International Journal of Horticultural Science and Technology
Journal homepage: http://ijhst.ut.ac.ir

Unraveling the Molecular Mechanisms and Evolutionary Significance of Raffinose Biosynthesis in *Vitis vinifera*: A Comprehensive Study of the *VvGolS* and *VvRfS* Gene Families

Zahra Shirazi[1]*, Mahsa Rostami[2], Abozar Ghorbani[2]*

1 Biotechnology Research Department, Research Institute of Forests and Rangelands, Agricultural Research Education and Extension Organization (AREEO), National Botanical Garden, Tehran Karaj Freeway, Tehran, Iran
2 Nuclear Agriculture Research School, Nuclear Science and Technology Research Institute (NSTRI), Karaj, Iran

ARTICLE INFO

Article history:

Received: 7 June 2023,
Received in revised form: 4 August 2023,
Accepted: 8 August 2023

Article type:

Research paper

Keywords:

Carbohydrates,
Galactinol synthase,
Gene expression,
Osmoprotectant,
Raffinose synthase

ABSTRACT

Here, we evaluated how *GolS* and *RfS* gene families can potentially assist with grapevine drought tolerance. Through comprehensive bioinformatics analysis, we identified a set of putative genes associated with drought response in grapes. The analysis revealed the conservation of key functional domains and conserved motifs within the gene sequences, suggesting their potential involvement in stress adaptation. Additionally, we performed expression analysis of the identified genes under drought stress conditions. Accordingly, the results indicated substantial up-regulation levels in several genes, suggesting their potential importance in how grapevines respond to water scarcity. We explored the co-expression network of these genes and identified possible interactions and regulatory relationships. The findings provide d valuable information on complex regulatory mechanisms underlying drought-response in grapevines. The results bear implications for the genetic improvement of grape varieties, particularly in enhancing drought tolerance. By understanding the genetic basis of the drought-response mechanism, we can develop targeted strategies to improve crop management and assist in future attempts to breed grape varieties, emphasizing an enhanced tolerance to water deficits. This study sheds light on the potential roles of the *GolS* and *RfS* gene families in grapevine drought tolerance. The findings underscore the importance of genetic adaptation in grapevines to drought stress conditions. The knowledge herein can guide future research and breeding efforts to develop drought-tolerant grape varieties, ultimately contributing to the sustainability of grape production and the agricultural industry.

Abbreviations

Abscisic acid (ABA), ABA-responsive transcription factor (ABF2), involved in the abscisic acid responsiveness (ABRE), essential for the anaerobic induction (ARE), galactinol synthase (*GolS*), involved in heat stress responsiveness (HSE), nonsynonymous mutations (Ka), synonym -ous mutations (Ks), marker-assisted selection (MAS), million years ago (MYA), raffinose families of oligosaccharides (RFOs), raffinose synthase (*RfS*), robust multi-array average (RMA), simple sequence repeats (SSR), estimated divergence times (T)

Introduction

Raffinose family oligosaccharides (RFOs) are α-1,6-galactosyl extensions of sucrose and have broad distributions among multiple plant species. They play critical roles in several physiological processes, including stress tolerance, seed

* Corresponding author's email: Ghorbani.abozar@gmail.com, z.shirazi@rifr-ac.ir

germination, carbon resource allocation, and signaling. For instance, during galactose metabolism in olive, raffinose functions as a viable carbon source for carbon acquisition during fruit development (Asadi et al., 2023). The biosynthesis of RFOs involves two key enzymes: galactinol synthase (*GolS*) and raffinose synthase (*RfS*). *GolS* catalyzes the conversion of myo-inositol to galactinol, while *RfS* transfers galactosyl units from galactinol to sucrose (Gu et al., 2018; Sengupta et al., 2015).

GolS and *RfS* genes have received attention in numerous studies, characterizing them in various plant species, including Arabidopsis, rice, tomato, maize, and sesame. These genes exhibit diverse expression patterns across different organs and usually function in response to environmental stimuli. For instance, in Arabidopsis, specific *GolS* isoforms are activated under conditions of drought, salt, heat, or cold stress (Cui et al., 2020; Downie et al., 2003; Li et al., 2020; Liu et al., 2022; Nishizawa et al., 2008; Saito and Yoshida, 2011; You et al., 2018; Zhou et al., 2012). The overexpression of *GolS* genes in Arabidopsis reportedly enhanced drought tolerance through increased accumulation of RFOs (de Gois et al., 2020).

Fruit tree crops are vital in agricultural systems, providing essential nutrition to the human diet and offering added value to the financial aspects of horticulture worldwide. Their cultivation and production are great pillars for food security, rural livelihoods, and environmental sustainability. Understanding the genetic basis and molecular mechanisms underlying plant growth, development, and fruit quality traits is a prerequisite to crop improvement and sustainable fruit production. In recent years, considerable progress has led to findings that unraveled the complex genetic architecture and regulatory networks associated with various fruit tree species. Numerous studies have focused on identifying key genes, pathways, and regulatory elements involved in fruit development, ripening, disease resistance, and other important agronomic traits. However, despite these advancements, there is still much to explore and comprehend before arriving at a fully developed ability to use the potential of fruit trees (Mathiazhagan et al., 2021). Numerous studies evaluated genetic modifications in trees to enhance their resilience to challenging environments. One relevant study focused on the genetic modification of Persian walnut trees using the BADH gene through Agrobacterium-mediated transformation. The findings revealed successful expression of the BADH gene in transgenic walnut plantlets, indicating its potential for breeding drought and salt-tolerant walnuts (Rezaei Qusheh Bolagh et al., 2021). Also, research on the genome-wide patterns of walnuts successfully identified 55 single nucleotide polymorphisms (SNPs) that exhibited significant associations with nut and kernel-related traits. The findings were valuable as they paved the way for future breeding programs in numerous countries (Arab et al., 2019).

Sugar content and other compatible solutes contribute to osmotic adjustment under drought stress. Therefore, previous researchers conducted a study on genome-wide association analysis and enrichment of metabolic pathways in Persian walnut trees under drought stress, showing several enriched metabolic pathways that shared similar features among two experimental conditions. These similarities included galactose metabolism, amino sugar, and nucleotide sugar metabolism under drought stress, recovery conditions, fructose and mannose metabolism, photosynthesis, and phenylpropanoid biosynthesis between the well-watered and drought-affected plants (Arab et al., 2022). Soluble sugars serve multiple functions beyond osmoregulation. They act as energy sources and safeguard the photosynthetic system through Na^+ sequencing compounds. Recent research on salinity tolerance in pistachio rootstocks has revealed several pistachio genotypes that exhibit tolerance by employing osmoprotectants like soluble sugars and proline. These genotypes effectively maintain cell osmotic pressure and stage a reasonable growth rate, even when confronted with osmotic stress by salinity (Behzadi et al., 2021).

In grapevine (*Vitis vinifera* L.), a widely cultivated fruit crop, *GolS* and *RfS* genes function on RFOs. However, the extent and mode of their function require further elucidation. Previous studies have provided partial sequences for *GolS* and *RfS* genes in grapevines, but the available literature lacks a comprehensive understanding of their functions using bioinformatics and relevant techniques (Elango et al., 2022; Shen et al., 2020).

This study aims to address this knowledge gap by conducting a whole-genome association study to identify the GolS and RfS gene families in grapes and provide detailed information about their sequences. By leveraging bioinformatics, relevant tools, and databases, we aimed to characterize putative *GolS* and *RfS* genes in grapes and evaluate their potential functions, particularly under drought-stress conditions. Furthermore, this study revealed how RFOs contribute significantly to vine growth and development. By unraveling the molecular mechanisms underlying

RFO biosynthesis and their role in stress response, we gained insights into the genetic background of stress tolerance in grapevine. This knowledge has practical implications for grapevine breeding programs, facilitating the development of improved grape varieties with enhanced tolerance to abiotic stress.

Materials and methods

Identification and phylogenetic analysis of the raffinose biosynthesis gene family in grapes

Water deficits are a critical environmental stress factor that causes considerable challenges to plant growth and development. In this study, we evaluated the function of the *GolS* and *RfS* gene families in the context of drought stress response in grapes.

The seven protein sequences of *GolSs* (At2g47180, At1g56600, At1g09350, At1g60470, At5g23790, At4g26250, At1g60450) and five *RfSs* (AT1G55740, AT3G57520, AT4G01970, AT5G40390, AT5G20250) from *Arabidopsis thaliana* (Nishizawa et al., 2008) were compared with the grape genome in the Phytozome database (https://phytozome.jgi.doe.gov) using the BLASTp method (Goodstein et al., 2012). Focusing on protein sequence analysis made a comprehensive understanding of the *GolS* and *RfS* gene families regarding their protein-level functions. The resultant sequences became validated by the presence of Glyco_trans_8 Pfam (PF01501) and raffinose_synthase Pfam (PF05691) domains in the putative *GolS* and *RfS* proteins, respectively, using the hmmscan tool. The identified *GolS* and *RfS* genes were designated as *VvGolS* and *VvRfS* genes based on their chromosomal position distribution from start to end (You et al., 2018) http://rice.plantbiology.msu.edu (You et al., 2018).

We used information from the specifics of *GolS* and *RfS* protein sequences about *A. thaliana* (TAIR), *Oryza sativa* subsp. Japonica (RGAP), *Brachypodium distachyon*, *Solanum lycopersicum*, *Populus trichocarpa*, and *Zea mays* (phytozome) projects. This information facilitated multiple sequence alignment with ClustalX 2.0.8 (Thompson et al., 1994). Phylogenetic relationships were determined using MEGA 7.0.9 by the neighbor-joining (NJ) method, with 1000 bootstraps (Felsenstein, 1985).

The selection of these plant species was based on several considerations. Firstly, *A. thaliana* has been extensively studied and is known to function as a model organism for plant molecular biology research. Other species represent diverse plant species with varying levels of genetic and evolutionary relatedness to grapevine. By including these species, research can aim to capture a broad range of genetic information and potential functional similarities or differences. It is important to note that the selection of species for comparative analysis is a complex decision influenced by multiple factors, including the availability of high-quality genome sequences, the relevance of the species to the research question, and the existing knowledge base. While species more closely related to grapes could provide valuable insights, researchers select particular species based on their comprehensive and well-annotated genome resources that facilitate robust comparative analysis.

Divergence times and Ka/Ks estimation

To ascertain the type of selection applied to genes derived from segmental duplication and tandem, the (Ks) and substitution rates (Ka) were estimated using the PAL2NAL web server (http://www.bork; Suyama et al., 2006). The Ka/Ks ratio was applied to determine selection pressure. The estimated divergence times (T) of duplicated genes were calculated using a relevant formula:

$T = Ks / 2\lambda$ (MYA), where $\lambda = 6.5 \times 10^{-9}$ for grapes (Cao et al., 2014).

Gene structure, chromosomal location, and duplication

The Gene Structure Presentation Server (GSDS, http://gsds.cbi.pku.edu.cn/) was used for analyzing the exon and intron structures of *VvGolS* and *VvRfS* in the grape genome. The MapChart (https://www.wur.nl/en/show/MapChart-2.32.htm) software established the positional imagery of *VvGolS* and *VvRfS* on each chromosomal location. Segmental duplication was identified using the PGDD server (http://chibba.agtec.uga.edu/duplication/ database of plants). Tandem duplication was considered present, with the highest spacing of twenty genes on the same chromosome (Liu et al., 2014; Ozyigit et al., 2016).

