Arch Biol Sci. 2016;68(4):863-876 DOI:10.2298/ABS151116076I

IN SILICO ANALYSIS OF TRANSCRIPTION FACTOR BINDING SITES IN PROMOTERS OF GERMIN-LIKE PROTEIN GENES IN RICE

Muhammad Ilyas[1], Syed Muhammad Saqlan Naqvi[2], Tariq Mahmood[1,*]

[1] *Plant Biochemistry and Molecular Biology Lab, Department of Plant Sciences, Quaid-i-Azam University*, Islamabad, 45320, Pakistan

[2] *Department of Biochemistry, PMAS Arid Agriculture University*, Rawalpindi, Pakistan

*__Corresponding author__: tmahmood.qau@gmail.com; tmahmood@qau.edu.pk

Received: November 16, 2015; **Revised**: January 12, 2016; **Accepted**: January 19, 2016; **Published online**: November 16, 2015

Abstract: Germins (GERs) and germin-like proteins (GLPs) play important roles in responses to various stresses; however, their function is still not fully understood. Significant insight into their function can be obtained by analyzing their promoters. In the present study, the 5' upstream promoters (1000 bp) of 43 Asian rice (*Oryza sativa* var. Japonica) GLP genes were retrieved from the Plant Ensemble, based on the Rice Annotation Project database (RAP-DB). Phylogenetic analysis via MEGA6 showed a narrow genetic background (0.2%) with a Tajima neutrality value (π) of 0.69. Overall, 4234 transcription factor (TF) binding sites (TFBSs) were found on chromosomes 1, 2, 3, 4, 5, 8, 9, 11 and 12 via "MatInspector" from 90 different TF families using a total of 444. Common TFs and DiAlign analyses showed that *Arabidopsis* homeobox protein (AHBP), MYB-like proteins (MYBL) and vertebrate TATA-box-binding protein (VTBP) were the most abundant, common and evolutionarily conserved elements in the upstream region from 0 to -800. Finding their mutual interaction via Farmworker analysis uncovered three new *cis*-regulatory modules (VTBP_VTBP, MYBS_MYBS, and AHBP_VTBP), which appear to be decisive for *Os*GLPs regulation. *In silico* functional analysis via ModelInspector revealed 77 *cis*-regulatory modules, each comprised of two elements, among which DOFF_OPAQ_03 and GTBX_MYCL_01 were the most frequent and mostly found on chromosome 8 and 12, indicating that the combinatorial interaction of these elements has a fundamental role in various biological processes. The study revealed the importance of these elements in regulating the expression of *Os*GLPs that will help in predicting the role of these genes in various stresses and can have application in biotechnology.

Key words: TFBS; *in silico*; rice; germin-like proteins (GLPs); promoter

INTRODUCTION

Germin (GER) was initially identified in the wheat embryo as a germination-specific marker [1] and later recognized as an oxalate oxidase. Proteins with an average similarity of 50% with GER were referred to as germin-like proteins (GLPs). GERs and GLPs constitute a diverse and ubiquitous families of plant glycoproteins known as the cupin superfamily which is involved in many developmental and stress related processes [2]. They possess a single or combination of enzymatic activities, including oxalate oxidase (OXO), superoxide dismutase (SOD), ADP glucose pyrophosphatase/phosphodiesterase (AGPPase) and polyphenol oxidase (PPO), which either act as structural proteins or participate in signal transduction through their receptor function [3]. A GLP from wheat leaf apoplast was reported for its ability to inhibit serine protease [4]. The roles of GERs and GLPs in the development of leaf, root, fruit, seed and floral senescence, defense against various biotic (bacteria, viruses, fungi, insects, nematodes, parasites), abiotic (salinity, drought, cold, heat, metal, nutritional) and physical stresses, have been validated [3,5,6], but their functions are not fully understood. However, modern bioinformatics tools provide an opportunity to obtain insight into their functions by analyzing important molecular components that control their spatiotemporal regulation.

In this context, promoter analysis is an important step towards an improved understanding of gene functioning and regulation and is considered a prerequisite for the development of resilient crops through genetic modification. Plant promoters that direct high level of gene expression induced by various stresses are critical

for the application of crop biotechnology [7]. Further, these promoters can be used for achieving tissue-specific expression against various stresses. The protein binding sites of promoter and corresponding transcriptional factors (TFs) are crucial for transcription and regulation [8]. Accurate spatiotemporal regulation of gene expression is vital for developmental and environmental adaptation of an organism, which is in large part accomplished by *cis*-elements acting as binding sites for TFs [9]. Analysis of these individual elements in the promoter and their combinatorial effects can improve our understanding of gene expression. Previously, numerous databases and softwares have been used for *in silico* promoter analysis to predict its role in gene regulation against various stresses. Chawade et al. [10] developed a putative cold acclimation network in *Arabidopsis* using microarray data, known promoter-binding sites and corresponding TFs. Similarly, analysis of *At*CHS7, *At*CHS8 [11] and sucrose transporter gene promoter families of *Arabidopsis* and rice [12] were performed to investigate potential TFBSs using PLANT CARE, PLACE and MatInspector (a Genomatix software suite). A similar approach was adapted for the *Os*Gα subunit (RGA1) [13] and *AtPrx* gene (*Arabidopsis thaliana* peroxidases) promoters. Similarly, using information about putative TFBSs, the role of the 276-bp promoter region in tissue-specific expression and development was predicted and verified for *HvGERB* and *HvGERF* gene promoters [14]. Likewise, the roles of *Zm*GLP1 and *Eg*GLP promoters in the control of circadian rhythm-oscillated pattern [15,16], *Pc*Ger1 in various hormonal stresses [17], *Ta*GLP3 in powdery mildew [18] and *Hv*Ger4c and *At*GLP13 in pathogenicity [19,20] were first predicted by TFBSs analysis using various bioinformatics tools and subsequently verified. Similarly, due to the presence of seed-specific TFBSs, the *BnGLP* gene promoter was used to direct and enhance the accumulation of omega-3 long chain molecules by achieving seed-specific expression in transgenic *Arabidopsis* [21].

