Pak. J. Bot., 57(4): DOI: http://dx.doi.org/10.30848/PJB2025-4(12)

COMPUTATIONAL ANALYSIS OF GERMIN-LIKE PROTEIN GENES IN *BRACHYPODIUM DISTACHYON* (BdGLPs)

MUHAMMAD ILYAS[1*], HAZRAT HUSSAIN[2], MUHAMMAD ADNAN SHEREEN[3*], SHAHEED ULLAH[4], MUHAMMAD SAFDAR HUSSAIN[5], NAZISH HUMA KHAN[6] AND ADIL A. MUJAWAH[7]

[1]Department of Botany, Kohsar University Murree, 47150, Punjab Pakistan
[2]Department of Biotechnology, University of Swabi, Swabi, Pakistan
[3]Department of Microbiology, Kohsar University Murree, 47150, Punjab Pakistan
[4]Department of Chemistry, Kohsar University, Murree, 47150, Punjab, Pakistan
[5]Department of Forestry and Range Management, Kohsar University, Murree,47150, Punjab, Pakistan
[6]Department of Environmental Sciences, University of Swabi, Swabi, Pakistan
[7]Department of Chemistry, College of Science, Qassim University, Buraidah 51452, Saudi Arabia
[*]Corresponding authors' email: dr.milyas@kum.edu.pk; msher017@uOttawa.ca

Abstract

Germin-like proteins (GLPs) are members of the Cupin superfamily playing an important role in plant development and inducing stress resistance. *Brachypodium distachyon's* genome contains 08 GLPs which are functionally unexplored. However, *in silico* analysis can provide better insights into their structural, functional, and phylogenetic properties. Presently, physicochemical properties, domain composition, phosphorylation & N-glycosylation sites, subcellular localization, 3D structures, phylogenetic relationships, and functional importance of the eight *BdGLPs* genes were studied using the latest computational tools. All the BdGLPs exhibited similar physicochemical properties, domain architecture, 3D structures, and subcellular localizations, except *BdGLP2-1*, and *BdGLP3-7*. The Molecular and Evolutionary Genetics Analysis (MEGA11) tool was used to assess their phylogenetic relationships while STRING (Ver. 10) was used for functional characterization. All the *BdGLPs* showed a close phylogenetic relationship due to their presence on the same chromosome (Chr 3) except *BdGLP2-1*, which was found on chromosome 2. Functionally, BdGLP2-1 exhibits NADH dehydrogenase activity, whereas other BdGLPs are associated with WPP domain-related functions. This analysis provides valuable insights into the potential utilization of *BdGLPs* in functional genomics to develop resistant cultivars against various stresses.

Key words: Stress, Oxalate oxidase, Germin-like protein, *Brachypodium distachyon*.

Introduction

Germins (GERs) and Germin-like proteins (GLPs) are members of the "Cupin" superfamily and represent a highly diverse group of plant proteins (Barman and Banerjee, 2015). Originally reported as a marker, specific to the germination process in wheat, the GER protein family was later found to be a glycoprotein-forming homohexamer with oxalate oxidase (OXO) activity (Dunwell *et al*., 2008). Proteins showing an average similarity of 50% to GERs were classified as GLPs. Research has shown that GLPs play a significant role by regulating various physiological and developmental processes of plants such as germination, pollen formation, and stress signaling (Davidson *et al*., 2010). GER and GLPs exhibit a wide array of enzymatic activities including oxalate oxidase (OXO), superoxide dismutase (SOD), polyphenol oxidase (PPO), ADP glucose pyrophosphatase/ phosphodiesterase, and cysteine peptidase etc. (Ilyas *et al*., 2019). While their roles in diverse stresses (mechanical, abiotic, biotic), physiological, biochemical, and developmental processes have been well-characterized (Ilyas *et al*., 2016b), further research is needed to understand their properties.

GERs and GLPs have been reported in diverse plant groups including mosses, monocots, dicots, gymnosperms, and angiosperms (Dunwell *et al*., 2008). Till now, 350 of them have been identified in diverse plant species (Ilyas *et al*., 2024 unpublished). Modern techniques such as whole genome sequencing and genome-wide analysis have identified many GLPs in diverse plant species such as *Glycine max* (Lu *et al*., 2010; Wang *et al*., 2014), *Arabidopsis* (Li *et al*., 2016), *Camelia sinensis* (Fu *et al*., 2018), *Triticum aestivum* (Yuan *et al*., 2021), *Vitis vinifera* (Ilyas *et al*., 2022), and *Zea mays* (Ilyas *et al*., 2023) etc. Further, they have been significantly used to provide resistance against various stresses in diverse crop species (Ilyas *et al*., 2019; 2016b) via activating Reactive oxygen species (ROS) i.e. H_2O_2 biosynthesis pathways. For example, among abiotic stresses they were found effective against heavy metals (Bernier *et al*., 2001), salt & drought stresses (Wang *et al*., 2013), heat (Gangadhar *et al*., 2021), desiccation (Giarola *et al*., 2020), and UV radiation (He *et al*., 2021), while among biotic stresses they protect plants against *Sclerotinia sclerotiorum* (Zhang *et al*., 2018), *Blumeria graminis* f. sp. *Tritici* (Yuan *et al*., 2021), and *Verticillium* and *Fusarium* wilt (Pei *et al*., 2023), etc. However, despite numerous studies regarding characterizing GLPs in various plant species, their function remains unexplored.

In recent days, various computational tools including Expasy's ProtParam server (https://web.expasy.org/ protparam/) (Artimo *et al*., 2012), Pse-in-One (http://bioinformatics.hitsz.edu.cn/Pse-in-One/) (Liu *et al*., 2015), Ensemble Plant (https://plants.ensembl.org/ index.html), BioSeq-Analysis (http://bioinformatics.hitsz. edu.cn/BioSeq-Analysis/) (Lin *et al*., 2014), Libd3C (http://lab.malab.cn/soft/LibD3C/) (Lin *et al*., 2014), Max-Relevance-Max-Distance (MRMD) (http://lab.malab.cn/ soft/ MRMD/index_en.html) (Hollebeek *et al*., 1999; Nagashima *et al*., 2018; Tong *et al*., 2012), Compartments database (https://compartments.jensenlab.org/Search), Pse-Analysis (http://bioinformatics.hitsz.edu.cn/Pse-Analysis/) (Liu *et al*., 2017), RiceXpro (http://ricexpro. dna.affrc.go.jp/) (Sato *et al*., 2010), and HMMER (http://hmmer.org/) etc., are

extensively used for the structural and functional validation of proteins or genes sequences. The information obtained through *in silico* analyses provides a base for designing future strategies to develop resilient cultivars with desirable modulation of gene expression.

Till now various *in silico* strategies have been adopted to understand the function of GER and GLPs in diverse plant species such as *Oryza sativa* (*OsGLPs*) (Ilyas *et al*., 2022; 2016a), wheat (*TaGLP-2b*) (Mahajan *et al*., 2018), potato (*StGLPs*) (Zaynab *et al*., 2022), *Astragalus sinicus* (*AsGLP1)* (Zeng *et al*., 2023) and *Zea mays* (*ZmGLPs*) (Ilyas *et al*., 2023) etc. The current study was specifically designed to comprehensively investigate the properties of GLP genes in *Brachypodium distachyon* using a range of *in silico* methodologies. Moreover, the goal was to provide valuable insights into the functional aspects of these previously uncharacterized genes, which could prove instrumental in their application in various biotechnological approaches. Additionally, this study aimed to predict the potential roles of 08 *BdGLP* genes in diverse plant processes, offering prospects for enhancing crop improvement strategies.

