

Genome-Wide Analysis of the GRAS Gene Family Exhibited Expansion Model and Functional Differentiation in Sea Buckthorn

Liyang Yu

Chinese Academy of Forestry

Guoyun Zhang

Chinese Academy of Forestry

Zhongrui Lyu

Chinese Academy of Forestry

Caiyun He (✉ hecy@caf.ac.cn)

Chinese Academy of Forestry <https://orcid.org/0000-0001-7155-9686>

Jianguo Zhang

Chinese Academy of Forestry

Research article

Keywords: Sea buckthorn, GRAS, Expansion models, Expression patterns

Posted Date: August 21st, 2020

DOI: <https://doi.org/10.21203/rs.3.rs-55445/v1>

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

[Read Full License](#)

Abstract

Background: GRAS proteins comprise a large family of transcription factors that experienced extensive replication, and play important roles in many aspects of growth regulatory and environmental signals. However, limited information was available about the GRAS genes in sea buckthorn of Elaeagnaceae. We perform a genome-wide investigation into the expansion model and tissue expression profiling of sea buckthorn GRAS genes based on the comparative genomes methods.

Results: In this study, 62 sea buckthorn GRAS (HrGRAS) genes were identified and renamed based on their respective chromosome distribution. Fifty-nine HrGRASs were further classified into nine subgroups and three HrGRASs did not belong to any of the subfamilies according to their phylogenetic features. HrGRAS genes tend to have a representative GRAS domain, few introns and unevenly distributed on chromosomes. Collinear homology dotplot and the median synonymous substitution sites of collinearity blocks distinguished gene pairs with different duplication models and found that segmental duplication was the main driver of the GRAS gene family expansion in sea buckthorn, followed by whole genome duplication and tandem duplication. The all gene pairs from the different duplication models may experience strong purifying selection pressure during evolution processes according to the Ka/Ks ratios. Expression profiles derived from transcriptome data exhibited distinct expression patterns of HrGRAS genes in various tissues and fruit developmental stages. The interaction network analysis of HrGRAS protein found that some HrGRAS proteins interacted with more proteins and retained more copies in sea buckthorn, which indicate that the products of these genes play more important roles in the growth and development of sea buckthorn.

Conclusions: Sea buckthorn GRAS gene family was first comprehensively analyzed in this study. This systematic analysis provided a foundation to further understand the expansion and potential functions of GRAS genes with an aim of sea buckthorn crop improvement.

Background

Sea buckthorn, a perennial woody plant, is a unique and valuable plant currently cultivated in various parts of the world, naturally distributed widely in China, Mongolia, Russia, and most parts of North Europe and also introduced in North and South America [1]. With the development of sea buckthorn resources, in addition to its ecological protection and food production, it has also begun to be widely used in medicine [2–4]. GRAS proteins play important roles in many aspects of growth regulatory and environmental signals [5, 6]. Considering the important role played by sea buckthorn in daily life and the importance of the GRAS gene family, it is necessary to investigate the GRAS genes in sea buckthorn to improve breeding for plant resistance and fruit yield. There have been no reports of sea buckthorn GRAS genes, although these proteins play important roles in the regulation and growth of plant.

The GRAS transcription factors are a large family in higher plants, whose name was derived from the three originally identified genes: GAI, RGA and SCR [7, 8]. An obvious feature of the GRAS proteins is that

there are very conserved sequence in their C-terminus, including the motifs of LHR I, VHIID, LHR II, PFYRE and SAW [9, 10]. The function of continuous conserved motifs in the C-terminal region has been reported in many studies, which may help maintain the common gene function of members of the GRAS gene family. In contrast, N-terminus of GRAS proteins varies in length and sequence except for two motifs (DELLA and TVHYNP) characterized for the proteins in DELLA subgroup, which may be a major factor in the functional specificity of each gene [9, 11].

GRAS genes have been characterized in a large number of species, and the many roles they play in plant growth and development have also been identified [6, 12–14]. The GRAS genes in plants such as plum, tomato, and corn are divided into eight to 16 subfamilies, although the members of this gene family in *Arabidopsis* and rice are classified into eight subfamilies, which indicating new subfamilies may exist in plants without GRAS gene identification [15]. In gene function, AtSCR may help maintain cellular status in radial pattern formation of root and shoot, and similar functions may also exist in the surrounding initial cells [16]. AtGAI, AtRGA and AtRGL have been found to inhibit the transfer of GA signals [17]. AtSCL13 has been shown to be involved in plant phytochrome-B (phyB) signal transduction [18], and three genes (PAT1, SCL5 and SCL21) from the same PAT1 subfamily can regulate plant growth and development by influencing the phytochrome-A (phyA) signaling pathway [6]. The OsMOC1 from the rice GRAS gene family has been shown to affect rice tillering, which is crucial for affecting rice yield [19]. The function of the GRAS genes in many plants needs to be further elucidated, although they have some functional characteristics in model plants.

Sea buckthorn is a unique and valuable plant which has tremendous value for medical researching, ecological protection and providing daily necessities [4, 20, 21]. To date, transcription factor families, such as WRKY have already been identified and characterized in sea buckthorn, although the gene family analysis based on transcriptome [22]. However, GRAS genes in sea buckthorn and their functional specificity have not been characterized. We conduct a genome-wide analysis for the GRAS gene family and explore the crucial function these genes play in the physiological processes and adaption to environment based on the genome we have sequenced. Here, we describe on the first characterization of the entire GRAS gene family of transcription factors in sea buckthorn. This work identified 62 HrGRAS genes, with analyzing the basic property, gene classification, phylogenetic comparison, gene structure and motif composition. In addition, we identified duplication models of the GRAS genes by colinearity analysis within the sea buckthorn genome and with *Arabidopsis* and grape. Finally, the expression pattern of sea buckthorn GRAS genes in different tissues and fruit developmental stages was investigated. This study provides details to further illuminate the expansion and functional characterization of GRAS genes in sea buckthorn.

Results

Genome-wide identification of GRAS gene family in sea buckthorn and sequence analysis

We found 62 non-redundant GRAS gene family members in sea buckthorn. All of GRAS genes could be mapped on the chromosomes and were renamed from HrGRAS1 to HrGRAS62 based on their order on the chromosomes (Additional file 1). The molecular mass and length of HrGRAS proteins varied greatly, with MWs ranging from 43.8 to 89.7 KDa and length from 384 to 806 aa. The average theoretical pI was 5.76, implying that most HrGRAS proteins were weakly acidic. Most (57 of 62) HrGRAS proteins had an instability index of more than 40, which means they were considered unstable, while a few (5 of 62) were predicted as stable. All HrGRASs were considered hydrophilic because the less GRAVY value (< 0) of each protein. The aliphatic indices of HrGRAS proteins range from 70.46 to 100.09, and most GRAS proteins contained many aliphatic amino acids.

Phylogenetic analysis, classification and functional characterization of HrGRAS family

The full-length GRAS proteins of three species (sea buckthorn, *Arabidopsis* and rice) were used to perform phylogenetic analysis in order to reveal the evolutionary relationship among HrGRAS proteins and their classifications. An unrooted phylogenetic tree was constructed (Fig. 1), demonstrating that most GRAS proteins could be classified into nine distinct subfamilies based on clade support values and classification from *Arabidopsis* and rice. They were termed as DELLA, PAT1, SCL3, SHR, SCR, LISCL, HAM, LAS and DLT, respectively. It was to be expected that no protein was classified into the Os4 subfamily, which was considered unique to rice in previous studies. Three HrGRAS proteins did not belong to any of the subfamilies.