In rice, *in silico* analysis of *Os*RGLP1 [22], *Os*-RGLP2 [23] and 52GLP gene promoters from various plant species including rice [24,25] has predicted the existence of numerous TFBSs that participate in responses to wounding, dehydration, light responsiveness, dark-induced senescence, stresses (pathogen and salt), pollen-specific expression, plant growth regulators and elements related to seed storage proteins, etc., with the roles of *Os*RGLP2 in the response to wounding, dehydration stress and pathogenicity confirmed [26,27]. Thus far, no comprehensive study has been conducted on rice GLPs promoters, TFBSs and their putative roles in various processes that could predict their functioning. In view of the importance of promoter analysis and its role in the functional predictability of genes, the current study was designed to analyze all monocupin GLP gene promoters of Asian rice (*Oryza sativa*. var Japonica) found in the Ensemble database(s) on 30 October 2015, with the aim of identifying the TFBSs in these gene regions and predicting GLP gene functions in rice by applying appropriate bioinformatics tools.

MATERIALS AND METHODS

Data retrieval

Forty three *Os*GLP gene promoters were retrieved from the Asian rice (*Oryza sativa* ssp. Japonica) genome using the online server of Plants Ensemble (http://plants.ensembl.org/index.html), based on the information obtained from the Rice Genome Annotation Project Database (http://rapdb.dna.affrc.go.jp/), including two already analyzed promoters of *Os*RGLP1 [28] and *Os*RGLP2 [23,26] for better comparison. The size of each promoter was purposely picked as a contact figure of "1000" for uniformity.

Phylogenetic analysis

Phylogenetic analysis of the above sequences were conducted using the Molecular Evolution Genetic Analysis 6 (MEGA6) tool [29] by the neighbor-joining tree-making method. Similarly, Tajima's neutrality test of selection was conducted using the same software to find nucleotide diversity.

Analysis of TFBSs

Promoters were searched for putative TFBSs using MatInspector (ver. 9.1) [8], with a core and matrix similarities of 1/1, identifying the most frequent and unique *cis*-elements. Common TFBSs were further searched using the online server of Common TFs with a core and matrix similarities of 0.75/0.75. The position analyses of common TFBSs were performed in Excel (2010).

The sequences were aligned and searched for TFBSs that were common to at least 10 sequences (23%) in the align regions using DiAlign software with a core and matrix similarities of 0.75/0.75.

Module analysis

The common pattern of TFBSs in all studied promoters and their role in gene regulation by mutual interaction were detected by Frameworker software (ver. 5.5.8), with a minimum number of two elements in each module. The resulting *cis*-regulatory promoter modules common to the studied sequences were identified with respect to the organization and relative position of TFBSs using data from MatInspector. *In silico* functional analysis of the studied promoters was performed by searching predefined, already reported and confirmed functional modules (Plant Modules, ver. 5.7) with ModelInspector (ver. 5.6.8.7) [30]. The different software used for TFBSs analysis (MatInspector, Common TF, DiAlign, Frameworker, and ModelInspector) were provided by the online server of the Genomatix software suite (http://www.genomatix.de/cgi-bin/eldorado/main.pl?s =78f50a57a64fa8ae8b6532b5fd0a410e) Genomatix Software, Munich, Germany).

RESULTS

Sequence retrieval

Forty-three 5′ upstream promoter regions of GLP genes of the Asian rice (*Oryza Sativa* ssp. Japonica) genome were retrieved using Plant Ensemble. The name of each sequence, number of base pairs, accession number, chromosomal position and associated reference are given in Table 1. The sequences included two already cloned and computationally analyzed promoter regions of *Os*RGLP1/*Os*GLP8-11 [22] and *Os*RGLP2/*Os*GLP8-10 [23] for better comparison. All the *Os*GLPs promoters belong to the monocupin domain subfamily and are mostly located on chromosomes 3 and 8.

Phylogenetic analysis

Phylogenetically, the promoters can be divided into clade 1 with 36 sequences, and clade 2 with the remainder of 7 sequences (*Os*GLP1-2, *Os*GLP1-4, *Os*GLP3-1, *Os*GLP5-1, LOC*Os*05g10830, *Os*GLP8-12, *Os*GLP8-13) (Fig. 1) with a narrow genetic background of about 0.2% and with a well-supported bootstrap values of 0 to 100. Clade 1 has 5 clusters, of which cluster 1 has 10 sequences (*Os*GLP1-3, *Os*GLP2-4, *Os*GLP3-2, *Os*GLP8-11/*Os*RGLP1, *Os*GLP9-1, *Os*GLP12-1, -2, -3, -4), while cluster 2 has 4 sequences (*Os*GLP1-1, *Os*GLP2-3, *Os*GLP3-3,

Table 1. List of selected germin-like protein genes promoters from Asian rice var. *Oryza Sativa* ssp. Japonica.