Sequence analysis of raffinose biosynthesis family proteins

The molecular weights and isoelectric points of *GolS* and *RfS* proteins were identified using ProtParam (http://web.expasy.org/protparam) (Gasteiger et al., 2005). The subcellular

localization of proteins was predicted using ProtComp (http://www.softberry.com/berry.phtml) (Vijayakumar et al., 2016). Conserved motifs were determined using MEME (meme-suite.org/tools/meme) with the following parameters: 5 ≤ width ≤ 60 and a maximum number of motifs 6 (BaileyTL and Buske, 2009). The function of these motifs was predicted using the tool hmmscan (https://www.ebi.ac.uk/Tools/hmmer/search/hmmscan).

Simple Sequence Repeats (SSR) markers and raffinose gene families-targeted miRNAs

SSR markers were detected in the *VvGolS* and *VvRfS* genes using the BatchPrimer 3v1.0 server (You et al., 2008). The gene CDS sequences were evaluated for miRNAs listed in the psRNA target database (Target), which target the raffinose gene families. This evaluation was performed using standard parameters against 186 published miRNAs of *V. vinifera*. The results were visualized using Cytoscape software (Draghici et al., 2007).

Promoter and gene expression analysis of VvGolS and VvRfS

The 1,500 bp upstream of each *VvGolS* and *VvRfS* gene were used for *cis*-regulatory element (CRE) detection and promoter analysis using the PlantCARE database (http://www.bioinformatics.psb.ugent.be/webtools/plantcare/html/). The developmental expression profile was performed by searching probe-set IDs for *VvGolS* and *VvRfS* using the Ensembl plant database (https://plants.ensembl.org/index.html) based on identifying the E-value of cDNA blast. Then, the array expression data of 54 plants from different green and tissue woody samples at different developmental stages were extracted using the Grape eFP Browser (hhttp://bar.utoronto.ca/efp_grape/cgi-bin/efpWeb.cgi) based on probe-set IDs (Fasoli et al., 2012; Wang et al., 2014). Probe-set IDs for *VvGolS* and *VvRfS* members were determined using BLAST searching in the Nimblegen Grape Whole-Genome Microarray 29K 090918-MD (nimbGrape) platform and were used for the effect of one hour of 20 µM ABA treatment and ABA-responsive transcription factor, known as ABF2, overexpression cell lines experiments microarray data that was obtained from the nimbGrape platform in PLEXdb. The RMA (Robust Multi-Array Average) algorithm was used to normalize the consequent unprocessed data, and gene fold change was calculated by comparing ABA and ABF data with the corresponding control data. The annotations were recorded based on information from well-curated databases such as NCBI, UniProt, and other relevant resources. Gene symbols, gene descriptions, gene ontology (GO) terms, and pathway annotations were among the recorded information. An extensive search into the literature gathered further functional annotations for the detected probe-set genes.

The Mev4.0 software worked with Pearson's correlation and the complete linkage algorithm to generate heat maps of developmental stage gene expression data, the ABA effect, ABF overexpression cell line, and hierarchical clustering (Saeed et al., 2003).

Results

Gene identification and phylogenetic analysis of GolS and RfS gene families in grapes

To evaluate the *GolS* and *RfS* gene families in grapevine, we performed a comprehensive analysis using protein sequences from *A. thaliana* (Vatansever et al., 2017) as references. The search occurred through BLASTp in the Phytozome database, identifying 19 *VvGolS and* 11 *VvRfS* genes in the grapevine genome. Subsequently, we validated the obtained sequences using the HMM database, confirming the presence of Glyco_trans_8 (PF01501) and Raffinose_synthase (PF05691) domains in the putative *GolS* and *RfS* proteins, respectively. Three protein sequences (GSVIVG01010036001, GSVIVG01024728001, and GSVIVG01006161001) assigned as putative *GolS* proteins did not exhibit the Glyco_trans_8 (PF01501) domain. Furthermore, we excluded one *VvGolS* (GSVIVG01031285001) and one *VvRfS* (GSVIVG01004662001) from the phylogenetic tree construction because of their short protein length and low E values within their respective families (Jiang et al., 2010). The remaining 15 *VvGolS* and 10 *VvRfS* genes were designated based on their chromosomal locations in the grapevine genome. Finally, we constructed a phylogenetic tree (Fig. 1) to elucidate the evolutionary relationships of the 15 *VvGolS* and 10 *VvRfS* genes within the grapevine and with other *GolS* and *RfS* species. Based on the phylogenetic analysis, we classified the *VvGolS* gene family into five distinct groups and the *VvRfS* gene family into six groups. This categorization provides insight into the diversity and evolutionary history of these gene families involved in grapevine raffinose biosynthesis.

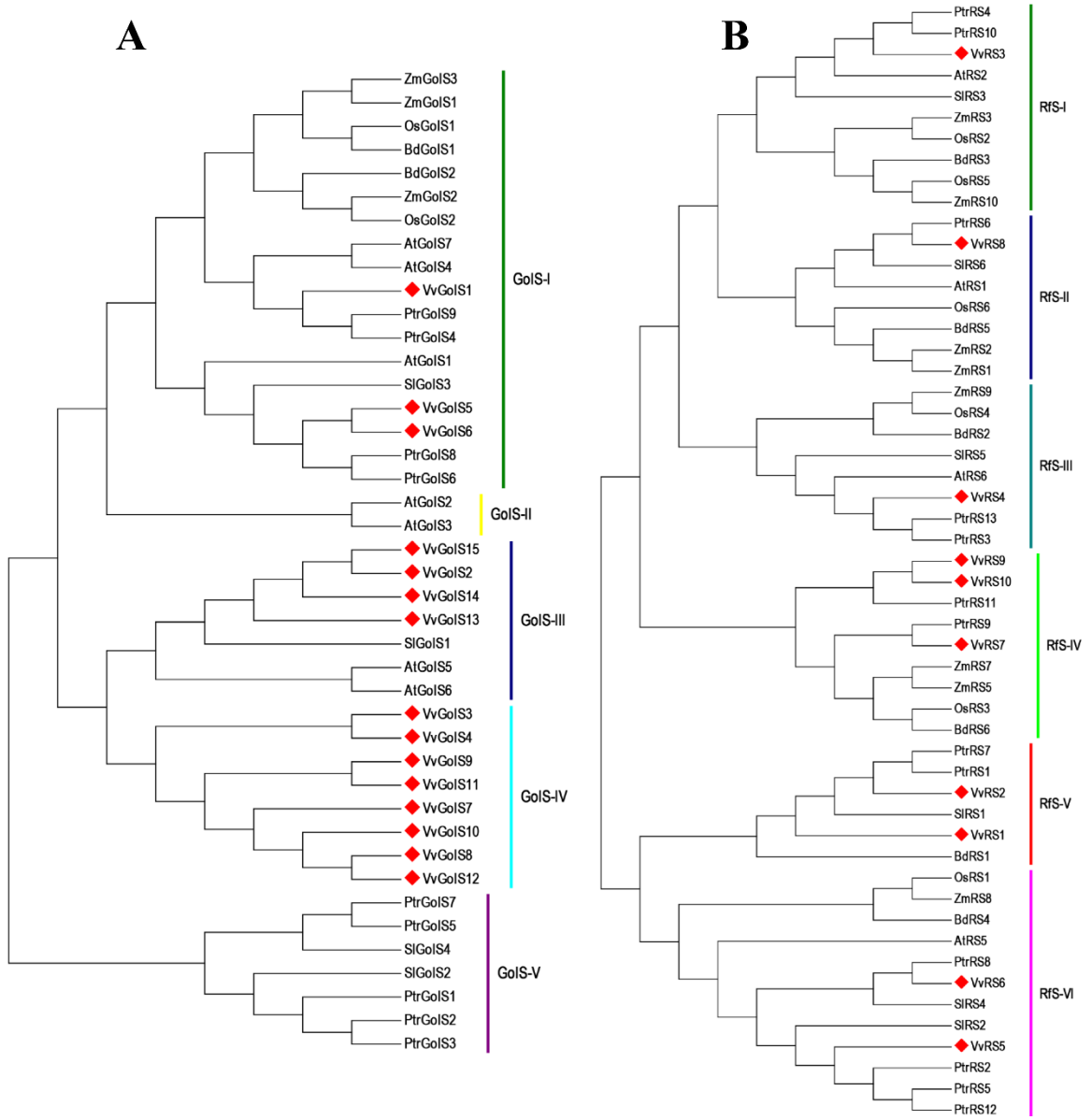

Fig. 1. Phylogenic trees of (A) GolS and (B) RfS protein sequences from Vitis vinifera, A. thaliana, Oryza sativa subsp. Japonica, Brachypodium distachyon, Solanum lycopersicum, Populus trichocarpa, and Zea mays were constructed using the neighbor-joining (NJ) method and MEGA 7.0.9 software with bootstrap 1,000 replicates. Grapevines are indicated with a red dot. Based on this evolutionary relationship, the GolS and RfS families are divided into five and six groups, respectively.

Analysis of evolutionary pressure in the raffinose biosynthesis pathway gene families of grapevine (Vitis vinifera L.), divergence times, and Ka/Ks estimation

To assess the evolutionary dynamics of the protein-coding sequences in gene families of the raffinose biosynthesis pathway in grapevine, we analyzed the ratio of non-synonymous (Ka) to synonymous (Ks) mutations (Wu et al., 2017). This ratio assists in measuring the selection pressure acting on these sequences. Where the Ka/Ks ratio exceeds one, it indicates adaptive (positive) selection. Ratios less than one signify purifying (negative) (Kayum et al., 2017) selection. A ratio equal to one suggests neutral evolution. We found that most duplicated gene pairs within the *VvRfS* families and *VvGolS* exhibited purifying selection pressure. However, specific gene pairs showed different patterns. For instance, the gene pair *VvGolS7* and *VvGolS10* experienced neutral evolution, indicating natural selection. However, the gene pairs *VvGolS9*, *VvGolS11*, *VvGolS7*,

and *VvGolS9* exhibited positive selection. By comparing the selection pressure between segmental and tandem duplications, we observed that genes derived from segmental duplication experienced more purifying selection (Ka/Ks ratio average of 0.0911) compared to genes originating from tandem duplication (Ka/Ks ratio average of 0.4566) (Table 1). This finding suggests that the duplicated genes resulting from segmental duplication have undergone more stringent selective constraints during their evolution. To estimate the divergence time between duplicated genes, we employed a divergence rate of 6.5610-9 mutations per synonymous site per year (Cao et al., 2014). Based on this rate, we determined that the divergence of these genes occurred approximately 347 million years ago (MYA) and persisted for up to 1 million years (MYA). Notably, the timing of divergence for the gene pairs originating from segmental duplication was significantly older than that of genes associated with a tandem duplication.

Table 1. Ks, Ka, and Ka/Ks calculation and divergent time of the duplicated grape VvGloS and VvRfS gene pairs.