Highly similar sequences tend to have highly similar structures and functions. Therefore, the function of HrGRAS proteins could be predicted based on the phylogenetic tree. Considering the important role of AtPAT1 and AtSCL13 in phytochrome A and phytochrome B signaling pathway, respectively, the HrGRAS proteins in this subfamily were predicted that associate with phytochrome signal transduction. In previous reports, AtSHR and AtSCR proteins played an important role in root and shoot radial patterning, which meant that HrGRAS proteins in this subfamily may possess the similar functions. Given that the function analysis of the proteins in LAS and DLT subfamilies, we reasonably speculate that the HrGRAS members in these subfamilies very likely participated in the regulation of the initiation of axillary meristems and brassinosteroid signal pathway, respectively.

Gene structure and motif composition of HrGRAS gene family

We examined the genetic structure of GRAS family to gain a deeper understanding of its evolution in sea buckthorn. As exhibited in Fig. 2, Most HrGRAS proteins contained very few introns (0 or 1). HrGRAS31 had up to seven introns, while the number of introns for other genes was no more than three. Of the 26 GRAS genes that contained introns, 17 had introns between UTRs, and no introns were adjacent to the DELLA proteins. The intron phase was also analyzed that almost all introns of the HrGRAS genes were phase 0, showing the extreme conserved. Most (55 of 62) of these proteins contained a representative GRAS domain (PF03514.11), while others contained two these domains.

According to the motif analysis results from the MEME (<http://meme-suite.org/>), we constructed a representation of all HrGRAS protein structures. As shown in Fig. 3B, most motifs were widely distributed in the HrGRAS genes, for example, motifs 1, 3, and 5 were distributed in all HrGRAS genes. At the same time, we also found that HrGRAS proteins in the same group share similar motif composition, for example, there was no motif 2 in each member of the HAM subfamily. The highly similar motif distributions were confirmed in clustered HrGRAS pairs, likely HrGRAS7 / 8, HrGRAS29 / 30. There were similar motif arrangements among HrGRAS proteins in the same subgroups which showed some conservation in a specific subfamily. The role these conserved motifs played in the growth of organisms was unclear and needed further research. The group's classification was strongly supported by the similarity of motif composition and structure in the same subfamily, as well as the results of phylogenetic analysis (Fig. 3A).

Collinearity and duplication models analysis of HrGRAS Gene

Considering the duplication events played an important role in the occurrence of novel functions and gene family expansion, we analyzed the duplication events of HrGRAS genes in the sea buckthorn genome in order to determine the duplication type of HrGRAS genes. The HrGRAS genes were unevenly distributed on the 10 sea buckthorn chromosomes except Chr8 and Chr9 (Fig. 4). The largest number of GRAS genes was mapped to Chr2. Three pairs of tandem duplicated genes located on Chr2, Chr5 and Chr11, while other HrGRAS genes considered to have originated from segmental duplications or whole genome duplication (WGD), apart from 15 HrGRAS genes were considered to be dispersed.

Furthermore, Microsynteny analysis was employed within the sea buckthorn genome to distinguish the HrGRAS genes that originated from whole-genome duplication or segmental duplication (Additional file 2). By marking the median value of synonymous substitution sites (K_s) of collinear blocks on the homologous dotplot and the complementary characteristics of the collinear blocks, we could find those blocks formed by the WGD, especially the longer blocks (Fig. 5). In fact, if allowing the longer blocks to absorb their nearby smaller blocks, which could have been produced by genomic fractionation, we would confirm more blocks originated from WGD, even if the WGD event of sea buckthorn was unclear.

By integrating the results of collinearity analysis, we finally identified the duplication types of GRAS genes, including tandem, WGD, and segmental duplication (Additional file 3), which showed that segmental duplication and WGD played a more prominent role in the expansion of sea buckthorn GRAS genes than tandem duplication. Considering the median K_s value of the collinear blocks where the HrGRAS gene is located, it is easier for us to distinguish the HrGRAS gene from WGD. The K_a/K_s ratios of all duplicated gene pairs were less than 0.5, suggesting these genes experienced strong purifying selection pressure during evolution processes (Additional file 3).

We also performed collinearity analysis between sea buckthorn, grape, and *Arabidopsis* genomes to further investigate the duplication model of HrGRAS genes (Fig. 6A). A total of 23 and 47 orthologous gene pairs containing HrGRAS were identified in the crosses of sea buckthorn and *Arabidopsis thaliana*,

sea buckthorn and grape, respectively (Additional file 4 and Additional file 5). We also found similar regions in collinearity analysis within sea buckthorn, which meant that GRAS gene family was highly conserved.

Some HrGRAS genes were found to be associated with at least three syntenic gene pairs in *A. thaliana* or grape, such as HrGRAS21 and HrGRAS24, guessed that these genes may play an important role in the evolution of GRAS gene family. Significantly, The GRAS collinear gene pairs identified between sea buckthorn and grape were located in longer blocks, with an average of more than 31 gene pairs per block (the longest block reached 191 gene pairs). The contrast was that there were an average of 20 gene pairs per block in the blocks where the GRAS collinear gene pairs identified between sea buckthorn and *A. thaliana* located (the longest block had only 61 gene pairs), although the number of collinear blocks containing HrGRAS between sea buckthorn and grapes was similar to the number of blocks containing HrGRAS between sea buckthorn and *A. thaliana* (Additional file 6 and Additional file 7). The difference in these statistical results may be related to the phylogenetic relationship between species.

The GRAS genes with microsynteny may evolve from the same ancestor. A slice of the genome-level alignment can be illustrated in a linear format, which can help us to observe the collinear relationship between HrGRAS gene of sea buckthorn with *A. thaliana* and grape more closely (Fig. 6B). We had also drawn a region colinearity map (Additional file 8 and Additional file 9) with the information on the positive and negative strands and the direction of the HrGRAS genes, which will further help us to study the GRAS genes of sea buckthorn based on existing research in *Arabidopsis* and grape.

Expression analysis of HrGRAS genes in various tissues and fruit developmental stages

The transcriptome data of three tissues (leaf, stem and root) and three developmental stages of fruit (46, 63, and 76 days post-anthesis) we sequenced was used to investigate the HrGRAS genes expression patterns (Fig. 7). There may be eight pseudogenes, because the transcripts of them were not detected in any tissues (RPKM < 0.001). Only five members (HrGRAS12, HrGRAS20, HrGRAS21, HrGRAS39 and HrGRAS60) showing high expression levels (RPKM > 50), and 49 HrGRAS genes were detected to express in all tissues and stages. Some HrGRAS members exhibit a certain degree of tissue specificity. For instance, the transcript of HrGRAS10 was only detected in leaf and stem, but not in root and all three fruit developments. HrGRAS20 was largely accumulated in the first stage of fruit development, and HrGRAS12 was very highly (RPKM > 100) expressed in leaf, stem, root, and the first stage of fruit development, but low in the third stage of fruit development (RPKM = 13.04). HrGRAS genes highly involved in the growth and development of various tissues, which can be showed the tissue-specific expression of these genes. Several genes, such as HrGRAS12, HrGRAS20, and HrGRAS39, were significantly reduced during the three stages of fruit ripening, indicating that they may play important roles in the fruit ripening process. HrGRAS21 has high expression levels in roots and stems, and is classified into the same subfamily with AtSCR, suggesting that it may play a radial pattern of sea buckthorn roots and stems. Furthermore, the expression patterns of gene pairs formed by tandem are very similar (HrGRAS7 and HrGRAS8, HrGRAS29

and HrGRAS30, HrGRAS50 and HrGRAS51), but the gene pairs formed by segmental duplication and WGD did not find this phenomenon. This may be related to the time of occurrence of different duplication models and the space in which the gene pairs are located. After all, the gene pairs formed by tandem are closer in space, and the gene expression adjustments from plants may be more similar.

Interaction network of HrGRAS proteins.

