits. R. Venn diagram of the orthogroups of proteins interacting with AUR62039362 in pale-yellow and dark-yellow C. quinoa fruits. The intersection represents a conserved orthogroup of proteins interacting with AUR62039362 in pale-yellow and dark-yellow C. quinoa fruits. S. Molecular function analysis of proteins interacting with AUR62039362 in two kinds of quinoa common region. T. Molecular function analysis of proteins interacting with AUR62039362 in dark-yellow quinoa specific region. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)
(P < 0.05) (Fig. 5H, Table S14). The core genes interacting with AUR62039362 were mainly enriched in metabolic process and phosphotransferase activity terms (Fig. 5I, J).

The above differentially expressed GRASs include PAT subfamily members involved in phytochrome-mediated light signal transduction and plant defense [77], LISCL subfamily members involved in auxin and stress-induced signals [19], and SHR subfamily members that regulate root development [78] (Fig. 5B). This leads us to speculate whether the differentially expressed GRASs might cross-talk with these pathways to regulate flavonoid synthesis. The correlation networks based on the expression of these GRASs and the auxin pathway, photosynthesis, flavonoid biosynthesis pathway genes and metabolites in pale-yellow and dark-yellow C. quinoa were further constructed (Fig. 5K, M), and these GRASs were closely related with these pathways in both C. quinoa types (Fig. 5L, N, Table S15). It was further found that there are two conserved genes
in the most robust correlation network of the two kinds of C. quinoa, which interact most closely with other genes (Fig. 5O, Table S15). The degree analysis of the close correlation network found that GRASs, photosynthesis, auxin and flavonoid biosynthesis pathways together regulate metabolite synthesis in pale-yellow C. quinoa (Fig. 5P), which is in sharp contrast to the absence of GRASs and photosynthesis in the close correlation network of dark-yellow C. quinoa (Fig. 5Q). The differential flavonoid biosynthesis pathway regulatory network may reveal the differential performance of the two kinds of C. quinoa responses to the environment.

Universal applicability correlation network of GRASs, photosynthesis, auxin and flavonoid biosynthesis pathways

In A. thaliana, the universality of the correlation network of GRAS, photosynthesis, auxin and flavonoid biosynthesis pathway genes was further validated. The correlation network between AtGRASs and photosynthesis, auxin and flavonoid biosynthesis pathway genes under different conditions was further validated (Fig. 6A, Table S17). Notably, the expression of AtGRASs is tightly correlated with photosynthesis, auxin and flavonoid biosynthesis pathway genes under all environmental stresses, and interaction networks under abiotic stress were more enriched compared to those under other stresses (P < 0.05) (Fig. 6A). Meanwhile, most AtGRASs, photosynthesis, auxin and flavonoid biosynthesis pathway genes were induced under nearly all abiotic stress conditions, indicating that they were extensively involved in responses to environmental stress (Fig. 6B, Table S16).

To explore the widespread nature of the network, a correlation network analysis of the expression levels of AtGRASs and photosynthesis, auxin and flavonoid biosynthesis pathway genes across 144 natural A. thaliana accessions [79] was performed (Fig. 6C). Notably, the expression of AtGRASs is closely related to the photosynthesis, auxin and flavonoid biosynthesis pathway gene network (Fig. 6C, Table S17), which is in sharp contrast to the absence of such a correlation in a random gene set (P < 0.01) (Fig. 6D, Table S17). The A. thaliana 1001 Genomes Project (GE: GSE80744) (Fig. 6E, F, Table S17) and 1000 Plants (OneKP) project [52] also validated this network (P < 0.01) (Fig. 6G, H, Table S17), suggesting that GRASs may regulate plant growth and development by participating in cross-talk among multiple pathways.

Discussion

GRASs play a central regulatory role in plant growth, development, and stress responses [80,81]. They are therefore promising targets for improved crop breeding. Here, 658 GRASs were identified from 15 species and divided into 22 orthogroups. The heterotetraploid nature of C. quinoa and the large scale of the GRAS family provide an ideal opportunity to explore the evolution of genes after gene duplication and polyploidization. WGD have events occurred in most species—particularly, C. quinoa, A. thaliana, S. tuberosum and F. tataricum—through independent WGD events (Fig. 2F). The evolution of GRASs after polyploidization was explored, which indicated that the widespread expansion of GRASs in most species is related to polyploidization, and these polyploidy-related GRASs may contribute to plant development and adaptation to environmental stress. The high number of GRASs observed in C. quinoa, A. thaliana, and F. tataricum is mainly a result of the significant expansion of the LISCL subfamily. This is consistent with reports on other species [82]. The expansion of the GRAS family is limited to a few subfamilies, and the number of GRASs in most subfamilies is constant. Moreover, several subfamilies lack members in one or more species. The absence of representative common orthologs in closely related species suggests that the loss predated the origin of some subfamilies. In particular, there was a lower number of 19 GRASs in B. vulgaris, and members of 13 orthogroups were missing (Fig. 1). In a similar study of other transcription factors, no complete loss of subfamily members in a given species was found [83]. Notably, most subfamilies missing members of one or more species belong to subfamilies containing multiple orthogroups [82]. In the unclosed subfamily, most species of multiple orthogroups lack GRASs, which suggests that these species may have lost subfamilies that are not important for their growth and development during evolution. The lack of members of some species in these orthogroups indicates that despite their ancient differentiation, the members of the GRAS family of close orthogroups remain redundant to a certain extent, and the complete loss of representative orthogroups in a given species seems to be compensated by the presence of members of the close orthogroups.