Gene 1	Gene 2	Duplication Type	Ka	Ks	Ka/Ks	Selection	Time (MYA)
VvGolS5	*VvGolS7*	Segmental	0.0594	4.5212	0.0131	Purifying	347.78
VvRS5	*VvRS6*	Segmental	0.1394	2.0094	0.0694	Purifying	154.57
VvGolS3	*VvGolS10*	Segmental	0.1782	2.1415	0.0832	Purifying	164.73
VvGolS3	*VvGolS5*	Segmental	0.1620	1.9117	0.0847	Purifying	147.05
VvGolS4	*VvGolS10*	Segmental	0.1807	2.1310	0.0848	Purifying	163.92
VvGolS4	*VvGolS5*	Segmental	0.1697	1.9327	0.0878	Purifying	148.67
VvRS1	*VvRS2*	Segmental	0.2467	2.1460	0.1150	Purifying	165.08
VvGolS5	*VvGolS6*	Tandem	0.0066	0.0316	0.2089	Purifying	2.43
VvGolS10	*VvGolS12*	Tandem	0.0273	0.1154	0.2366	Purifying	8.88
VvGolS8	*VvGols12*	Tandem	0.0257	0.0897	0.2865	Purifying	6.90
VvGolS7	*VvGolS8*	Tandem	0.0149	0.0491	0.3035	Purifying	3.78
VvGolS8	*VvGols10*	Tandem	0.0236	0.0686	0.3440	Purifying	5.28
VvGolS9	*VvGolS10*	Tandem	0.1045	0.2500	0.4180	Purifying	19.23
VvGolS8	*VvGols9*	Tandem	0.1013	0.2065	0.4906	Purifying	15.88
VvGolS11	*VvGolS12*	Tandem	0.1244	0.2404	0.5175	Purifying	18.49
VvGolS8	*VvGols11*	Tandem	0.1355	0.2617	0.5178	Purifying	20.13
VvGolS7	*VvGolS12*	Tandem	0.0164	0.0292	0.5616	Purifying	2.25
VvGolS9	*VvGolS12*	Tandem	0.0835	0.1462	0.5711	Purifying	11.25
VvGolS10	*VvGolS11*	Tandem	0.1317	0.2198	0.5992	Purifying	16.91
VvGolS7	*VvGolS11*	Tandem	0.0744	0.0912	0.8158	Purifying	7.02
VvGolS7	*VvGolS10*	Tandem	0.0150	0.0150	1.0000	Neutral	1.15
VvGolS9	*VvGolS11*	Tandem	0.0741	0.0652	1.1365	Positive	5.02
VvGolS7	*VvGolS9*	Tandem	0.0845	0.0681	1.2408	Positive	5.24

Ks is the number of synonymous substitutions per synonymous site, Ka the number of nonsynonymous substitutions per nonsynonymous site, MYA million years ago.

Structural and localization analysis of raffinose biosynthesis pathway gene families in grapevines

To elucidate the organization and distribution of genes involved in the raffinose biosynthesis pathway in grapevine, we conducted a comprehensive analysis of the VvGolS and VvRfS gene families.

Exon-intron arrangement: Fig. 2 illustrates the exon-intron arrangement of the VvGolS and VvRfS family members. The VvGolS family exhibited a range of two to ten exons, with most family members containing four exons. In contrast, the VvRfS family displayed a more diverse exon count, ranging from four to sixteen exons.

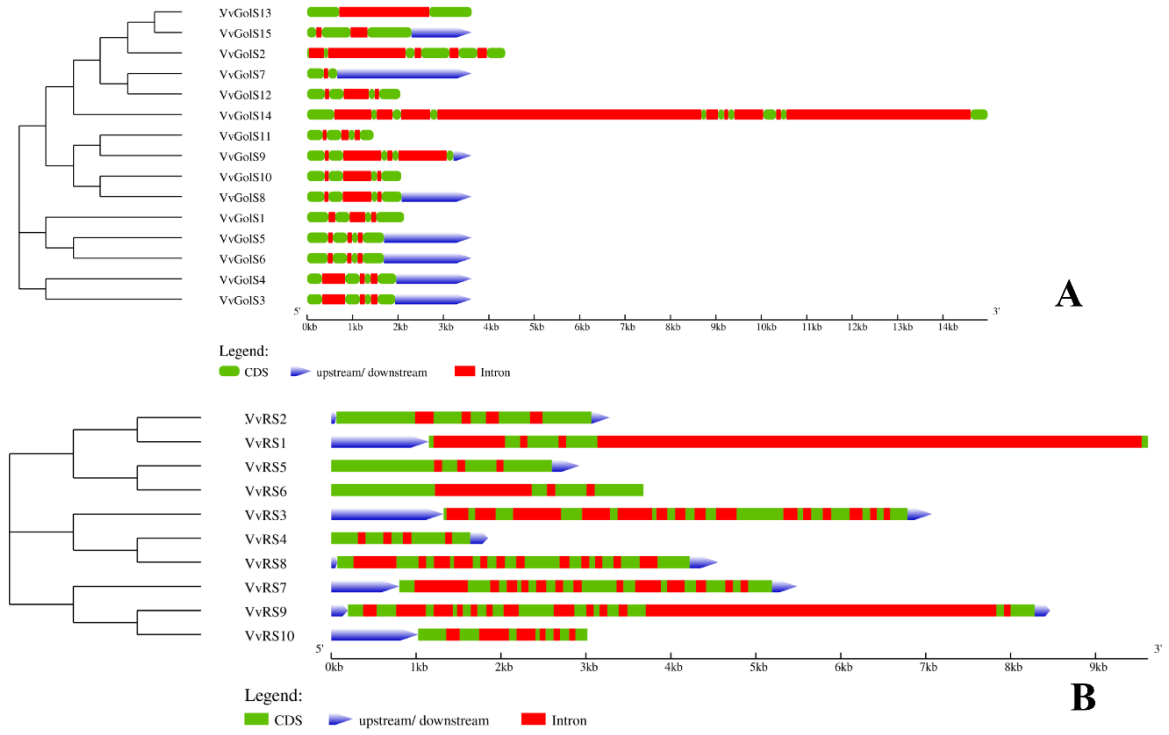

Fig. 2. Gene structure of the (A) GolS and (B) RfS proteins according to the phylogenies tree based on the neighbor-joining (NJ) method. CDSs are presented with green boxes. Arrow blue lines at either terminal of the genes show UTRs (untranslated regions) and the red boxes show introns.

Chromosomal localization: The *VvGolS* and *VvRfS* gene families had a distribution across ten linkage groups on nineteen chromosomes of the grape genome. Notably, two genes of the *VvRfS* family, *VvRfS9* and *VvRfS10*, were not located on the same chromosome as the others, indicating potential genetic rearrangements or translocation events.

Duplication events: We identified segmental and tandem duplications within the *VvGolS* and *VvRfS* gene families. Four linkage groups in the grape genome exhibited evidence of duplication events involving *VvGolS* and *VvRfS* genes. Specifically, the *VvGolS* gene family displayed two tandem duplication groups and five segmental duplication groups, as depicted in Fig. 3. In contrast, the *VvRfS* gene family showed only two segmental duplications.

Chromosomal clustering: Chromosome 14 contained a cluster of six consecutive genes, namely *VvGolS7, VvGolS8, VvGolS9, VvGolS10, VvGolS11,* and *VvGolS12*, involved in the raffinose biosynthesis pathway. These genes were close to each other without any intergenic spacing or gaps. Additionally, chromosome 7 harbored VvGolS5 and VvGolS6, which were mutually adjacent and exhibited high sequence similarity and identity (>98%). These genes also displayed evidence of segmental duplication. Furthermore, VvGolS5 showed segmental duplications with *VvGolS7, VvGolS3, and VvGolS4*, while *VvGolS10* had segmental duplications with *VvGolS3* and *VvGolS4.*

By investigating the structural characteristics, chromosomal distribution, and duplication patterns of the *VvGolS* and *VvRfS* gene families, we provide valuable insights into the genomic organization and evolution of the raffinose biosynthesis pathway in grapevine (*V. vinifera* L.). These findings contribute to our knowledge of the genetic basis underlying this valuable metabolic pathway and its potential implications for grapevine physiology and fruit quality.

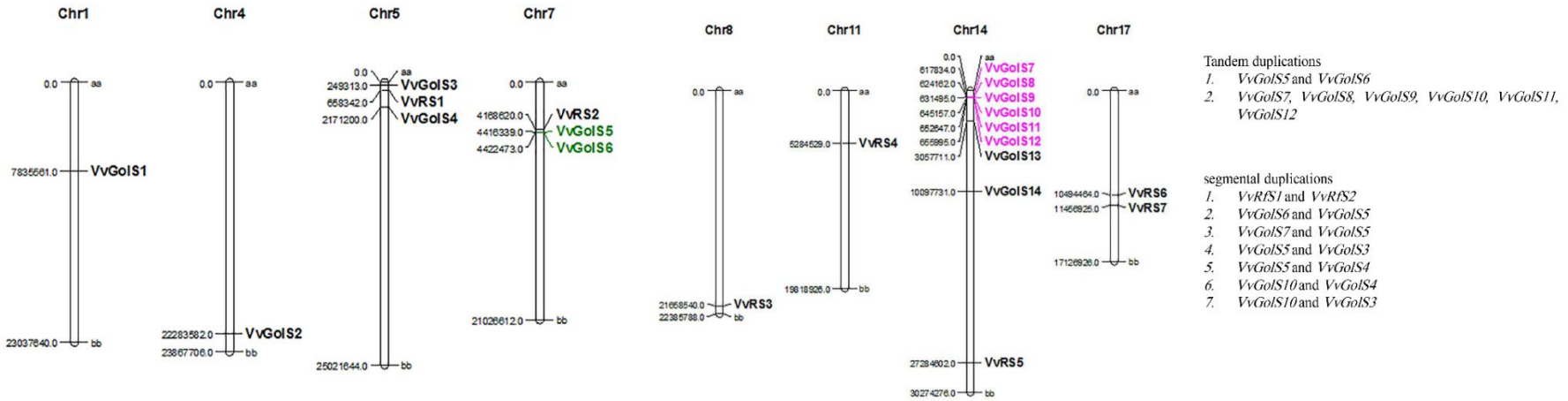

Fig. 3. Chromosomal position, segmental and tandem duplication of *VvGolS* and *VvRfS* genes on ten linkage groups of the grape genome.

Sequence analysis of VvGolS and VvRfS proteins

We conducted a detailed sequence analysis of the *VvGolS* and *VvRfS* proteins as vital components of the raffinose biosynthesis pathway in grapevine (*V. vinifera* L.).

The evaluation created insight into the characteristics and functional motifs in these protein families, supporting our assessment of the raffinose biosynthesis pathway in grapevine. Table 2 presents the sequence analysis results of the raffinose biosynthesis family proteins. The *VvRfS* member lengths supposedly ranged from 346 to 792 amino acids, while the *VvGolS* family member lengths varied from 162 to 590 amino acid residues. The isoelectric points (pIs) of *VvRfS* members mostly fell within a similar range, ranging from 5.26 to 6.25, with molecular weights ranging from 38.43 to 87.85 kDa. However, the pIs of *VvGolS* varied widely among family members, ranging from 4.83 to 9.02, with molecular weights ranging from 18.24 to 68.72 kDa. Subcellular localization predictions indicated that all *VvRfS* family members occurred primarily in the cytoplasm. Most members of the *VvGolS* family also exhibited cytoplasmic localization, while some appeared in the Golgi apparatus.