The blastp program was performed between the GRAS proteins of sea buckthorn with *Arabidopsis* GRAS proteins, every sea buckthorn GRAS protein matched the best matching GRAS protein in *Arabidopsis*. Finally, 22 *Arabidopsis* GRAS proteins were found to represent 62 sea buckthorn GRAS proteins and five predicted functional partners of the input proteins were used to construct the network. The sea buckthorn GRAS proteins corresponded with the *Arabidopsis* GRAS proteins are listed in Additional file 10. The online program ran with default parameters.

Discussion

With the rapid development of bioinformatics and sequencing technologies, more and more information stored in genomic sequences has been excavated to help explore plant growth and development mechanisms. GRAS transcription factors play an important role in growth and development, which has been extensively performed in many species whose genomes have been sequenced [13, 23–25]. In the current study, we comprehensively analyzed this important transcription factor family in sea buckthorn, including genome-wide identification, chromosomal localization, intron-exon structure, physical-chemical features, phylogenetic analysis, duplication analysis, microsyntenic mapping and expression profiles in various sea buckthorn tissues and fruit developmental stages.

Gene duplication is an important reason for new functions of genes and gene family expansion [15]. Compared with *Arabidopsis* and grape, rice contains more GRAS genes, which indicates that more GRAS genes are replicated in rice or the GRAS genes have a higher retention rate after replication [10]. Tandem replication events have played a role in the expansion of the GRAS gene in *Arabidopsis*, tomato, rice, pepper and other species, as previously reported [6, 15, 25, 26]. However, the detailed analysis of the effect of segmental duplication and WGD on gene family expansion is rarely due to the fragmentation and fusion of chromosomes after the WGD, which makes it difficult to distinguish gene pairs from the segment duplication and WGD. Colinearity analysis is an important method to analyze plant genome duplication. In recent years, this method has been used to analyze the genome duplication of multiple species [27, 28]. In this study, colinear analysis was used for the first time to correlate the GRAS gene family in sea buckthorn with WGD events. The specific method involves comprehensively analyzing the size of the collinearity blocks of the homologous gene dotplot within the sea buckthorn genome and between the sea buckthorn and grape genome, the complementarity between them, and the median Ks of collinearity blocks, etc., and finally distinguishing the HrGRAS members formed by segmental duplication and WGD. The gene pairs formed by the WGD were identified, the contribution of WGD (15 gene pairs) on the GRAS gene family is slightly smaller than that of segmental duplication (23 gene pairs), but that is far

greater than tandem (three gene pairs). These results show that WGD contributed significantly to gene family expansion, although they are mostly thought occurred millions of years ago. SCL14, GAI, SCL3, RGL1 and RGL2 interact with more proteins, and they also have more best matching genes in sea buckthorn GRAS proteins in the interaction network analysis (Additional file 11). These results are consistent with the idea of the gene dose hypothesis, which holds that genes whose products are dose-sensitive or involved in other protein network interactions are preferentially retained [29, 30]. Therefore, we speculated that the products of these genes and their interologs in sea buckthorn may play more roles in the growth and development of sea buckthorn.

Fifty-nine HrGRAS members were classified into nine subfamilies based on clade support values and classification from *Arabidopsis* and rice. Notably, three HrGRAS proteins did not belong to any of the subfamilies and there is no GRAS gene member of *Arabidopsis* and sea buckthorn in the rice-specific subfamily (Os4). In fact, species-specific GRAS subfamilies also exist widely in other plant genomes, which may be determined by the highly specialized needs of GRAS genes in species evolution. Whether the three HrGRAS genes not classified into any subfamily is unique to sea buckthorn needs further analysis. We will include more species of GRAS genes in the future to determine the status of these three HrGRAS genes.

Gene structure is an important feature of genes and an important basis for determining gene functions. We found that 58.1% of HrGRAS genes were intronless in sea buckthorn according to the exon-intron tissue analysis. In many plants, a significant proportion of the GRAS genes is intronless, such as *Arabidopsis*(67.6%), rice(55%), tomato(74%) [6, 26]. In fact, the proportion of intronless GRAS genes in many plants is very large, showing the close evolutionary relationship. Intronless genes also exist in some gene families in eukaryotes, such as DEAD box RNA helicases [31] and auxin-up RNAs (SAUR) gene family [32], although intronless genes are typical in prokaryotic genomes [6]. Previous studies have suggested that the GRAS genes in plants originate from the bacterial prokaryotic genome, which is consistent with a large number of intronless genes in this gene family [33]. It suggested that the GRAS genes originated from the horizontal gene transfer of ancient prokaryotes, and then experienced extensive replication such as WGD and segmental duplication.

Gene expression patterns are an important part of assessing the possible functions of genes, in addition to gene structure. Eight HrGRAS genes may be pseudogenes, because their expression is not detected in any tissue / organ, and they may be degraded after gene replication or evolution, lose their original function, and fail to evolve new functions. Some HrGRAS genes from segment duplication or WGD are not only in the different branch of the phylogenetic tree, but also have very different expression patterns. These results suggest that GRAS genes may undergo neo-functionalization or sub-functionalization in sea buckthorn. Yet still, some HrGRAS genes from segment duplication or WGD maintain extremely high sequence similarity and exhibited conserved expression patterns, showing that they may play similar functions after duplication. The duplication model of genes and the functional differentiation of new genes formed are important aspects of plant evolution, which need to be studied in more detail later.

Overall, we conducted a genome-wide comprehensive analysis for GRAS gene family members in sea buckthorn. These analyses provide insight into the GRAS genes expansion with evolution and the potential functional roles of sea buckthorn GRAS genes, which would be helpful to further understand the expansion and functional differentiation of sea buckthorn GRAS genes.

Conclusions

A system analysis of GRAS gene family in sea buckthorn was carried out in this study. Sixty two full-length GRAS genes were characterized. The genome comparison method was used to distinguish gene pairs from segment duplication and WGD, and it was found that both played a considerable role in GRAS gene expansion. The comparative analyses of the gene structure, phylogenetic, and expression patterns will be useful to characterizing the GRAS gene family, and to better understanding their possible roles in many aspects of growth regulatory and environmental signals. These results provide a valuable resource for better understanding the biological roles of individual GRAS genes in sea buckthorn.

Methods

Gene identification

Whole genome data *H. rhamnoides* subsp. *mongolia* Rousi 'sunny' was used for this study. *Arabidopsis* GRAS genes and rice GRAS genes were downloaded from Phytozome (phytozome.jgi.doe.gov/) and RGAP (<http://rice.plantbiology.msu.edu>), respectively. The latest Hidden Markov Model (HMM) file corresponding to the GRAS domain (PF03514.11) (<http://pfam.sanger.ac.uk/>) was used as a query to search against the entire protein datasets of sea buckthorn with an E-value of $1e-5$ using HMMER 3.0. Meanwhile, all AtGRAS and OsGRAS proteins were used as BLAST queries to search against the sea buckthorn databases using default parameters. The length of the hit out of the range from 350 to 820 aa was rejected. All candidate genes from the above steps were further examined by confirming the existence of the GRAS core domain using the PFAM and CDD program. Finally, the outputs of different approaches were aligned and those had exactly the same sequence were deemed as the same gene. Sixty-two GRAS models were finally identified in the sea buckthorn genome after comprehensive curation. Protein molecular weight (MW), isoelectric point (pI), aliphatic index, instability index, grand average of hydropathicity (GRAVY), the subcellular localization, the lengths of the protein sequence and coding sequence obtained by using tools from ExPasy website(<http://web.expasy.org/protparam/>).

Phylogenetic and sequence analysis

The GRAS proteins from sea buckthorn, *Arabidopsis* and rice were used for multiple alignments by ClustalW program. All Gene IDs of GRAS genes used in this study were listed in Additional file 12. *Arabidopsis* and rice are the most commonly used model plants for studying genetic evolution. An unrooted phylogenetic tree was generated using neighbor-joining method of MEGA7.0 with the following parameters: Poisson model, pair wise deletion, and 1000 bootstrap replications based on alignment

results. GRAS genes of sea buckthorn were categorized into different subfamilies based on well-established classifications in *Arabidopsis*.

