Brazilian Journal of Biology®

ISSN 1519-6984 (Print)
ISSN 1678-4375 (Online)

Original Article

Analysis of Germin-like protein genes family in *Vitis vinifera* (*VvGLPs*) using various *in silico* approaches

Análise da família de genes de proteínas semelhantes a germin em *Vitis vinifera* (VvGLPs) usando várias abordagens *in silico*

M. Ilyas[a]*, A. Rahman[a], N. H. Khan[b], M. Haroon[c], H. Hussain[d], L. Rehman[d], M. Alam[c], A. Rauf[e], D. S. Waggas[f] and S. Bawazeer[g]

[a]University of Swabi, Department of Botany, Swabi, Khyber Pakhtunkhwa, Pakistan
[b]University of Swabi, Department of Environmental Sciences, Swabi Khyber Pakhtunkhwa, Pakistan
[c]University of Swabi, Department of Agriculture, Swabi, Khyber Pakhtunkhwa, Pakistan
[d]University of Swabi, Department of Biotechnology, Swabi, Khyber Pakhtunkhwa, Pakistan
[e]University of Swabi, Department of Chemistry, Swabi, Khyber Pakhtunkhwa, Pakistan
[f]Fakeeh College of Medical Sciences, Department of Pharmacology, Jeddah, Saudi Arabia
[g]Umm Al-Qura University, Faculty of Pharmacy, Department of Pharmacology, Mecca, Kingdom of Saudi Arabia

Abstract

Germin-like proteins (GLPs) play an important role against various stresses. *Vitis vinifera* L. genome contains 7 GLPs; many of them are functionally unexplored. However, the computational analysis may provide important new insight into their function. Currently, physicochemical properties, subcellular localization, domain architectures, 3D structures, N-glycosylation & phosphorylation sites, and phylogeney of the VvGLPs were investigated using the latest computational tools. Their functions were predicted using the Search tool for the retrieval of interacting genes/proteins (STRING) and Blast2Go servers. Most of the VvGLPs were extracellular (43%) in nature but also showed periplasmic (29%), plasma membrane (14%), and mitochondrial- or chloroplast-specific (14%) expression. The functional analysis predicted unique enzymatic activities for these proteins including terpene synthase, isoprenoid synthase, lipoxygenase, phosphate permease, receptor kinase, and hydrolases generally mediated by Mn^+ cation. VvGLPs showed similarity in the overall structure, shape, and position of the cupin domain. Functionally, *VvGLPs* control and regulate the production of secondary metabolites to cope with various stresses. Phylogenetically VvGLP1, -3, -4, -5, and VvGLP7 showed greater similarity due to duplication while VvGLP2 and VvGLP6 revealed a distant relationship. Promoter analysis revealed the presence of diverse *cis*-regulatory elements among which CAAT box, MYB, MYC, unnamed-4 were common to all of them. The analysis will help to utilize *VvGLPs* and their promoters in future food programs by developing resistant cultivars against various biotic (*Erysiphe necator* and in Powdery Mildew etc.) and abiotic (Salt, drought, heat, dehydration, etc.) stresses.

Keywords: VvGLPs, *In silico*, genes, function, protein.

Resumo

As proteínas do tipo germin (GLPs) desempenham um papel importante contra vários estresses. O genoma de *Vitis vinifera* L. contém 7 GLPs; muitos deles são funcionalmente inexplorados. No entanto, a análise computacional pode fornecer informações importantes sobre sua função. Atualmente, as propriedades físico-químicas, localização subcelular, arquitetura de domínio, estruturas 3D, sítios de N-glicosilação e fosforilação e estudos filogenéticos dos VvGLPs foram conduzidos usando as ferramentas computacionais mais recentes. Suas funções foram previstas usando a ferramenta Search para recuperação de genes/proteínas em interação (STRING) e servidores Blast2Go. A maioria dos VvGLPs são extracelulares (43%) na natureza, mas também mostraram expressão periplasmática (29%), na membrana plasmática (14%) e específica para mitocôndrias ou cloroplastos (14%). A análise funcional previu atividades enzimáticas únicas para essas proteínas, incluindo terpeno sintase, isoprenoide sintase, lipoxigenase, fosfato permease, receptor quinase e hidrolases geralmente mediadas por cátion Mn +. VvGLPs mostraram similaridade na estrutura geral, forma e posição do domínio cupin. Funcionalmente, os VvGLPs controlam e regulam a produção de metabólitos secundários para lidar com vários estresses. Filogeneticamente, VvGLP1, -3, -4, -5 e VvGLP7 mostraram maior similaridade devido à duplicação, enquanto VvGLP2 e VvGLP6 revelaram uma relação distante. A análise do promotor revelou a presença de diversos elementos *cis*-reguladores, entre os quais CAAT box, MYB, MYC, sem nome-4, sendo comum a todos eles. A análise ajudará a utilizar VvGLPs e seus promotores em programas alimentares futuros, desenvolvendo cultivares resistentes contra vários estresses bióticos (*Erysiphe necator* e no oídio, etc.) e abióticos (sal, seca, calor, estresse hídrico, etc.).

Palavras-chave: VvGLPs, *In silico*, genes, função, proteína.

*e-mail: mohammadilyas1000@gmail.com; Dr.ilyas@uoswabi.edu.pk
Received: September 27, 2021 – Accepted: December 28, 2021

1. Introduction

Germins (GERs) and Germin-like proteins (GLPs) formed a ubiquitous family of plant glycoproteins known as the "cupin superfamily". They are characterized by a jelly roll β–barrel structure playing important role in various developmental and stress- related processes (Ilyas et al., 2016b). First identified in wheat embryo; GER was recognized as a homohexamer glycoprotein having oxalate oxidase (OXO) activity. Proteins with an average resemblance of 50% with it were called Germin-like proteins or GLPs (Dunwell et al., 2004). They possess diverse enzymatic activities which include oxalate oxidase (OXO), superoxide dismutase (SOD), ADP glucose pyrophosphatase/phosphodiesterase, polyphenol oxidase (PPO), and cysteine peptidase (Ilyas et al., 2020) due to which they have been widely used in providing potential resistance against diverse environmental stresses (Ilyas et al., 2016b). Recently, a number of GLPs were found effective in desiccation (*Craterostigma plantagineum CpGLP1*) (Giarola et al., 2020), thermotolerance (*Solanum tuberosum* L. *StGLP*) (Gangadhar et al., 2021), UV-B radiation 's acclimation (*OsGLP1*) (He et al., 2021), *Blumeria graminis* f. sp. *tritici* (*Bgt*) (258 wheat *TaGLPs*) (Yuan et al., 2021), growth, and development (38 *Cucumis sativus* GLPs) (Liao et al., 2021) and salt & drought stresses (8 rice *OsGLPs*), (Anum et al., 2022). However, further study is necessary to explore their novel enzymatic activities in diverse plant species by analyzing their properties at a molecular and cellular level.

Modern computational tools such as Ensemble Plant (Bolser et al., 2017), BioSeq-Analysis (Kearse et al., 2012), Pse-in-One (Liu et al., 2015), Pse-Analysis (Liu et al., 2017), Libd3C (Lin et al., 2014), MRMD (Hollebeek et al., 1999; Nagy et al., 2014; Tong et al., 2012), RiceXpro (Sato et al., 2010) and PlantPAN3.0 (Chow et al., 2019) etc. provide an opportunity to explore various structural and functional properties of a DNA or peptide sequence. Such studies help to predict the structure, function, and enzymatic properties of a gene or promoter with great certainty. Functional genomics techniques coupled with modern computational tools for functional analysis of genes is a widely used practice in plant molecular biology. These studies are not only economical but may also provide significant new insight into the function and regulation of genes which may then be used in various transgenic approaches to cope with various stresses.

So far, numerous (almost 350) Germins (GERs) and Germin-like proteins (GLPs) genes have been recognized in diverse plant genomes (Ilyas et al., 2016b). Many of them and their promoters were subjected to various computational studies to predict their function and regulatory mechanism. Previously, several GLPs from *Oryza sativa* (12 genes) (Davidson et al., 2010), *Glycine max* (21 genes) (Lu et al., 2010), and *Eucalyptus Grandis* (*EgGLP*) (Sassaki et al., 2015), were first characterized using various computational methods by studying their domain architectures, enzymatic activities, and phylogenetic relationships, etc. and then functionally confirmed through different experimental approaches. The same methodology was adopted for 43 *OsGLPs* genes (Ilyas et al., 2020) and their promoters (Das et al., 2019; Ilyas et al., 2016a) which provided important new insights into their structure and function against diverse stresses. Recently, 258 GLP genes were identified in wheat genome through various *in silico* approaches and their role was confirmed against *Blumeria graminis* f. sp. *tritici* (*Bgt*) (Yuan et al., 2021).

Vitis vinifera contains seven GLP genes among which function of the *VvGLP3* was previously predicted against Powdery mildew using various computational tools (Ahmad et al., 2019). Similarly, in a separate study, they were also proved effective against *Erysiphe necator* and in Powdery Mildew (Godfrey et al., 2007). So based on these studies we deduced that these genes may carry additional potential abilities against various stresses which need to be explored. So due to the importance of *VvGLPs* in various stress-related processes, the present approach was adopted to analyze and investigate various structural and functional properties of the 7 *VvGLPs* genes and their promoters using various computational tools. The analysis will help to utilize them in various crop improvement programs.

2. Methods

2.1. Database search & sequence retrieval

The nucleotides sequence of the 7 *VvGLPs* genes were downloaded from the Plant Ensemble database (Release 52 - Dec 2021 © EMBL-EBI) (Bolser et al., 2017) and translated into protein sequence through the EMBOSS Transeq website accessed via EMBL (European Molecular Biology Laboratory) server. The protein domain of the selected genes was authenticated using the NCBI conserved domain search (CD Search) tool (Lu et al., 2020).

2.2. Multiple sequence Alignment & Motives analysis

Multiple sequence alignment of the selected protein sequences was carried out using the Clustal Omega program (Sievers and Higgins, 2018) to uncover their common features. The aligned sequences were searched for GER motives (GER box), peptide signal, and KGD motives, etc. However, to get further insights into the motives structure and organization; the sequences were scanned with the "Multiple Em for Motif Elicitation" (MEME) software (Version 5.4.1) for 5 possible motives (Bailey et al., 2015). The function of each motif was predicted with Motif scan and Motif search softwares.

2.3. Protein sequence analysis

Different Physicochemical properties like Molecular weight (M.wt), Isoelectric point (IP), total Positive (+R) and Negative (-R) residues, Aliphatic index (AI), Extinction coefficient (EC), and GRAVY (Grand average of hydropathicity) of the selected sequences were searched using the "ExPASY-ProtParam" server (operated by SIB or Swiss Institute of Bioinformatics) (Duvaud et al., 2021). Subcellular localization study was carried out using CELLO (sub CELlular LOcalization predictor Ver.2.5) (Yu et al., 2006) and WoLP PSORT (Horton et al., 2007). N-glycosylation- & Phosphorylation-sites (N-sites & P-sites) were examined

with NetNglyc - 1.0 (Gupta and Brunak, 2002) and NetPhos – 3.1 (Blom et al., 2004) servers respectively. Similarly, the Peptide signals were predicted using the "SignalP 4.1" server (Armenteros et al., 2019). Swiss Modeling server (Waterhouse et al., 2018) was used to predict structural models (3D) for each peptide. The quality of each model was confirmed with the Rampage (RPA) server (DasGupta et al., 2015). Similarly, MEGAX (Molecular and evolutionary genetics analysis tool Ver.10) was used to investigate their phylogenetic relationship by employing the neighbor-joining tree-making method.

2.4. Functional analysis

Various possible functional roles for each gene were deduced through the Tool for the Retrieval of Interacting Genes/Proteins (STRING Ver. 11) (Szklarczyk et al., 2019) server which considers its homology and co-expression pattern with the related previously reported genes. In a similar approach, the genes were also scanned with Blast2GO (Ver. 4.1) server to predict their possible functions (Conesa and Götz, 2008).

2.5. Promoter analysis

The upstream promoter region of 1000 base pairs (1 kb) was downloaded for each *VvGLP* gene using the plant Ensemble server so that more insight into their regulatory mechanism can be obtained. The promoter sequences were scanned for putative regulatory elements with the Plant Pan (Ver. 3) software (Chow et al., 2019). Abundant, common, and unique elements were identified. Similarly, the detail of each element in each promoter along with its number of copies, sequence, position, and proposed function was identified and the data were presented in graphical as well as in tabulated form.