Two members supposedly appeared in the chloroplast and plasma membrane, suggesting potential diverse roles for *VvGolS* in different cellular compartments.

By further describing the functional motifs within *VvGolS* and *VvRfS* proteins, we employed the MEME tool, which predicted a total of six conserved motifs in each protein family (Fig. 4). Motif 1 in *VvRfS* exhibited a raffinose synthase or seed imbibition protein Sip1 domain, which appeared in all members except VvRfS10. The remaining five raffinose synthase motifs were 50 amino acids wide, while the sixth motif consisted of 44 amino acid residues.

Notably, seven members of the *VvRfS* family contained all six motifs, whereas *VvRS4* and *VvRS1* possessed three motifs (1, 2, and 4), and *VvRS10* contained three motifs (3, 5, and 6). These shared motifs suggest similar protein functions among *VvRfS* family members. Ten members in the *VvGolS* family had the glycosyltransferase family eight domain, indicating their involvement in glycosylation processes. The residues of these motifs ranged from 15 to 50 amino acids. While *VvGolS7* contained motifs 1 and 4, only the fourth motif appeared in *VvGolS2* and *VvGolS15*. Overall, our sequence analysis of the *VvGolS* and *VvRfS* proteins provides valuable insights into the structural and functional characteristics of these protein families in grapevine. These findings contribute to our understanding of the raffinose biosynthesis pathway and its gene families in *V. vinifera* L., suggesting further research possibilities on the metabolic processes and regulatory mechanisms underlying raffinose metabolism in grapevine.

Identification of SSR markers and prediction of raffinose gene family-targeted miRNAs

To evaluate the presence of Simple Sequence Repeat (SSR) markers within the raffinose biosynthesis pathway gene families, we analyzed the grapevine (*V. vinifera* L.) genome. A total of 27 SSRs were identified, distributed among 12 of the 25 raffinose gene families, with 7 SSRs found in *GolS* and 5 SSRs in *RfS* (Table 3).

Notably, most genes within these families contained a single SSR, with the highest number of SSRs observed in VvRS1 (8 SSRs) and VvGolS14 (4 SSRs). The identified SSRs displayed a range of repeat motifs, with tri-nucleotide repeats being the most frequent (9 SSRs), followed by di-nucleotide repeats (8 SSRs), tetra-nucleotide repeats (6 SSRs), and pentanucleotide and hexanucleotide repeat (2 SSRs each).

Furthermore, we predicted miRNAs that targeted the identified raffinose gene families. Our analysis of 36 vvimiRNAs revealed that 9 *VvRS* and 8 *VvGolS* genes were the targets of miRNAs (Fig. 5). Interestingly, all genes, except *VvRS2* and *VvRS5* (1 SSR each), were found to be repressed by at least one vvimiRNA. We observed a complex network of interactions between vvimiRNAs and raffinose gene families, highlighting their intertwined regulatory mechanisms. Notably, specific miRNAs acted as co-targets for multiple genes, such as vvi-miR398c and vvi-miR398b, which interacted with *VvRS3, VvGolS5*, and *VvGolS6.*

Table 2. Protein information of raffinose biosynthesis families in grape

Gene	Accession number	Peptide length	PI	MW(kDa)	Subcellular localization
RfS					
VvRS1	GSVIVT01034980001	347	5.54	39.01	Cytoplasmic
VvRS2	GSVIVT01028143001	792	5.48	87.85	Cytoplasmic
VvRS3	GSVIVT01033305001	792	6.02	87.12	Cytoplasmic
VvRS4	GSVIVT01015589001	422	6.25	46.14	Cytoplasmic
VvRS5	GSVIVT01032425001	775	5.54	85.24	Cytoplasmic
VvRS6	GSVIVT01007681001	780	5.54	86.26	Cytoplasmic
VvRS7	GSVIVT01007597001	750	5.99	83.07	Cytoplasmic
VvRS8	GSVIVT01014778001	758	6.04	83.48	Cytoplasmic
VvRS9	GSVIVT01002784001	739	5.26	81.47	Cytoplasmic
VvRS10	GSVIVT01005620001	346	5.53	38.43	Cytoplasmic
GolS					
VvGolS1	GSVIVG01013763001	239	4.87	27.82	Chloroplast
VvGolS2	GSVIVG01026525001	568	7.22	56.42	Golgi
VvGolS3	GSVIVG01034938001	240	4.89	28.13	Cytoplasmic
VvGolS4	GSVIVG01017634001	239	4.95	27.99	Cytoplasmic
VvGolS5	GSVIVG01028174001	247	4.93	29.02	Cytoplasmic
VvGolS6	GSVIVG01028176001	247	4.94	29.11	Cytoplasmic
VvGolS7	GSVIVG01031284001	162	6.10	18.24	Cytoplasmic
VvGolS8	GSVIVG01031282001	234	4.85	27.40	Cytoplasmic
VvGolS9	GSVIVG01031280001	243	5.05	28.25	Cytoplasmic
VvGolS10	GSVIVG01031278001	234	4.83	27.30	Cytoplasmic
VvGolS11	GSVIVG01031276001	197	4.86	23.61	Cytoplasmic
VvGolS12	GSVIVG01031274001	324	5.25	36.85	Cytoplasmic
VvGolS13	GSVIVG01000046001	546	8.98	63.31	Golgi
VvGolS14	GSVIVG01023535001	546	9.02	61.24	Plasma membrane
VvGolS15	GSVIVG01009501001	590	8.85	68.72	Golgi

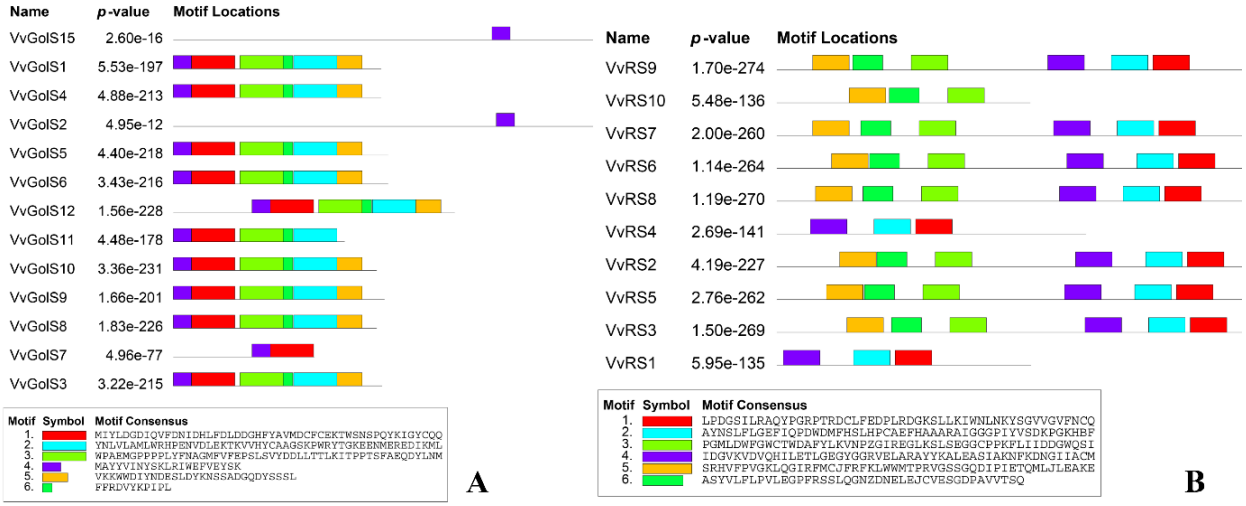

Fig. 4. Six conserved motifs were determined using the MEME and function verified by the hmmscan tool. (A) Ten *VvGolS* family members had the glycosyltransferase family 8 domain. VvGolS7 had motifs 1 and 4, whereas only the fourth motif was found in *VvGolS2* and *VvGolS15* (B) Motif 1 in *VvRfS* had a raffinose synthase or seed imbibition protein Sip1 domain which was detected in all members except *Vv*RfS10. Seven members of the *VvRfS* family contained all six motifs, *VvRS4* and *VvRS1* had three motifs 1, 2 and 4, whereas *VvRS10* contained three motifs 3, 5 and 6.

Table 3. Twenty-seven SSRs were identified in twelve raffinose gene family sequences in *Vitis vinifera*.

Gene Name	count	SSR Motives
VvGolS13	3	(AT)9, (TTA)10, (TTTG)4
VvGolS15	1	(CCT)5
VvGolS1	1	(TGA)8
VvGolS4	2	(AT)10, (AAAT)3
VvGolS14	4	(TC)10, (AT)15, (TA)9, (ACAA)3
VvGolS12	1	(TTC)5
VvGolS3	1	(AT)11
VvRS7	1	(ATTTTT)3
VvRS6	2	(ATAA)3, (TTTCT)4
VvRS5	1	(ATTTTT)3
VvRS3	2	(TGT)4, (AAG)4
VvRS1	8	(AT)11, (AT)17, (CTC)5, (GAA)9, (ATT)6, (TTCC)5, (TAAT)3, (TTTTA)4

Furthermore, we predicted miRNAs that targeted the identified raffinose gene families. Our analysis of 36 vvimiRNAs revealed that 9 *VvRS* and 8 *VvGolS* genes were the targets of miRNAs (Fig. 5). Interestingly, all genes, except *VvRS2* and *VvRS5* (1 SSR each), were found to be repressed by at least one vvimiRNA. We observed a complex network of interactions between vvimiRNAs and raffinose gene families, highlighting their intertwined regulatory mechanisms. Notably, specific miRNAs acted as co-targets for multiple genes, such as vvi-miR398c and vvi-miR398b, which interacted with *VvRS3, VvGolS5*, and *VvGolS6.*