We also used the GRAS genes from sea buckthorn alone for multiple alignments by ClustalW program and generated an unrooted phylogenetic tree using maximum likelihood method. The conserved motifs in the identified sea buckthorn GRAS proteins were identified using the MEME online program (<http://meme.nbcr.net/meme/intro.html>) for protein sequence analysis. The parameters were employed as the following: the number of repetitions, zero or one; and the maximum number of motifs, 10. The exon-intron organization and intron phase of sea buckthorn GRAS genes was determined by comparing predicted coding sequences with their corresponding full-length sequences using TBtools [17].

Chromosomal distribution and Gene duplication

The HrGRAS genes were mapped to sea buckthorn chromosomes based on physical location information from the information of sea buckthorn genome using TBtools [17]. We renamed each HrGRAS gene according to its ascending chromosomal distribution. Multiple Collinearity Scan toolkit (MCScanX) was adopted to analyze the synteny within sea buckthorn, sea buckthorn with *Arabidopsis*, sea buckthorn with grape to distinguish the duplication models of HrGRAS. The syntenic analysis maps were constructed using the Multiple Synteny Plotter software and Genome Gene-dotplot (<https://github.com/CJ-Chen/TBtools>) to exhibit the synteny relationship of the orthologous GRAS genes of sea buckthorn and other selected species. Non-synonymous (k_a) and synonymous substitution sites (K_s) of each duplicated GRAS genes were calculated using KaKs_Calculator.

Expression analysis of HrGRAS genes in various tissues and fruit developmental stages

The transcriptome data of leaf, stem, root and fruit at three developmental stages of sea buckthorn had been previously generated. We retrieved the fragments per kilobase per million reads value representing the expression level of each HrGRAS gene and displayed the result using TBtools [17].

Prediction of HrGRAS protein–protein interaction network

We predicted a protein–protein interaction network based on the interologs of HrGRAS genes in *Arabidopsis*. First, the blastp program was performed between the GRAS proteins of sea buckthorn with *Arabidopsis* GRAS proteins, every sea buckthorn GRAS protein matched the best matching GRAS protein in *Arabidopsis*. Then, we used the best matching genes of HrGRAS in *Arabidopsis* to predict the functional interaction between proteins on the string website (<http://string-db.org/cgi/input.pl>). Finally, implement the interactive network construction between HrGRAS and export the image on the string website.

Abbreviations

HrGRAS

Sea buckthorn (*Hippophae rhamnoides* L.) GRAS;

AtGRAS

Arabidopsis thaliana GRAS;

OsGRAS

Rice (*Oryza sativa* L.) GRAS;

PhyA

phytochrome-A;
PhyB
phytochrome-B;
WGD
the whole-genome duplication;
Ks
synonymous substitution sites;

Declarations

Ethics approval and consent to participate

Hippophae rhamnoides L. cv. Sunny is widely cultivated in Liaoning and Nei Monggol province, China. *Hippophae rhamnoides* L. is not listed in the appendices I, II and III of the Convention on the Trade in Endangered Species of Wild Fauna and Flora, that has been valid from 4 October 2017 (<https://cites.org/eng/app/appendices.php>). The sea buckthorn samples used in this study were collected from State Forestry Administration, Research Institute of Forestry, Chinese Academy of Forestry. Collection of plant materials complied with the institutional, national and international guidelines. No specific permits were required.

Conflict of Interest

The authors declare they have no conflict of interest.

Funding

This work was supported by the Beijing Natural Science Foundation [No. 6192027] and the Fundamental Research Funds for the Central Non-profit Research Institution of Chinese Academy of Forestry [No. ZDRIF2019].

Availability of data and materials

All data analyzed during this study are included in this article and its Additional files.

Acknowledgments

We would like to thank Mrs. H.M. Luo for their assistance with sample collection. We would also like to thank everyone who contributed to this manuscript in various capacities.

Consent for publication

Not applicable.

Authors' Contributions

CYH and JGZ supervised the research. LYY performed the experiments, analyzed the data, prepared figures and captions, and wrote the manuscript. GYZ collected field samples. ZRL analyzed RNA-seq data. All of the authors have read and approve of the submitted manuscript.

References

1. Yang Y, Yao Y, Xu G, Li C. Growth and physiological responses to drought and elevated ultraviolet-B in two contrasting populations of *Hippophae rhamnoides*. *Physiol Plant*. 2005;124(4):431–40.
2. Boca AN, Ilies RF, Saccomanno J, Pop R, Vesa S, Tataru AD, Buzoianu AD. Sea buckthorn extract in the treatment of psoriasis. *Exp Ther Med*. 2019;17(2):1020–3.
3. Ren Q, Li X, Li Q, Yang H, Wang H, Zhang H, Zhao L, Jiang Y, Meng X, Zhang Y, Shen X. Total flavonoids from sea buckthorn ameliorates lipopolysaccharide/cigarette smoke-induced airway inflammation. *Phytother. Res*. 2019; 33(8).
4. Serban M, Serban A, Ursoniu S. Dragan. Systematic review on the potential of sea buckthorn *Hippophae Rhamnoides* L. for a possible novel enriched bread for the patients with cardiovascular diseases. *Atherosclerosis*. 2019;287:e285.
5. Sun X, Jones WT, Rikkerink EH. GRAS proteins: the versatile roles of intrinsically disordered proteins in plant signalling. *Biochem J*. 2012;442(1):1–12.
6. 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(1):209.
7. Di Laurenzio L, Wysocka-Diller J, Malamy JE, Pysh L, Helariutta Y, Freshour G, Hahn MG, Feldmann KA, Benfey PN. The SCARECROW gene regulates an asymmetric cell division that is essential for generating the radial organization of the *Arabidopsis* root. *Cell*. 1996;86(3):423–33.
8. Silverstone AL, Ciampaglio CN, Sun T. The *Arabidopsis* RGA gene encodes a transcriptional regulator repressing the gibberellin signal transduction pathway. *Plant Cell*. 1998;10(2):155–69.
9. Bolle C. The role of GRAS proteins in plant signal transduction and development. *Planta*. 2004;218(5):683–92.
10. Tian CG, Wan P, Sun SH, Li JY, Chen MS. Genome-wide analysis of the GRAS gene family in rice and *Arabidopsis*. *Plant Mol Biol*. 2004;54(4):519–32.
11. Sun XL, Xue B, Jones WT, Rikkerink E, Dunker AK, Uversky VN. A functionally required unfoldome from the plant kingdom: intrinsically disordered N-terminal domains of GRAS proteins are involved in molecular recognition during plant development. *Plant Mol Biol*. 2011;77(3):205–23.
12. Lu J, Wang T, Xu Z, Sun L, Zhang Q. Genome-wide analysis of the GRAS gene family in *Prunus mume*. *Mol Genet Genomics*. 2015;290(1):303–17.
13. Zhang H, Mi L, Xu L, Yu C, Li C, Chen C. Genome-wide identification, characterization, interaction network and expression profile of GRAS gene family in sweet orange (*Citrus sinensis*). *Sci Rep*. 2019;9(1):2156.