S.N.	Promoter Name	Accession No	Location	References
1	*Os*RGLP1	EU742684	**8:**5263245:5264244	[28]
2	*Os*RGLP2	AAD43972	**8:**5259152:5260151	[23]
3	*Os*GLP1-1	LOC_Os01g18170	**1:**10169317:10170316	[5]
4	*Os*GLP1-2	LOC_Os01g50900	**1:**29230900:29231899	-
5	*Os*GLP1-3	LOC_Os01g72290	**1:**41913684:41914683	-
6	*Os*GLP1-4	LOC_Os01g72300	**1:**41916699:41917698	-
7	*Os*GLP2-1	LOC_Os02g29000	**2:**17165998:17166997	-
8	*Os*GLP2-2	LOC_Os02g29010	**2:**17166417:17167416	-
9	*Os*GLP2-3	LOC_Os02g29020	**2:**17174222:17175221	-
10	*Os*GLP2-4	LOC_Os02g32980	**2:**19598616:19599615	-
11	*Os*GLP3-1	LOC_Os03g08150	**3:**4156589:4157588	-
12	*Os*GLP3-2	LOC_Os03g44880	**3:**25314697:25315696	-
13	*Os*GLP3-3	LOC_Os03g48750	**3:**27780097:27781096	-
14	*Os*GLP3-4	LOC_Os03g48760	**3:**27784582:27785581	-
15	*Os*GLP3-5	LOC_Os03g48770	**3:**27788914:27789913	-
16	*Os*GLP3-6	LOC_Os03g48780	**3:**27792048:27793047	-
17	*Os*GLP3-7	LOC_Os03g58980	**3:**33585465:33586464	-
18	*Os*GLP3-8	LOC_Os03g59010	**3:**33590195:33591194	-
19	Loc_*Os*04g52720	LOC_Os04g52720	**4:**31395186:31396185	-
20	Loc_*Os*05g10830	LOC_Os05g10830	**5:**5985014:5986013	-
21	*Os*GLP5-1	LOC_Os05g19670	**5:**11466161:11467160	-
22	*Os*GLP8-1	LOC_Os08g08920	**8:**5185875:5186874	[34]
23	*Os*GLP8-2	LOC_Os08g08960	**8:**5207385:5208384	-
24	*Os*GLP8-3	LOC_Os08g08970	**8:**5221214:5222213	-
25	*Os*GLP8-4	LOC_Os08g08980	**8:**5227822:5228821	-
26	*Os*GLP8-5	LOC_Os08g08990	**8:**5232768:5233767	-
27	*Os*GLP8-6	LOC_Os08g09000	**8:**5237999:5238998	-
28	*Os*GLP8-7	LOC_Os08g09010	**8:**5241495:5242494	-
29	*Os*GLP8-8	LOC_Os08g09020	**8:**5247666:5248665	-
30	*Os*GLP8-9	LOC_Os08g09040	**8:**5253286:5254285	-
31	*Os*GLP8-10	LOC_Os08g09060	**8:**5259152:5260151	-
32	*Os*GLP8-11	LOC_Os08g09080	**8:**5263245:5264244	-
33	*Os*GLP8-12	LOC_Os08g13440	**8:**7995718:7996717	-
34	*Os*GLP8-13	LOC_Os08g35750	**8:**22554033:22555032	[5]
35	*Os*GLP8-14	LOC_Os08g35760	**8:**22559098:22560097	[2]
36	*Os*GLP9-1	LOC_Os09g39510	**9:**22696146:22697145	[5]
37	*Os*GLP9-2	LOC_Os09g39520	**9:**22698390:22699389	-
38	*Os*GLP9-3	LOC_Os09g39530	**9:**22700960:22701959	-
39	*Os*GLP11-1	LOC_Os11g33110	**11:**19584667:19585666	-
40	*Os*GLP12-1	LOC_Os12g05840	**12:**2689827:2690826	[34]
41	*Os*GLP12-2	LOC_Os12g05860	**12:**2693512:2694511	-
42	*Os*GLP12-3	LOC_Os12g05870	**12:**2697215:2698214	-
43	*Os*GLP12-4	LOC_Os12g05880	**12:**2700273:2701272	-

S.N. – serial number; Location – coordinates of respective promoter; chromosome number is shown in bold; hyphenation represents the same text as in the upper cell

LOC*Os*04g52720), mostly in chromosomes 1, 2 and 12. Cluster 3 has 3 (*Os*GLP3-7, *Os*GLP9-2, -3) and cluster 4 has 12 sequences (*Os*GLP8-1, -2, -3, -4, -5, -6, -7, -8, -9, -10/*Os*RGLP2, *Os*GLP8-14), mostly in chromosomes 9 and 8, respectively. Cluster 5 has 5 sequences (*Os*GLP2-1, *Os*GLP2-2, *Os*GLP3-4, -5, -6) that form a sister cluster with all other clusters. Generally, promoters on the same chromosome displayed the highest similarity with each other, as was distinctly observed in the promoters of chromosomes 3, 8 and 12. Promoters with the highest sequence similarity included *Os*GLP2-1/*Os*GLP2-2, *Os*GLP3-4/*Os*GLP3-5, *Os*GLP8-10/*Os*RGLP2, *Os*GLP12-1/*Os*GLP12-2 and *Os*GLP1-3/*Os*GLP3-2 pairs, with well supported bootstrap value (100%). However, some promoters on chromosomes 1 (*Os*GLP1-1, -2, -3, -4, represented by black circles), 2 (*Os*GLP2-3, -4, represented by pink circles), 3 (*Os*GLP3-2, -3, -7, -8, represented by light green circles), 8 (*Os*GLP8-11, -12, -13, represented by blue circles) and 9 (*Os*GLP9-1, represented by a red circle) exhibited distant relationships irrespective of their close chromosomal positions. Contrary to this, a close relationship was established between *Os*GLP1-1/*Os*GLP2-3, *Os*GLP1-3/*Os*GLP3-2, *Os*GLP3-7/*Os*GLP9-2, *Os*GLP3-8/*Os*GLP11-1, *Os*GLP3-1/*Os*GLP8-13 and *Os*GLP1-2/Loc*Os*05g10830 pairs of promoters, even though they belong to different chromosomes. Tajima's neutrality test showed a nucleotide diversity (π) of 0.69 in all of the studied promoters.