Material and Methods

Sequences retrieval: Ensemble plants (https://plants. ensembl.org/index.html, accessed on 05 November 2023) (Cunningham *et al.,* 2022), and National Centre for Biotechnology Information (https://www. ncbi.nlm.nih. gov/, accessed on 05 November 2023) were used to retrieve coding sequences of the eight *BdGLP*s genes. The sequences were translated via the GENSCAN web server at MIT (http://hollywood.mit.edu/ GENSCAN. html, accessed on 06 November 2023) and searched for the Cupin domain presence via NCBI CDD-search (conserved domain database-search (https://www. ncbi.nlm.nih.gov/ structure/cdd/wrpsb.cgi, accessed on 06 November 2023).

Protein sequence analysis: The peptide sequences of the *BdGLPs* were aligned using the Clustal Omega Multiple sequence alignment tool of EMBL-EBI (https://www. ebi.ac.uk/jdispatcher/msa/clustalo, accessed on 8 November 2023) and identified three conserved GER motives. Various physicochemical properties of the BdGLPs, encompassing residue charges (+R, -R), molecular weight (M. Wt), grand average of hydropathicity (GRAVY), aliphatic index (AI), instability index (II), extinction coefficient (EC), and isoelectric point (pI) were predicted utilizing the EXPASY Protparam server (http://web.expasy.org/protparam/, accessed on 10 November 2023) (Gasteiger *et al*., 2003). Subcellular localization was determined using WoLF PSORT (http://www.genescript.com/wolf-psort.html, accessed on 10 November 2023) (Horton *et al*., 2007), and CELLO Ver. 2.5 (http://cello.life.nctu.edu.tw/, accessed on 11 November 2023). Signal peptides were predicted with SignalP (Ver. 4.1), (http://www.cbs.dtu.dk/services/signal, accessed 12 November 2023) while N-glycosylation (N-site) and phosphorylation sites (P-site) were identified with NetNglyc (Ver. 1.0) (http://www.cbs.dtu.dk/service/ NetNGlyc/) and Netphos (Ver. 2.0) (http://www.cbs.dtu. dk/services/NetPhos/, accessed on 13 November 2023) respectively. Structural models (3D) were built with the help of a Swiss modeling server (https://swissmodel. expasy.org/interactive, accessed on 15 November 2023). Ramachandran plot analysis (RPA) was subsequently carried out to confirm their quality using the Ramachandran plot server (https://swift.cmbi.umcn.nl/ servers/html/ramaplot.html, accessed on 15 November 2023) (Laskowski *et al.,* 2013). The neighbor-joining method was employed to construct a phylogenetic tree using Molecular and evolutionary genetics analysis (Ver. 10) software.

Functional analysis: Functional analysis was conducted by searching five conserved functional domains using Multiple Em for Motif Elicitation (MEME) (http:// meme.nbcr.net/meme/cgi-bin/meme.cgi, accessed on 20 November 2023) (Bailey *et al*., 2015), and their functions were predicted using Motif Scan software (https://myhits.sib.swiss/cgi-bin/motifscan accessed on 25 November 2023). Further, the identified BdGLPs were analyzed for their possible functions and co-expression network using the Search Tool for the Retrieval of Interacting Genes/Proteins (STRING ver. 10) available at https://string-db.org/ which was accessed on 25 November 2023 (Szklarczyk *et al*., 2015).

Results and Discussion

Data retrieval: The Results revealed that the genome of the *Brachypodium distachyon* possesses 09 GLP (Breen *et al*., 2010) genes among which sequence data for *BdGLP3-6* was not found in both Ensemble and Phytozome databases. The number of GLPs varies greatly in different plant species. Till now, they have been reported in diverse plant species in various numbers such as *Brassica napus* (14 *BnGLP* genes) (Rietz *et al.,* 2012), *Arabidopsis thaliana* (32 *AtGLP* genes) (Li *et al.,* 2016), and *Musa acuminata* (44 *MaGLP* genes) (Liu *et al.,* 2020), etc. Currently, all the identified *BdGLP*s (except *BdGLP2-1*) are located on chromosome 3, which is similar to the chromosomal distribution observed in rice (*OsGLP*s), barley (*HvGLPs*), maize (*ZmGLPs*), etc. where *GLPs* clusters are prevalent on specific chromosomes (3, 8, and 12 etc.) suggesting potential gene duplication events on these chromosomes (Ilyas *et al.,* 2019). Gene duplication and clustering is a common phenomenon among *GLP* genes, and it has previously been observed in rice, wheat, barley, maize, and grapes and are often associated with disease resistance (Ilyas *et al*., 2016b). In *Brachypodium*, they were initially reported by Breen *et al*., (2010) but no further study was conducted using these genes. Such duplication events might be attributed to various factors, including transposable elements, epigenetic regulation, and ectopic recombination events (Gangadhar *et al.,* 2021; Liu *et al.,* 2021). Sequence analysis revealed that *BdGLP2-1* and *BdGLP3-4* possess the longest nucleotide sequences (680 bp each), while *BdGLP3-7* had the shortest sequences (530 bp each (Supplementary table S1). This variation extends to the encoded amino acid sequences where *BdGLP2-1* and *BdGLP3-4* possess the highest number of amino acids (226 each) and *BdGLP3-7* possess the fewest (173 amino acids) of them.

Protein sequence analysis

Multiple sequence alignment: Multiple sequence alignments of the BdGLPs were performed to identify their conserved regions and shared features (Fig. 1). Overall the sequences showed greater similarity with little variation suggesting their common origin mainly through duplication on specific loci. However, certain mutational gaps were observed at the start of the sequences particularly in BdGLP3-3 and BdGLP3-7. The highest mutation rate was observed in BdGLP2-1 which may be due to environmental constraints. Germin Motives 1, 2, and 3 (represented with yellow shades) represent the Germin box (shown with red box); the fundamental structural feature of the Germin-like protein. It is part of the basic domain having a consensus sequence of GxxxxHxHPxAxEh (where "x" represents amino acid while "h" is a hydrophobic one) (Dunwell *et al*., 2004). A peptide signal was found at the start of the sequence which plays an important role in various metabolic processes. GER motif 1 exists ≈24 to 25 amino acids, GER motif 2 ≈102 to 104 amino acids while GER motif 3 was found 148 to 150 amino acids away from the start of the N-end. The KGD motif was found before the GER motif 3, which is important for protein-protein interaction and helps plants cope with pathogenicity. In some cases, the KGD motif was mutated to KGE, as observed in OsGLPs or VvGLPs (Ilyas *et al*., 2020; 2022).

SUPPLEMENTARY DATA

Table S1. Description of the Germin-like protein genes in *Brachypodium distachyon* (*BdGLPs*).