3. Result and Discussion

3.1. Genes sequence retrieval

The nucleotide sequences of the 7 monocupin *VvGLPs* genes were retrieved from the Ensemble database and their accession no, amino acid, & nucleotide count, and chromosomal coordinates were presented as Table S1 in supplementary data (Supplementary Material). GLPs play important role in various plant processes but their number varies greatly among different plant species for example 48, 43, 32, 26, and 258 GLPs were found in barley, rice, *Arabidopsis*, maize, and wheat, respectively. *Vitis vinifera* possess 7 GLPs and all of them were situated close to each other on chromosome 14 except *VvGLP6* which was found on chromosome 11. It may be due to duplication on this specific chromosome through the course of evolution during which multiple copies of the same gene were generated on a specific locus to cope with various stresses. Duplication is common among GLPs of many crops as previously reported in several studies (Ilyas et al., 2020; Zimmermann et al., 2006; Lu et al., 2010; Davidson et al., 2010). Recently, 258 GLPs were found in the wheat genome forming tandemly duplicated genes clusters on specific chromosomes (specifically on 4B) (Yuan et al., 2021). Genes with the highest and smallest number of nucleotides (nt) were *VvGLP3* (1,023 nt) and *VvGLP2* (217) respectively. Similarly, the highest and smallest number of amino acids were noted for *VvGLP3* (225) and *VvGLP5* (207) respectively showing very little difference which may be due to a similar exon structure that arose due to tandem duplication events as already reported for GLPs in *Arabidopsis* (Membré et al., 2000), barley (Zimmermann et al., 2006), soybean (Lu et al., 2010), rice (Davidson et al., 2010; Ilyas et al., 2020), and wheat (Yuan et al., 2021).

3.2. Multiple sequence alignment & Motives Analysis

The peptide sequences of the *VvGLPs* genes were aligned to uncover similarities and variations in the domain architecture as well as in the functional features of these genes (Figures 1 and 2). Peptide signal was located at the start of each sequence which is essential for entry into the secondary pathways of various metabolic processes in both prokaryotes and eukaryotes (von Heijne, 1990). The basic cupin domain or Germin box was represented by three Germin motives (A, B, and C) which are considered as the fundamental structural feature of the GLPs. It has a consensus sequence of GxxxxHxHPxAxEh where "x" represents any aa (amino acid) while "h" is a hydrophobic aa (Dunwell et al., 2004). The motif A existed at the beginning of the Germin box (≈23-36 aa) while motif B and C were found around ≈105-120 and ≈152-167 aa away from the amino-terminal end (N-end) of each protein respectively. The putative KGE/KGD or sometimes RGD sequence existed upstream to the GER motif C. It is one of the essential features of GLPs structure shared by almost half of the GLP sequences, but not found in GER genes (Barman and Banerjee, 2015). It is not only involved in protein-protein interaction but also acts as a receptor in the ECM (Extracellular matrix) by helping plant protein to interact with several other proteins or exchange information to perfom function (Bernier and Berna, 2001). Such mechanisms play a fundamental role in protecting plants from pests and pathogens's attack, acting either by cell wall adhesion, plasma membrane strengthening or by preventing the penetration of fungal toxin. Its presence in GLPs of different plant species such as *Medicago truncatula* (Doll et al., 2003), sunflower (*Helianthus annuus*) (Beracochea et al., 2015), *Prunus salicina* (El-Sharkawy et al., 2010), soybean (*Glycine max*) (Langenbach et al., 2016), cotton (*Gossypium hirsutum*) (Kim and Triplett, 2004; Pei et al., 2019), *Oryza sativa* (Ilyas et al., 2020) and *Craterostigma plantagineum* (Giarola et al., 2020), etc. is widely reported. All the VvGLP proteins have KGD motif except VvGLP2 where it is replaced by KGE which may be due to mutation. The presence of this motif in VvGLPs indicates their role as plant defense protein. It also shows its evolutionary importance in the structure and function of these genes. It not only suggests the functional similarity of VvGLPs but also their similar phylogenetic history. It also indicates their better interactive ability with other membrane proteins.

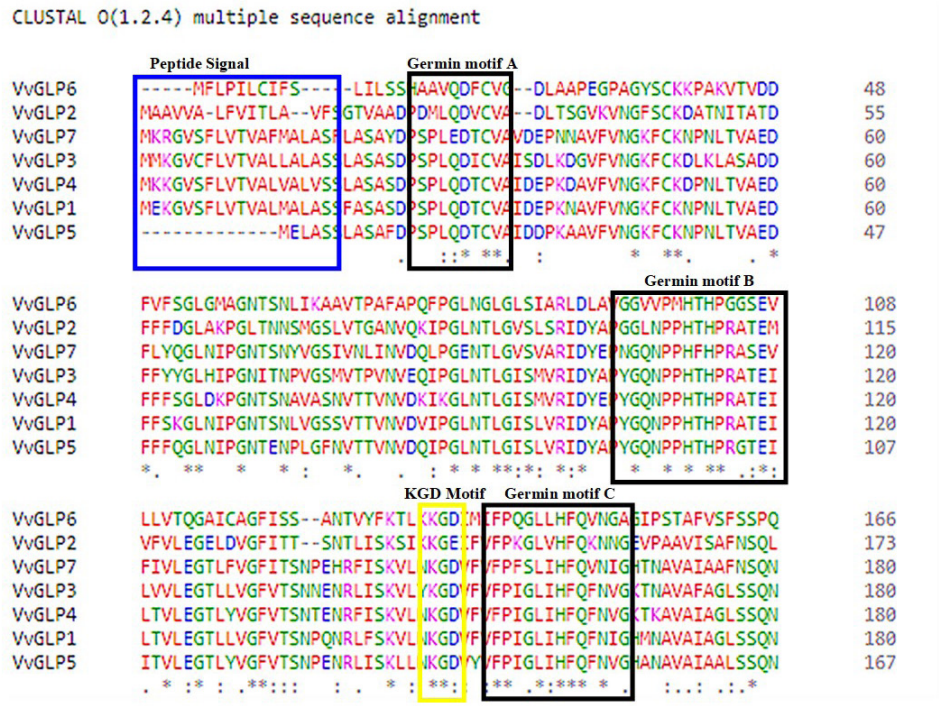
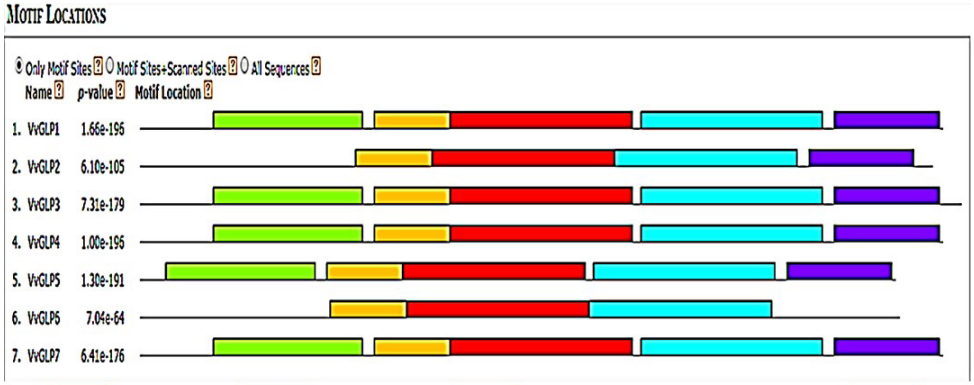


Figure 1. Multiple sequence alignment of the *Vitis vinifera* Germin-like proteins (VvGLPs) genes family. The analysis was conducted with the Clustal Omega program (Sievers and Higgins, 2018). Each Germin motif is represented with a black box. The peptide signal region and KGD/KGE motif are represented with blue and yellow boxes respectively.



S.No	Motives	Width	Motif scan	Motif search
M1	DKIKGLNTLGISMVRIDYEPYGQNPPHTHP RATEILTVLEGTLYVGFVTS	50	Cupin Domain	Cupin Domain
M2	ENRLISKVLYKGDVFVFPIGLIHFQFNVGK TNAVAFAGLSSQNPGVITIA	50	Cupin Domain	Cupin Domain
M3	ASASDPSPLQDTCVAIDEPKDAVFVNGKFC KDPNLTVAEDF	41	N-glycosylation site, Casein kinase II, phosphorylation site	S4 Domain
M4	FGSNPPINPDFLARAFQLDKRVVEYLQAR	29	NA	Cupin Domain
M5	GLNIPGNTSNLVGSSVTTVNV	21	NA	NA

Figure 2. Domain analysis of the *Vitis vinifera* Germin-Like proteins (VvGLPs) genes family. Multiple Em for Motif Elicitation (MEME) suit (Version 5.4.1) (Bailey et al., 2015) was used for the analysis to predict 05 possible motives in the VvGLPs as represented with colored blocks. The conserved motives were confirmed for their functional role via "Motif scan" as well as "Motif Search" softwares. Motives width and sequence is given in the figure. **M:** Motif, **S4:** Small RNA-binding protein domain, **N/A:** Not available.

MEME analysis revealed 5 motives in VvGLPs in which motif (M) 1, 2, and 4 represent the basic Germin motives which are the fundamental structural feature of GLPs. The length of motives ranged from 21 to 50 amino acids whereas their number in each protein was from 3 to 5. Overall, VvGLP1, 3, 4, 5, and 7 showed similarity in motives arrangement while VvGLP2 lacks motif 1 (M1) and VvGLP6 lacks M1 and M5. Various phenomena such as duplication, mutation, deletion, and insertion, etc. may be responsible for variation in the motives number and their position. Cupin domain possesses a variety of functions against different biotic (fungal, bacterial, viruses etc.) and abiotic (salt, drought, heat, light, wounding etc.) stresses (Ilyas et al., 2016b). However, besides the basic cupin domain, additional functional domains were also found among which M3 is common to all peptide sequences. It is involved in N-Glycosylation, Phosphorylation, Casein kinase II-, and S4 domain (small RNA-binding protein domain) -related activities. Protein N-glycosylation mainly operated by N-glycans perform major function in the quality control and protein's folding inside the lumen of the endoplasmic reticulum (ER). It is also important in plant development, cellulose biosynthesis, immune responses, and growth under various biotic or abiotic stresses (Nagashima et al., 2018). Similarly, phosphorylation of protein is an important and well-studied modification at a post-translational level that controls and regulates almost all the cellular processes using a variety of mechanisms such as altering protein interactions or localization, and conformations (Strumillo et al., 2019). In the same way, casein kinase II regulates circadian rhythm, hormone signaling as well as various plastids-related functions in plants (Ogiso et al., 2010). Similarly, the S4 domain is a multifunctional protein domain playing important role in transcription and translation. It is also part of an important transmembrane protein playing a significant role in various processes by regulating the movement of various molecules and ions across the cell membrane (Jegla et al., 2018). The presence of these motives in VvGLPs suggested their importance and role in these processes.

3.3. Protein sequence analysis

3.3.1. Physicochemical properties

VvGLPs exhibited a notable variations in various physicochemical properties (Table 1). Their molecular weight (M.wt) and size showed comparable variations along with other studied properties. The observed values for M.wt ranged from 21525.13 to 243226.95 Dalton (Da) noted for VvGLP3 and VvGLP6 respectively showing variation in the structure and physicochemical properties of these proteins. Similarly, the lowest and highest recorded values of isoelectric point (pI) were 4.90 (VvGLP3) and 8.53 (VvGLP2) which is important for the estimation of protein solubility, electrophoresis, and electrophoretic separation (Audain et al., 2016). In the case of Extinction coefficient (EC), the lowest value was shown by VvGLP2 (1490 M^{-1} cm^{-1}) while the highest by VvGLP3 (18450 M^{-1} cm^{-1}) at 280 nm. High EC value is largely due to a large amount of Tryptophan (Trp), phenylalanine (Phe), and tyrosine (Tyr) residues in a protein structure (Adeloye and Ajibade, 2011). The instability index (II) ranged from 17.51 (VvGLP4) to 29.21 (VvGLP6) which is an important indicator for protein stability. Protein stability in the cellular environment is crucial to its proper function (Gamage et al., 2019). All the VvGLPs are stable at the cellular level due to their low II values (below 40) which makes them suitable candidates for in vitro expression studies. The aliphatic index (AI) values ranged from 93.00 (VvGLP7) to 105.05 (VvGLP6). It indicates higher thermal stability of the protein in the cellular environment, largely depend on the presence of percentage (%) of aliphatic amino acids in each protein (Ikai, 1980). VvGLP1 (100), VvGLP3 (103), and VvGLP6 (105) showed the highest AI values which indicate their stability at a wide range of temperatures. Nevertheless, proteins that showed low AI values showed their structural flexibility which is largely contributed to the presence of aliphatic amino acids with aliphatic side chains. The highest and lowest values of the negatively charged residues were noted for VvGLP6 (12) and VvGLP4 (19) respectively. Similarly, total positive charged residues were ranged from 11 to 19

Table 1. Computational analysis based on various structural and functional properties of the *Vitis vinifera* Germin-like proteins (VvGLPs) genes family.