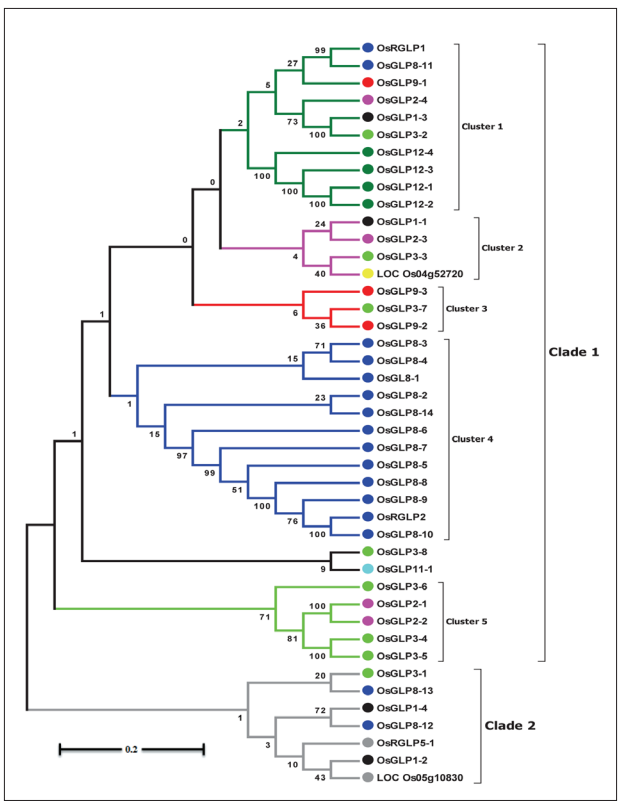

Fig. 1. Phylogenetic analysis of Asian rice (*Oryza sativa* ssp. Japonica) germin-like protein (GLPs) gene promoters using MEGA6 through Neighbor's joining methods. Numerical values indicate bootstrap support for each node. Bootstrap support values were based on 1000 replicates and are given as percentages. Clade-1 can be distinguished into 5 clusters; each is in a different color. Promoters located on the same chromosome are shown with circles of the same color. Labelled parentheses are used to represent each clade and cluster.

Analysis of TFBSs

A total of 4234 TFBSs from 90 different TF families were found to be distributed throughout the promoter regions on both strands, and were mostly located in the upstream region of -100 to -900 bp, with the relatively highest incidence of occurrence in the upstream region -200 to -300 bp (11%) and -400 to -500 bp from the transcription start site, while the lowest occurrence was observed at 0 to -100 bp and -900 to -1000 bp. Most of the TFBSs belonged to the AHBP family (275 copies), followed by MYBL (236 copies), while the iron-dependent response element (IDRS) and the ethylene response element (EREF) had lowest number (1) of copies. The highest number of TFBSs (141) was found in *Os*GLP8-14, while the smallest number (74) was in *Os*GLP9-2. Description of the first ten most frequently occurring TFBSs of the total detected elements is provided in Table 2. Most of these elements were found in promoters residing in chromosomes 8 (1421), 3 (760), 1 (413) and 12 (391). Common TFBSs included AHBP, VTBP and MYBL which were distributed throughout the promoter regions (Fig. 2). Graphical analysis revealed that in these promoters most of the AHBP elements were located in the upstream region from -200 to -800 bp, while the lowest numbers of elements were found on the first and last 200 bp (Fig. 3). Similarly, most of the VTBP and MYBL elements were found at upstream positions from 0 to -400 bp and -600 to -800 bp (Figs. 4 and 5). Further, DiAlign analysis showed that MYBL, PTBP and VTBP were conserved in the aligned regions of *Os*GLPs promoters. The conserved position lies upstream from the transcriptional start site between 0 to -800 bp, and is mostly located on chromosomes 8 and 12. The common TFBSs in the aligned region are

shown as colored boxes in Fig. 6. MYBL was found conserved in the aligned region of *Os*RGLP1, -2, *Os*GLP8-1, -2, -3, -4, -5, -6, -7, -8, -9, -10, -11, -12, -13 and *Os*GLP9-1, while VTBP and PTBP were found in the aligned region of *Os*RGLP1, 2, *Os*GLP1-4, Loc_*Os*04g52720, Loc_*Os*05g10830, *Os*GLP8-1, -2, -3, -4, -5, -6, -7, -8, -9, -10, -11, -14, *Os*GLP11-1, *Os*GLP12-1, -2, -3 and *Os*GLP12-4 promoters.