S. No.	Plant name	Gene name	Acc. No.	Chr. no	Chr. Position	Bp	aa	Reference
1.	*Brachypodium distachyon*	*BdGLP2-1*	Bradi_2g21010v3	Chr no 2	2:18393515-18394672	681	226	Breen *et al*., 2010
2.	*Brachypodium distachyon*	*BdGLP3-1*	Bradi_3g15190v3	Chr no 3	3:13504258-13505260	678	225	Breen *et al*., 2010
3.	*Brachypodium distachyon*	*BdGLP3-2*	Bradi_3g15200v3	Chr no 3	3:13507473-13508475	678	225	Breen *et al*., 2010
4.	*Brachypodium distachyon*	*BdGLP3-3*	Bradi_3g15210v3	Chr no 3	3:13511002-13512125	675	224	Breen *et al*., 2010
5.	*Brachypodium distachyon*	*BdGLP3-4*	Bradi_3g15220v3	Chr no 3	3:13513339-13514481	681	226	Breen *et al*., 2010
6.	*Brachypodium distachyon*	*BdGLP3-5*	Bradi_3g15240v3	Chr no 3	3:13525866-13526904	678	225	Breen *et al*., 2010
7.	*Brachypodium distachyon*	*BdGLP3-6*	Bradi3g17310.1	Chr no 3	Not Found	?	?	Breen *et al*., 2011
8.	*Brachypodium distachyon*	*BdGLP3-7*	Bradi_3g17320v3	Chr no 3	3:15510771-15511415	522	173	Breen *et al*., 2010
9.	*Brachypodium distachyon*	*BdGLP3-8*	Bradi_3g17330v3	Chr no 3	3:15517276-15518205	678	225	Breen *et al*., 2010

Acc. No: Accession number, **Chr:** Chromosome, **bp:** Base pair, **aa:** Amino acids

The sequences were further searched for five possible conserved motives using the Multiple Em for Motif Elicitation (MEME) server followed by its scanning with Motif scan software to get further insights into their structure and possible function (Fig. 2). Overall, the proteins showed greater similarity. However, variation was observed in motives position, particularly in BdGLP2-1 and BdGLP3-7 which may be due to various evolutionary forces such as mutation or duplication, etc. Among the eight putative GLPs, five contained all five recorded motifs. Motif 1 (M1) and 2 were present in all proteins except *BdGLP3-7*. Similarly, all proteins except *Bd*GLP2-1 contained motifs 3 and 4 (M3, M4). In the same way, BdGLP3-4 lacks M5. Functional analysis with Motif scan revealed that Motifs 1, 2, and 3 are part of the cupin domain with a width ranging from 41 to 50 amino acids. The Cupin domain possesses a wide range of diverse functions against both biotic and abiotic stresses in diverse plant species (Ilyas *et al*., 2016b). No function was predicted for M4 while M5 possessed a phospho-casein kinase II (CK2) activity: a plant pleiotropic enzyme involved in plant growth, development, cell cycle regulation, hormone, and stress responses (DA Sol *et al.*, 2022). The presence of the same motives within members of the same protein family indicates that the homologs have partially remained conserved during the course of evolution (Ahmed *et al.*, 2023).

Physiochemical properties: Analysis of the BdGLPs' physicochemical Properties revealed significant variations in their molecular weight (M. Wt) (Table 1). BdGLP2-1 emerged as the heaviest protein with a M. Wt of about 24576.37 AMU (Atomic mass unit), while BdGLP3-7 possessed the lowest M. Wt (18570.55 amu). The isoelectric point (pI) also displayed variation, with BdGLP3-7 exhibiting the highest pI (9.10). Conversely, BdGLP3-1 and BdGLP3-3 shared the lowest pI (5.90) values suggesting that BdGLPs are both acidic as well as basic in nature indicating a high degree of variability in their amino acid composition (Ahmed *et al.*, 2024). Finally, the instability index (II) of BdGLPs ranged from 20.53 (*Bd*GLP3-4) to 25.79 (*Bd*GLP3-8). Interestingly, all the *Bd*GLPs displayed an II below 40, suggesting they might be stable within the cellular environment which is crucial for the normal functioning of the proteins. At the same time, Aliphatic Index (AI) analysis suggests potential thermostability of the BdGLPs. The AI ranged from 92.97 (BdGLP2-1) to 107.69 (BdGLP3-7) where the higher values indicate greater thermal stability (Ikai, 1980). In a similar study, GLPs of *Vitis vinifera* (*Vv*GLPs) exhibit thermostability with an AI value of 102.18 (Ilyas *et al*., 2022). Genes encoding BdGLP3-7, BdGLP3-2, BdGLP3-4, BdGLP3-5, and BdGLP3-1 possessed the highest AI values suggesting their potential for increased thermal stability. Conversely, lower AI values indicate protein flexibility in response to temperature changes due to the presence of aliphatic amino acids (Leucine, Isoleucine, Valine, and Alanine) which agrees with the fact that Ala, Val, and Leu were the most frequently occurring amino acids in these proteins (Fig. 3) Analysis of the Grand Average of Hydropathy (GRAVY) values revealed all BdGLPs as positive confirming their hydrophobic nature. Their GRAVY scores ranged from 0.111 (BdGLP2-1) to 0.303 (BdGLP3-7), with all the values exceeding zero. This result aligns with the previous findings in rice, where 86.95% (40) of the *Os*GLPs were identified as hydrophobic (Ilyas *et al.*, 2019). Conversely, lower GRAVY values indicate better interaction with water molecules.