Gene Name	M.wt	pI	-R	+R	EC	II	AI	GRAVY	Pfam domain	SL (CELLO)	SL (PSORT)	N-sites	p-sites
VvGLP1	23456.99	6.40	15	14	3105	23.15	100.09	0.273	Cupin	Extracellular	Periplasmic	2	6
VvGLP2	22732.27	8.53	16	18	1490	21.00	95.67	0.241	Cupin	Mitochondria	Extracellular	2	7
VvGLP3	24326.95	4.90	21	14	18450	25.39	103.47	0.261	Cupin	Extracellular	Plasma membrane	1	7
VvGLP4	23607.06	6.82	19	19	6085	17.51	96.05	0.138	Cupin	Extracellular	Periplasmic	3	10
VvGLP5	22294.39	5.20	17	11	7450	18.89	97.49	0.123	Cupin	Extracellular	Periplasmic	2	6
VvGLP6	21525.13	6.38	12	11	4470	29.21	105.05	0.576	Cupin	Plasma membrane	Extracellular	1	3
VvGLP7	24078.37	5.45	18	13	7450	20.92	93.00	0.160	Cupin	Periplasmic	Chloroplast	2	9

M. wt: Molecular weight, **pI:** Isoelectric point, **+R:** Positively charged residues, **-R:** Negatively charged residues. **EC:** Extinction coefficient, **II:** Instability index, **AI:** Aliphatic index values. **GRAVY:** Grand average of hydropathcity, **SL:** Subcellular localization, **N- site:** N-Glycosylation site, **P-site:** Phosphorylation site.

for VvGLP5 and VvGLP4 respectively. Charged residues (positive or negative) affect multiple properties such as stability, interaction with other molecules, conformation, and thermal adaptation of natural proteins which is helpful in protein functional analysis and designing (Berezovsky et al., 2007). All VvGLPs showed positive GRAVY values revealing that all of them are hydrophobic in nature. The importance of hydrophobic protein in deriving various biological processes is universally accepted (Rego et al., 2021). However, the observation is in contrast to the previous study of GLPs in rice where all proteins were hydrophilic (Ilyas et al., 2020). The analysis revealed that all VvGLPs exhibit comparable variation in physicochemical properties except VvGLP6 which showed high M.wt, II, AI, and negatively charged residues which may be due to its presence on a separate chromosome which gave rise to its more distinct properties.

3.3.2. Sub-cellular localizations

At the cellular level, *VvGLPs* are mainly expressed in the extracellular (43%) (*VvGLP1*, *-3*, *-4*, *-5*, *-6*) or periplasmic (29%) (*VvGLP1, -4, -5*, *-7*) regions. However, they may also be expressed in mitochondria (*VvGLP2*), chloroplast, or plasma membrane (*VvGLP7*). It indicates the possible role of these genes in these regions which are mainly related to photosynthesis, lipid or protein metabolism, and various other biochemical processes of the cell. Such diverse expression patterns may have evolved through the course of evolution due to the specific function of these genes in these cellular parts. The finding is similar to the earlier study of *OsGLPs* where most of the genes showed either extracellular - (57%) or plasma membrane-specific (33%) expression. However, it is contradictory to the study of 258 *TaGLPs* which were mostly expressed in secretory pathways (Yuan et al., 2021). Similarly, such diverse expression pattern was also noted for GLP families of rice, soybean, and barley (Dunwell et al., 2008; Lu et al., 2010; Zimmermann et al., 2006). However, some of the genes showed organelles-specific expression which is similar to the previous results in *Arachis hypogea's* GLPs where various members of the family showed cytoplasm - (*AhGLP4*), cell wall - (*AhGLP2, AhGLP6*), both cell wall & cytoplasm - (*AhGLP3, AhGLP5*) or cytoplasm -, plasma membrane and cell wall (*AhGLP1,* and *AhGLP6*) -specific expression during transient expression studies in onion cells (Wang et al., 2013). Similarly, in another study, *AtGLP4* showed the highest expression in the Golgi complex (Yin et al., 2009). Recently, a similar expression pattern was also detected during *in silico* study of the *OsGLPs* where *OsGLP1-2* showed chloroplast-, *OsGLP3-7* endoplasmic reticulum-, and *OsGLP3-1* mitochondrial-specific expression (Ilyas et al., 2020).

3.3.3. N-glycosylation and phosphorylation sites

Variable numbers of N-sites (from 1-3) and P-sites (from 3-10) were observed in VvGLPs (Table 1). N-sites were mostly found between 18-26 amino acids away from the amino-terminus (N-terminal end) of protein. The highest number of N-sites (3 sites) were shown by VvGLP4 while the lowest (1 site) by VvGLP3 and VvGLP6. Variation in N-sites indicates structural and functional diversity. Glycosylation of amino-terminus affects conformation, stability, interaction, and various enzymatic functions of the peptide in various parts of the plant (Rayon et al., 1998). Similarly, VvGLPs also showed variability in P-sites (from 3-10) as well, the highest being found in VvGLP4 (10) and lowest in VvGLP6 (3). VvGLP2 and VvGLP3 have 7 while VvGLP1 and VvGLP5 contained 6 P-sites. Various plant processes depend on cell signaling including cell propagation, metabolism, cell differentiation, and program cell death (apoptosis) which is controlled and regulated by phosphorylation either at serine, tyrosine, or threonine residue of the protein (Blom et al., 2004). The presenceof N-glycosylation and Phosphorylation sites in VvGLPs suggested their role in several cellulars, physiological, and metabolic-related processes. Recent studies have greatly emphasized on the role of glycosylation and phosphorylation on the structure, properties, and overall importance of protein in various plant processes (Liu et al., 2021; Wu et al., 2021).

3.3.4. 3D structural analysis

Protein structure determination through alignment and blast search with known proteins structures using various online databases is one of the common strategies in plant computational biology (McGuffin et al., 2013). Swiss modeling server gave 3 models for each *VvGLP* protein (Figure S1 in supplementary data). GMQE (Global model quality estimation) and Z-score (Qmean estimation scores) were used as a criteria for the selection of the best model. Ramachandran plot analysis (RPA) showed that on average approximately above 90% of amino acids of each protein exist in the favored region, 8.90% in the allowed while 1% in the outliner regions which confirmed the best quality of each model. RPA is considered as a standard procedure for confirmation of protein models and it is vastly used in such studies (DasGupta et al., 2015). It can also be used to explore various additional characteristics of protein models (Abel et al., 2020). Overall, VvGLPs are highly conserved in their structure showing greater similarity in the overall position and shape of the cupin domains. Each protein is comprised of six Germin monomers bind by six Mn^{+2} ions forming a hexameric structure. The result is similar to the previously studied GLPs in various crops (Breen and Bellgard, 2010; Ilyas et al., 2020). Their similar structure revealed their common evolutionary history. Many of these genes (*VvGLP1, 2, 3, 4, 5,* and *7*) are located close to each other on chromosome 14 showing that they might originate via tandem duplication along the course of evolution. Such findings have been previously reported for GLPs in various plants including rice, maize, barley, and soybean (Davidson et al., 2010; Ilyas et al., 2020; Lu et al., 2010; Zimmermann et al., 2006). Even though, VvGLP6 is located on chromosome 11 but its structure is similar to other members of the family which may be due to their common origin due to retrotransposon activity on this chromosome. However, it showed variation in other properties.

3.3.5. Phylogenetic analysis

The phylogenetic study revealed greater similarity among VvGLPs with a narrow genetic background by showing a branch length scale of 0.05% (Figure 3)

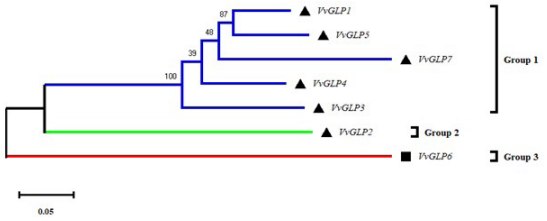


Figure 3. Phylogenetic analysis of the *Vitis vinifera* Germin-like proteins (VvGLPs) genes family. Groups with distinct colors were highlighted with parenthesis. Genes located on chromosomes 14 and 11 were labeled as triangles and rectangles respectively. The evolutionary history was inferred using the Neighbor-Joining method. The percentage of replicate trees in which the associated taxa clustered together in the bootstrap test (1000 replicates) is shown next to the branches. The tree is drawn to scale, with branch lengths in the same units as those of the evolutionary distances used to infer the phylogenetic tree. The evolutionary distances were computed using the p-distance method and are in the units of the number of amino acid differences per site. This analysis involved 7 amino acid sequences. All ambiguous positions were removed for each sequence pair (pairwise deletion option). There were a total of 225 positions in the final dataset. Evolutionary analysis was conducted with Molecular and evolutionary genetic analysis tool Ver.10 (MEGAX).

suggesting little variation in their protein sequences. It shows that these genes might have evolved very recently on an evolutionary timescale mainly through duplication. VvGLP1 and VvGLP5 showed greater similarity while the most distant relationship was observed for VvGLP2 and VvGLP6. The distant relationship of VvGLP6 can be justified by its presence on a separate chromosome (Chr 11) which may be due to retrotransposon activity through the course of evolution. However, the distant relationship of VvGLP2 may be due to a greater mutation rate which gave rise to changes in an overall domains structures. Similarly, a close relationship was observed for VvGLP1, -3, -4, -5, and -7 which may be due to multiple duplication events on chromosome 14. Comparative analysis of their domain architecture revealed similarity in their exon structure, which suggest their evolutionary relatedness which further gives rise to their close phylogenetic relationships. All the genes showed similar domain architecture (except VvGLP2 and -6) that may be due to gene tandem duplication events from an ancestral DNA sequence. Previously, similar results were also reported in rice, barley, maize, etc. and recently in wheat (Yuan et al., 2021). GLPs often exist in clusters on specific chromosomes which arose through multiple tandem duplication events. Sometimes such clustered loci offer significant resistance against diverse environmental constraints. A cluster of GLPs on chromosome 8 was found effective against *B. Graminis* in barely (Zimmermann et al., 2006). Similarly, 12 *OsGLPs* provided resistance against *Magnaporthe oryzae* and *Rhizoctonia solani* in rice (Manosalva et al., 2009). Such phenomena were also reported in soybean (Lu et al., 2010), maize (Breen and Bellgard, 2010), and peanut (Wang et al., 2013) etc. Previously *VvGLP3* was proved effective against Powdery mildew both at computational study (Ahmad et al., 2019) and experimental level (Godfrey et al., 2007). Its close relationship with VvGLP1, -4, -5, and -7 showed their similar properties which need to be examined against fungal pathogenicity.