Module analysis

Common TFBSs sharing the same framework of *cis*-regulatory elements were investigated, revealing three novel *cis*-regulatory modules, of which VTBP_VTBP was the most frequent module, having 56 copies covering 58% of the sequences (25); this was followed by MYBS_MYBS (with 49 copies) and AHBP_VTBP (33 copies). The name, element type, element orientation, parameter used and distance to the next element for each module are shown in Table 3. *In silico* functional analysis via ModelInspector revealed 77 modules in 33 sequences. Detailed information about modules, including their names, number of copies in each strand, start and end positions, frequency of occurrence and promoters with the highest number of these modules, is presented in Tables 4 and 5. The most frequently occurring modules were DOFF_OPAQ_03 and GTBX_MYCL_01, occurring in 19 and 11 sequences, respec-

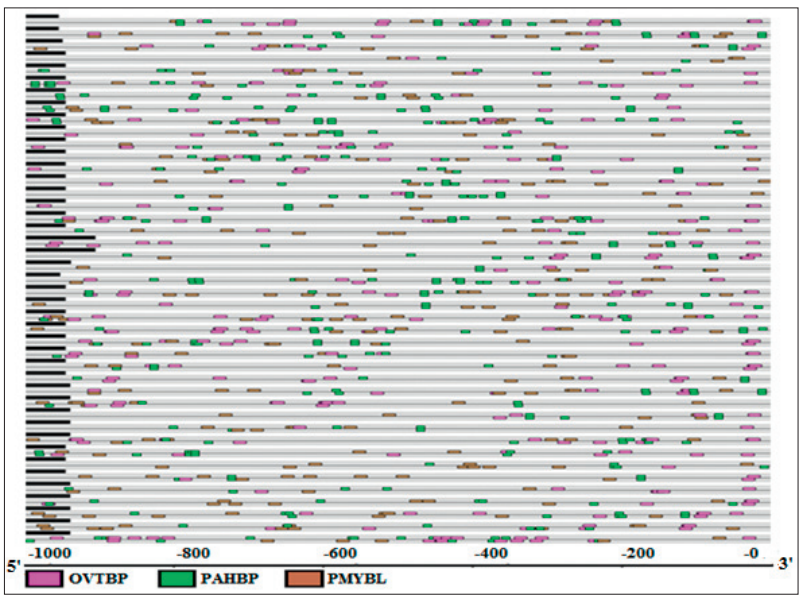

Fig. 2. Positions of common TFBSs found in rice Germin-like protein (*Os*GLPs) promoters. Each element is represented by a different color.

Table 2. Description of the first ten most frequently occurring TFBSs found in rice germin-like protein gene (*Os*GLPs) promoters.

S.N.	Family	Complete Name	*p*-value	Total	Seqn	Copies	Chrom. No	Common
1	AHBP	*Arabidopsis* homeobox protein	0.00120599	275	*Os*GLP2-3	15	8, 3, 12	43
2	MYBL	MYB-like proteins	0.00663691	236	*Os*GLP8-4	11	8, 3, 12	43
3	GTBX	GT-box elements	0.20721	229	*Os*GLP3-7	14	8, 3, 12	42
4	VTBP	Vertebrate TATA binding protein factor	0.00018636	217	*Os*GLP12-4	14	8, 3, 12	43
5	YTBP	Yeast TATA binding protein factor	7.71253E-15	179	*Os*RGLP1	18	8, 3, 12	30
6	MYBS	MYB proteins with single DNA binding repeat	0.197507	171	*Os*GLP12-1	12	8, 3, 12	40
7	NACF	Plant-specific NAC transcription factors	0.0199598	148	*Os*GLP1-2	10	8, 3, 12	41
8	PTBP	Plant TATA binding protein factor	2.60438E-15	142	*Os*GLP8-11	13	8, 3, 12	38
9	LEGB	Legumin Box family	2.05575E-12	139	*Os*GLP12-3	9	8, 3, 12	38
10	MYCL	Myc-like basic helix-loop-helix binding factors	0.000107576	134	*Os*RGLP1	10	8, 3, 12	41

S.N. – serial number; Family – family of element; Total – total copies of elements; Seq – promoter with the highest number of elements; Copies – number of copies found in specific promoter; Chrom. – chromosome number having the highest number of elements; Common – number of promoters in which a common specific element was found

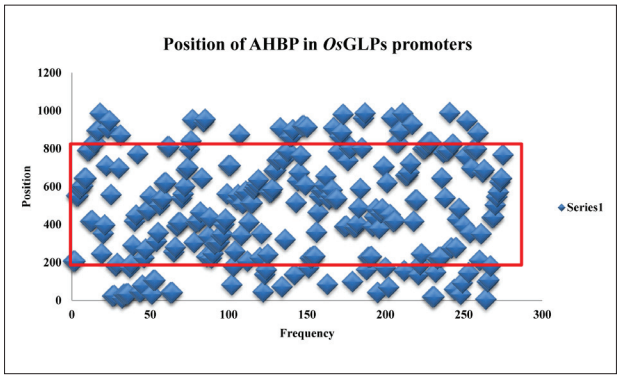

Fig 3. The distribution of *Arabidopsis* homeobox protein (AHBP) *cis*-elements in rice germin-like protein gene (*Os*GLPs) promoters. The position in the graph starts from the 5' end. The red box represents promoter regions with the highest number of AHBP elements equivalent to the upstream region -200 to -800 bp relative to the transcription start site. The frequency is given as the number of the elements in all promoters.

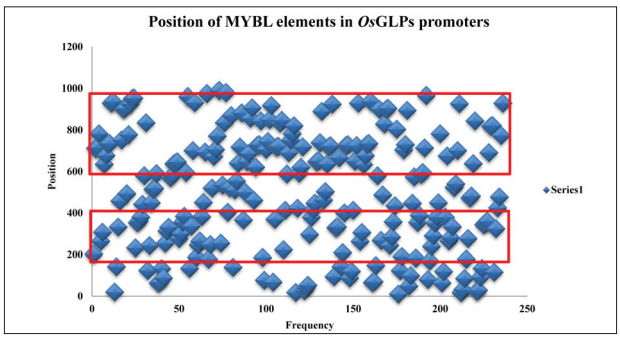

Fig. 4. Position analysis of vertebrate TATA box binding protein (VTBP) regulatory elements in *Os*GLPs promoters. The positions in the graph starts from the 5' end. The red box represents the highest number of elements at two positions of *Os*GLPs promoters, which is equivalent to the upstream region from -0 to -400 bp and -600 to -800 bp relative to the transcription start site. The frequency is given as the number of element in all sequences.