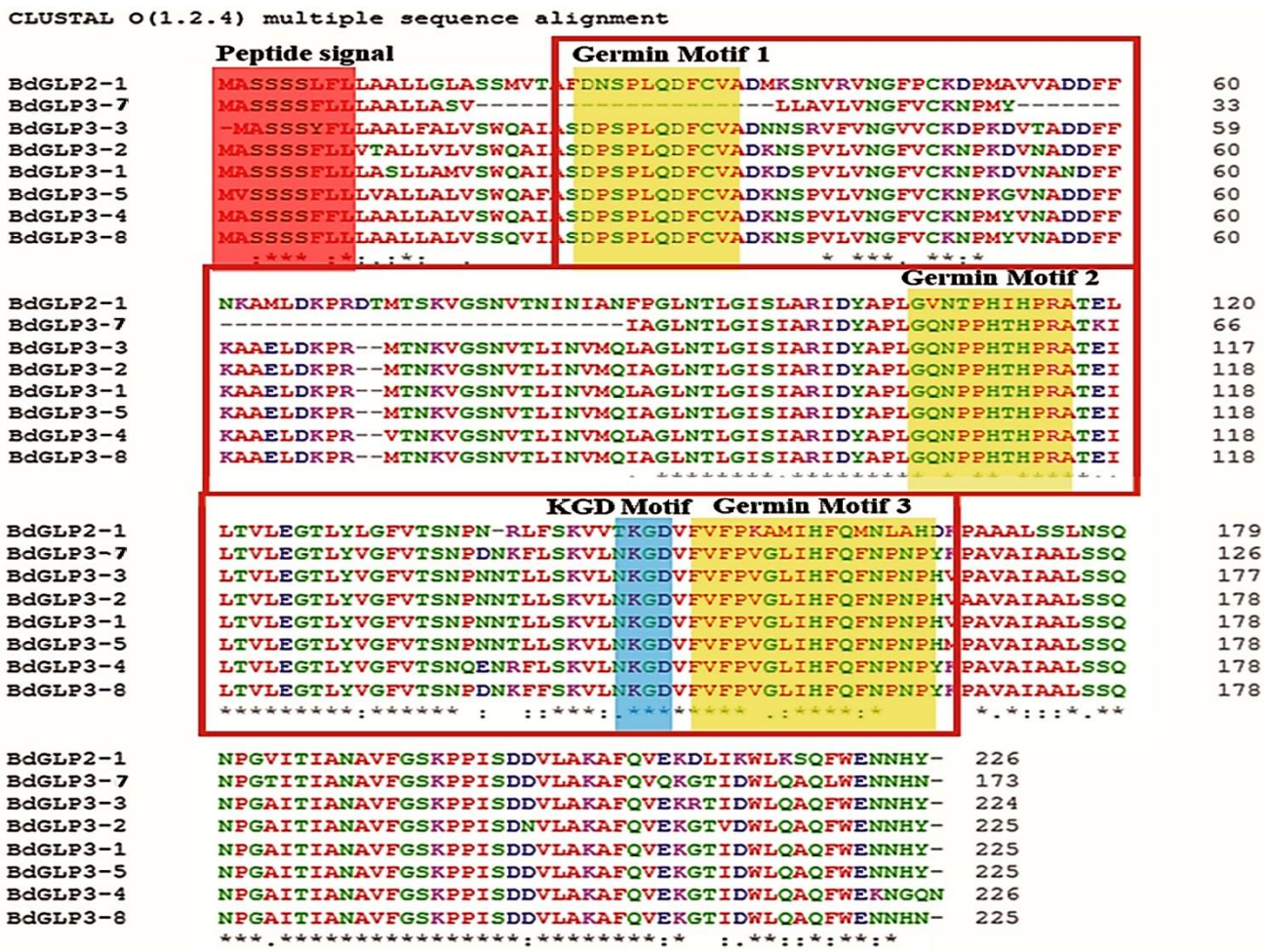
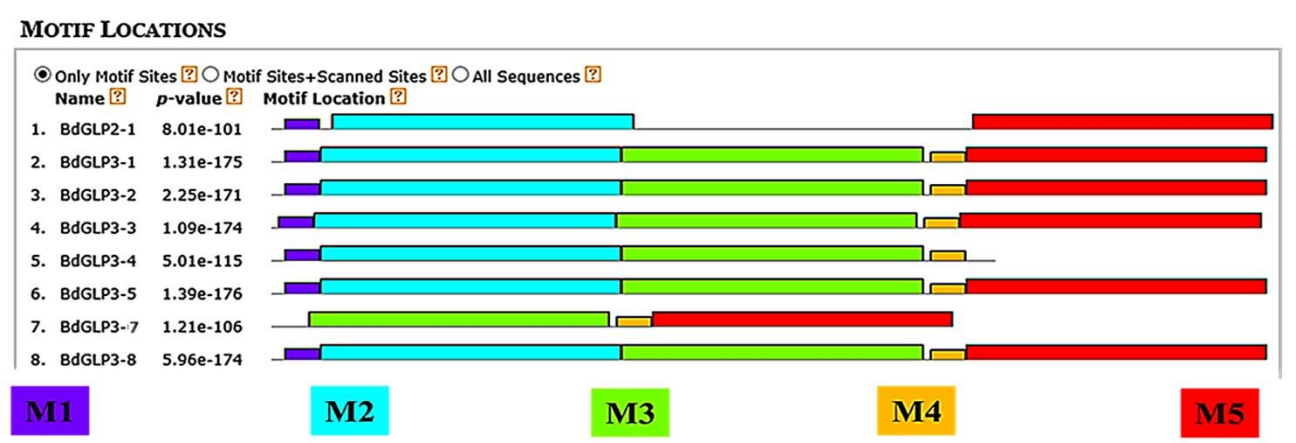


Fig. 1. Multiple sequence alignment of the Germin-like protein in *Brachypodium distachyon* (*BdGLPs)*. The analysis was accomplished using the Clustal Omega server (https://www.ebi.ac. uk/Tools/msa/clustalo/) of EMBL-EBI. Germin's motives are highlighted with a yellow shade. The peptide signal and KGD/KGE motives are highlighted with red and blue shades respectively. Germin Box is represented with red box.



Motif	Motives (Best possible match)	Width (aa)	Site No	Motif scan
1	MQIAGLNTLGISIARIDYAPLGQNPPHTHPRATEILTVLEGTLYVGFVTS	50	8	Cupin
2	TIANAVFGSKPPISDDVLAKAFQVEKGTIDWLQAQFWENN	41	8	Cupin
3	LSKVLNKGDVFVFPVGLIHFQFNPNPHMPAVAIAALSSQN	41	8	Cupin
4	SDPSPLQDFCVADKNSPVLVNGFVCKNPMYVNADDFFKAAELDKPRMTNK	50	7	N/A
5	SSSSFFLLAALLALVSWQAIA	21	8	Phospho CKII

Fig. 2. Domain analysis of the Germin-like protein in *Brachypodium distachyon* (BdGLPs*)*. Multiple Em for Motif Elicitation (MEME) (http://meme.nbcr.net/meme/cgi-bin/meme.cgi) software was employed to predict 05 possible motives in the *BdGLPs* while Motif scan (https://myhits.sib.swiss/cgi-bin/motifscan/software) was used for functional analysis. **M:** Motif, **Cupin:** Cupin domain, **N/A:** Not available, **Phospho:** Phosphorylates, **CKII:** Casein kinase 2 (CK2).

Table 1. Physiochemical properties of the Germin-like protein in *Brachypodium distachyon* (*BdGLPs*).

Gene name	M. Wt	PI	+R	-R	EC	II	AI	Gravy	Pfam domain	Subcell prediction (Cello)	Subcell prediction (Psort)	N-site	P-site
BdGLP2-1	24576.37	7.79	19	18	15595	24.5	92.97	0.111	Cupin	Periplasmic	Extracellular	1	7
BdGLP3-1	24227.77	5.9	14	17	0.871	25.48	97.96	0.134	Cupin	Periplasmic	Extracellular	2	6
BdGLP3-2	24196.74	6.17	14	16	21045	21.49	100.09	0.172	Cupin	Periplasmic	Extracellular	2	5
BdGLP3-3	24308.79	5.9	15	18	22460	23.83	96.65	0.116	Cupin	Periplasmic	Extracellular	3	6
BdGLP3-4	24532.26	6.63	16	17	22460	20.53	97.1	0.178	Cupin	Periplasmic	Extracellular	1	7
BdGLP3-5	24355.00	6.17	14	16	20970	24.23	97.07	0.167	Cupin	Periplasmic	Extracellular	2	6
BdGLP3-7	18570.55	9.1	11	8	16960	23.53	107.69	0.303	Cupin	Periplasmic	Chloroplast	NA	5
BdGLP3-8	23765.22	6.04	15	17	16960	25.79	93.47	0.077	Cupin	Periplasmic	Extracellular	1	7

PM: Plasma membrane**, M. Wt:** Molecular weight, **EC:** Extinction coefficient**, II:** Instability index, **GRAVY:** Grand average of hydropathicity, **AI:** Aliphatic index**, PI:** Isoelectric point, **+R:** Positive charge residues, **-R**: Negative charge residues, **p-site:** Phosphorylation site, **N-site:** N glycosylation site, **NA:** Not available