14. Liu M, Huang L, Ma Z, Sun W, Wu Q, Tang Z, Bu T, Li C, Chen H. Genome-wide identification, expression analysis and functional study of the GRAS gene family in Tartary buckwheat (*Fagopyrum tataricum*). BMC Plant Biol. 2019;19(1):1–17.
15. Liu B, Sun Y, Xue J, Jia X, Li R. Genome-wide characterization and expression analysis of GRAS gene family in pepper (*Capsicum annuum* L.). Peer J. 2018;6(1):e4796.
16. Cui H, Levesque MP, Vernoux T, Jung JW, Paquette AJ, Gallagher KL, Wang JY, Blilou I, Scheres B, Benfey PN. An evolutionarily conserved mechanism delimiting SHR movement defines a single layer of endodermis in plants. Science. 2007;316(5823):421–5.
17. Wang Y, He S, Wei Y, Dong C, Liu L, Jue D, Shi S, Li W. Molecular and functional characterization of two DELLA protein-coding genes in litchi. Gene. 738.
18. Torres-Galea P, Hirtreiter B, Bolle C. Two GRAS proteins, SCARECROW-LIKE21 and PHYTOCHROME A SIGNAL TRANSDUCTION1, function cooperatively in phytochrome A signal transduction. Plant Physiol. 2013;161(1):291–304.
19. Li X, Qian Q, Fu Z, Wang Y, Xiong G, Zeng D, Wang X, Liu X, Teng S, Hiroshi F. Control of tillering in rice. Nature. 2003;422(6932):618.
20. Sharma A, Singh V, Sharma A, Negi N. Seabuckthorn a new approach in ecological restoration of Himalayan Ecosystem: A review. Int J Chem Stud. 2019;7(1):1219–26.
21. Ivanova G, Nikulina E, Kolman O, Ivanova A. Products of sea-buckthorn berries processing in parapharmaceutical production. In: IOP Conference Series: Earth and Environmental Science: 2019: IOP Publishing; 2019: 052020.
22. Wang Z, Feng R, Zhang X, Su Z, Wei J, Liu J. Characterization of the *Hippophae rhamnoides* WRKY gene family and functional analysis of the HrWRKY21 gene in resistance to abiotic stresses. Genome. 2019;62(10):689–703.
23. Zeng X, Ling H, Chen X, Guo S. Genome-wide identification, phylogeny and function analysis of GRAS gene family in *Dendrobium catenatum* (*Orchidaceae*). Gene. 2019;705:5–15.
24. Guo P, Wen J, Yang J, Ke Y, Wang M, Liu M, Ran F, Wu Y, Li P, Li J. Genome-wide survey and expression analyses of the GRAS gene family in *Brassica napus* reveals their roles in root development and stress response. Planta. 2019;250:1051–70.
25. Quan S, Niu J, Zhou L, Xu H, Ma L, Qin Y. Genome-wide Identification, Classification, Expression and Duplication Analysis of GRAS Family Genes in *Juglans regia* L. Sci Rep. 2019;9(1):1–14.
26. Liu X, Widmer A. Genome-wide comparative analysis of the GRAS gene family in *Populus*, *Arabidopsis* and rice. Plant Mol Biol Rep. 2014;32(6):1129–45.
27. Wang J, Yu J, Li J, Sun P, Wang L, Yuan J, Meng F, Sun S, Li Y, Lei T. Two Likely Auto-Tetraploidization Events Shaped Kiwifruit Genome and Contributed to Establishment of the Actinidiaceae Family. iScience. 2018; 7:230–240.
28. Wang J, Sun P, Li Y, Liu Y, Yang N, Yu J, Ma X, Sun S, Xia R, Liu X. An Overlooked Paleotetraploidization in Cucurbitaceae. Mol Biol Evol. 2018;35(1):16–26.

29. Thomas BC, Pedersen B, Freeling M. Following tetraploidy in an *Arabidopsis* ancestor, genes were removed preferentially from one homeolog leaving clusters enriched in dose-sensitive genes. *Genome Res.* 2006;16(7):934–46.
30. Birchler JA, Veitia RA. The gene balance hypothesis: from classical genetics to modern genomics. *Plant Cell.* 2007;19(2):395–402.
31. Aubourg S, Kreis M, Lecharny A. The DEAD box RNA helicase family in *Arabidopsis thaliana*. *Nucleic Acids Res.* 1999;27(2):628–36.
32. Jain M, Tyagi AK, Khurana JP. Genome-wide analysis, evolutionary expansion, and expression of early auxin-responsive SAUR gene family in rice (*Oryza sativa*). *Genomics.* 2006;88(3):360–71.
33. Zhang D, Iyer LM, Aravind L. Bacterial GRAS domain proteins throw new light on gibberellic acid response mechanisms. *Bioinformatics.* 2012;28(19):2407–11.

Figures

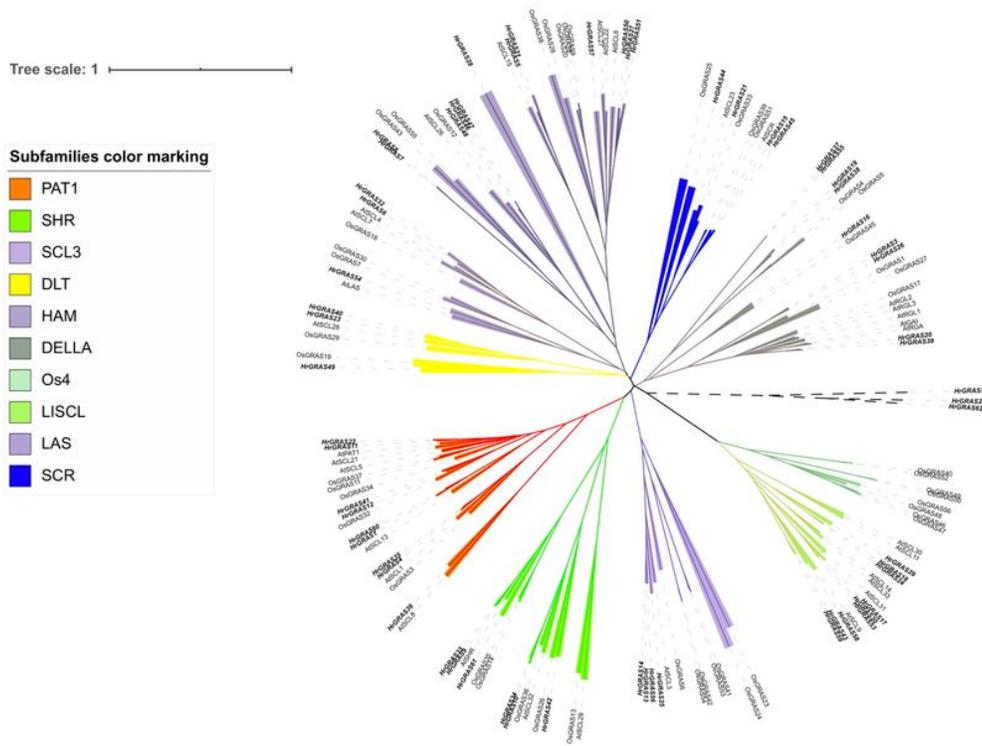


Figure 1

Unrooted phylogenetic tree representing relationships among GRAS proteins of Arabidopsis, rice and sea buckthorn. The different-colored arcs indicate different groups (or subgroups) of GRAS genes