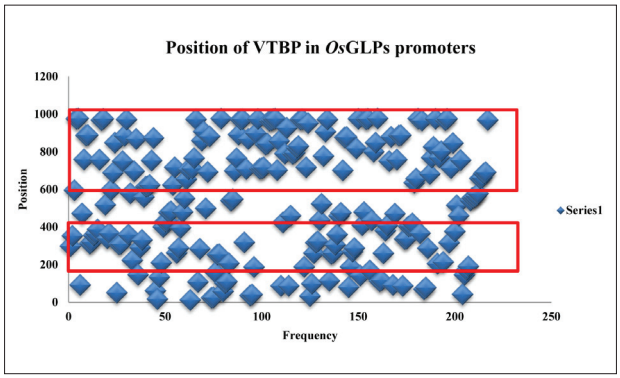

Fig. 5. The distribution of MYBL elements in rice germin-like protein gene (*Os*GLPs) promoters. The positions of elements in graph starts from the 5' end. The red box indicates the highest number of occurrence at two positions, which is equivalent to 0 to -400 bp and -600 to -800 bp relative to the transcription start site. The frequency is given as the number of elements in all sequences.

tively. The DOFF_OPAQ_03 has a distribution value of 13.92%, showing its overall distribution in ModelInspector database promoters (which consists of 71,000 plant and 311,000 vertebrate promoters), and currently found in *Os*GLP2-1, *Os*GLP3-1, -3, -5, -6, -7, -8, *Os*GLP8-4, -8, -13, *Os*GLP12-1, -2 and *Os*RGLP12-3. However, NACF_LEGB_01 was a novel functional module found only in *Os*GLP8-12. Other novel modules include OCSE_DOFF_01, AREF_MYCL_01 and GARP_GARP_01 that occur with distribution values of 0.34%, 0.42% and 1.26%, respectively, and are found in *Os*GLP3-5, -4 and *Os*GLP3-3, respectively. The highest number of modules was found in *Os*GLP3-5 and *Os*GLP3-6 (6), followed by *Os*GLP3-1, *Os*GLP8-1, -6, *Os*GLP12-1 and *Os*GLP12-2, each containing 4 functional modules. Similarly, no modules were found in *Os*GLP1-1, -2, -3, *Os*GLP2-3, -4, *Os*GLP3-2, *Os*GLP8-2 and *Os*GLP9-3 promoters. Typically, functional modules were located in the promoters residing in chromo-

Table 3. Description of three novel common cis-regulatory modules found in rice germin-like protein (OsGLPs) genes promoters using Frameworker analysis.

S.N.	Element Type	Strand	Matrix sim	Distance to next element	Copies	Sequences	Nature
Model 1	MYBS	-	Optimized (min. 0.79)	5-13 bp	49	23 (53%)	35 non-overlapping
	MYBS	+	Optimized (min. 0.79)	5-13 bp			
Model 2	VTBP	-	Optimized (min. 0.80)	5-16 bp	56	25 (58%)	37 non-overlapping
	VTBP	+	Optimized (min. 0.82)	5-16 bp			
Model 3	AHBP	-	Optimized (min. 0.90)	47-66 bp	33	22 (51%)	24 non-overlapping
	VTBP	+	Optimized (min. 0.78)	47-66 bp			

S.N. – serial number; Element type – *cis*-regulatory elements in each module; Matrix sim – conditions used for analysis; Copies – number of copies of each module; Sequences – number of promoters in which a particular module was found (also given as a percentage); Nature – nature of modules: whether they overlap or not.

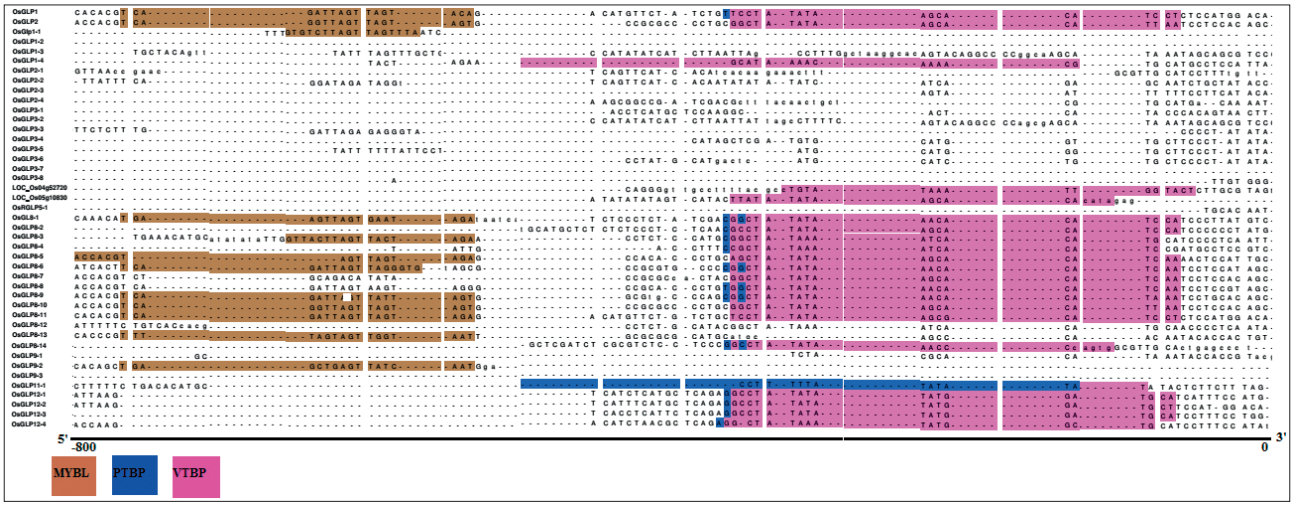

Fig. 6. Conserved positions of MYB (grey), PTBP (blue) and VTBP (pink) in the aligned region of *Os*GLPs promoters. All conserved elements were found in the upstream region of 0 to -800 bp.