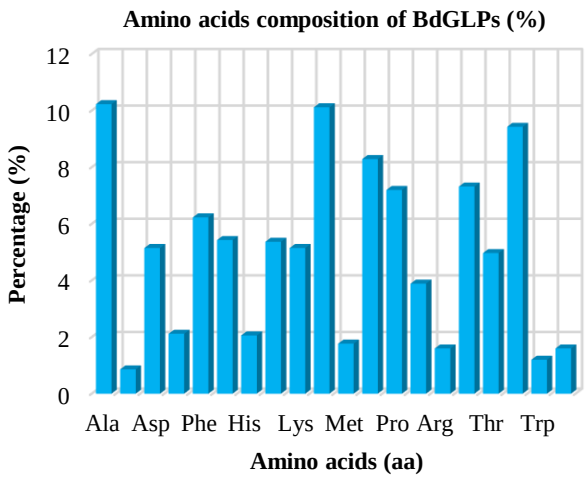


Fig. 3. Amino acids composition (%) of the Germin-like proteins in *Brachypodium distachyon* (BdGLPs). The analysis was conducted using the Molecular and evolutionary Genetic analysis tool (MEGA11) and Excel 365.

Subcellular localization & N & P Sites analysis: The Subcellular localization study revealed that most of the *BdGLPs* showed expression in the extracellular (~87.5% proteins) or periplasmic (100% proteins) regions suggesting their functional similarity due to similar evolutionary history. It also suggested their pivotal role in these regions by showing response to various external cellular stimuli (Bernier and Berna, 2001; Breen and Bellgard, 2010; Manosalva *et al.,* 2009). However, BdGLP3-7 showed expression within the chloroplast (12.5%) showing its role in photosynthesis, indicating that BdGLPs may participate in diverse biochemical and physiological processes within the cell organelles.

N-glycosylation sites (N-sites) analysis revealed variations in their number (1-3) and location (between 29-42 amino acids) flanking the right side of the QDFCVAD motif. This variation suggests functional diversity among *Bd*GLPs, potentially influencing their enzymatic activities, stability, and structural conformations. Notably, BdGLP3-3 possessed the highest number of N-sites, whereas BdGLP3-7 lacked such sites. Similarly, the number of potential phosphorylation sites (P-sites) varied (5-7) across BdGLPs. The highest being (7) was observed in the BdGLP2-1, BdGLP3-4, and BdGLP3-8, while BdGLP3-2 and BdGLP3-7 displayed the lowest number (5) of them. Protein phosphorylation, particularly at tyrosine, serine, and threonine residues, plays a crucial role in regulating various cellular processes like apoptosis, differentiation, proliferation, and metabolism. Therefore, the presence of both P-sites and N-sites in BdGLPs suggests their potential involvement in a diverse range of cellular processes.

3D Structures analysis: Recent advancements in bioinformatics tools have revolutionized the prediction of protein structure by offering a cost-effective alternative to traditional experimental techniques like nuclear magnetic resonance (NMR) spectroscopy and X-ray crystallography. In the current study, various bioinformatic tools were utilized to generate three candidate three-dimensional (3D) structural models for each BdGLP (Supplementary Fig. S1). The current *In-silico* approach relies on aligning & blasting the target BdGLP amino acid sequence with the known protein structures deposited in public databases. The quality of the generated models was then actively assessed using two key parameters: Qmean Z-score and Global Model Quality Estimation (GMQE). High values for both matrices relate to increased reliability and accuracy of the predicted protein structures. The authenticity and quality of the BdGLPs were confirmed by Ramachandran plot analysis (RPA). More than 80% of the RP (Ramachandran plot) values in the favored region indicate protein stability regarding their functions whereas values below 80% are considered unreliable or characterize the protein as non-functional. In the current result, all BdGLPs showed RP values which range from 85 to 90% confirming them as functional proteins. These findings provide additional validation for the high-quality nature of their predicted protein structures. The analysis revealed that most of the proteins (BdGLP2-1, BdGLP3-1, BdGLP3-3, BdGLP3-4, BdGLP3-5, and BdGLP3-8) showed similarity in structure which showed their similar evolutionary history and function. However, some variations related to protein cavities and overall structures were noted in the BdGLP3-4 and BdGLP3-7 which may be due to evolutionary changes over time causing variation in other properties as well due to environmental effects. Despite similarities in the structures of BdGLPs, variations in the distribution of protein cavities were observed. This variation may be the likely reason behind the functional diversity of these structurally similar proteins because the interaction of a substrate is hugely dependent on the surface and cavities of the protein. It is pertinent to note that structural compatibility is a prerequisite for molecular interactions such as protein-ligand, protein-DNA, or protein-protein interaction (Berezovsky *et al.,* 2007).

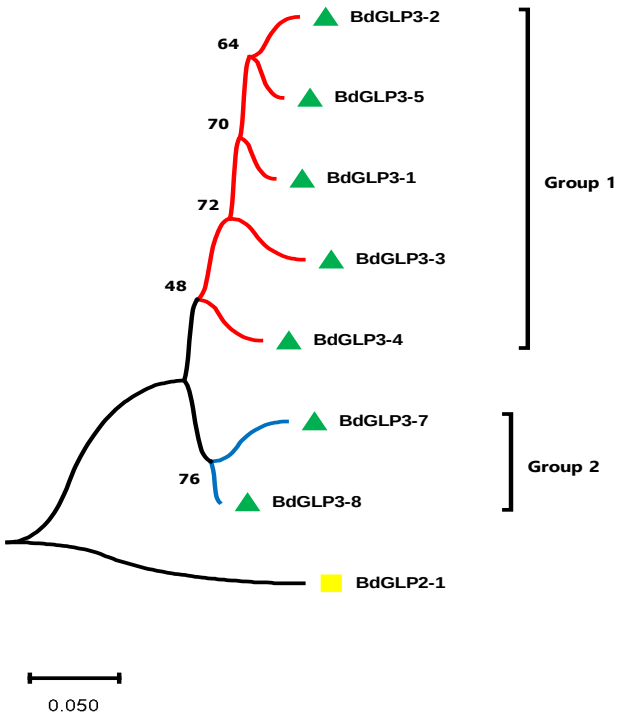


Fig. 4. Phylogenetic analysis of the Germin-like protein in *Brachypodium distachyon* (BdGLPs). The evolutionary history was inferred using the Neighbor-Joining method. The optimal tree is shown. The percentage of replicate trees in which the associated taxa clustered together in the bootstrap test (500 replicates) are 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 Poisson correction method and are in the units of the number of amino acid substitutions per site. This analysis involved 8 amino acid sequences. All ambiguous positions were removed for each sequence pair (pairwise deletion option). There was a total of 228 positions in the final dataset. Genes on the same chromosome were labeled with distinct shapes. Phylogenetic analyses were conducted in Molecular Evolutionary Genetics analysis tool (MEGA ver. 11).