3.4. Functional analysis

Enzymatic activities and the predicted roles of VvGLPs in diverse plant processes are presented in Table 2. STRING predicted various possible roles for VvGLP1, -2, -3, and -6 but no function was predicted for VvGLP4, -5, and -7. VvGLP1 and VvGLP3 showed similar functions which are mainly related to terpenoid biosynthesis. In plants, terpenoids play important role in numerous cellular, physiological and biochemical processes for example electron transport chain, photosynthesis, membrane architecture, and development (Pichersky and Raguso, 2018). They also showed Lipoxygenase activity which is important in lipid oxidation that is crucial for many developmentally-regulated processes as well as plays a vital role in numerous abiotic and biotic stresses (Andreou and Feussner, 2009). Similarly, VvGLP2 exhibit enzymatic activities which mediate the transport of various proteins, molecules, or ions such as phosphate, zinc, iron, peptides, etc. [Phosphate/Zinc/Iron permeases, oligopeptide transporter (OPT) protein], and hydrolysis (Calcineurin-like phosphodiesterase) (Castañeda-García et al., 2009; Shin et al., 2008). VvGLP2 also showed a co-occurrence pattern with the thaumatin (THN) domain (InterPro ID: SM00205) family which plays important role in plant pathogenesis (Ruiz-Medrano et al., 1992). Such observation is in accordance with the role of GLPs against diverse stresses including pests and pathogens protection employing a variety of mechanisms (Ilyas et al., 2016b). Similarly, VvGLP6 showed enzymatic activities (Receptor kinases) related to the protection of plants against pests and pathogens attacks (Stegmann et al., 2017). It also has a prominent role in response to various biotic and abiotic stresses [Alpha/beta hydrolase (ABH)] (Mindrebo et al., 2016), as well as in plant immunity, growth and development, self-incompatibility, and disease resistance (Receptor kinases) etc. (Guo et al., 2011). VvGLP6 also showed a co-expression pattern with tetratricopeptide repeat (TPR) domain possessing genes which are strongly involved in protein-protein interaction (Perez-Riba and Itzhaki, 2019). Engineering tolerance in various crops against diverse stresses is a highly desirable practice in recent days (Fahad et al., 2021b) and VvGLPs play important role in defense against various biotic and abiotic stresses. At the cellular level, they also help other proteins to transport various biomolecules inside the plant body.

Previously, expression analysis via reverse transcriptase-polymerase chain reaction (RT-PCR) showed that VvGLPs provide significant resistance against *Erysiphe necator, Plasmopara viticola,* and *Botrytis cinerea*. The highest induction was noted for VvGLP3 against *E. necator* by showing infection-site specific expression. At the cellular level, the highest expression was noted in the the cell wall when analyzed through transient expression studies in onion cells using the VvGLP3:GFP fusion construct. Protein isolated from transformed *Arabidopsis thaliana* showed strong SOD activity (Godfrey et al., 2007). Similarly, the role

Table 2. *In silico* functional analysis of the *Vitis vinifera* Germin-like proteins (VvGLPs) genes. The analysis was conducted using Search Tool for the Retrieval of Interacting Genes/Proteins (STRING) (Szklarczyk et al., 2019) and Blast2go Servers (Conesa and Götz, 2008).

Gene Name	STRING Analysis	Blast2Go
VvGLP1	Terpene synthase/Terpene cyclase-like 1/Mn^{+}/ Lipoxygenase/Cytochrome P450/Monoterpenoid biosynthesis/IPT Domain/C & N Terminal domain/ Isoprenoid synthase	Mn^{+} ion binding activity, Noncytoplasmic
VvGLP2	Phosphate permease, and Zinc/iron permease fungal/plant/ OPT oligopeptide transporter protein, and Natural resistance-associated macrophage protein/Achaete-scute transcription factor-related, and Haemerythrine-like/ Calcineurin-like phosphoesterase/ Mn^{+}/ Major facilitator superfamily (MFS)/	Mn^{+} ion binding activity, Noncytoplasmic
VvGLP3	Terpene synthase/Terpene cyclase-like 1/Mn^{+}/ Lipoxygenase/Monoterpenoid biosynthesis/ Isoprenoid synthase/IPT Domain/Cytochrome P450/N & C Terminal domain	Mn^{+} ion binding activity, Noncytoplasmic
VvGLP4	NIF	Mn^{+} ion binding activity, Noncytoplasmic
VvGLP5	NIF	Mn^{+} ion binding activity, Noncytoplasmic
VvGLP6	Tetratricopeptide repeat (TPR) domain/ S-locus, Receptor kinase, and Alpha/beta hydrolase fold-1/ C-terminal	Mn^{+} ion binding activity, Noncytoplasmic
VvGLP7	NIF	Mn^{+} ion binding activity, Noncytoplasmic

NIF: No interaction found, **Mn^{+}:** Manganese ion; **IPt:** Ig-like, plexins, transcription factors, **N & C terminal:** Amino and Carboxyl terminal.

of the VvGLP3 was also predicted against Powdery mildew disease via *in silico* analysis (Ahmad et al., 2019). Though the current study reveals some novel enzymatic activities for these proteins but it needs further experimental confirmation. However, Blast2Go results revealed their metal ion binding nature and non-cytoplasmic activities which is similar to the earlier findings related to Germin-like protein in rice (Ilyas et al., 2020). Metal ion binding activity is mainly related to the antimicrobial nature of protein with the main role in plant defense against pests and pathogens (Dunwell et al., 2008). In a recent study, GLPs located on the fourth homologous group chromosomes in wheat were found effective against *Blumeria graminis* f. sp. *tritici* (*Bgt*) invasion (Yuan et al., 2021). Similar findings were also reported for some of the GLPs in cucumber against Downy mildew (DM) disease caused by *Pseudoperonospora cubensis* infection (Liao et al., 2021). The analysis revealed that VvGLP1, -2, -3, and -6 are functionally more important due to their diverse enzymatic activities which may seem suitable candidates for further studies.

3.5. Promoter analysis

Transcriptional factors binding sites (TFBSs) analysis of the *VvGLPs* promoters revealed a total of 393 diverse regulatory elements distributed throughout the promoter regions on both sense and anti-sense strands (Figure 4). Many of these elements were involved in the regulation of light response, anaerobic induction, hormonal (Auxin, gibberellin, methyl jasmonate and salicylic acid) stresses, cell cycle, low temperature response, seed-specific activities, and circadian rhythm, etc. Detailed information related to the sequence, position and function of these *cis*-regulatory elements in each promoter is presented in Table 3 as well as in the supplementary data file (Table VvGLP1 to VvGLP7). The highest number of elements were found in *VvGLP6* (68 elements), lowest in *VvGLP1* (44 elements), while *VvGLP2, -3, -4, -5,* and *-7* contained 55, 64, 57, 53, and 52 TFBSs respectively. Previously, similar *in silico* approaches were adopted for the promoter study of 05 different plant species (Mahmood et al., 2010), *EgGLP* (Sassaki et al., 2015), and *OsGLPs* (Das et al., 2019; Ilyas et al., 2016a), etc. to get insight into their function and regulatory mechanism. Recently, *in silico* promoter analysis of 3 pathogenesis-related (PR1, 2, 3) and *OsRGLP1* genes promoters revealed several common *cis*-elements (Ohwofasa et al., 2021). The highest number of *cis*-elements in *VvGLP6* represents its importance in various processes.

3.5.1. Abundant elements

CAAT box was the most abundant element in every promoter showing the importance of this element in *VvGLPs* regulation. Other abundant elements included the TATA box, STRE, unnamed-4, and Box4. It is similar to the previous study of *OsRGLP2* promoter of different rice accession where CAAT and TATA box elements were found abundant (Mahmood et al., 2018), but comparable to the study of 43 *OsGLPs* promoters where *Arabidopsis* homeobox protein or AHBP and TATA box were the most abundant elements (Ilyas et al., 2016a). Based on the TATA box; VvGLPs promoters can be divided into TATA-containing (*VvGLP1, -4,* and *-6*) and TATA-less (*VvGLP2, -3, -5,* and *-7*) sequences. The previous group is important in

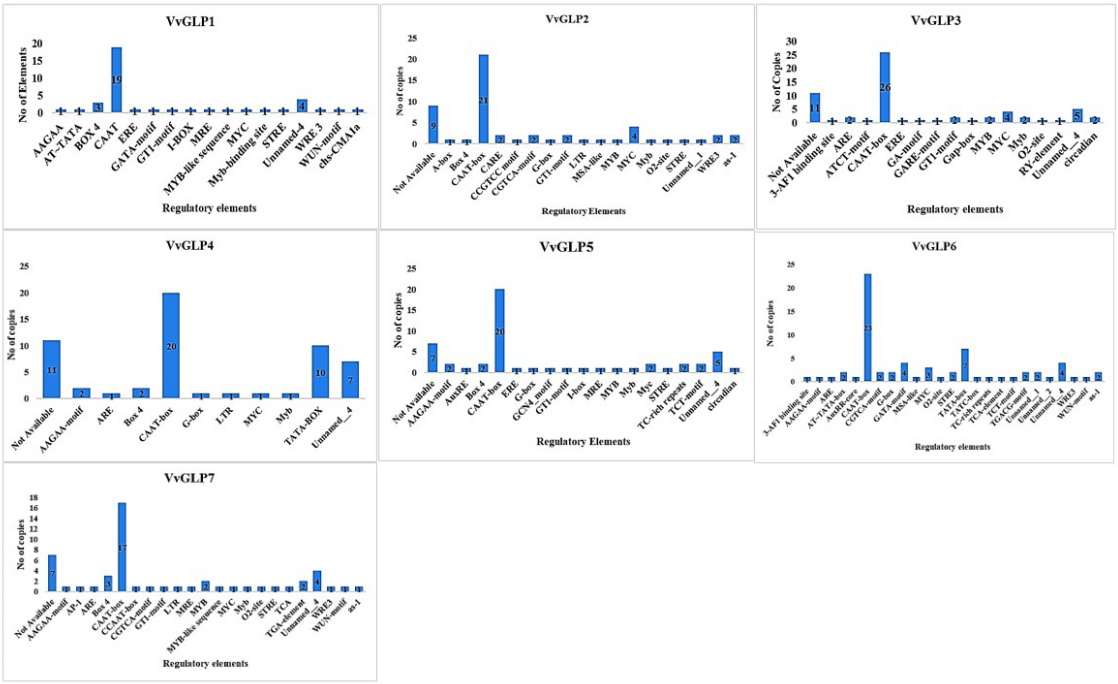


Figure 4. Regulatory element analysis of the *Vitis vinifera* Germin-like proteins (*VvGLPs*) genes promoters. Detail of the elements in each promoter is given in each graph labelled with respective promoter name. The analysis was conducted with the PlantPAN (Ver.3) server (Chow et al., 2019). The number of each element is given on each bar. Graphs were made using Excel 2016.

stress-related responses while the latter is important in cell growth or housekeeping genes functions (Bae et al., 2015). The highest number of these elements were found in *VvGLP4* and *-6* promoters which indicate their importance. Stress response elements (STRE) (consensus sequence of AGGGG) were found in *VvGLP2, -5, -6,* and *-7* which was important in heat shock induction of the *AtHsp90-1* gene in transgenic *Arabidopsis* (Haralampidis et al., 2002) as well as in oxidative, heat, and osmotic stresses due to its ability to bind with Msn2 and Msn4 (multicopy suppressor of SNF1 mutation proteins 2 and 4) proteins (Gao et al., 2013). It shows the importance of these promoters in the development of various agricultural strategies to increase crop production under heat and drought stresses (Fahad et al., 2017). Similarly, Box4 play important role in light response regulation (Kaur et al., 2017). It was found in all promoters except *VvGLP3* and *-6,* the highest being found in *VvGLP7* (3 copies) showing the importance of these genes in these processes.

3.5.2. Common cis-regulatory elements

Common *cis*-elements included CAAT box, MYB, MYC, and unnamed-4 showing their importance in various biological processes related to the overall function of these genes. Previously, AHBP, vertebrate TATA-box-binding protein (VTBP) and MYB-like proteins (MYBL) were also found common in *OsGLPs* promoters (Ilyas et al., 2016a). Among these *OsRGLP1 & -2* were highly induced by salt, drought, ABA, wounding and pathogenic [*Fusarium solani* (Mart.) Sacc. and *Alternaria solani* Sorauer] stresses (Ilyas et al., 2019; Munir et al., 2016). However, it is contradictory to the study of Mahmood et al., (2010) where ARA, W-box, GT-element and ACGT Sequence were common to 10 GLPs promoters of 5 plant species. CAAT box is part of the core promoter element with a consensus sequence of GGCCAATCT located upstream to the start codon of most eukaryotic genes and enhancer region playing important role in transcription (Etminan et al., 2018). Assembly of the CAAT box-binding factor is regulated by cytokinin, light, and stages of the plastids in spinach photosynthestic gene (*AtpC*) promoter (Kusnetsov et al., 1999). CAAT box is also important against various abiotic stress responses such as chilling stress etc. (Zhang et al., 2016). Such observation are important in crop improvement against various abiotic stresses (Fahad et al., 2021a). In another study, the CAAT box controls the expression level of *GW6* (grain width 6), weight, and width of the grain in rice (Shi et al., 2020). Similarly, the CAAT box and TATA box may also be important against drought and salt stresses as previously reported in banana MaTIP1;2 promoter in transgenic *Arabidopsis* (Song et al., 2018). The highest number of CAAT box elements were found in *VvGLP3* (26), *-6* (23), and *-2* (21 copies) showing the importance of these promoters. Similarly, MYB transcriptional factors belong to one of the diverse gene family playing important role in development, plant growth, cell morphology, cellular and physiological processes, metabolism, primary and secondary metabolic reactions as well as secondary metabolite biosynthesis (Cao et al., 2020), phenylpropanoid biosynthesis (Ma and Constabel, 2019), drought stress (Zhang et al., 2019),

Table 3. Transcriptional factor binding sites (TFBSs) analysis of the *Vitis vinifera* Germin-like protein genes (*VvGLPs*) promoter. The *cis*-regulatory elements were found using PlantPAN (Version 3) software (Chow et al., 2019). The elements may exist on both strands (sense or antisense) of the promoters.