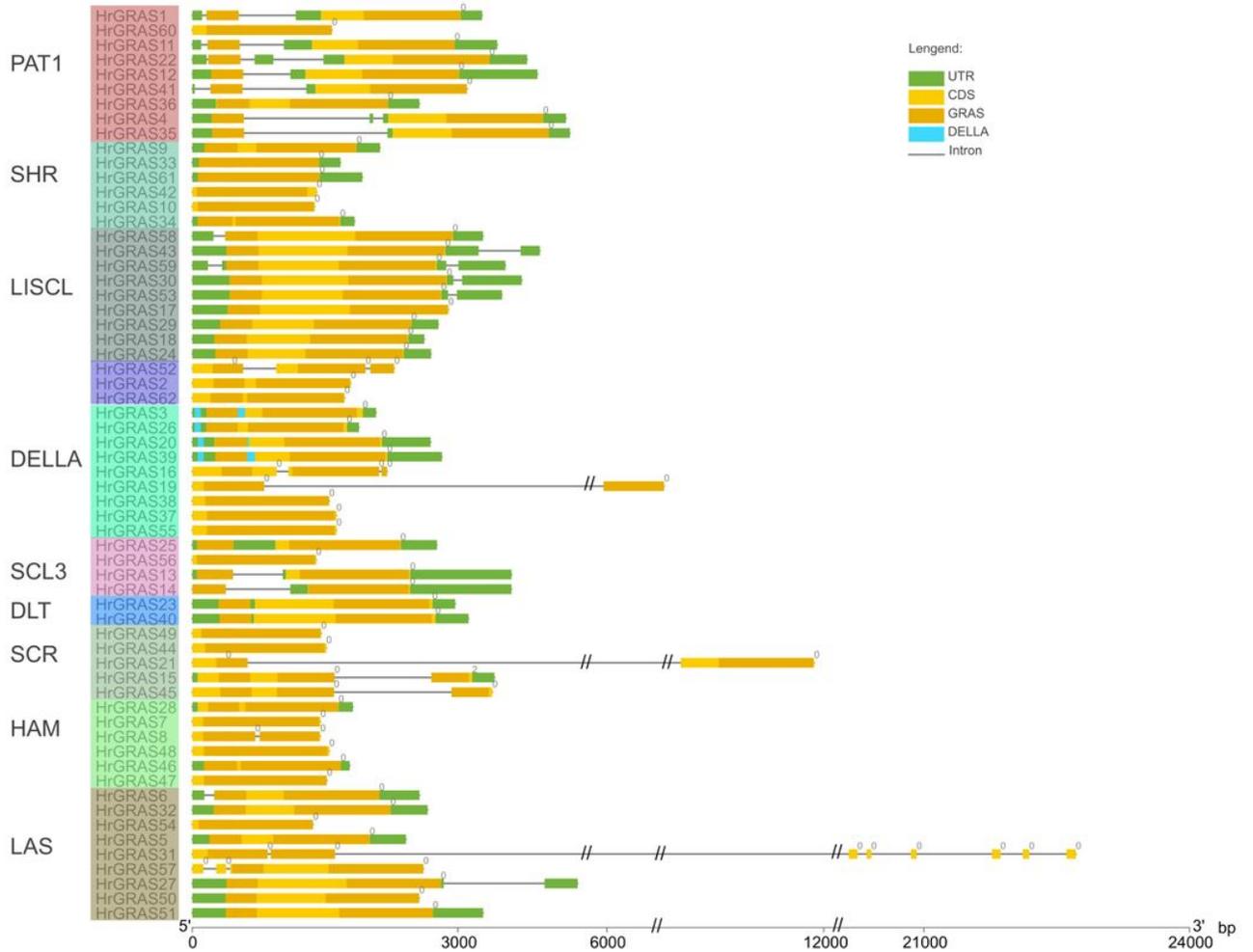


Figure 2

Gene structure of HrGRAS genes. Black line indicates intron, boxes of different colors are used to represent UTR, CDS, GRAS domain and DELLA domain. Y-axis represents the subfamily name of each HrGRAS genes. The lengths of the elements were drawn to scale

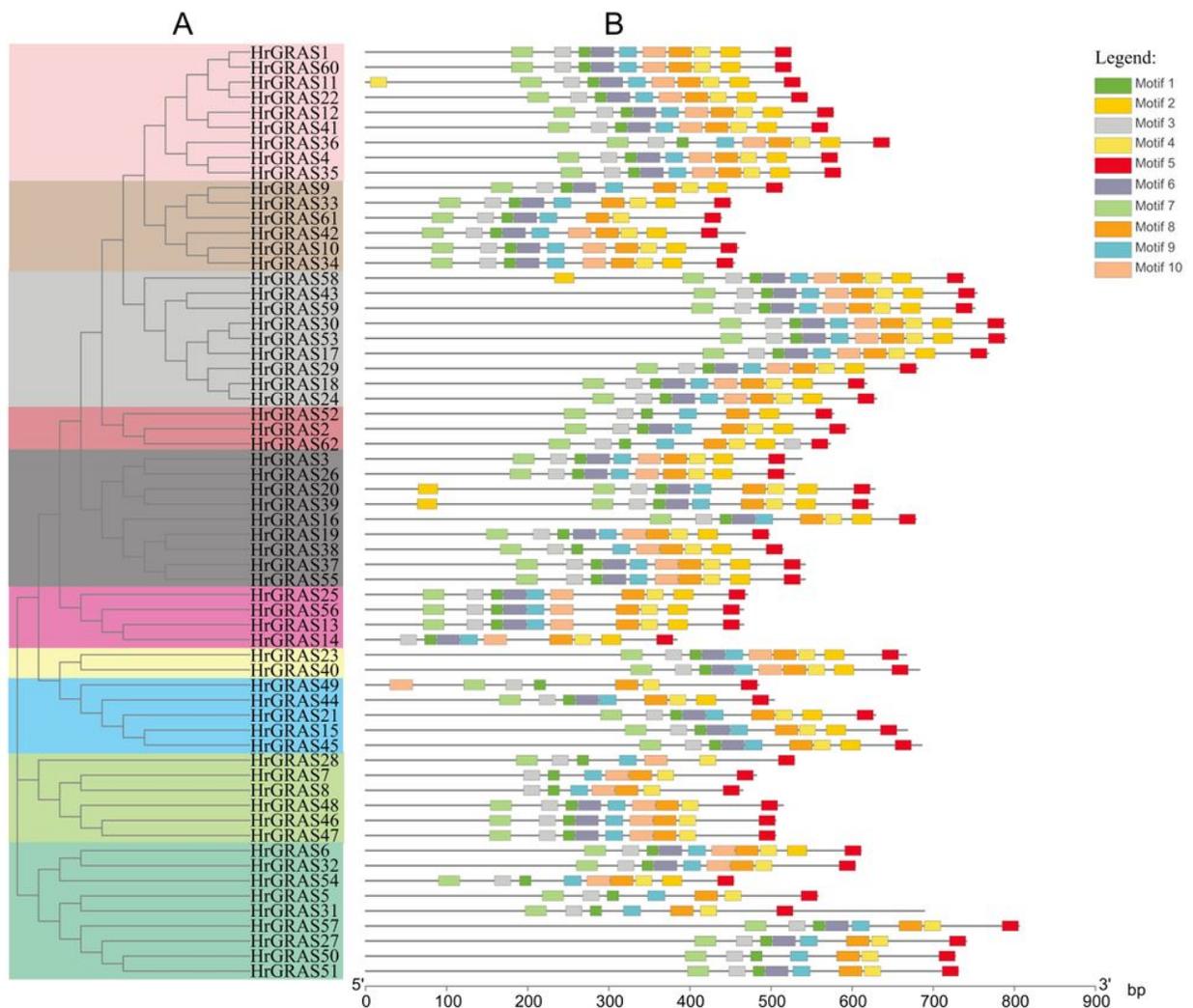


Figure 3

Distribution of conserved motifs in HrGRAS proteins. A The phylogenetic tree was constructed based on the full-length sequences of sea buckthorn GRAS proteins using MEGA 7.0 software. Details of clusters are shown in different colors. B The motif composition of sea buckthorn GRAS proteins. The motifs, numbers 1–10, are displayed in different colored boxes. The length of protein can be estimated using the scale at the bottom