Phylogenetic analysis: Phylogenetic analysis unveiled a limited genetic diversity among BdGLPs as shown by the bootstrap's values (Fig. 4). The proteins were divided into two groups. Group 1 consists of BdGLP3-1, BdGLP3-2, BdGLP3-3, BdGLP3-4, and BdGLP3-5 which showed a close relationship as supported by their presence on the same chromosome. Similarly, in Group 2 BdGLP3-7 and BdGLP3-8 were found constituting a separate group from the rest of the BdGLPs on chromosome 3 which may be due to mutation and evolutionary changes with time. However, BdGLP2-1 forms a side branch suggesting its distinct position in the tree, possibly due to its presence on a separate chromosome. Based on these bootstrap values the closest relationship was observed between BdGLP3-7 and BdGLP-8 suggesting their close occurrence on chromosome 3, caused by duplication. However, a distant relationship between BdGLP2-1 and chromosome 3 GLPs suggests variation with time due to mutation, duplication, or transposon activity etc. due to environmental stresses. Divergence via duplication is common in GLPs and has been previously observed in many families of GLPs such as rice, maize, and grapes (Ilyas *et al*., 2020; 2022; 2023).

Functional analysis: Functional analysis revealed a more preserved and similar functional pattern (Table 2). However, BdGLP2-1 has more diverse functional properties related to NADH dehydrogenase activity which plays an important role in energy production mostly in mitochondria. In chloroplast, it is a multiple-subunit complex in the thylakoid membranes mediating cyclic electron transport and is essential for plant growth and development during stress periods (Ma *et al*., 2021). However, GLPs located on chromosome 3 have similar functional properties which are related to WPP (tryptophan-proline-proline) domain-containing protein-related activity which is a novel protein and acts by helping other proteins to interact with the nuclear envelope and carries out their transportation thus providing help in disease resistance (Zhang *et al*., 2023). It was observed that BdGLP2-1 has a more diverse and specific function due to mutation while GLPs located on chromosome 3 have a similar function which may be due to duplication. However, it's noteworthy to mention, that most of the time GLPs located on the same chromosome act in co-regulated manure or act as QTL (Quantitative trait loci) to protect against various diseases such as those previously reported in rice (Davidson *et al*., 2010) and barley (Himmelbach *et al*., 2010), etc.

Table 2. Functional analysis of the Germin-like protein in *Brachypodium distachyon* (*BdGLPs)*.

S. No.	Gene name	Function prediction (GO)
1.	*BdGLP*2-1	Mitochondrial electron transport, NADH to ubiquinone, NADH dehydrogenase (ubiquinone) activity, Mitochondrial respiratory chain complex-I related activity
2.	*BdGLP*3-1	WPP domain-containing protein-related activities
3.	*BdGLP*3-2	WPP domain-containing protein-related activities
4.	*BdGLP*3-3	WPP domain-containing protein-related activities
5.	*BdGLP*3-4	WPP domain-containing protein-related activities
6.	*BdGLP*3-5	WPP domain-containing protein-related activities
7.	*BdGLP*3-7	WPP domain-containing protein-related activities
8.	*BdGLP*3-8	WPP domain-containing protein-related activities

The analysis was conducted using Search Tool for the Retrieval of Interacting Genes/Proteins (STRING ver. 10) available at https://string-db.org/. **S. No:** Serial number, **GO:** Gene ontology, **WPP:** Tryptophan-proline-proline domain, **NADH:** NADH dehydrogenase activity

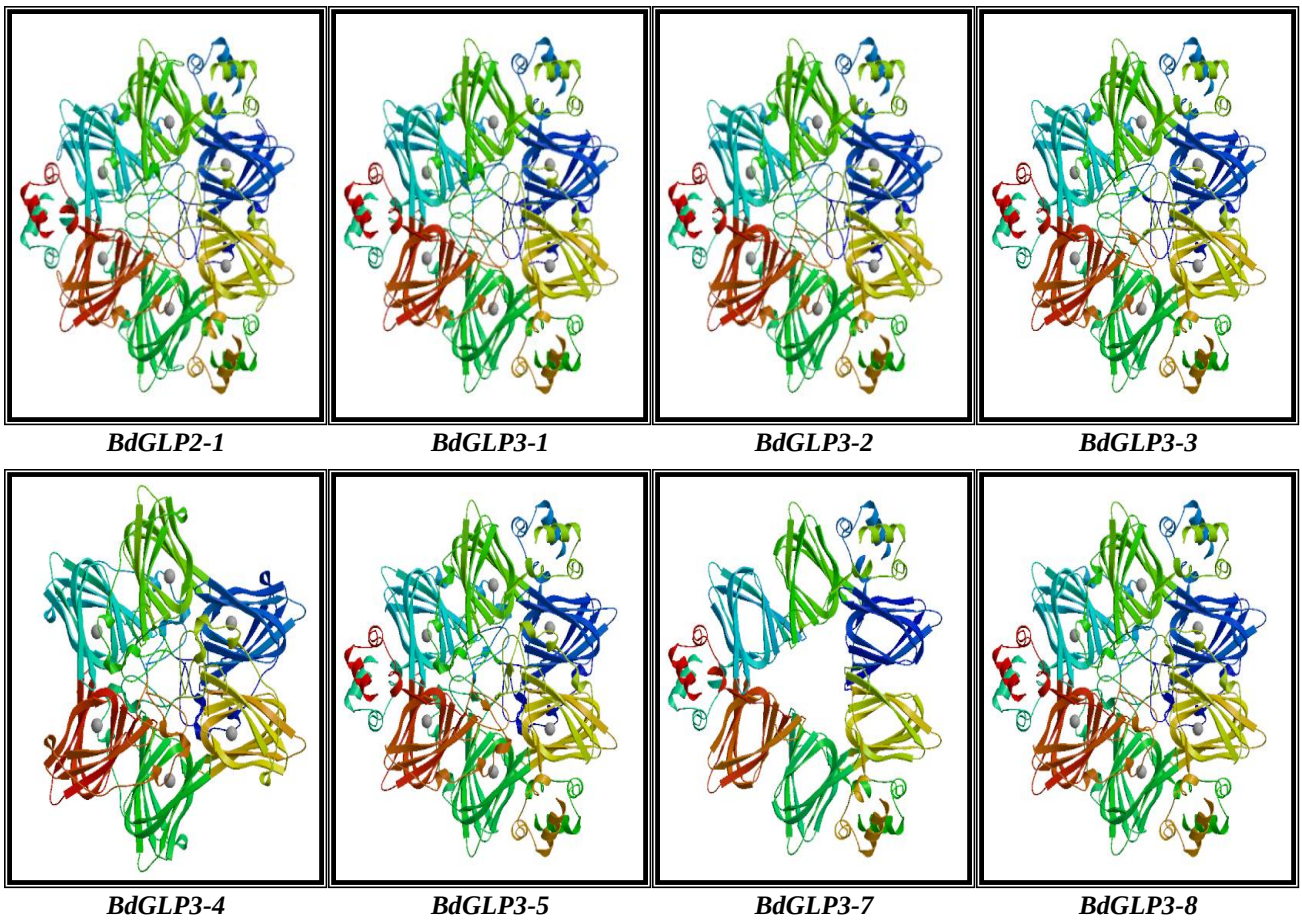


Fig. S1. 3D structural analysis of the Germin-like protein in *Brachypodium distachyon* (BdGLPs). Swiss modeling server (https://swissmodel.expasy.org/) was used to predict the 3D structure of the BdGLPs protein. The structure was authenticated using the Ramachandran plot analysis server (https://swift.cmbi.umcn.nl/servers/html/ramaplot.html).