Element name	Sequence	Function & Description
3-AF1 binding site	AAAGAGAGGAA	Light-responsive element
AAGAA-motif	AAGAAAG	Not available
ARE	AAACCA	A *cis*-acting regulatory element essential for the anaerobic induction
AT~TATA-box	TATATA	Transcription regulation
AuxRR-core	GGTCCAT	*Cis*-acting regulatory element involved in auxin responsiveness
CAAT-box	CAAAT	Common *cis*-acting element in promoter and enhancer regions
CGTCA-motif	ACTGC	*Cis*-acting regulatory element involved in the MeJA-responsiveness
G-box	AGCAC	*Cis*-acting regulatory element involved in light responsiveness
GATA-motif	AAGGATAA	Part of a light-responsive element
MSA-like	ATTGGCAATCGTAAAACTAC	*Cis*-acting element involved in cell cycle regulation
MYC	CATTTG	Jasmonic acid signaling, abiotic & biotic stress responsiveness
O2-site	GTTGTCGTGA	*Cis*-acting regulatory element involved in Zn metabolism regulation
STRE	GAGGGG	Heat shock, oxidative, and osmotic response element
TATA-box	TATAGA	Core promoter element around -30 of the transcription start site
TATC-box	TATCCCA	*Cis*-acting element involved in Gibberellin-responsiveness
TC-rich repeats	CAATCACTTA	*Cis*-acting element involved in defense and stress responsiveness
TCA-element	GTAGAAGACT	*Cis*-acting element involved in Salicylic acid responsiveness
TCT-motif	CATTCT	Part of a light-responsive element
Unnamed__1	GGTGC	Not available
Unnamed__2	GGCCCC	Not available
Unnamed__4	GTGG	Promoter development and gene regulation
WRE3	TCCACC	Not available
WUN-motif	TAATTAC	Wound response element
as-1	TGACG	Not available
Box 4	ATTACA	Part of a conserved DNA module involved in light responsiveness
chs-CMA1a	TTACTTAA	Part of a light-responsive element
MRE	AATCCAA	MYB binding site involved in light responsiveness
ERE	ATTTCATA	Drought, submergence, fruit ripening & secondary metabolite production
MYC	GTTTAC	Biotic and abiotic stresses
I-BOX	CCATCTT	Part of a light-responsive element
GT1-motif	AATTGG	Light responsive element, light responsive element
LTR	CCGAAA	*Cis*-acting element involved in low-temperature responsiveness
MYB	TAACCA	Propanoid biosynthesis, drought and salt stress, etc.
CARE	CACTCAAC	Not available
CCGTCC motif	CCGTCC	Not available
GA-motif	ATAGATAA	Part of a light-responsive element
GARE-motif	GTTGTCT	Gibberellin-responsive element
Gap-box	CAAATGAAAAACTA	Part of a light responsive element
RY-element	GTACGTAC	*Cis*-acting regulatory element involved in seed-specific regulation
circadian	CATAGATATC	*Cis*-acting regulatory element involved in circadian control
AP-1	TGAGTTAG	Not available
CCAAT-box	CAACG	MYBHv1 binding site
TGA-element	AACGAC	Auxin-responsive element

flavonoid biosynthesis, and salt stress (Li et al., 2019a), etc. Previously, MYB was also found frequent in *OsGLPs* promoters (Das et al., 2019) and it is involved in the transactivation of *OsRGLP2* gene expression (Deeba et al., 2017). Similarly, MYC *cis*-elements regulate diverse biological, cellular, physiological processes, flavonoid biosynthesis, secondary metabolite biosynthesis thereby providing resistance against a wide range of biotic and abiotic (drought, salinity, cold, etc.) stresses (Xie et al., 2020). The highest number of these elements were present in *VvGLP2* and *-3* promoter. In the same way, unnamed-4 is involved in the regulation of genes controlling various plant processes (Gupta and Ranjan, 2017). The element was largely found in *VvGLP4* promoter. The presence of these elements in all promoter represents their evolutionary significance.

3.5.3. Unique Cis-regulatory elements

Some unique elements were also identified in *VvGLPs* promoters which include TGA-element, RY-element, AuxRR-core (Auxin responsive element), WUN-motif (Wound responsive element), ERE (Ethylene responsive elements), and MRE (MYB responsive element), etc. TGA-elements were specific to *VvGLP7* which play important role in various biological processes such as growth regulation, development, pathogens response, hormonal and abiotic (salt & drought) stresses (Li et al., 2019b). In the same way, RY-element is involved in seed-specific regulation, embryogenesis, and seed development (Reidt et al., 2000). One copy of this element was found in the *VvGLP3* promoter. WUN elements play important roles in wounding and abiotic (salt, drought etc) stresses (Hayashi et al., 2003; Valifard et al., 2015) while AuxR-core

is important in hormonel response (auxin) regulation (Mironova et al., 2014). Previous studies have shown that phytohormone play important role in acclimatizing plants to various environmental stresses largely mediated by these elements in promoter (Fahad et al., 2015). A single copy of these elements can be found in *VvGLP1* and *-5* promoters respectively suggesting their role in these processes. They also contained a single copy of the ERE elements which play important role in abiotic stresses (drought, salinity, submergence), fruit ripening, and secondary metabolite production (Srivastava and Kumar, 2018). MYB response element (MRE) is crucial for light responsiveness and auxin regulation (Hartmann et al., 2005) which was found in *VvGLP5*. The presence of unique elements in these promoters suggest that these genes have adapted to multiple challenges of the natural environment by inserting novel *cis*-elements in their promoter through the course of evolution.

4. Conclusion

VvGLPs are similar in structure but showed significant variation in their functions. They showed similar physicochemical properties, domain architectures, expression pattern at the subcellular level, 3D structures, and functional properties. The phylogenetic study further revealed a narrow genetic background suggesting their origin via recent duplication events that gave rise to great resemblance among them. Through the course of evolution, these genes have adopted diverse enzymatic activities mainly related to the production of secondary metabolites to better cope with environmental stresses. *VvGLPs* play important role in plant defense through increase immunity, better growth, regulating diverse physiological and biochemical processes both in the cellular and extracellular environment largely mediated by the presence of CAAT, MYB, MYC, and unnamed-4 *cis*-regulatory elements in their promoters. Among all, *VvGLP2* and *VvGLP6* showed more distinctive features which make them more suitable candidates for future use. In the future, the data can be used to develop agronomically important cultivars against a diverse range of stresses.

Human and animal rights

No humans/animals were used in the current research project.

Acknowledgments

This research project was funded by grant number 14-MED333-10 from National Science, Technology and Innovation Plan (MAARIFAH), King Abdul Aziz City for Science and Technology (KACST), Saudi Arabia.

References

ABEL, J., KADEN, M., BOHNSACK, K.S., WEBER, M., LEBERECHT, C. and VILLMANN, T., 2020. Detection of native and mirror protein structures based on Ramachandran plot analysis by interpretable machine learning models. *bioRxiv*, vol. 15, pp. 222-240.

ADELOYE, A.O. and AJIBADE, P.A., 2011. A high molar extinction coefficient mono-anthracenyl bipyridyl heteroleptic ruthenium (II) complex: Synthesis, photophysical and electrochemical properties. *Molecules*, vol. 16, no. 6, pp. 4615-4631. http://dx.doi.org/10.3390/molecules16064615. PMid:21642936.

AHMAD, A., AHMED, A., ESSA, R., BABER, S., JAMSHED, A., SHOUKAT, A., YOUNAS, T., SALEEM, M.A. and NAVEED, M., 2019. *In-Silico* analysis of grapevine Germin like Protein (*VvGLP3*) and its probable role in defense Against Powdery Mildew disease. *Life Science Journal of Pakistan*, vol. 1, pp. 17-23.

ANDREOU, A. and FEUSSNER, I., 2009. Lipoxygenases–structure and reaction mechanism. *Phytochemistry*, vol. 70, no. 13-14, pp. 1504-1510. http://dx.doi.org/10.1016/j.phytochem.2009.05.008. PMid:19767040.

ANUM, J., O'SHEA, C., ZEESHAN HYDER, M., FARRUKH, S., SKRIVER, K., MALIK, S.I. and YASMIN, T., 2022. Germin like protein genes exhibit modular expression during salt and drought stress in elite rice cultivars. *Molecular Biology Reports*, vol. 49, no. 1, pp. 293-302. http://dx.doi.org/10.1007/s11033-021-06871-3. PMid:34725746.

ARMENTEROS, J.A., TSIRIGOS, K.D., SØNDERBY, C.K., PETERSEN, T.N., WINTHER, O., BRUNAK, S., VON HEIJNE, G. and NIELSEN, H., 2019. SignalP 5.0 improves signal peptide predictions using deep neural networks. *Nature Biotechnology*, vol. 37, no. 4, pp. 420-423. http://dx.doi.org/10.1038/s41587-019-0036-z. PMid:30778233.

AUDAIN, E., RAMOS, Y., HERMJAKOB, H., FLOWER, D.R. and PEREZ-RIVEROL, Y., 2016. Accurate estimation of isoelectric point of protein and peptide based on amino acid sequences. *Bioinformatics*, vol. 32, no. 6, pp. 821-827. http://dx.doi.org/10.1093/bioinformatics/btv674. PMid:26568629.

BAE, S.-H., HAN, H.W. and MOON, J., 2015. Functional analysis of the molecular interactions of TATA box-containing genes and essential genes. *PLoS One*, vol. 10, no. 3, pp. e0120848. http://dx.doi.org/10.1371/journal.pone.0120848. PMid:25789484.

BAILEY, T.L., JOHNSON, J., GRANT, C.E. and NOBLE, W.S., 2015. The MEME suite. *Nucleic Acids Research*, vol. 43, no. W1, pp. W39-W49. http://dx.doi.org/10.1093/nar/gkv416. PMid:25953851.

BARMAN, A.R. and BANERJEE, J., 2015. Versatility of germin-like proteins in their sequences, expressions, and functions. *Functional & Integrative Genomics*, vol. 15, no. 5, pp. 533-548. http://dx.doi.org/10.1007/s10142-015-0454-z. PMid:26174051.

BERACOCHEA, V.C., ALMASIA, N.I., PELUFFO, L., NAHIRÑAK, V., HOPP, E., PANIEGO, N., HEINZ, R.A., VAZQUEZ ROVERE, C. and LIA, V.V., 2015. Sunflower germin-like protein *HaGLP1* promotes ROS accumulation and enhances protection against fungal pathogens in transgenic *Arabidopsis thaliana*. *Plant Cell Reports*, vol. 34, no. 10, pp. 1717-1733. http://dx.doi.org/10.1007/s00299-015-1819-4. PMid:26070410.

BEREZOVSKY, I.N., ZELDOVICH, K.B. and SHAKHNOVICH, E.I., 2007. Positive and negative design in stability and thermal adaptation of natural proteins. *PLoS Computational Biology*, vol. 3, no. 3, pp. e52. http://dx.doi.org/10.1371/journal.pcbi.0030052. PMid:17381236.