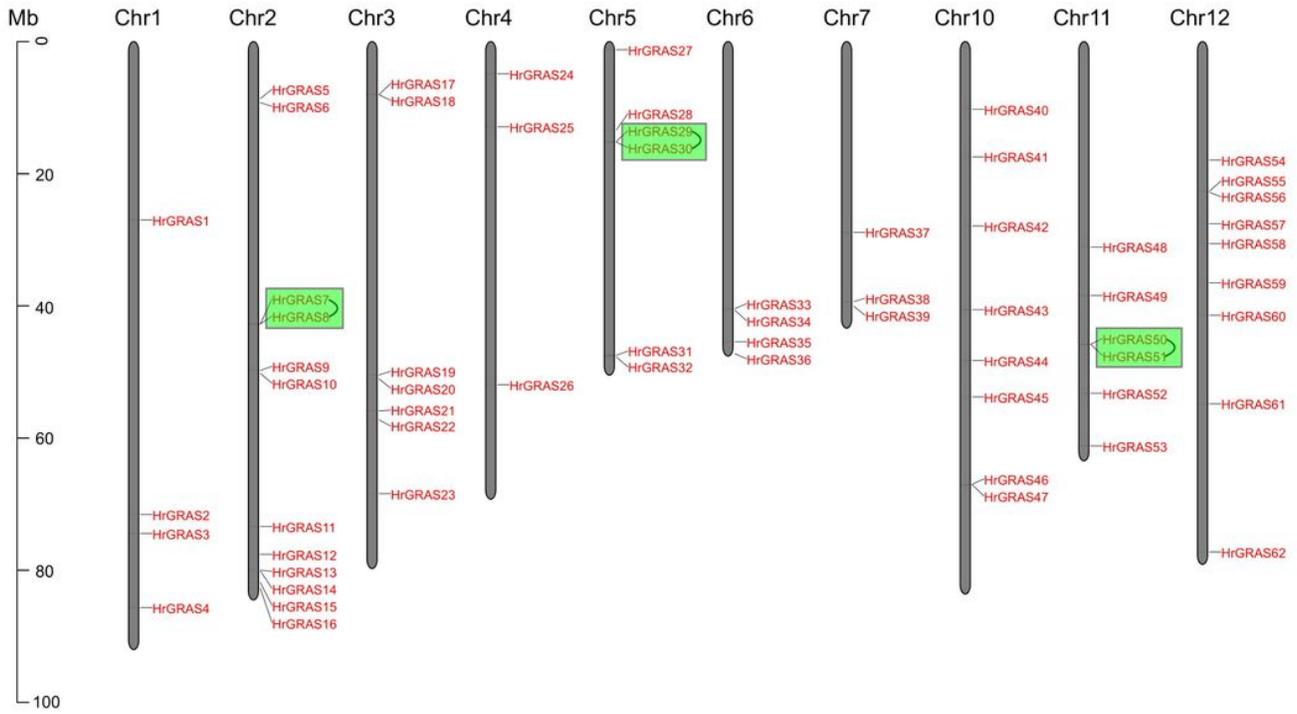


Figure 4

Positions of GRAS gene family members on the sea buckthorn chromosomes. Tandemly duplicated genes were indicated in green color

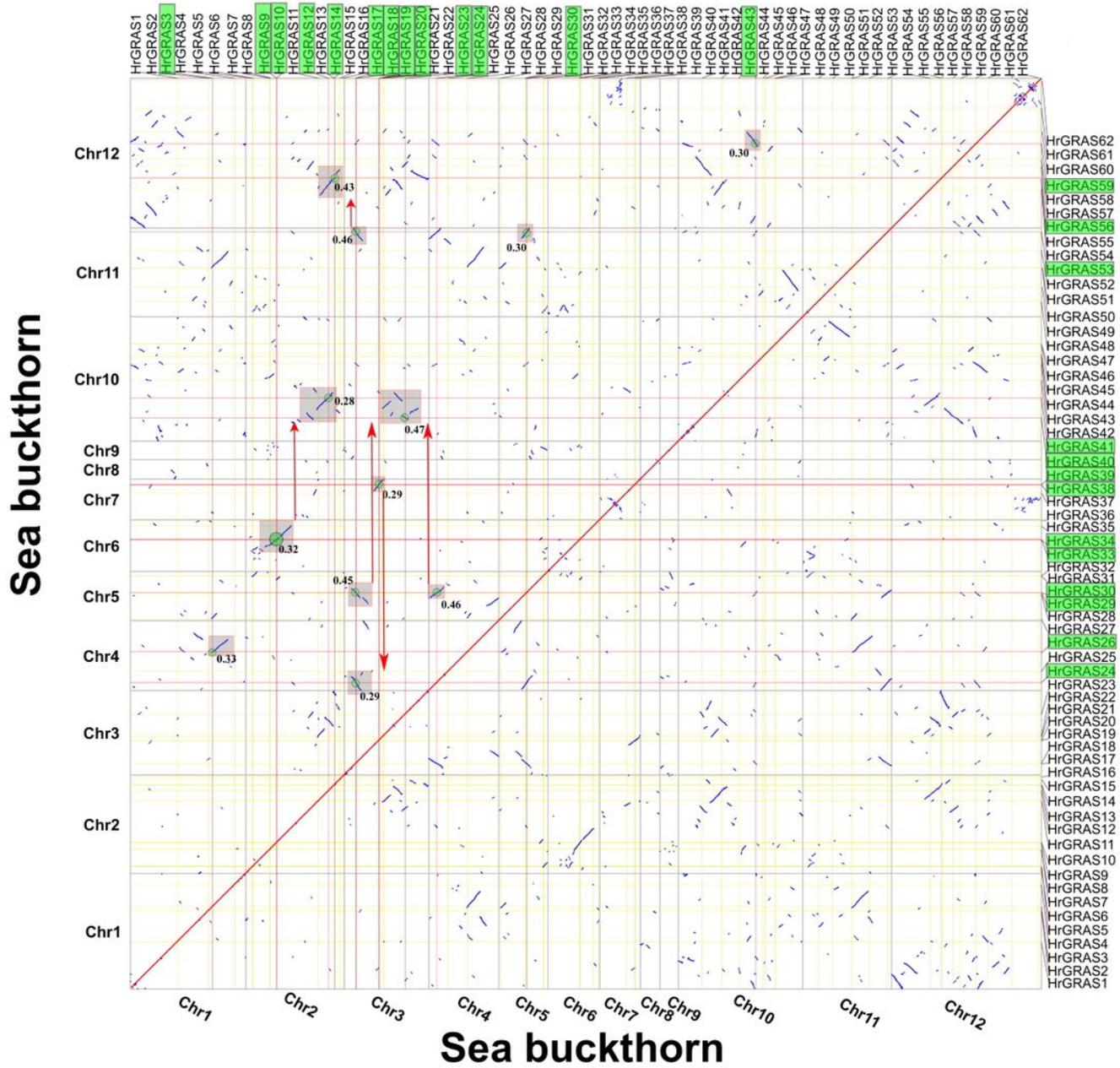


Figure 5

Homologous dotplot within sea buckthorn genome. The IDs of the GRAS genes are displayed at the edge. The GRAS genes derived from the WGD are in green boxes. The collinearity blocks formed by the WGD are in gray boxes in the figure, and the median value of Ks of the collinearity blocks are marked.