Conclusions and Recommendations

The study revealed that *BdGLPs* originate via multiple gene duplication events on chromosome 3 due to which they showed a high degree of similarity in physicochemical properties, subcellular localization, phosphorylation & glycosylation sites occurrence, and 3D structures. However, *BdGLP2-1* exhibits distinct physiochemical properties, phylogeny, and functions (NADH Dehydrogenase activity) as compared to other GLPs which may be due to its presence on a separate chromosome (2) caused by mutation etc. The findings suggest that structurally and functionally similar *BdGLPs* may play a crucial role in plant defense against various biotic and abiotic stresses by performing NADH Dehydrogenase or WPP domain-related activities. The identified properties of these proteins hold potential applications in future research endeavors. Furthermore, experimental analyses are needed to confirm enzymatic activities and specific functions of these individual *BdGLPs* genes.

References

Ahmed, J., Y. Sajjad, A. Latif, M.S. Lodhi, M. Huzafa, C. Situ and A. Hassan. 2024. Genome-wide identification and characterization of wall-associated kinases, molecular docking, and polysaccharide elicitation of monoterpenoid indole alkaloids in micro-propagated Catharanthus roseus. *J. Plant Res*., 137(1): 125-142.

Ahmed, J., Y. Sajjad, M.K. Gatasheh, K.E. Ibrahim, M. Huzafa, S.A. Khan, A.M. Abbasi and A. Hassan. 2023. Genome-wide identification of NAC transcription factors and regulation of monoterpenoid indole alkaloid biosynthesis in *Catharanthus roseus*. *Front. Plant Sci*., 14: 1286584.

Artimo, P., M. Jonnalagedda, K. Arnold, D. Baratin, G. Csardi, E. De Castro, S. Duvaud, V. Flegel, A. Fortier and E. Gasteiger. 2012. ExPASy: SIB bioinformatics resource portal. *Nucl. Acids Res.,* 40: W597-W603.

Bailey, T.L., J. Johnson, C.E. Grant and W.S. Noble. 2015. The MEME suite. *Nucl. Acids Res.*, 43(W1): W39-W49.

Barman, A.R. and J. Banerjee. 2015. Versatility of germin-like proteins in their sequences, expressions, and functions. *Func. Integ. Genom.,* 15: 533-48.

Berezovsky, I.N., K.B. Zeldovich and E.I. Shakhnovich. 2007. Positive and negative design in stability and thermal adaptation of natural proteins. *PLOS Comput. Biol.,* 3: e52.

Bernier, F. and A. Berna. 2001. Germins and germin-like proteins: plant do-all proteins. But what do they do exactly? *Plant Physiol. Biochem.,* 39: 545-554.

Breen, J. and M. Bellgard. 2010. Germin-like proteins (*GLPs*) in cereal genomes: gene clustering and dynamic roles in plant defense. *Func. Integ. Genom.,* 10: 463-476.

Cunningham, F., J. E. Allen, J. Allen, J. Alvarez-Jarreta, M.R. Amode, I.M. Armean and P. Flicek. 2022. Ensembl 2022. *Nucl. Acids Res.*, 50(D1): D988-D995.

DA SOL, L., S. Lee, C. Kim, D. Kim, K.P. Kim and C. Yoo. 2022. The casein kinase 2 inhibitor CX-4945 promotes cholangiocarcinoma cell death through PLK1. *Anticancer Res.,* 42: 3435-3443.

Davidson, R.M., P.M. Manosalva, J. Snelling, M. Bruce, H. Leung and J.E. Leach. 2010. Rice germin-like proteins: Allelic diversity and relationships to early stress responses. *Rice*, 3: 43-55.

Dunwell J.M., A. Purvis and S. Khuri. 2004. Cupins: the most functionally diverse protein superfamily? *Phytochem.,* 65(1): 7-17.

Dunwell, J.M., J.G. Gibbings, T. Mahmood and S.S. Naqvi. 2008. Germin and germin-like proteins: evolution, structure, and function. *Crit. Rev. Plant Sci.,* 27: 342-375.

Fu, J.Y., X. Wang, C. Mao, T.F. Cheng, H. Chen and F. Yang. 2018. Identification and functional analysis of germin-like protein gene family in tea plant (*Camellia sinensis*). *Sci. Hort*., 234: 166-175.

Gangadhar, B.H., R.K. Mishra, S. Kappachery, V. Baskar, J. Venkatesh, A. Nookaraju and M. Thiruvengadam. 2021. Enhanced thermo-tolerance in transgenic potato (*Solanum tuberosum* L.) overexpressing hydrogen peroxide-producing germin-like protein (GLP). *Genomics,* 113: 3224-3234.

Gasteiger, E., A. Gattiker, C. Hoogland, I. Ivanyi, R.D. Appel and A. Bairoch. 2003. ExPASy: The proteomics server for in-depth protein knowledge and analysis. *Nucl. Acids Res.*, 31(13): 3784-3788.

Giarola, V., P. Hen, S.J. Dulitz, M. König, S. Manduzio and D. Bartels. 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*, 252: 84.

He, Z.D., M.L. Gao, D.W.M. Leung, X.Y. Yan, L. Chen, X.X. Peng and E.E. Liu. 2021. The rice germin-like protein *OsGLP1* participates in acclimation to UV-B radiation. *Plant Physiol.*, 186: 1254-1268.

Himmelbach, A., L. Liu, U. Zierold, L. Altschmied, H. Maucher, F. Beier, D. Müller, G. Hensel, A. Heise, A. Schützendübel and J. Kumlehn. 2010. Promoters of the barley germin-like *GER4* gene cluster enable strong transgene expression in response to pathogen attack. *The Plant Cell.,* 22(3): pp. 937-952.

Hollebeek, T., T.S. Ho and H. Rabitz. 1999. Constructing multidimensional molecular potential energy surfaces from ab initio data. *Ann. Rev. Physic. Chem.,* 50: 537-570.

Horton, P., K.J. Park, T. Obayashi, N. Fujita, H. Harada, C.J. Adams-Collier and K. Nakai. 2007. WoLF PSORT: Protein localization predictor. *Nucl. Acids Res.*, 35(2): 585-587.

Ikai, A. 1980. Thermostability and aliphatic index of globular proteins. *J. Biochem.,* 88: 1895-1898.

Ilyas, M., A. Rasheed and T. Mahmood. 2016b. Functional characterization of germin and germin-like protein genes in various plant species using transgenic approaches. *Biotech. Lett.,* 38: 1405-1421.

Ilyas, M., I. Ali, D.N. Binjawhar, S. Ullah, S.M. Eldin, B. Ali and T. Mahmood. 2023. Molecular characterization of Germin-like protein genes in *Zea mays* (*ZmGLPs*) using various in silico approaches. *ACS Omega*, 8(18): 16327-16344.

Ilyas, M., M. Irfan, T. Mahmood, H. Hussain, L.-ur-Rehman, I. Naeem and R.-ur-Rahman. 2020. Analysis of germin-like protein genes (*OsGLPS*) family in rice using various *In silico* approaches. *Curr. Bioinfor.*, 15: 17-33.