Table 4. Description of the total number of predefined functional modules found in rice germin-like protein (*Os*GLPs) gene promoters using ModelInspector.

S.N.	Module Name	Total	Copies		No of Seq	Distribution	Function	Reference
			+	-				
1	AHBP_DOFF_01	5	3	2	5	5.55%	Regulate companion/source-specific gene expression	[61]
2	AREF_MYCL_01	2	1	1	2	0.42%	Auxin and brassinosteroid-induced expression	[56]
3	DOFF_MYBL_01	3	2	1	3	4.63%	Gibberellin-induced expression	[64]
4	DOFF_OPAQ_03	19	13	6	14	13.92%	Endosperm specificity of *GluD-1* gene	[62]
5	GARP_GARP_01	2	2	0	2	1.26%	Cytokinin inducible expression	[65]
6	GBOX_GBOX_01	6	4	2	5	5.62%	Regulate phaseolin/expression in cotyledons	[66]
7	GTBX_GTBX_01	5	2	3	5	1.88%	Critical for light responsive transcriptional activity	[67]
8	GTBX_MYCL_01	11	3	8	11	5.56%	Dehydration-response activation of the *erd1* promoter.	[63]
9	MYBL_DOFF_02	1	1	0	1	2.86%	Endosperm-specific expression	[40]
10	MYCL_GCCF_01	2	2	0	2	3.30%	Anthocyanin regulated expression	[42]
11	MYCL_MIIG_01	4	1	3	4	1.67%	Dehydration/ABA induced expression of the *rd22* gene	[68]
12	MYCL_MYBL_01	7	2	5	7	2.74%	Promote brassinosteroid (BR) targeted gene expression.	[44]
13	NACF_LEGB_01	1	0	1	1	0.00%	Regulated Fe homeostasis gene expression	[52]
14	OCSE_DOFF_01	1	0	1	1	0.34%	Glutathione S-transferase (GST) induced expression	[69]
15	OPAQ_DOFF_01	1	1	0	1	5.26%	Induced endosperm/seed storage protein expression	[70]
16	TALE_TALE_01	7	4	3	7	4.07%	Tuberization, rooting and vegetative development	[71]
	Total	77	41	36	71	0.5908		

S.N. – serial number; Total – total number of each module found in all promoters; plus/minus (+/-) – strand orientation ("+"is the sense strand, "–" is the antisense strand); No of seq – number of *Os*GLPs promoters; Distribution – percentage distribution of the modules in promoters of the plant ModelInspector database.

Table 5. Description of the predefined already reported functional modules found in *Os*GLPs promoters using ModelInspector.