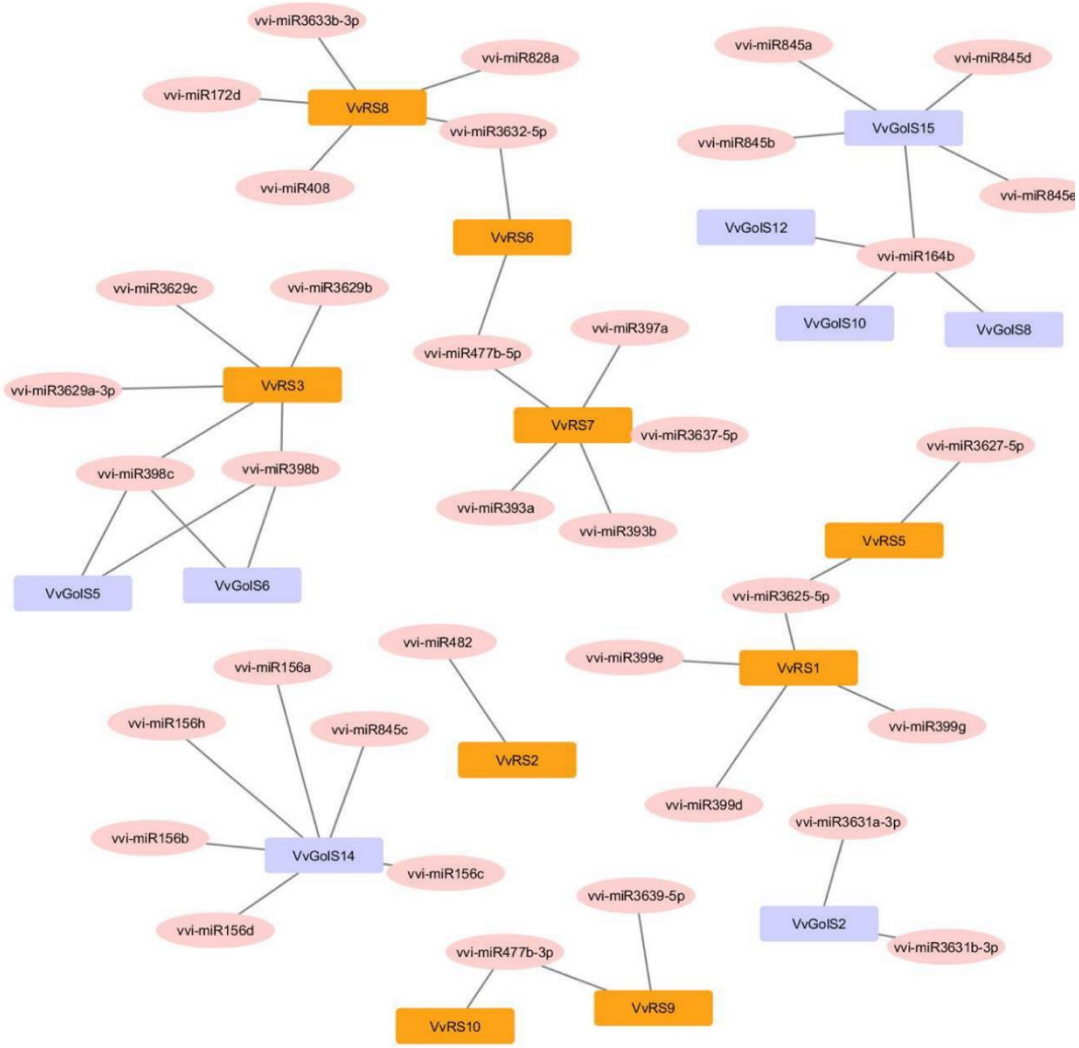

Fig. 5. Interaction network of raffinose gene families-targeted miRNAs. Thirty-six miRNAs (pink) from *Vitis vinifera* were identified for the *VvGolS* (violet) and *VvRfS* (orange) genes.

Promoter characteristics and gene expression patterns

To evaluate the regulatory mechanisms underlying raffinose biosynthesis in grapevine, we analyzed the promoter regions and gene expression profiles of the *RfS* and *GolS* gene families. We identified four abiotic stress-responsive elements in the promoter regions, namely ABRE (involved in abscisic acid responsiveness), LTR (involved in low-temperature responsiveness), HSE (involved in heat stress responsiveness; Zhang et al., 2005), and ARE (essential for anaerobic induction). We quantified the occurrence of these elements in the promoter region of *RfS* and *GolS* genes.

In the *RfS* gene family, all members had at least one ARE element in their upstream regions, while ABRE, LTR, and HSE elements were present in almost all members. Notably, *VvRS3* exhibited the highest number of these four cis-regulatory elements, whereas *VvRS9* was only associated with the ARE element. In contrast, the *GolS* gene family displayed greater diversity in cis-regulatory elements. ABRE was the most prevalent cis-element in this family, with *VvGolS14* and VvGolS8 containing all four *cis*-elements, while *VvGolS7* had the LTR element only (Table 4).

Table 4. Four cis-regulatory elements ABRE, LTR, HSE, and ARE, upstream of the *RafS* and *GolS* genes.

Genes	*Cis* regulatory elements			
RFS	**ABRE**	**LTR**	**HSE**	**ARE**
VvRS1	3	0	1	4
VvRS2	3	2	1	1
VvRS3	4	2	3	4
VvRS4	3	4	1	2
VvRS5	1	0	3	1
VvRS6	0	0	4	1
VvRS7	0	1	2	5
VvRS8	2	3	0	2
VvRS9	0	0	0	4
VvRS10	1	0	0	1
GolS				
VvGolS1	2	0	0	1
VvGolS2	0	0	1	2
VvGolS3	4	0	1	1
VvGolS4	2	0	0	1
VvGolS5	3	0	5	2
VvGolS6	3	0	4	0
VvGolS7	0	0	0	3
VvGolS8	3	2	1	2
VvGolS9	1	2	1	0
VvGolS10	3	2	1	0
VvGolS11	2	1	0	3
VvGolS12	4	1	0	0
VvGolS13	0	1	5	0
VvGolS14	1	1	2	1
VvGolS15	2	1	0	0

ABRE is involved in the abscisic acid responsiveness, LTR is involved in low-temperature responsiveness, HSE is involved in heat stress responsiveness, and ARE is essential for anaerobic induction.

To gain insights into the expression patterns of *VvGolS* and *VvRfS* genes, we constructed a heat map based on their gene expression levels across 54 different tissues and developmental stages of grapevine (Fig. 6). Among the *VvGolS* genes, *VvGolS14* had the highest expression pattern across all tissues and developmental stages, whereas *VvGolS3* and *VvGolS4* displayed the lowest expression levels. Interestingly, *VvGolS3* and *VvGolS4* had similar gene expression patterns, with up-regulation observed only at the leaf-senescence stage. *VvGolS11* and *VvGolS13* demonstrated lower expression patterns compared to other gene members. *VvGolS2*, *VvGolS15*, *VvGolS5*, and *VvGolS6* exhibited similar and intermediate expression patterns in all analyzed situations. *VvGolS7, VvGolS8, VvGolS9, VvGolS10,* and *VvGolS12* displayed consistently high expression levels.

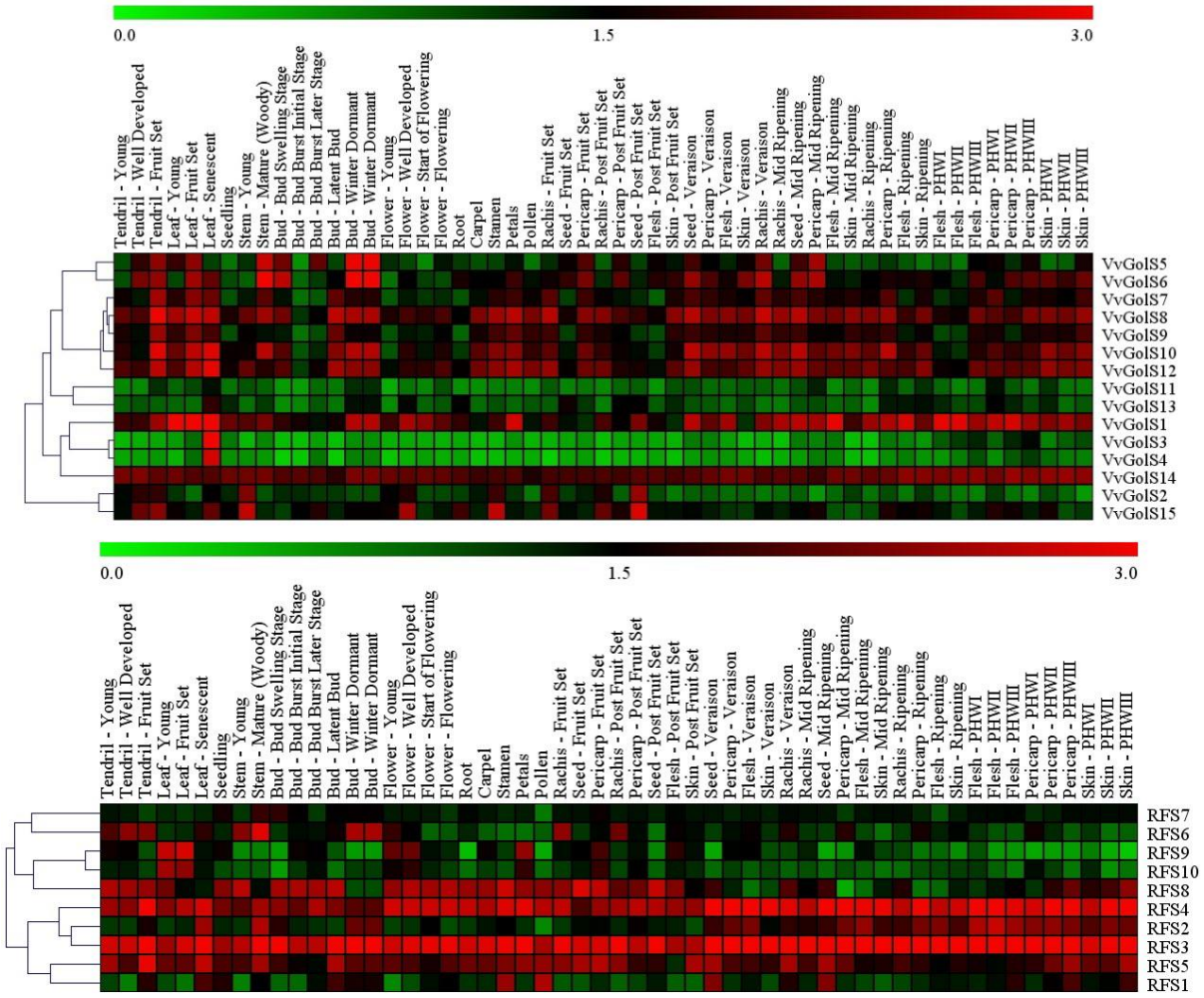

Fig. 6. (A) *VvGolS* and (B) *VvRfS* gene expression profiles in 54 different herbaceous and woody tissues and organs at various developmental phases. The genes that have similar profiles are grouped in hierarchical clustering on the left of the heat map. The value of the expression is shown in a colored bar at the top of the chart. The lowest, medium, and the highest are illustrated with green, black and red colors, respectively, and the scale bar is in log10 values from 0 to 3.

In the *VvRfS* family, the genes *VvRfS3* and *VvRfS4* had the highest expression patterns across all tissues and developmental stages, followed by *VvRfS5, VvRfS8,* and *VvRfS2*. Other genes demonstrated distinct expression patterns, with up-regulation in some tissues and stages and down-regulation in others. *VvRfS9* and *VvRfS10* displayed nearly identical up- and down-regulation patterns.

To explore the effects of treatment with 20 µM ABA and overexpression of ABF2 cells on gene expression in the *VvGolS* and *VvRfS* families, we analyzed their expression profiles (Fig. 7). The highest up-regulation in response to 20 µM ABA treatment occurred in *VvRfS6, VvRfS1*, and *VvRfS5*. However, *VvRfS7* and *VvRfS9* exhibited the weakest response to this hormone. The weakest effect of this hormone occurred by e*VvRfS7* and *VvRfS9*. Also, *VvRfS6* was up-regulated by the overexpression of ABF2 compared to other genes, but the least up-regulation occurred by *VvRfS9* and *VvRfS1*. The strongest effect of 20 µM ABA on the *VvGolS* gene family occurred by *VvGolS11*, *VvGolS7*, and *VvGolS3*, respectively, and the weakest effect by *VvGolS13*. Overexpression of ABF2 negatively affected the partial expression of the *VvGolS* gene family. Overall, treatments with 20 µM ABA were more effective than the ABF2 overexpression strategy in up-regulating the *VvRfS* and *VvGolS* gene families.