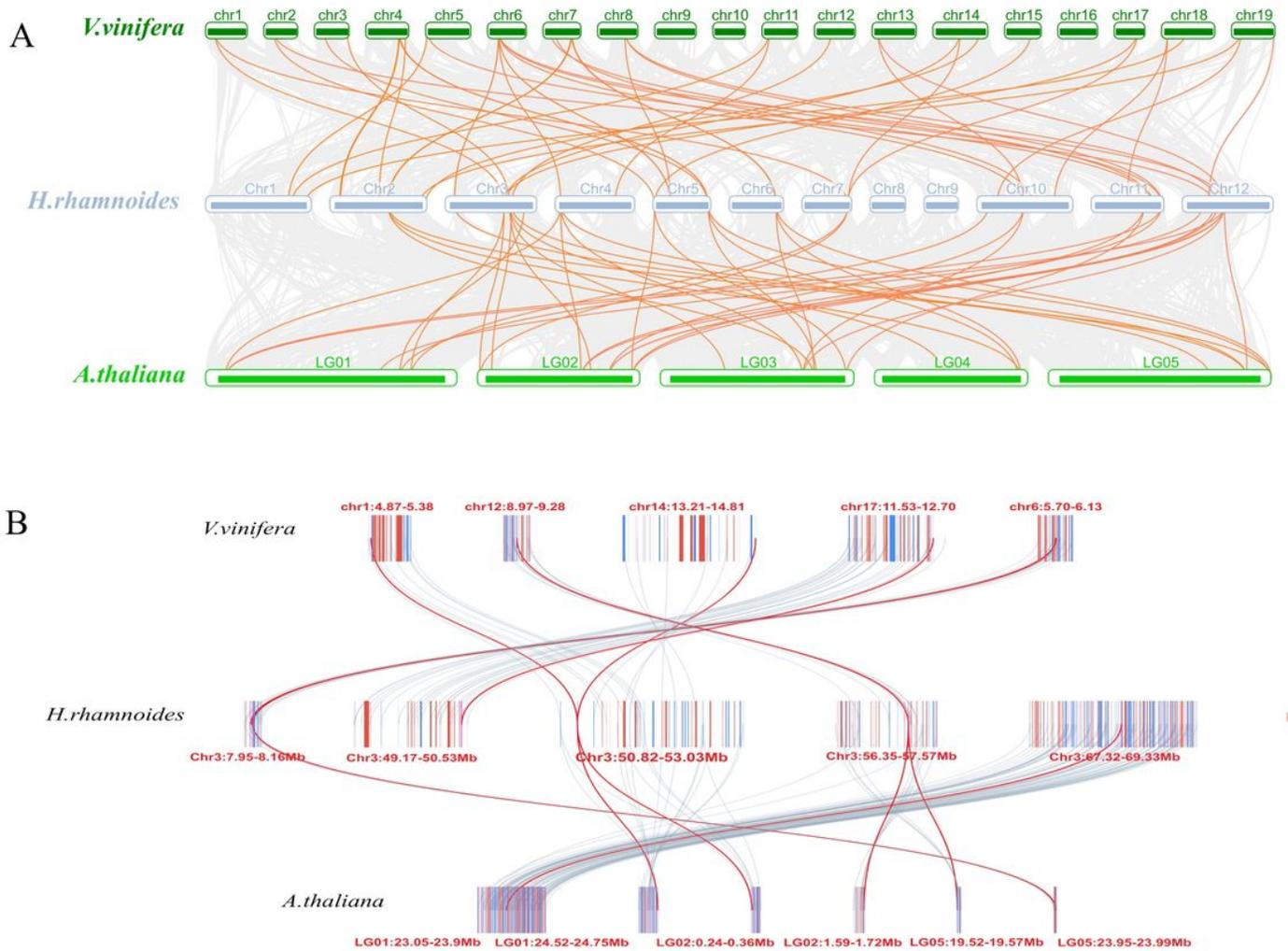


Figure 6

Synteny analysis of GRAS genes between sea buckthorn and two representative plant species. A The diagram of collinearity analysis of grape, sea buckthorn and Arabidopsis genomes. Gray lines in the background indicate the collinear blocks within sea buckthorn and other plant genomes, while the red lines highlight the syntenic GRAS gene pairs. B Local alignment of grape, sea buckthorn and Arabidopsis genomes. Species are shown in different colors, and homologous genes between adjacent chromosomes are ligated with Bezier curves

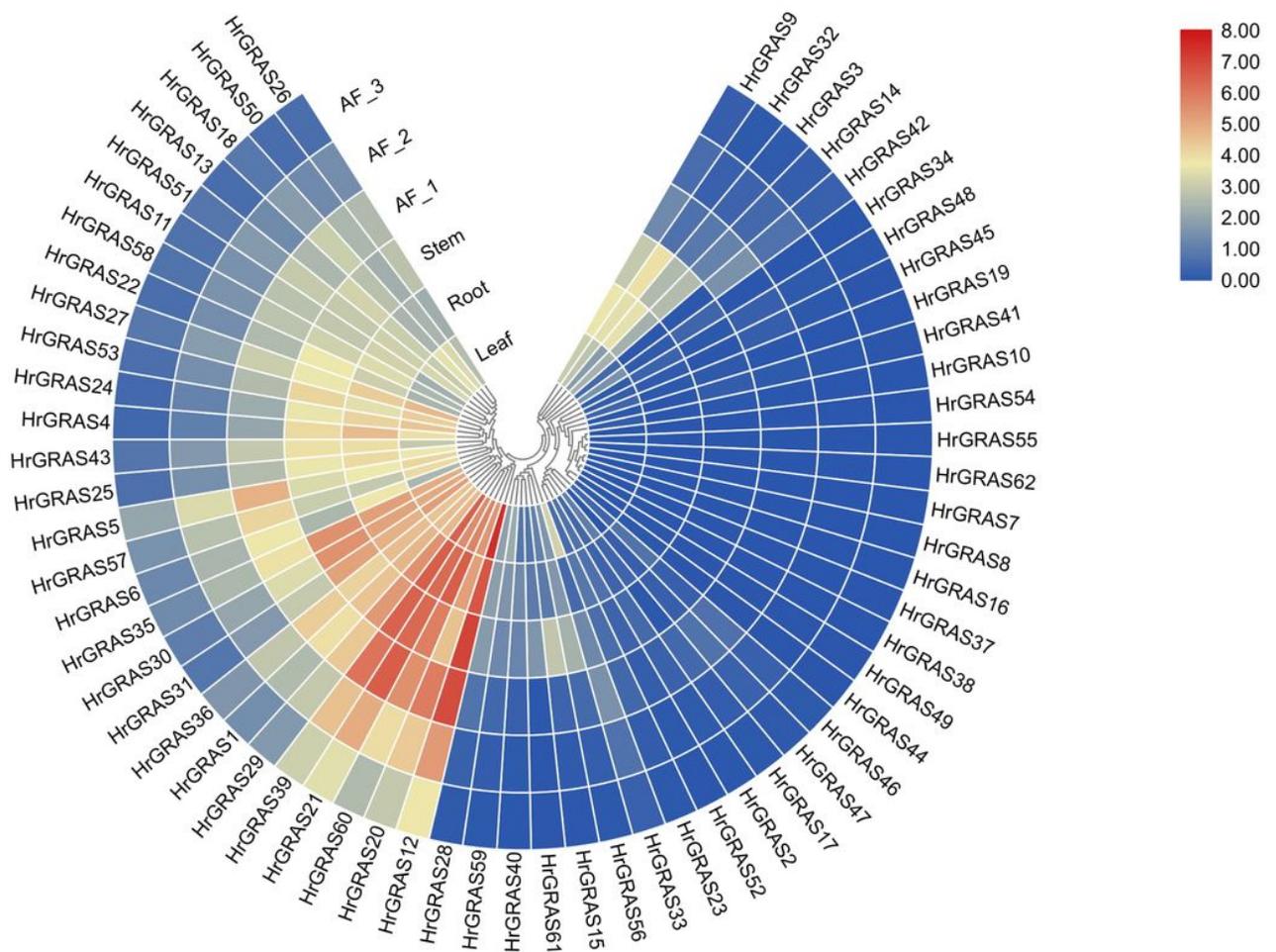


Figure 7

Heatmap of HrGRAS genes in leaf, stem, root and 46, 63, and 76 days post-anthesis (AF_1, AF_2, AF_3). The expression values were calculated by RPKM measure and then were log₂ transformed before generating heat maps

Supplementary Files

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

- [Additionalfile12.xls](#)
- [Additionalfile11.pdf](#)
- [Additionalfile10.xls](#)
- [Additionalfile9.pdf](#)

- [Additionalfile8.pdf](#)
- [Additionalfile7.rtf](#)
- [Additionalfile6.rtf](#)
- [Additionalfile5.rtf](#)
- [Additionalfile4.rtf](#)
- [Additionalfile3.xls](#)
- [Additionalfile2.pdf](#)
- [Additionalfile1.xls](#)