Previous studies have identified the GRAS family in a single or several species but have not systematically explored the evolutionary mechanisms of plant GRASs after polyploidization [84–86]. By exploring the genomic collinearity among different species, it was found that the functions of polyploid-related GRASs in some plants were more conserved. Furthermore, the evolution rates of different species showed that the functions of these genes were relatively conserved. The higher evolution rate of homologous genes in C. quinoa suggested that CqGRASs might be more conducive to adapting to environmental changes. As plants grow in changing environments, crosstalk between pathways is an absolute requirement for robust growth. Studies have shown that there is crosstalk between different signaling pathways or between hor-
mones and developmental pathways [87–89]. Light signaling plays a critical regulatory role in plant morphogenesis, metabolism, growth and development, and the core process of the light signaling pathway is to establish the relationship between light signaling and related gene expressions [90]. Flavonoids play important roles in a variety of physiological processes [91], and the accumulation of flavonoids in plant tissues depends on the availability of light signals, but the potential regulatory network of light signals controlling flavonoid biosynthesis is rarely known. The biosynthetic process of flavonoids is complex and usually activated by a variety of flavonoid biosynthetic genes, including early biosynthetic genes [92], while flavonoid biosynthesis is also affected by various plant hormones, such as auxin [93], methyl jasmonate [94,95], salicylic acid [94,95], etc. Both the light signaling pathway and the auxin pathway play pivotal roles in regulating flavonoid synthesis. As a positive regulator, AtPAT1 from the PAT subfamily participates in the phytochrome A (phy A)-specific signaling pathway and plays an important role in the early stage of the phy A signaling cascade [77]. Moreover, LISCL subfamily members are reported to play important regulatory roles in auxin signaling [19]. The differentially expressed GRASs identified in two kinds of C. quinoa fruits belong to the LISCL and PAT subfamilies, which raised an interesting question concerning whether these GRASs might regulate flavonoid synthesis by cross-talk between these pathways.
This study identified differentially expressed GRASs, which, in combination with other interacting genes, result in differences in the flavonoid contents of two kinds of C. quinoa fruits (Fig. 5). Furthermore, a correlation network of differentially expressed GRASs with the light signaling pathway, auxin signaling pathway and flavonoid biosynthesis pathway was constructed, which confirmed that GRASs can regulate flavonoid synthesis by crosstalk between multiple pathways. Interestingly, GRASs in A. thaliana can regulate the synthesis of flavonoids by crosstalk between multiple pathways under various environmental stresses. Meanwhile, GRASs have formed a robust correlation network with these pathways in A. thaliana populations and 1000 plants, which confirms that GRASs, as a core component, may regulate the synthesis of plant flavonoids through multiple cross-talk pathways. These in-depth studies will identify and confirm interactions and further clarify the role of interactions in molecular networks and biological processes. Nevertheless, despite our growing understanding of the role of TFs in these processes, the transcriptional regulation of genes and their complex qualitative and quantitative cooperativity between proteins and DNA remain to be investigated.

Conclusions

Our work not only provides a panoramic view of the evolution and expansion of GRASs in important species but also provides the first step for more detailed investigations of the functional diversity of GRASs of important nutritional crops after polyploidization. Our data highlight the universality of the robust correlation network between polyploid-related GRASs achieved by crosstalk between multiple pathways and flavonoid pathways. Furthermore, the population transcriptome emphasizes the universality of the key features of polyploidy-related GRASs in the flavonoid synthesis pathway, which is a role implying that they regulate plant metabolism by being involved in multiple pathways to achieve plant environmental adaptation. The interactive proteins that we identified may share functions, and although further identification of GRAS roles is required, information on their stable interaction with other proteins or genes will guide future research. Comprehensive and systematic analysis of the regulatory network of GRAS adaptation to the environment after polyploidization through multomics and phenotype may contribute to understanding the gene expansion mechanisms of important crops under polyploidy and provide valuable resources for important crop breeding.

Compliance with ethics requirements

This article does not contain any studies with human or animal subjects.

Author Contributions

M.-Y.L. planned and designed the research. M.-Y.L., C.-R.L., J.-H. L. and G.-L.Y. performed the experiments and analyzed the results. W.-J.S. and X.W. wrote the original manuscript. W.-J.S. and Y.-D.W. revised the manuscript. M.-Y.L. and X.W. supervised the research. All authors read and approved the final manuscript.

Declaration of Competing Interest

The authors have declared no conflict of interest.

Acknowledgements

This research was sponsored by National Natural Science Foundation of China (32071160), Shanghai Pujiang Program (19PJ1406600), Shanghai Sailing Program (20YF1422000), Natural Science Foundation of Shanghai (20ZR1427900) and the Shanghai Jiao Tong University (WF220415005 & 20X100040052).

Appendix A. Supplementary material

Supplementary data to this article can be found online at https://doi.org/10.1016/j.jare.2020.10.004.

References

[1] Zhu J-K. Abiotic stress signaling and responses in plants. Cell 2016;167:313–24.
[2] Forrester N, Rebolledo-Gómez M, Sachs J, Ashman T-L. Polyploid plants obtain greater fitness benefits from a nutrient acquisition mutualism. New Phytol 2020;227.
[3] Van de Peer Y, Mizrachi E, Marchal K. The evolutionary significance of polyploidy. Nat Rev Genet 2017;18.
[4] Barker M, Arrigo N, Baniaga A, Li Z, Levin D. On the relative abundance of autopolyploids and allopolyploids. New Phytologist 2015;210.
[5] Felfern G, Casacuberta J, Wendel J. Genomics of evolutionary novelty in hybrids and polyploids. Front Genet 2020;11.
[6] Freeing M, Scanlon M, Fowler J. Fractionation and subfunctionalization following genome duplications: Mechanisms that drive gene content and their consequences. Curr Opin Genet Dev 2015;35:110–8.