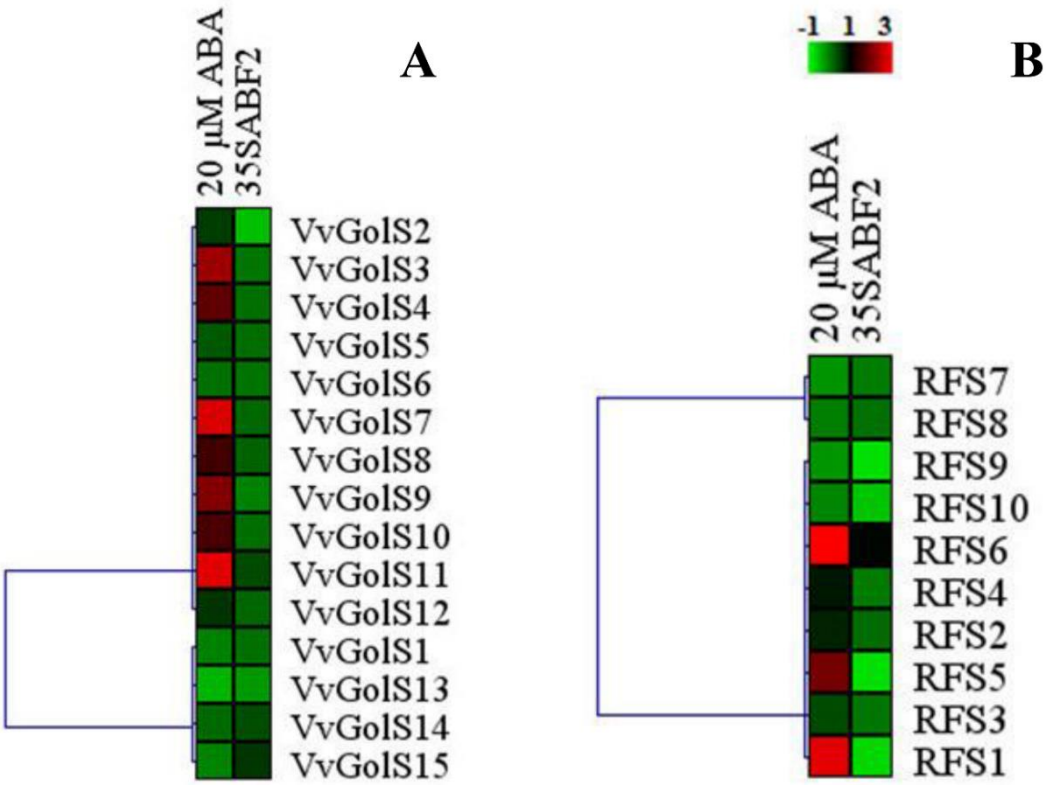

Fig. 7. Effect of one hour, 20 µM ABA and ABF2 overexpression cell lines on (A) VvGolS and (B) VvRfS gene expression profiles. Genes that have similar profiles are grouped in hierarchical clustering on the left of the heat map. The value of the expression is shown in a colored bar at the top of the chart. The lowest, medium, and highest are illustrated with green, black and red colors, respectively, and the scale bar is based on fold change compared to the control condition value.

Discussion

Fifteen *VvGolS* and 10 *VvRfS* were identified and classified into five and six groups, respectively, based on a phylogenetic tree in consistency with the classification of the GolS family in *P. trichocarpa* and Sesamum indicum, and the *VvRfS* family in *S. indicum* (You et al., 2018; Zhou et al., 2014). The members of the *VvRfS* protein family were supposedly located in the cytoplasm, which is consistent with the localization of raffinose synthase enzyme activity in the cytoplasmic leaves of cucumber (Li et al., 2018). *GolS* was predominantly located in the cytosol, likely associated with its osmoregulatory function in the cytoplasm. However, in some group members, *GolS* was found in different subcellular locations, as observed in the subcellular localization of *S. lycopersicum* and *B. distachyon* (Filiz et al., 2015). *VvGolS13, VvGolS15*, and *VvGolS2* are localized in the Golgi apparatus and form the same group in the phylogenetic tree. All three members only have the fourth motif but exhibit varying numbers of cis-elements and gene structures. *VvGolS15* and *VvGolS2* show similar expression patterns across different organs and developmental stages. It appears that *VvGolS* and *VvRfS* predominantly occur in the cytoplasm. However, raffinose oligosaccharides in leaves can also originate from the vacuole, chloroplast, and, in some cases, storage in mesophyll cells and epidermal tissues (Gu et al., 2018; Mayer et al., 2021). *VvGolS15* and *VvGolS2*, located in the same group in the phylogenetic tree, share only the fourth motif, suggesting a conserved role in the evaluation process (Filiz et al., 2015).

Furthermore, both *VvGolS15* and *VvGolS2* lacked motif 1. In the clustering analysis of *VvGolS* gene expression across different organs and tissues, *VvGolS15* and *VvGolS2* were grouped, while the remaining members formed another cluster. A similar pattern occurred in the *VvRfS* family, where seven members had all six functional motifs. However, *VvRfS10* contained only three motifs, while *VvRfS1* and *VvRfS4* exhibited the other three motifs. Analyzing the expression of *VvRfS* genes in various tissues and organs

revealed that *VvRfS1, VvRfS4, VvRfS5, VvRfS3,* and *VvRfS2* belonged to one group, whereas *VvRfS10, VvRfS7, VvRfS6, VvRfS9,* and *VvRfS8* formed another group. These findings confirm a correlation between functional motifs in family members and gene expression levels.

The *VvGolS* family exhibited a consistent exon structure, with almost all members having four exons. On the other hand, the *VvRfS* family displayed variable exon counts, ranging from four to sixteen. Similar exon-intron patterns were identified in *Gol*S genes across different plant species, such as cotton with three exons (Zhou et al., 2012), *S. lycopersicum* and *B. distachyon* with two to three exons (Filiz et al., 2015), and Musa acuminata with three to four exons (Dolcimasculo et al., 2018). The disparity in gene structure may emanate from gene mutations during the monocot-dicot split (de Gois et al., 2020; Filiz et al., 2015).

Chromosome 14 contained the highest number (9) of RFO genes, and the abundance of these RFO genes was associated with *GolS* gene tandem duplications. Both *VvRfS* and *GolS* genes experienced multiple tandem and segmental duplications. Functional segregation in duplicated genes could lead to new molecular variations in organisms. This finding suggests gene duplication can lead to functional divergence, enabling tolerance to various abiotic stress treatments (Filiz et al., 2015; Kuzmin et al., 2021).

This suggestion is supported by two groups of segmental duplications between *VvRfS1* and *VvRfS2, VvRfS5,* and *VvRfS6,* although each segmental group occurred in the same phylogenetic groups. Tandem duplications are usual in cells. Perhaps tandem duplications in plants are more dependent on stress responses (Lu et al., 2022). SSRs are repetitive sequences of 1-6 nucleotides that vitally orchestrate gene regulation (Hasan et al., 2017). For example, tri-nucleotides (33.34% in the current study) cause gene silencing in non-translated regions. However, in coding regions, they alter gene transcription, translation, and expression. The predominant SSRs are variants of different types. The parent type of SSR is taxon-dependent, and the SSR frequency of the AT type is higher in dicotyledons than in monocotyledons (Qin et al., 2015). The SSRs obtained in *VvRfS* and *GolS* can identify polymorphic varieties and their relationship to the raffinose biosynthesis pathway. They can screen more resistant varieties to abiotic and biotic stresses based on marker-assisted selection (MAS) in breeding programs. Also, the analysis of the genetic structure of grapevine germplasm using SSR markers has unveiled the presence of two distinct subpopulations within the association panel. This finding carries significant implications for determining future breeding values (Razi et al., 2023).

MiRNAs are non-coding small RNAs (19–24 bp) that have a crucial function in the plant growth process and environmental stresses with post-transcriptional regulation (Wang et al., 2017). The miRNAs function through a regulatory cause by targeting specific mRNAs for degradation and repressing the expression of the target gene (Song et al., 2019). Bioinformatic techniques can usefully determine the potential of miRNA interactions associated with specific gene families (Šečić et al., 2021). The study on the expression of vvi-miR156b/c/d showed generic temporal, spatial, and particulate expression. Expression levels were higher in vegetative organs than in reproductive ones (Wang et al., 2016). Moreover, vvi-miR156b/c/d/e interacted with carbohydrate metabolism genes at the biological level, confirming their role in coordinating the raffinose biosynthesis pathway (Palumbo et al., 2014). Each *VvRfS* and *GolS* gene contained a type and number of a specific cis-regulatory element, indicating that a form is up- or down-regulated under particular stress conditions. Several tandem duplications occurred in the *GolS* gene family. All these tandem groups were in the same phylogenetic tree group, and almost all had the same peptide length, PI, MW (kDa), and gene structure, especially *GolS5* and *GolS6, GolS3* and *GolS4*. Hypothetically, tandem gene copies should perform better than a single gene copy (Loehlin and Carroll, 2016). This theory applied to tissue- and development-specific gene expression, similar to tandem groups clustered in the same gene expression profile group. This condition suggests that the plant may require additional amounts of an external product of this gene; therefore, the mechanism of gene duplication may contribute to the overexpression of the gene, while the number and type of predicted upstream regulatory elements of these genes were different and may be activated under different stress conditions.

Duplicated genes, based on the type of evolutionary pressure, undergo selective processes such as hypo-functionalization, neo-functionalization, sub-functionalization, and non-functionalization. However, many duplicated genes are extinguished within millions of years, and some duplicated genes preserved during evolution undergo very intense purification selection (Loehlin and Carroll, 2016). A study

on *VvRfS* and *VvGolS* gene selection showed that many were subject to intensive purifying selection, much more intensive than those of genes created by segmental duplication. Purifying selection limits the functional divergence of genes during evolution (Wang et al., 2018). Thus, two duplicated genes, *VvGolS* and *VvRfS*, have essentially maintained their roles in function during the evolutionary stages.

An estimation of duplication timing for the *VvGolS* and *VvRfS* families revealed that segmental duplication is much older than tandem duplication in both families. Except for the pair of *VvGolS5* and *VvGolS7* genes, which duplicated 347 million years ago (MYA), other segmental duplications occurred after the separation of monocotyledons and dicotyledons, which are older than 170 MYA (Li et al., 2021; Yang et al., 1999). The timing of divergence of the gene pairs resulting from segmental duplication and the robust selection pressure that led to purification indicate the importance of the functional role of these genes (Wang et al., 2018).

Two essential elements in transcriptional control are transcription factors and cis-regulatory elements in the upstream region of genes (Zemlyanskaya et al., 2021). This line of argument led to an understanding of the effect of these two elements, particularly the major cis-acting abiotic stress response elements such as ABRE, LTR, HSE, ARE, and WRKY, on the promoter of the raffinose biosynthesis family gene, as well as the effect of ABA treatment and ABA-responsive transcription factors on gene expression profiling. Abiotic stress triggers a series of responses, including the reception of stress signals, signal transduction, and modulation of gene expression profiles, ultimately negatively affecting plant growth and development. One of the cellular mechanisms protecting plants from abiotic stress is the accumulation of osmoprotectants such as raffinose (He et al., 2018). Rejŝková et al. (2007) conducted a study to examine the alterations in the soluble carbohydrate composition of shoot segments in olives exposed to chilling temperatures using *in vitro* conditions. Researchers noted a rise in the overall concentration of endogenous carbohydrates, with notable variations observed in mannitol and raffinose RFO levels. Abiotic stress can regulate the expression of the RFOs gene family by binding to these cis-regulatory elements upstream of the *VvGolS* and *VvRfS* genes in response to abiotic stress (Zhou et al., 2014). These four major cis-regulatory elements suggest that the RFOs gene family responds effectively to cold, heat, and anaerobic conditions.