BERNIER, F. and BERNA, A., 2001. Germins and germin-like proteins: plant do-all proteins. But what do they do exactly? *Plant Physiology and Biochemistry*, vol. 39, no. 7-8, pp. 545-554. http://dx.doi.org/10.1016/S0981-9428(01)01285-2.

BLOM, N., SICHERITZ-PONTÉN, T., GUPTA, R., GAMMELTOFT, S. and BRUNAK, S., 2004. Prediction of post-translational glycosylation

and phosphorylation of proteins from the amino acid sequence. *Proteomics*, vol. 4, no. 6, pp. 1633-1649. http://dx.doi.org/10.1002/pmic.200300771. PMid:15174133.

BOLSER, D.M., STAINES, D.M., PERRY, E. and KERSEY, P.J. (2017). Ensembl plants: integrating tools for visualizing, mining, and analyzing plant genomic data. In A. van Dijk, ed. *Plant genomics databases*. New York: Springer, pp. 1-31. http://dx.doi.org/10.1007/978-1-4939-6658-5_1.

BREEN, J. and BELLGARD, M., 2010. Germin-like proteins (*GLPs*) in cereal genomes: gene clustering and dynamic roles in plant defense. *Functional & Integrative Genomics*, vol. 10, no. 4, pp. 463-476. http://dx.doi.org/10.1007/s10142-010-0184-1. PMid:20683632.

CAO, Y., LI, K., LI, Y., ZHAO, X. and WANG, L., 2020. MYB transcription factors as regulators of secondary metabolism in plants. *Biology (Basel)*, vol. 9, no. 3, pp. 55-61. http://dx.doi.org/10.3390/biology9030061. PMid:32213912.

CASTAÑEDA-GARCÍA, A., RODRÍGUEZ-ROJAS, A., GUELFO, J.R. and BLÁZQUEZ, J., 2009. The glycerol-3-phosphate permease GlpT is the only fosfomycin transporter in *Pseudomonas aeruginosa*. *Journal of Bacteriology*, vol. 191, no. 22, pp. 6968-6974. http://dx.doi.org/10.1128/JB.00748-09. PMid:19734311.

CHOW, C.N., LEE, T.Y., HUNG, Y.C., LI, G.Z., TSENG, K.C., LIU, Y.H., KUO, P.L., ZHENG, H.Q. and CHANG, W.C., 2019. PlantPAN3.0: a new and updated resource for reconstructing transcriptional regulatory networks from ChIP-seq experiments in plants. *Nucleic Acids Research*, vol. 47, no. D1, pp. D1155-D1163. http://dx.doi.org/10.1093/nar/gky1081. PMid:30395277.

CONESA, A. and GÖTZ, S., 2008. Blast2GO: a comprehensive suite for functional analysis in plant genomics. *International Journal of Plant Genomics*, vol. 2008, pp. 619832. http://dx.doi.org/10.1155/2008/619832. PMid:18483572.

DAS, A., PRAMANIK, K., SHARMA, R., GANTAIT, S. and BANERJEE, J., 2019. *In silico* study of biotic and abiotic stress-related transcription factor binding sites in the promoter regions of rice germin-like protein genes. *PLoS One*, vol. 14, no. 2, pp. e0211887. http://dx.doi.org/10.1371/journal.pone.0211887. PMid:30763346.

DASGUPTA, D., KAUSHIK, R. and JAYARAM, B., 2015. From Ramachandran maps to tertiary structures of proteins. *The Journal of Physical Chemistry B*, vol. 119, no. 34, pp. 11136-11145. http://dx.doi.org/10.1021/acs.jpcb.5b02999. PMid:26098815.

DAVIDSON, R.M., MANOSALVA, P.M., SNELLING, J., BRUCE, M., LEUNG, H. and LEACH, J.E., 2010. Rice germin-like proteins: allelic diversity and relationships to early stress responses. *Rice (New York, N.Y.)*, vol. 3, no. 1, pp. 43-55. http://dx.doi.org/10.1007/s12284-010-9038-7.

DEEBA, F., SULTANA, T., MAHMOOD, T., O'SHEA, C., SKRIVER, K. and NAQVI, S.M.S., 2017. Involvement of WRKY, MYB and DOF DNA-binding proteins in interaction with a rice germin-like protein gene promoter. *Acta Physiologiae Plantarum*, vol. 39, no. 8, pp. 189. http://dx.doi.org/10.1007/s11738-017-2488-4.

DOLL, J., HAUSE, B., DEMCHENKO, K., PAWLOWSKI, K. and KRAJINSKI, F., 2003. A member of the germin-like protein family is a highly conserved mycorrhiza-specific induced gene. *Plant & Cell Physiology*, vol. 44, no. 11, pp. 1208-1214. http://dx.doi.org/10.1093/pcp/pcg153. PMid:14634158.

DUNWELL, J.M., GIBBINGS, J.G., MAHMOOD, T. and SAQLAN NAQVI, S., 2008. Germin and germin-like proteins: evolution, structure, and function. *Critical Reviews in Plant Sciences*, vol. 27, no. 5, pp. 342-375. http://dx.doi.org/10.1080/07352680802333938.

DUNWELL, J.M., PURVIS, A. and KHURI, S., 2004. Cupins: the most functionally diverse protein superfamily? *Phytochemistry*, vol. 65, no. 1, pp. 7-17. http://dx.doi.org/10.1016/j.phytochem.2003.08.016. PMid:14697267.

DUVAUD, S., GABELLA, C., LISACEK, F., STOCKINGER, H., IOANNIDIS, V. and DURINX, C., 2021. Expasy, the Swiss Bioinformatics Resource Portal, as designed by its users. *Nucleic Acids Research*, vol. 49, no. W1, pp. W216-W227. http://dx.doi.org/10.1093/nar/gkab225. PMid:33849055.

EL-SHARKAWY, I., MILA, I., BOUZAYEN, M. and JAYASANKAR, S., 2010. Regulation of two germin-like protein genes during plum fruit development. *Journal of Experimental Botany*, vol. 61, no. 6, pp. 1761-1770. http://dx.doi.org/10.1093/jxb/erq043. PMid:20202999.

ETMINAN, A., POUR-ABOUGHADAREH, A., MOHAMMADI, R., NOORI, A. and AHMADI-RAD, A., 2018. Applicability of CAAT Box-derived polymorphism (CBDP) markers for analysis of genetic diversity in durum wheat. *Cereal Research Communications*, vol. 46, no. 1, pp. 1-9. http://dx.doi.org/10.1556/0806.45.2017.054.

FAHAD, S., BAJWA, A.A., NAZIR, U., ANJUM, S.A., FAROOQ, A., ZOHAIB, A., SADIA, S., NASIM, W., ADKINS, S., SAUD, S., IHSAN, M.Z., ALHARBY, H., WU, C., WANG, D. and HUANG, J., 2017. Crop production under drought and heat stress: plant responses and management options. *Frontiers in Plant Science*, vol. 8, pp. 1147-1175. http://dx.doi.org/10.3389/fpls.2017.01147. PMid:28706531.

FAHAD, S., NIE, L., CHEN, Y., WU, C., XIONG, D., SAUD, S., HONGYAN, L., CUI, K. and HUANG, J., 2015. Crop plant hormones and environmental stress. *Sustainable Agriculture Research*, vol. 15, pp. 371-400.

FAHAD, S., SAUD, S., CHEN, Y., WU, C. and WANG, D. (2021a). *Abiotic stress in plants*. London, United Kingdom: BoD–Books on Demand. http://dx.doi.org/10.5772/intechopen.91549.

FAHAD, S., SÖNMEZ, O., SAUD, S., WANG, D., WU, C., ADNAN, M. and ARIF, M. (2021b). *Engineering tolerance in crop plants against abiotic stress*. Boca Raton: CRC Press. http://dx.doi.org/10.1201/9781003160717.

GAMAGE, D.G., GUNARATNE, A., PERIYANNAN, G.R. and RUSSELL, T.G., 2019. Applicability of instability index for in vitro protein stability prediction. *Protein and Peptide Letters*, vol. 26, no. 5, pp. 339-347. http://dx.doi.org/10.2174/0929866526666190228144219. PMid:30816075.

GANGADHAR, B.H., MISHRA, R.K., KAPPACHERY, S., BASKAR, V., VENKATESH, J., NOOKARAJU, A. and THIRUVENGADAM, M., 2021. Enhanced thermo-tolerance in transgenic potato (*Solanum tuberosum* L.) overexpressing hydrogen peroxide-producing germin-like protein (GLP). *Genomics*, vol. 113, no. 5, pp. 3224-3234. http://dx.doi.org/10.1016/j.ygeno.2021.07.013. PMid:34273496.

GAO, Z., ZHAO, R. and RUAN, J., 2013. A genome-wide cis-regulatory element discovery method based on promoter sequences and gene co-expression networks. *BMC Genomics*, vol. 14, no. S1, suppl. 1, pp. S4-S4. http://dx.doi.org/10.1186/1471-2164-14-S1-S4. PMid:23368633.

GIAROLA, V., CHEN, P., DULITZ, S.J., KÖNIG, M., MANDUZIO, S. and BARTELS, D., 2020. The dehydration- and ABA-inducible germin-like protein *CpGLP1* from *Craterostigma plantagineum* has SOD activity and may contribute to cell wall integrity during desiccation. *Planta*, vol. 252, no. 5, pp. 84-111. http://dx.doi.org/10.1007/s00425-020-03485-0. PMid:33044571.

GODFREY, D., ABLE, A.J. and DRY, I.B., 2007. Induction of a grapevine germin-like protein (*VvGLP3*) gene is closely linked to the site of *Erysiphe necator* infection: A possible role in defense? *Molecular Plant-Microbe Interactions*, vol. 20, no. 9, pp. 1112-1125. http://dx.doi.org/10.1094/MPMI-20-9-1112. PMid:17849714.

GUO, Y.L., ZHAO, X., LANZ, C. and WEIGEL, D., 2011. Evolution of the S-Locus region in *Arabidopsis* relatives. *Plant Physiology*, vol. 157, no. 2, pp. 937-946. http://dx.doi.org/10.1104/pp.111.174912. PMid:21810962.

GUPTA, D. and RANJAN, R., 2017. *In silico* comparative analysis of promoters derived from plant pararetroviruses. *Virusdisease*, vol. 28, no. 4, pp. 416-421. http://dx.doi.org/10.1007/s13337-017-0410-8. PMid:29291233.

GUPTA, R. and BRUNAK, S. 2002. Prediction of glycosylation across the human proteome and the correlation to protein function. In: *Pacific Symposium on Biocomputing*, 2002, Washington, DC. Washington, DC: United States. Dept. of Energy. Office of Science, pp. 310-322.

HARALAMPIDIS, K., MILIONI, D., RIGAS, S. and HATZOPOULOS, P., 2002. Combinatorial interaction of cis-elements specifies the expression of the *Arabidopsis AtHsp90-1* gene. *Plant Physiology*, vol. 129, no. 3, pp. 1138-1149. http://dx.doi.org/10.1104/pp.004044. PMid:12114568.

HARTMANN, U., SAGASSER, M., MEHRTENS, F., STRACKE, R. and WEISSHAAR, B., 2005. Differential combinatorial interactions of cis-acting elements recognized by R2R3-MYB, BZIP, and BHLH factors control light-responsive and tissue-specific activation of phenylpropanoid biosynthesis genes. *Plant Molecular Biology*, vol. 57, no. 2, pp. 155-171. http://dx.doi.org/10.1007/s11103-004-6910-0. PMid:15821875.

HAYASHI, T., KOBAYASHI, D., KARIU, T., TAHARA, M., HADA, K., KOUZUMA, Y. and KIMURA, M., 2003. Genomic cloning of ribonucleases in *Nicotiana glutinosa* leaves, as induced in response to wounding or to TMV-infection, and characterization of their promoters. *Bioscience, Biotechnology, and Biochemistry*, vol. 67, no. 12, pp. 2574-2583. http://dx.doi.org/10.1271/bbb.67.2574. PMid:14730135.

HE, Z.-D., TAO, M.-L., LEUNG, D.W.M., YAN, X.-Y., CHEN, L., PENG, X.-X. and LIU, E.E., 2021. The rice germin-like protein *OsGLP1* participates in acclimation to UV-B radiation. *Plant Physiology*, vol. 186, no. 2, pp. 1254-1268. http://dx.doi.org/10.1093/plphys/kiab125. PMid:33713137.