Ilyas, M., S.M.S. Naqvi and T. Mahmood. 2016a. *In silico* analysis of transcription factor binding sites in promoters of germin-like protein genes in rice. *Arch. Biol. Sci.,* 68: 863-876.

Ilyas, M., W. Akhtar, S. Rehman, S.M.S. Naqvi and T. Mahmood. 2019. Functional characterization of the rice root Germin-like protein gene-1 (*OsRGLP1*) promoter in *Nicotiana tabacum*. *3 Biotech.,* 9: 130.

Ilyas, M.A., Rahman, N.H. Khan, M. Haroon, H. Hussain, L. Rehman, M. Alam, A. Rauf, W. Waggas and S.D. Bawazeer. 2022. Analysis of Germin-like protein genes family in *Vitis vinifera* (*VvGLPs*) using various in silico approaches. *Braz. J. Biol.,* 84: e256732.

Laskowski, R.A., N. Furnham and J.M. Thornton. 2013. The Ramachandran plot and protein structure validation. *In Biomolecular forms and functions:* a celebration of 50 years of the ramachandran map (pp. 62-75).

Li, L., X. Xu, C. Chen and Z. Shen. 2016. Genome-wide characterization and expression analysis of the germin-like protein family in rice and *Arabidopsis*. *Int. J. Mol. Sci.*, 17(10): 1622.

Lin, C., W. Chen, C. Qiu, Y. Wu, S. Krishnan and Q. Zou. 2014. LibD3C: Ensemble classifiers with a clustering and dynamic selection strategy. *Neurocomputing.,* 123: 424-435.

Liu, B., F. Liu, X. Wang, J. Chen, L. Fang and K.C. Chou. 2015. Pse-in-One: A web server for generating various modes of pseudo components of DNA, RNA, and protein sequences. *Nucl. Acids Res.,* 43: W65-W71.

Liu, B., H. Wu, D. Zhang, X. Wang and K.C. Chou. 2017. Pse-Analysis: a Python package for DNA/RNA and protein/peptide sequence analysis based on pseudo components and kernel methods. *Oncotarget,* 8: 13338.

Liu, C., N.J. Talbot and X.L. Chen. 2021. Protein glycosylation during infection by plant pathogenic fungi. *New Phytol.,* 230: 1329-1335.

Liu, F., N. Tian and X.L. Sun. 2020. Genome-wide identification and expression analysis of banana GLP gene family. *J. Hort.*, 47(10): 1930-46.

Lu, M., Y.P. Han, J.G. Gao, X.J. Wang and W.B. Li. 2010. Identification and analysis of the germin-like gene family in soybean. *BMC Genom.*, 11: 620.

Ma, M., Y. Liu, C. Bai and J.W.H. Yong. 2021. The significance of chloroplast NAD (P) H dehydrogenase complex and its dependent cyclic electron transport in photosynthesis. *Front. Plant Sci.*, 12: 661863.

Mahajan, M. and G.R. Rathod. 2018. *In silico* characterization and molecular modeling of GLP-2b protein from (*Triticum aestivum* L.). *Indian Res. J. Genet. Biotech.*, 10: 120-126.

Manosalva, P.M., R.M. Davidson, B. Liu, X. Zhu, S.H. Hulbert, H. Leung and J.E. Leach. 2009. A germin-like protein gene family functions as a complex quantitative trait locus conferring broad-spectrum disease resistance in rice. *Plant Physiol.,* 149: 286-296.

Nagashima, Y., A. von Schaewen and H. Koiwa. 2018. Function of N-glycosylation in plants. *Plant Sci.,* 274: 70-79.

Pei, Y., X. Li, Y. Zhu, X. Ge, Y. Sun, N. Liu, Y. Jia, F. Li and Y. Hou. 2023. *GHABP19*, a novel germin-like protein from *Gossypium hirsutum*, plays an important role in the regulation of resistance to *Verticillium* and *Fusarium* wilt pathogens. *Front. Plant Sci.*, 10: 789-801.

Rietz, S., F.E. Bernsdorff and D. Cai. 2012. Members of the germin-like protein family in *Brassica napus* are candidates for the initiation of an oxidative burst that impedes pathogenesis of *Sclerotinia sclerotiorum*. *J. Exp. Bot.*, 63(15): 5507-5519.

Sato, Y., B.A. Antonio, N. Namiki, H. Takehisa, H. Minami, K. Kamatsuki, K. Sugimoto, Y. Shimizu, H. Hirochika and Y. Nagamura. 2010. RiceXPro: A platform for monitoring gene expression in japonica rice grown under natural field conditions. *Nucl. Acids Res.,* 39: D1141-D1148.

Szklarczyk, D., A. Franceschini, S. Wyder, K. Forslund, D. Heller, J. Huerta-Cepas and C. Von Mering. 2015. STRING v10: Protein-protein interaction networks, integrated over the tree of life. *Nucl. Acids Res.*, 43(D1): D447-D452.

Tong, X., T. Nagy, J.Y. Reyes, M. Germann, M. Meuwly and S. Willitsch. 2012. State-selected ion–molecule reactions with Coulomb-crystallized molecular ions in traps. *Chem. Phys. Lett.,* 547: 1-8.

Wang, T.X. Chen, F. Zhu, H. Li, L. Li, Q. Yang, X. Chi, S. Yu and X. Liang. 2013. Characterization of peanut germin-like proteins, *AhGLPs* in plant development and defense. *PLoS One*., 8: e61722.

Wang, X.H. Zhang, Y. Gao, G. Sun, W. Zhang and L. Qiu. 2014. A comprehensive analysis of the Cupin gene family in soybean (*Glycine max*). *PLoS One*., 9: e110092.

Yuan, B.Y. Yang, P. Fan, J. Liu, H. Xing, Y. Liu and D. Feng. 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 Mol. Biol. Rep.*, 39: 821-832.

Zaynab, M.J. Peng, Y. Sharif, M. Fatima, M. Albaqami, R. Al-Yahyai, A. Raza, K. Khan, S.S. Alotaibi, I.A. Alaraidh, H.O. Shaikhaldein and S. Li. 2022. Genome-wide identification and expression profiling of germin-like proteins reveals their role in regulating abiotic stress response in potatoes. *Front. Plant Sci.*, 12: 3450.

Zeng, X.D. Li, Y. Lv, Y. Lu, L. Mei, D. Zhou, D. Chen, F. Xie, H. Lin and Y. Li. 2023. A germin-like protein GLP1 of legumes mediates symbiotic nodulation by interacting with an outer membrane protein of rhizobia. *Microbiol. Spect*., e03350-e03322.

Zhang, H., Y. Liu, X. Zhang, W. Ji and Z. Kang. 2023. A necessary considering factor for breeding: Growth-defense tradeoff in plants. *Stress Biol*., 3(1): p.6.

Zhang, Y., X. Wang, X. Chang, M. Sun, Y. Zhang, W. Li and Y. Li. 2018. Overexpression of germin-like protein *GmGLP10* enhances resistance to *Sclerotinia sclerotiorum* in transgenic tobacco. *Biochem. Biophys. Res. Comm.*, 497: 160-166.

(Received for publication 18 March 2024)