In this study, ABA treatment and ABA-responsive transcription factors, particularly ABA, significantly affected the up-regulation of raffinose biosynthesis family genes. ABA upregulated *PtrGolS2* and *PtrGolS5-PtrGolS7*, but their expression levels were lower than those observed under osmotic and salt stress conditions (Zhou et al., 2014). ABA treatment on chickpeas increased raffinose content and induced tolerance to dehydration (Salvi et al., 2020). During the cold acclimation process, there is a notable buildup of ABA in woody tissues. This increase in ABA levels triggers the up-regulation of *VivRafS5*, which subsequently stimulates the synthesis of raffinose (Noronha et al., 2022). These results suggest the role of ABA in raffinose-oligosaccharide metabolism, and one of the factors regulating raffinose accumulation is ABA. The concentration of raffinose and seed longevity was lower than the wild type in mutant abi3 and aba1/abi3 seeds (Zhou et al., 2020). Several research cases reported an increase in raffinose oligosaccharides upon low-temperature stress, such as in cucumber (Gu et al., 2018), rice, Arabidopsis, kidney bean (*Phaseolus vulgaris*), tomato (*S. lycopersicum*) (ElSayed et al., 2014), maize (Li et al., 2020), and poplar (Liu et al., 2021). In genetic research, the overexpression of *GolS* members has reportedly increased tolerance to low temperatures, e.g., overexpression of *CsGolS1* in cucumber improves cold tolerance in this target plant along with increased levels of RFOs (Dai et al., 2020). The location of RFO synthesis and related gene expression is known to control RFO biosynthesis. Raffinose biosynthesis gene families such as *GolS* and *RfS* have more than one isoform, and each member has different expression levels in various tissues and under different stress conditions (Sengupta et al., 2015). A similar case resulted from gene family expressions of raffinose biosynthesis in previous research. Transcriptional profiling of *GolS* in Panicum virgatum and *P. hallii* showed that the genes are expressed in different tissues and have distinctive tissue-specific expression patterns. The enzymes of these genes have diverse physiological functions in multiple tissues (de Gois et al., 2020). Expression in different tissues is due to the particular situation in which the gene expresses itself, creating a barrier between their metabolites. In *Z. mays*, three isoforms of *GolS* correlated with seed development and callus formation. Three GolS hybrids in poplar showed daily and annual spatial and temporal expression profiles (dos Santos and Vieira, 2020). Transgenic *Gols* from cotton (*GhGolS*) were also expressed in specific patterns in anthers, fibers, and leaves and localized in the cell membrane

(dos Santos & Vieira, 2020; Zhou et al., 2012). Furthermore, comparing metabolites during cold acclimation between *V. amurensis* and *V. vinifera* revealed distinct accumulation patterns of galactinol, indicating the presence of different *GolS* members in the two grape species (Chai et al., 2019). Besides, the heat stress-responsive factor *VvHsfA2* activates the promoter of *VvGOLS1* in a manner dependent on heat stress conditions. The expression pattern of *VvGOLS1* demonstrates a positive correlation with the accumulation of galactinol in grapevine berries experiencing heat stress. This observation highlights the signaling role of galactinol in these particular responses (Pillet et al., 2012). Overexpression of the *GolS* gene resulted in the accumulation of galactinol and raffinose, which increased tolerance to abiotic stress in the transgenic plant. Therefore, the *GolS* gene is a suitable target for molecular engineering and breeding programs to aid in plant resistance to abiotic stress (dos Santos and Vieira, 2020).

The observed up-regulation of *GolS* and *RfS* genes under drought stress further supports their putative involvement in the drought stress response mechanism. The *GolS* genes are known to be involved in the biosynthesis of compatible solutes such as trehalose, which play a crucial role in osmotic adjustment and protection against cellular damage caused by drought stress. Similarly, the *RfS* genes correlated with regulating reactive oxygen species (ROS) and antioxidant defense pathways, essential for mitigating oxidative stress induced by drought conditions (Jing et al., 2023).

Conclusion

The current research provided valuable insights into the genetic characteristics and potential functions of the *GolS* and *RfS* gene families in grapes under drought-stress conditions. Using bioinformatics analysis, we identified a set of putative genes that may vitally assist in the adaptive response of grapevines to water scarcity. These findings have important implications for the genetic improvement of grapes, particularly in overcoming barriers related to drought tolerance and water-use efficiency. Understanding the molecular mechanisms underlying the response to drought stress is essential for developing targeted strategies for crop improvement. By elucidating the potential functions of the *GolS* and *RfS* gene families, our study contributes to the existing knowledge of the genetic basis of drought adaptation in grapes. These findings can serve as a foundation for further investigations and experimentation to confirm how these genes are involved in the response to drought stress.

Identifying candidate genes in drought response opens avenues for future research and application in grape breeding programs. This knowledge can be leveraged to develop new grape cultivars with enhanced tolerance to water scarcity, ultimately contributing to sustainable agriculture and high-quality grape cultivars. Our study thoroughly analyzed the *GolS* and *RfS* gene families in grapes, highlighting their potential significance in drought tolerance. We emphasize the need for further experimental evaluations and validations to confirm the functional roles of these genes. The outcomes of this research are promising for grape breeding programs and the development of tolerant grape varieties capable of withstanding challenging environmental conditions.

Conflict of Interest

The authors indicate no conflict of interest for this work.

References

Arab MM, Brown PJ, Abdollahi-Arpanahi R, Sohrabi SS, Askari H, Aliniaeifard S, Vahdati K. 2022. Genome-wide association analysis and pathway enrichment provide insights into the genetic basis of photosynthetic responses to drought stress in Persian walnut. Horticulture Research 9, uhac124.

Arab MM, Marrano A, Abdollahi-Arpanahi R, Leslie CA, Askari H, Neale DB, Vahdati K. 2019. Genome-wide patterns of population structure and association mapping of nut-related traits in Persian walnut populations from Iran using the Axiom J. regia 700K SNP array. Scientific reports 9(1), 6376.

Asadi A, Shariati V, Mousavi S, Mariotti R, Mazinani MH. 2023. Meta-analysis of transcriptome reveals key genes relating to oil quality in olive. Authorea, 1-11.

Bailey TLB, Buske F. 2009. Memesuite: tools for motif discovery and searching. Nucleic Acids Research 37, W202rW208.

Behzadi Rad P, Roozban MR, Karimi S, Ghahremani R, Vahdati K. 2021. Osmolyte accumulation and sodium compartmentation has a key role in salinity tolerance of pistachios rootstocks. Agriculture 11(8), 708.

Cao J, Han X, Zhang T, Yang Y, Huang J, Hu X. 2014. Genome-wide and molecular evolution analysis of the subtilase gene family in *Vitis vinifera*. BMC genomics 15(1), 1-15.

Chai F, Liu W, Xiang Y, Meng X, Sun X, Cheng C, Li S. 2019. Comparative metabolic profiling of *Vitis amurensis* and *Vitis vinifera* during cold acclimation. Horticulture Research 6, 1-12.

Cui LH, Byun MY, Oh HG, Kim SJ, Lee J, Park H, Kim WT. 2020. Poaceae Type II galactinol synthase 2 from Antarctic flowering plant *Deschampsia antarctica* and

rice improves cold and drought tolerance by accumulation of raffinose family oligosaccharides in transgenic rice plants. Plant and Cell Physiology 61(1), 88-104.

Dai H-B, Zhu Z-H, Wang Z-G, Zhang Z, Kong W, Miao M. 2020. Cold-induced intermediary cell-specific overexpression of galactinol synthase 1 revealed a unique strategy improving stress adaptability of cucumber. Authorea Preprints 3, 1-7.

de Gois EHB, Menegazzo RF, dos Santos TB, de Souza SGH. 2020. Identification, evolutionary and expression analysis of the galactinol synthase (*GolS*) genes in *Panicum virgatum* L. and *Panicum hallii*: an in silico approach. Plant Gene 24, 100262.

Dolcimasculo F, Ribas AF, Vieira LGE, dos Santos TB. 2018. A genome-wide analysis of the galactinol synthase gene family in banana (*Musa acuminata*). Colloquium Agrariae 14(3), 01-11.

dos Santos TB, Vieira LGE. 2020. Involvement of the galactinol synthase gene in abiotic and biotic stress responses: a review on current knowledge. Plant Gene 24, 100258.

Downie B, Gurusinghe S, Dahal P, Thacker RR, Snyder JC, Nonogaki H, Bradford KJ. 2003. Expression of a galactinol synthase gene in tomato seeds is up-regulated before maturation desiccation and again after imbibition whenever radicle protrusion is prevented. Plant Physiology 131(3), 1347-1359.

Draghici S, Khatri P, Tarca AL, Amin K, Done A, Voichita C, Romero R. 2007. A systems biology approach for pathway level analysis. Genome Research 17(10), 1537-1545.

Elango D, Rajendran K, Van der Laan L, Sebastiar S, Raigne J, Thaiparambil NA, Ferela A. 2022. Raffinose family oligosaccharides: friend or foe for human and plant health? Frontiers in Plant Science 13, 829118.

ElSayed AI, Rafudeen MS, Golldack D. 2014. Physiological aspects of raffinose family oligosaccharides in plants: protection against abiotic stress. Plant Biology 16(1), 1-8.

Fasoli M, Dal Santo S, Zenoni S, Tornielli GB, Farina L, Zamboni A, Murino V. 2012. The grapevine expression atlas reveals a deep transcriptome shift driving the entire plant into a maturation program. The Plant Cell 24(9), 3489-3505.

Felsenstein J. 1985. Confidence limits on phylogenies: a justification. Evolution 39, 783-791.

Filiz E, Ozyigit II, Vatansever R. 2015. Genome-wide identification of galactinol synthase (*GolS*) genes in *Solanum lycopersicum* and *Brachypodium distachyon*. Computational Biology and Chemistry 58, 149-157.

Gasteiger E, Hoogland C, Gattiker A, Wilkins MR, Appel RD, Bairoch A. 2005. Protein identification and analysis tools on the ExPASy server. The proteomics protocols handbook, Humana Press. 571-607.

Goodstein DM, Shu S, Howson R, Neupane R, Hayes RD, Fazo J, Putnam N. 2012. Phytozome: a comparative platform for green plant genomics. Nucleic Acids Research 40(D1), D1178-D1186.

Gu H, Lu M, Zhang Z, Xu J, Cao W, Miao M. 2018. Metabolic process of raffinose family oligosaccharides during cold stress and recovery in cucumber leaves. Journal of Plant Physiology 224, 112-120.

Hasan M, Cheng Y, Kanwar MK, Chu X-Y, Ahammed GJ, Qi Z-Y. 2017. Responses of plant proteins to heavy metal stress—a review. Frontiers in Plant Science 8, 1492.