HOLLEBEEK, T., HO, T.-S. and RABITZ, H., 1999. Constructing multidimensional molecular potential energy surfaces from ab initio data. *Annual Review of Physical Chemistry*, vol. 50, no. 1, pp. 537-570. http://dx.doi.org/10.1146/annurev.physchem.50.1.537. PMid:15012421.

HORTON, P., PARK, K.-J., OBAYASHI, T., FUJITA, N., HARADA, H., ADAMS-COLLIER, C. and NAKAI, K., 2007. WoLF PSORT: protein localization predictor. *Nucleic Acids Research*, vol. 35, no. Web Server, pp. W585-W587. http://dx.doi.org/10.1093/nar/gkm259. PMid:17517783.

IKAI, A., 1980. Thermostability and aliphatic index of globular proteins. *Journal of Biochemistry*, vol. 88, no. 6, pp. 1895-1898. PMid:7462208.

ILYAS, M., AKHTAR, W., REHMAN, S., NAQVI, S.M.S. and MAHMOOD, T., 2019. Functional characterization of the rice root Germin-like protein gene-1 (OsRGLP1) promoter in Nicotiana tabacum. *3 Biotech*, vol. 9, pp. 130.

ILYAS, M., IRFAN, M., MAHMOOD, T., HUSSAIN, H., LATIF-UR-REHMAN., NAEEM, I. and KHALIQ-UR-RAHMAN., 2020. Analysis of germin-like protein genes (*OsGLPS*) family in rice using various *in silico* approaches. *Current Bioinformatics*, vol. 15, no. 1, pp. 17-33. http://dx.doi.org/10.2174/1574893614666190722165130.

ILYAS, M., NAQVI, S.M.S. and MAHMOOD, T., 2016a. *In silico* analysis of transcription factor binding sites in promoters of germin-like protein genes in rice. *Archives of Biological Sciences*, vol. 68, no. 4, pp. 863-876. http://dx.doi.org/10.2298/ABS151116076I.

ILYAS, M., RASHEED, A. and MAHMOOD, T., 2016b. Functional characterization of germin and germin-like protein genes in various plant species using transgenic approaches. *Biotechnology Letters*, vol. 38, no. 9, pp. 1405-1421. http://dx.doi.org/10.1007/s10529-016-2129-9. PMid:27230937.

JEGLA, T., BUSEY, G. and ASSMANN, S.M., 2018. Evolution and structural characteristics of plant voltage-gated K^+ channels. *The Plant Cell*, vol. 30, no. 12, pp. 2898-2909. http://dx.doi.org/10.1105/tpc.18.00523. PMid:30389753.

KAUR, A., PATI, P.K., PATI, A.M. and NAGPAL, A.K., 2017. *In silico* analysis of cis-acting regulatory elements of pathogenesis-related proteins of *Arabidopsis thaliana* and *Oryza sativa*. *PLoS One*, vol. 12, no. 9, pp. e0184523. http://dx.doi.org/10.1371/journal.pone.0184523. PMid:28910327.

KEARSE, M., MOIR, R., WILSON, A., STONES-HAVAS, S., CHEUNG, M., STURROCK, S., BUXTON, S., COOPER, A., MARKOWITZ, S., DURAN, C., THIERER, T., ASHTON, B., MEINTJES, P. and DRUMMOND, A., 2012. Geneious Basic: an integrated and extendable desktop software platform for the organization and analysis of sequence data. *Bioinformatics*, vol. 28, no. 12, pp. 1647-1649. http://dx.doi.org/10.1093/bioinformatics/bts199. PMid:22543367.

KIM, H.J. and TRIPLETT, B.A., 2004. Cotton fiber germin-like protein. I. Molecular cloning and gene expression. *Planta*, vol. 218, no. 4, pp. 516-524. http://dx.doi.org/10.1007/s00425-003-1133-1. PMid:14648117.

KUSNETSOV, V., LANDSBERGER, M., MEURER, J. and OELMÜLLER, R., 1999. The assembly of the CAAT-box binding complex at a photosynthesis gene promoter is regulated by light, cytokinin, and the stage of the plastids. *The Journal of Biological Chemistry*, vol. 274, no. 50, pp. 36009-36014. http://dx.doi.org/10.1074/jbc.274.50.36009. PMid:10585491.

LANGENBACH, C., SCHULTHEISS, H., ROSENDAHL, M., TRESCH, N., CONRATH, U. and GOELLNER, K., 2016. Interspecies gene transfer provides soybean resistance to a fungal pathogen. *Plant Biotechnology Journal*, vol. 14, no. 2, pp. 699-708. http://dx.doi.org/10.1111/pbi.12418. PMid:26096357.

LI, B., FAN, R., GUO, S., WANG, P., ZHU, X., FAN, Y., CHEN, Y., HE, K., KUMAR, A., SHI, J., WANG, Y., LI, L., HU, Z. and SONG, C.-P., 2019a. The *Arabidopsis* MYB transcription factor, MYB111 modulates salt responses by regulating flavonoid biosynthesis. *Environmental and Experimental Botany*, vol. 166, pp. 103807. http://dx.doi.org/10.1016/j.envexpbot.2019.103807.

LI, B., LIU, Y., CUI, X.-Y., FU, J.-D., ZHOU, Y.-B., ZHENG, W.-J., LAN, J.-H., JIN, L.-G., CHEN, M., MA, Y.-Z., XU, Z.-S. and MIN, D.-H., 2019b. Genome-wide characterization and expression analysis of soybean TGA transcription factors identified a novel TGA gene involved in drought and salt tolerance. *Frontiers in Plant Science*, vol. 10, pp. 549. http://dx.doi.org/10.3389/fpls.2019.00549. PMid:31156656.

LIAO, L., HU, Z., LIU, S., YANG, Y. and ZHOU, Y., 2021. Characterization of germin-like proteins (*GLPs*) and their expression in response to abiotic and biotic stresses in cucumber. *Horticulturae*, vol. 7, no. 10, pp. 412. http://dx.doi.org/10.3390/horticulturae7100412.

LIN, C., CHEN, W., QIU, C., WU, Y., KRISHNAN, S. and ZOU, Q., 2014. LibD3C: ensemble classifiers with a clustering and dynamic selection strategy. *Neurocomputing*, vol. 123, pp. 424-435. http://dx.doi.org/10.1016/j.neucom.2013.08.004.

LIU, B., LIU, F., WANG, X., CHEN, J., FANG, L. and CHOU, K.-C., 2015. Pse-in-One: A web server for generating various modes of pseudo components of DNA, RNA, and protein sequences. *Nucleic Acids Research*, vol. 43, no. W1, pp. W65-W71. http://dx.doi.org/10.1093/nar/gkv458. PMid:25958395.

LIU, B., WU, H., ZHANG, D., WANG, X. and CHOU, K.-C., 2017. Pse-Analysis: a python package for DNA/RNA and protein/peptide sequence analysis based on pseudo components and kernel methods. *Oncotarget*, vol. 8, no. 8, pp. 13338-13343. http://dx.doi.org/10.18632/oncotarget.14524. PMid:28076851.

LIU, C., TALBOT, N.J. and CHEN, X.L., 2021. Protein glycosylation during infection by plant pathogenic fungi. *The New Phytologist*, vol. 230, no. 4, pp. 1329-1335. http://dx.doi.org/10.1111/nph.17207. PMid:33454977.

LU, M., HAN, Y.P., GAO, J.G., WANG, X.-J. and LI, W.-B., 2010. Identification and analysis of the germin-like gene family in soybean. *BMC Genomics*, vol. 11, no. 1, pp. 620. http://dx.doi.org/10.1186/1471-2164-11-620. PMid:21059215.

LU, S., WANG, J., CHITSAZ, F., DERBYSHIRE, M.K., GEER, R.C., GONZALES, N.R., GWADZ, M., HURWITZ, D.I., MARCHLER, G.H., SONG, J.S., THANKI, N., YAMASHITA, R.A., YANG, M., ZHANG, D., ZHENG, C., LANCZYCKI, C.J. and MARCHLER-BAUER, A., 2020. CDD/SPARCLE: the conserved domain database in 2020. *Nucleic Acids Research*, vol. 48, no. D1, pp. D265-D268. http://dx.doi.org/10.1093/nar/gkz991. PMid:31777944.

MA, D. and CONSTABEL, C.P., 2019. MYB repressors as regulators of phenylpropanoid metabolism in plants. *Trends in Plant Science*, vol. 24, no. 3, pp. 275-289. http://dx.doi.org/10.1016/j.tplants.2018.12.003. PMid:30704824.

MAHMOOD, T., NAZAR, N., YASMIN, T., ABBASI, B.H., AHMAD, M. and NAQVI, S.M.S., 2010. Comparative analysis of regulatory elements in different germin-like protein gene promoters. *African Journal of Biotechnology*, vol. 9, no. 13, pp. 1871-1880. http://dx.doi.org/10.5897/AJB09.1803.

MAHMOOD, T., TAHIR, T., MUNIR, F. and SHINWARI, Z.K., 2018. Characterization of regulatory elements in *OsRGLP2* gene promoter from different rice accessions through sequencing and *in silico* evaluation. *Computational Biology and Chemistry*, vol. 73, pp. 206-212. http://dx.doi.org/10.1016/j.compbiolchem.2018.02.015. PMid:29501997.

MANOSALVA, P.M., DAVIDSON, R.M., LIU, B., ZHU, X., HULBERT, S.H., LEUNG, H. and LEACH, J.E., 2009. A germin-like protein gene family functions as a complex quantitative trait locus conferring broad-spectrum disease resistance in rice. *Plant Physiology*, vol. 149, no. 1, pp. 286-296. http://dx.doi.org/10.1104/pp.108.128348. PMid:19011003.

MCGUFFIN, L.J., BUENAVISTA, M.T. and ROCHE, D.B., 2013. The ModFOLD4 server for the quality assessment of 3D protein models. *Nucleic Acids Research*, vol. 41, no. W368-72, pp. W368-W372. http://dx.doi.org/10.1093/nar/gkt294. PMid:23620298.

MEMBRÉ, N., BERNIER, F., STAIGER, D. and BERNA, A., 2000. *Arabidopsis thaliana* germin-like proteins: common and specific features point to a variety of functions. *Planta*, vol. 211, no. 3, pp. 345-354. http://dx.doi.org/10.1007/s004250000277. PMid:10987552.

MINDREBO, J.T., NARTEY, C.M., SETO, Y., BURKART, M.D. and NOEL, J.P., 2016. Unveiling the functional diversity of the alpha/beta hydrolase superfamily in the plant kingdom. *Current Opinion in Structural Biology*, vol. 41, pp. 233-246. http://dx.doi.org/10.1016/j.sbi.2016.08.005. PMid:27662376.

MIRONOVA, V.V., OMELYANCHUK, N.A., WIEBE, D.S. and LEVITSKY, V.G., 2014. Computational analysis of auxin responsive elements in the *Arabidopsis thaliana* L. genome. *BMC Genomics*, vol. 15, no. S12, suppl. 12, pp. S4-S4. http://dx.doi.org/10.1186/1471-2164-15-S12-S4. PMid:25563792.

MUNIR, F., HAYASHI, S., BATLEY, J., NAQVI, S.M.S. and MAHMOOD, T., 2016. Germin-like protein 2 gene promoter from rice is responsive to fungal pathogens in transgenic potato plants. *Functional & Integrative Genomics*, vol. 16, no. 1, pp. 19-27. http://dx.doi.org/10.1007/s10142-015-0463-y. PMid:26277722.

NAGASHIMA, Y., VON SCHAEWEN, A. and KOIWA, H., 2018. Function of N-glycosylation in plants. *Plant Science*, vol. 274, pp. 70-79. http://dx.doi.org/10.1016/j.plantsci.2018.05.007. PMid:30080642.

NAGY, T., YOSA REYES, J. and MEUWLY, M., 2014. Multisurface adiabatic reactive molecular dynamics. *Journal of Chemical Theory and Computation*, vol. 10, no. 4, pp. 1366-1375. http://dx.doi.org/10.1021/ct400953f. PMid:26580356.

OGISO, E., TAKAHASHI, Y., SASAKI, T., YANO, M. and IZAWA, T., 2010. The role of casein kinase-II in flowering time regulation has diversified during evolution. *Plant Physiology*, vol. 152, no. 2, pp. 808-820. http://dx.doi.org/10.1104/pp.109.148908. PMid:20007447.

OHWOFASA, A., PERVEEN, S., KHAN, T.A., IJAZ, B. and YASMIN, T., 2021. In silico promoter characterization of pr genes and expression analysis in transgenic tomato plants expressing *OsRGLP1*. *Pakistan Journal of Botany*, vol. 53, no. 5, pp. 1707-1716. http://dx.doi.org/10.30848/PJB2021-5(36).

PEI, Y., LI, X., ZHU, Y., GE, X., SUN, Y., LIU, N., JIA, Y., LI, F. and HOU, Y., 2019. GHABP19, a novel germin-like protein from *Gossypium Hirsutum*, plays an important role in the regulation of resistance to verticillium and fusarium wilt pathogens. *Frontiers in Plant Science*, vol. 10, pp. 789-801. http://dx.doi.org/10.3389/fpls.2019.00583. PMid:31134119.

PEREZ-RIBA, A. and ITZHAKI, L.S., 2019. The tetratricopeptide-repeat motif is a versatile platform that enables diverse modes of molecular recognition. *Current Opinion in Structural Biology*, vol. 54, pp. 43-49. http://dx.doi.org/10.1016/j.sbi.2018.12.004. PMid:30708253.

PICHERSKY, E. and RAGUSO, R.A., 2018. Why do plants produce so many terpenoid compounds? *The New Phytologist*, vol. 220, no. 3, pp. 692-702. http://dx.doi.org/10.1111/nph.14178. PMid:27604856.

RAYON, C., LEROUGE, P. and FAYE, L., 1998. The protein N-glycosylation in plants. *Journal of Experimental Botany*, vol. 49, no. 326, pp. 1463-1472. http://dx.doi.org/10.1093/jxb/49.326.1463.

REGO, N.B., XI, E. and PATEL, A.J., 2021. Identifying hydrophobic protein patches to inform protein interaction interfaces. *Proceedings of the National Academy of Sciences of the United States of America*, vol. 118, no. 6, pp. e2018234118. http://dx.doi.org/10.1073/pnas.2018234118. PMid:33526682.

REIDT, W., WOHLFARTH, T., ELLERSTRÖM, M., CZIHAL, A., TEWES, A., EZCURRA, I., RASK, L. and BÄUMLEIN, H., 2000. Gene regulation during late embryogenesis: the RY motif of maturation-specific gene promoters is a direct target of the *FUS3* gene product. *The Plant Journal*, vol. 21, no. 5, pp. 401-408. http://dx.doi.org/10.1046/j.1365-313x.2000.00686.x. PMid:10758492.

RUIZ-MEDRANO, R., JIMENEZ-MORAILA, B., HERRERA-ESTRELLA, L. and RIVERA-BUSTAMANTE, R.F., 1992. Nucleotide sequence of an osmotin-like cDNA induced in tomato during viroid infection. *Plant Molecular Biology*, vol. 20, no. 6, pp. 1199-1202. http://dx.doi.org/10.1007/BF00028909. PMid:1463856.

SASSAKI, F.T., BRAVO, J.P., GONZÁLEZ, E.R. and MAIA, I.G., 2015. Expression pattern and promoter analysis of a *Eucalyptus grandis* germin-like gene. *Plant Molecular Biology Reporter*, vol. 33, no. 1, pp. 12-21. http://dx.doi.org/10.1007/s11105-014-0734-0.

SATO, Y., ANTONIO, B.A., NAMIKI, N., TAKEHISA, H., MINAMI, H., KAMATSUKI, K., SUGIMOTO, K., SHIMIZU, Y., HIROCHIKA, H. and NAGAMURA, Y., 2010. RiceXPro: a platform for monitoring

gene expression in japonica rice grown under natural field conditions. *Nucleic Acids Research*, vol. 39, no. Database, pp. D1141-D1148. PMid:21045061.

SHI, C.-L., DONG, N.-Q., GUO, T., YE, W.-W., SHAN, J.-X. and LIN, H.-X., 2020. A quantitative trait locus GW6 controls rice grain size and yield through the gibberellin pathway. *The Plant Journal*, vol. 103, no. 3, pp. 1174-1188. http://dx.doi.org/10.1111/tpj.14793. PMid:32365409.

SHIN, D.H., PROUDFOOT, M., LIM, H.J., CHOI, I.K., YOKOTA, H., YAKUNIN, A.F., KIM, R. and KIM, S.H., 2008. Structural and enzymatic characterization of DR1281: a calcineurin-like phosphoesterase from *Deinococcus radiodurans. Proteins*, vol. 70, no. 3, pp. 1000-1009. http://dx.doi.org/10.1002/prot.21584. PMid:17847097.

SIEVERS, F. and HIGGINS, D.G., 2018. Clustal Omega for making accurate alignments of many protein sequences. *Protein Science*, vol. 27, no. 1, pp. 135-145. http://dx.doi.org/10.1002/pro.3290. PMid:28884485.

SONG, S., XU, Y., HUANG, D., MIAO, H., LIU, J., JIA, C., HU, W., VALAREZO, A.V., XU, B. and JIN, Z., 2018. Identification of a novel promoter from banana aquaporin family gene (*MaTIP1;2*) which responses to drought and salt-stress in transgenic *Arabidopsis thaliana. Plant Physiology and Biochemistry*, vol. 128, pp. 163-169. http://dx.doi.org/10.1016/j.plaphy.2018.05.003. PMid:29778840.

SRIVASTAVA, R. and KUMAR, R., 2018. The expanding roles of APETALA2/Ethylene Responsive Factors and their potential applications in crop improvement. *Briefings in Functional Genomics*, vol. 18, no. 4, pp. 240-254. http://dx.doi.org/10.1093/bfgp/elz001. PMid:30783669.

STEGMANN, M., MONAGHAN, J., SMAKOWSKA-LUZAN, E., ROVENICH, H., LEHNER, A., HOLTON, N., BELKHADIR, Y. and ZIPFEL, C., 2017. The receptor kinase FER is a RALF-regulated scaffold controlling plant immune signaling. *Science*, vol. 355, no. 6322, pp. 287-289. http://dx.doi.org/10.1126/science.aal2541. PMid:28104890.

STRUMILLO, M.J., OPLOVÁ, M., VIÉITEZ, C., OCHOA, D., SHAHRAZ, M., BUSBY, B.P., SOPKO, R., STUDER, R.A., PERRIMON, N., PANSE, V.G. and BELTRAO, P., 2019. Conserved phosphorylation hotspots in eukaryotic protein domain families. *Nature Communications*, vol. 10, no. 1, pp. 1977-1977. http://dx.doi.org/10.1038/s41467-019-09952-x. PMid:31036831.

SZKLARCZYK, D., GABLE, A.L., LYON, D., JUNGE, A., WYDER, S., HUERTA-CEPAS, J., SIMONOVIC, M., DONCHEVA, N.T., MORRIS, J.H., BORK, P., JENSEN, L.J. and MERING, C.V., 2019. STRING v11: protein–protein association networks with increased coverage, supporting functional discovery in genome-wide experimental datasets. *Nucleic Acids Research*, vol. 47, no. D1, pp. D607-D613. http://dx.doi.org/10.1093/nar/gky1131. PMid:30476243.

TONG, X., NAGY, T., REYES, J.Y., GERMANN, M., MEUWLY, M. and WILLITSCH, S., 2012. State-selected ion–molecule reactions with Coulomb-crystallized molecular ions in traps. *Chemical Physics Letters*, vol. 547, pp. 1-8. http://dx.doi.org/10.1016/j.cplett.2012.06.042.

VALIFARD, M., MOHSENZADEH, S., NIAZI, A. and MOGHADAM, A. (2015). "Phenylalanine ammonia lyase isolation and functional analysis of phenylpropanoid pathway under salinity stress in "*Salvia*" species," Southern Cross Journals.

VON HEIJNE, G., 1990. The signal peptide. *The Journal of Membrane Biology*, vol. 115, no. 3, pp. 195-201. http://dx.doi.org/10.1007/BF01868635. PMid:2197415.

WANG, T., CHEN, X., ZHU, F., LI, H., LI, L., YANG, Q., CHI, X., YU, S. and LIANG, X., 2013. Characterization of peanut germin-like proteins, *AhGLPs* in plant development and defense. *PLoS One*, vol. 8, no. 4, pp. e61722. http://dx.doi.org/10.1371/journal.pone.0061722. PMid:23626720.

WATERHOUSE, A., BERTONI, M., BIENERT, S., STUDER, G., TAURIELLO, G., GUMIENNY, R., HEER, F.T., DE BEER, T.A.P., REMPFER, C., BORDOLI, L., LEPORE, R. and SCHWEDE, T., 2018. SWISS-MODEL: homology modelling of protein structures and complexes. *Nucleic Acids Research*, vol. 46, no. W1, pp. W296-W303. http://dx.doi.org/10.1093/nar/gky427. PMid:29788355.

WU, C., BA, Q., LU, D., LI, W., SALOVSKA, B., HOU, P., MUELLER, T., ROSENBERGER, G., GAO, E., DI, Y., ZHOU, H., FORNASIERO, E.F. and LIU, Y., 2021. Global and site-specific effect of Phosphorylation on protein turnover. *Developmental Cell*, vol. 56, no. 1, pp. 111-124.e6. http://dx.doi.org/10.1016/j.devcel.2020.10.025. PMid:33238149.

XIE, P.F., ZHU, L., FENG, L., WU, J.C. and LIU, J.-L., 2020. Research progress in transcription factor MYC2 mediating plant resistance to biological stress. *Yingyong Kunchong Xuebao*, vol. 57, pp. 781.

YIN, K., HAN, X., XU, Z. and XUE, H., 2009. *Arabidopsis* GLP4 is localized to the Golgi and binds auxin in vitro. *Acta Biochimica et Biophysica Sinica*, vol. 41, no. 6, pp. 478-487. http://dx.doi.org/10.1093/abbs/gmp036. PMid:19499151.

YU, C.S., CHEN, Y.C., LU, C.H. and HWANG, J.K., 2006. Prediction of protein subcellular localization. *Proteins*, vol. 64, no. 3, pp. 643-651. http://dx.doi.org/10.1002/prot.21018. PMid:16752418.

YUAN, B., YANG, Y., FAN, P., LIU, J., XING, H., LIU, Y. and FENG, D., 2021. Genome-wide identification and characterization of Germin and Germin-Like Proteins (*GLPs*) and their response under Powdery Mildew Stress in Wheat (*Triticum aestivum* L.). *Plant Molecular Biology Reporter*, vol. 10, no. 4, pp. 1-12. http://dx.doi.org/10.1007/s11105-021-01291-w.

ZHANG, L., SONG, Z., LI, F., LI, X., JI, H. and YANG, S., 2019. The specific MYB binding sites bound by Ta MYB in the GAPCp2/3 promoters are involved in the drought stress response in wheat. *BMC Plant Biology*, vol. 19, pp. 1-14.

ZHANG, Y., SUN, T., LIU, S., DONG, L., LIU, C., SONG, W., LIU, J. and GAI, S., 2016. MYC cis-elements in psmpt promoter is involved in chilling response of *Paeonia Suffruticosa. PLoS One*, vol. 11, no. 5, pp. e0155780. http://dx.doi.org/10.1371/journal.pone.0155780. PMid:27228117.

ZIMMERMANN, G., BÄUMLEIN, H., MOCK, H.-P., HIMMELBACH, A. and SCHWEIZER, P., 2006. The multigene family encoding germin-like proteins of barley. Regulation and function in basal host resistance. *Plant Physiology*, vol. 142, no. 1, pp. 181-192. http://dx.doi.org/10.1104/pp.106.083824. PMid:16844832.

Supplementary Material

Supplementary data is provided along with the article as excel files (S1).

Supplementary material accompanies this paper.

Figure S1. 3D structure of the *Vitis vinifera* Germin-like proteins (VvGLPs) genes. Different domains are represented with different colours. The data were obtained with Swiss modeling server.

Table S1. Description of the *Vitis vinifera* Germin like protein genes (*VvGLPs*) sequences used in the analysis.

Table *VvGLP1* to *VvGLP7*. Description of the *cis*-elements in the promoters of *Vitis vinifera* Germin like proteins genes.

This material is available as part of the online article from https://www.scielo.br/j/bjb