He M, He C-Q, Ding N-Z. 2018. Abiotic stresses: general defenses of land plants and chances for engineering multistress tolerance. Frontiers in Plant Science 9, 1771.

Jiang S-Y, Ma Z, Ramachandran S. 2010. Evolutionary history and stress regulation of the lectin superfamily in higher plants. BMC Evolutionary Biology 10(1), 79.

Jing Q, Chen A, Lv Z, Dong Z, Wang L, Meng X, Zhu X. 2023. Systematic analysis of galactinol synthase and raffinose synthase gene families in potato and their expression patterns in development and abiotic stress responses. Genes 14(7), 1344.

Kayum M, Park J-I, Nath UK, Saha G, Biswas MK, Kim H-T, Nou I-S. 2017. Genome-wide characterization and expression profiling of PDI family gene reveals function as abiotic and biotic stress tolerance in Chinese cabbage (*Brassica rapa* ssp. pekinensis). BMC Genomics 18(1), 1-20.

Kuzmin E, Taylor JS, Boone C. 2021. Retention of duplicated genes in evolution. Trends in Genetics 38(1), 59-72.

Li T, Zhang Y, Liu Y, Li X, Hao G, Han Q, Wang J. 2020. Raffinose synthase enhances drought tolerance through raffinose synthesis or galactinol hydrolysis in maize and Arabidopsis plants. Journal of Biological Chemistry 295(23), 8064-8077.

Li W-H. 2021. Adventures of a mathematician in evolutionary biology (1st ed.). Oxford University Press.

Li X, Du J, Guo J, Wang H, Ma S, Lü J, Zhang Z. 2018. The functions of cucumber sucrose phosphate synthases 4 (CsSPS4) in carbon metabolism and transport in sucrose-and stachyose-transporting plants. Journal of Plant Physiology 228, 150-157.

Liu L, Wu X, Sun W, Yu X, Demura T, Li D, Zhuge Q. 2021. Galactinol synthase confers salt-stress tolerance by regulating the synthesis of galactinol and raffinose family oligosaccharides in poplar. Industrial Crops and Products 165, 113432.

Liu Y, Zhang L, Ma J, Meng S, Pang C, Zhao X, He Y. 2022. Application of galactinol to tomato enhances tolerance to cold and heat stresses. Horticulture, Environment, and Biotechnology, 1-13.

Liu Z, Zhang M, Kong L, Lv Y, Zou M, Lu G, Yu X. 2014. Genome-wide identification, phylogeny, duplication, and expression analyses of two-component system genes in Chinese cabbage (*Brassica rapa* ssp. pekinensis). DNA Research 21(4), 379-396.

Loehlin DW, Carroll SB. 2016. Expression of tandem gene duplicates is often greater than twofold. Proceedings of the National Academy of Sciences 113(21), 5988-5992.

Lu Y, Zhao P, Zhang A, Wang J, Ha M. 2022. Genome-wide analysis of HSP70s in hexaploid wheat: tandem duplication, heat response, and regulation. Cells 11(5), 818.

Mathiazhagan M, Chidambara B, Hunashikatti LR, Ravishankar KV. 2021. Genomic approaches for improvement of tropical fruits: fruit quality, shelf life and nutrient content. Genes 12(12), 1881.

Mayer JA, Wone BW, Alexander DC, Guo L, Ryals JA, Cushman JC. 2021. Metabolic profiling of epidermal and mesophyll tissues under water-deficit stress in *Opuntia ficus-indica* reveals stress-adaptive metabolic responses. Functional Plant Biology 48(7), 717-731.

Nishizawa A, Yabuta Y, Shigeoka S. 2008. Galactinol and raffinose constitute a novel function to protect plants from oxidative damage. Plant Physiology 147(3), 1251-1263.

Noronha H, Silva A, Silva T, Frusciante S, Diretto G, Gerós H. 2022. VviRafS5 is a raffinose synthase involved in cold acclimation in grapevine woody tissues. Frontiers in Plant Science 12, 3453.

Ozyigit II, Filiz E, Vatansever R, Kurtoglu KY, Koc I, Öztürk MX, Anjum NA. 2016. Identification and comparative analysis of H2O2-scavenging enzymes (ascorbate peroxidase and glutathione peroxidase) in selected plants employing bioinformatics approaches. Frontiers in Plant Science 7, 301.

Palumbo MC, Zenoni S, Fasoli M, Massonnet M, Farina L, Castiglione F, Paci P. 2014. Integrated network analysis identifies fight-club nodes as a class of hubs encompassing key putative switch genes that induce major transcriptome reprogramming during grapevine development. The Plant Cell 26(12), 4617-4635.

Pillet J, Egert A, Pieri P, Lecourieux F, Kappel C, Charon J, Lecourieux D. 2012. *VvGOLS1* and *VvHsfA2* are involved in the heat stress responses in grapevine berries. Plant and Cell Physiology 53(10), 1776-1792.

Qin Z, Wang Y, Wang Q, Li A, Hou F, Zhang L. 2015. Evolution analysis of simple sequence repeats in plant genome. PLoS One 10(12), e0144108.

Razi M, Darvishzadeh R, Alipour H. 2023. Association analysis and estimation of breeding values of table grapevine germplasm by integrating pomological and molecular data. International Journal of Horticultural Science and Technology 10(4), 491-512.

Rejšková A, Patková L, Stodůlkov E, Lipavská H. 2007. The effect of abiotic stresses on carbohydrate status of olive shoots (Olea europaea L.) under in vitro conditions. Journal of Plant Physiology 164(2), 174-184.

Rezaei Qusheh Bolagh F, Solouki A, Tohidfar M, Zare Mehrjerdi M, Izadi-Darbandi A, Vahdati K. 2021. *Agrobacterium*-mediated transformation of Persian walnut using BADH gene for salt and drought tolerance. The Journal of Horticultural Science and Biotechnology 96(2), 162-171.

Saeed A, Sharov V, White J, Li J, Liang W, Bhagabati N, Thiagarajan M. 2003. TM4: a free, open-source system for microarray data management and analysis. Biotechniques 34(2), 374-378.

Saito M, Yoshida M. 2011. Expression analysis of the gene family associated with raffinose accumulation in rice seedlings under cold stress. Journal of Plant Physiology 168(18), 2268-2271.

Salvi P, Kamble NU, Majee M. 2020. Ectopic over-expression of ABA-responsive chickpea galactinol synthase (*CaGolS*) gene results in improved tolerance to dehydration stress by modulating ROS scavenging. Environmental and Experimental Botany 171, 103957.

Šečić E, Kogel K-H, Ladera-Carmona MJ. 2021. Biotic stress-associated microRNA families in plants. Journal of Plant Physiology 263, 153451.

Sengupta S, Mukherjee S, Basak P, Majumder AL. 2015. Significance of galactinol and raffinose family oligosaccharide synthesis in plants. Frontiers in Plant Science 6, 656.

Shen Y, Jia B, Wang J, Cai X, Hu B, Wang Y, Sun X. 2020. Functional analysis of *Arabidopsis thaliana* galactinol synthase AtGolS2 in response to abiotic stress. Molecular Plant Breeding 14, 1-11

Song X, Li Y, Cao X, Qi Y. 2019. MicroRNAs and their regulatory roles in plant–environment interactions. Annual Review of Plant Biology 70, 489-525.

Thompson JD, Higgins DG, Gibson TJ. 1994. CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. Nucleic Acids Research 22(22), 4673-4680.

Vatansever R, Filiz E, Eroglu S. 2017. Genome-wide exploration of metal tolerance protein (MTP) genes in common wheat (*Triticum aestivum*): insights into metal homeostasis and biofortification. Biometals 30(2), 217-235.

Vijayakumar H, Thamilarasan SK, Shanmugam A, Natarajan S, Jung H-J, Park J-I, Nou S. 2016. Glutathione transferases superfamily: cold-inducible expression of distinct GST genes in *Brassica oleracea*. International Journal of Molecular Sciences 17(8), 1211.

Wang B, Wang J, Wang C, Shen W, Jia H, Zhu X, Li X. 2016. Study on expression modes and cleavage role of miR156b/c/d and its target gene Vv-SPL9 during the whole growth stage of grapevine. Journal of Heredity 107(7), 626-634.

Wang G, Lovato A, Polverari A, Wang M, Liang Y-H, Ma Y-C, Cheng Z-M. 2014. Genome-wide identification and analysis of mitogen activated protein kinase kinase kinase gene family in grapevine (*Vitis vinifera*). BMC Plant Biology 14(1), 1-19.

Wang X, Shi X, Chen S, Ma C, Xu S. 2018. Evolutionary origin, gradual accumulation and functional divergence

of heat shock factor gene family with plant evolution. Frontiers in Plant Science 9, 71.

Wang Z, Qiao Y, Zhang J, Shi W, Zhang J. 2017. Genome wide identification of microRNAs involved in fatty acid and lipid metabolism of *Brassica napus* by small RNA and degradome sequencing. Gene 619, 61-70.

Wu M, Li Q, Hu Z, Li X, Chen S. 2017. The complete *Amomum kravanh* chloroplast genome sequence and phylogenetic analysis of the commelinids. Molecules 22(11), 1875.

Yang Y-W, Lai K-N, Tai P-Y, Li W-H. 1999. Rates of nucleotide substitution in angiosperm mitochondrial DNA sequences and dates of divergence between Brassica and other angiosperm lineages. Journal of Molecular Evolution 48(5), 597-604.

You FM, Huo N, Gu YQ, Luo M-c, Ma Y, Hane D, Anderson OD. 2008. BatchPrimer3: a high throughput web application for PCR and sequencing primer design. BMC Bioinformatics 9(1), 1-13.

You J, Wang Y, Zhang Y, Dossa K, Li D, Zhou R, Zhang X. 2018. Genome-wide identification and expression analyses of genes involved in raffinose accumulation in sesame. Scientific Reports 8(1), 1-11.

Zemlyanskaya EV, Dolgikh VA, Levitsky VG, Mironova V. 2021. Transcriptional regulation in plants: using omics data to crack the cis-regulatory code. Current Opinion in Plant Biology 63, 102058.

Zhang W, Ruan J, Ho T-h D, You Y, Yu T, Quatrano RS. 2005. Cis-regulatory element based targeted gene finding: genome-wide identification of abscisic acid- and abiotic stress-responsive genes in *Arabidopsis thaliana*. Bioinformatics 21(14), 3074-3081.

Zhou J, Yang Y, Yu J, Wang L, Yu X, Ohtani M, Zhuge Q. 2014. Responses of *Populus trichocarpa* galactinol synthase genes to abiotic stresses. Journal of Plant Research 127(2), 347-358.

Zhou M-L, Zhang Q, Zhou M, Sun Z-M, Zhu X-M, Shao J-R, Wu Y-M. 2012. Genome-wide identification of genes involved in raffinose metabolism in maize. Glycobiology 22(12), 1775-1785.

Zhou T, Zhang R, Guo S. 2012. Molecular cloning and characterization of GhGolS1, a novel gene encoding galactinol synthase from cotton (*Gossypium hirsutum*). Plant Molecular Biology Reporter 30(3), 699-709.

Zhou W, Chen F, Luo X, Dai Y, Yang Y, Zheng C, Shu K. 2020. A matter of life and death: molecular, physiological, and environmental regulation of seed longevity. Plant Cell and Environment 43(2), 293-302.