S.N.	Seq. name	No of modules	Module	Start pos	End pos	Strand Ori	Chrom. No	Function
1	*Os*RGLP1	2	GTBX_MYCL_01	798	659	-	8	Dehydration response activation
2	-		TALE_TALE_01	865	895	+	-	Tuberization, rooting and vegetative development
3	*Os*RGLP2	3	TALE_TALE_01	251	227	-	8	-
4	-		GTBX_MYCL_01	378	236	-	-	Dehydration-response activation
5	-		GTBX_GTBX_01	484	444	-	-	Light-responsive transcriptional activity
6	*Os*GLP1-4	3	GTBX_GTBX_01	395	439	+	1	-
7	-		MYCL_MIIG_01	613	544	-	-	Dehydration/ABA inducible expression
8	-		MYCL_MYBL_01	613	587	-	-	Brassinosteroid (BR) inducible expression.
9	*Os*GLP2-1	1	DOFF_OPAQ_03	244	203	-	2	Endosperm-specific expression
10	*Os*GLP2-2	1	-	338	379	+	-	-
11	*Os*GLP3-1	4	-	123	158	+	3	-
12	-		GTBX_MYCL_01	237	97	-	-	Dehydration-response activation
13	-		MYCL_MYBL_01	687	709	+	-	Brassinosteroid (BR) inducible expression.
14	-		MYCL_MIIG_01	958	896	-	-	Dehydration/ABA-induced expression
15	*Os*GLP3-3	2	GARP_GARP_01	467	496	+	-	Cytokinin inducible expression
16	-		DOFF_OPAQ_03	855	817	-	-	Endosperm-specific expression
17	*Os*GLP3-4	2	MYCL_MYBL_01	129	107	-	-	Brassinosteroid (BR) inducible expression.
18	-		AREF_MYCL_01	795	726	-	-	-
19	*Os*GLP3-5	6	AHBP_DOFF_01	509	567	+	-	Regulate companion /Source-specific gene expression
20	-		DOFF_MYBL_01	551	575	+	-	Gibberellin-induced expression
21	-		DOFF_OPAQ_03	566	604	+	-	Endosperm-specific expression
22	-		GBOX_GBOX_01	804	911	+	-	Regulate phaseolin/expression in cotyledons
23	-		DOFF_OPAQ_03	872	910	+	-	Endosperm-specific expression
24	-		OCSE_DOFF_01	958	920	-	-	Glutathione S-transferase (GST)-induced expression
25	*Os*GLP3-6	6	DOFF_OPAQ_03	548	510	-	-	Endosperm-specific expression
26	-		-	641	608	-	-	-
27	-		GTBX_MYCL_01	748	894	+	-	Dehydration response activation
28	-		GTBX_GTBX_01	843	889	+	-	Light responsive transcriptional activity
29	-		GBOX_GBOX_01	894	793	-	-	Regulate phaseolin/expression in cotyledons
30	-		-	894	787	-	-	-
31	*Os*GLP3-7	2	AHBP_DOFF_01	735	797	+	-	Regulation of companion/source-specific gene expression
32	-		DOFF_OPAQ_03	929	964	+	-	Endosperm-specific expression
33	*Os*GLP3-8	1	DOFF_OPAQ_03	324	357	+	-	-
34	LOC_*Os*05g10830	1	GARP_GARP_01	199	223	+	5	Cytokinin inducible expression
35	*Os*RGLP5-1	3	TALE_TALE_01	555	531	-	5	Tuberization, rooting and vegetative development
36	-		MYCL_GCCF_01	805	823	+	-	Anthocyanin regulated expression
37	-		GTBX_MYCL_01	945	805	-	-	Dehydration-response activation
38	*Os*GL8-1	4	DOFF_MYBL_01	48	76	+	8	Gibberellins-induced expression
39	-		GTBX_MYCL_01	216	355	+	-	Dehydration-response activation
40	-		GBOX_GBOX_01	789	895	+	-	Regulate phaseolin/expression in cotyledons
41	-		AREF_MYCL_01	800	862	+	-	Auxin and brassinosteroid (BR)-induced expression
42	*Os*GLP8-3	1	AHBP_DOFF_01	780	842	+	-	Regulate companion/source-specific gene expression
43	*Os*GLP8-4	2	DOFF_OPAQ_03	558	525	-	-	Endosperm-specific expression
44	-		GTBX_GTBX_01	982	936	-	-	Light-responsive transcriptional activity
45	*Os*GLP8-5	2	MYBL_DOFF_02	454	597	+	-	Endosperm-specific expression
46	-		AHBP_DOFF_01	966	907	-	-	Regulate companion/source-specific gene expression
47	*Os*GLP8-6	4	MYCL_MIIG_01	267	202	-	-	Dehydration/ABA induced expression

Table 5. continued:

48	-		GTBX_MYCL_01	387	247	-	-	Dehydration-response activation
49	-		TALE_TALE_01	391	416	+	-	Tuberization, rooting and vegetative development
50	-		AHBP_DOFF_01	396	336	-	-	Regulate companion /source-specific gene expression
51	*Os*GLP8-7	1	OPAQ_DOFF_01	814	858	+	-	Induced Endosperm/seed storage protein expression
52	*Os*GLP8-8	2	DOFF_OPAQ_03	228	262	+	-	Endosperm-specific expression
53	-		MYCL_GCCF_01	304	323	+	-	Anthocyanin-regulated expression
54	*Os*GLP8-10	3	TALE_TALE_01	250	226	-	-	Tuberization, rooting and vegetative development
55	-		GTBX_MYCL_01	377	235	-	-	Dehydration response activation
56	-		GTBX_GTBX_01	483	443	-	-	Light-responsive transcriptional activity
57	*Os*GLP8-11	2	TALE_TALE_01	221	250	+	-	Tuberization, rooting and vegetative development
58	-		GBOX_GBOX_01	813	915	+	-	Regulate phaseolin/expression in cotyledons
59	*Os*GLP8-12	1	NACF_LEGB_01	908	764	-	-	Regulated Fe homeostasis gene expression
60	*Os*GLP8-13	1	DOFF_OPAQ_03	852	891	+	-	Endosperm-specific expression
61	*Os*GLP8-14	2	DOFF_MYBL_01	391	370	-	-	Gibberellin-induced expression
62	-		GBOX_GBOX_01	573	680	+	-	Regulation of phaseolin expression in cotyledons
63	*Os*GLP9-1	1	MYCL_MYBL_01	431	455	+	9	Brassinosteroid (BR)-targeted gene expression
64	*Os*GLP9-2	1	TALE_TALE_01	399	424	+	-	Tuberization, rooting and vegetative development
65	*Os*GLP11-1	1	MYCL_MIIG_01	611	681	+	11	Dehydration/ABA-induced expression
66	*Os*GLP12-1	4	GTBX_MYCL_01	242	384	+	12	Dehydration response activation
67	-		DOFF_OPAQ_03	247	210	-	-	Endosperm-specific expression
68	-		MYCL_MYBL_01	441	419	-	-	Brassinosteroid (BR) inducible expression.
69	-		DOFF_OPAQ_03	449	487	+	-	Endosperm-specific expression
70	*Os*GLP12-2	4	MYCL_MYBL_01	192	170	-	-	Brassinosteroid (BR) inducible expression.
71	-		DOFF_OPAQ_03	200	238	+	-	Endosperm-specific expression
72	-		GTBX_MYCL_01	315	172	-	-	Dehydration-response activation
73	-		DOFF_OPAQ_03	446	485	+	-	Endosperm-specific expression
74	*Os*GLP12-3	2	DOFF_OPAQ_03	534	575	+	-	Endosperm-specific expression
75	-		-	542	579	+	-	Endosperm-specific expression
76	*Os*GLP12-4	2	MYCL_MYBL_01	491	469	-	-	Brassinosteroid (BR) inducible expression.
77	-		GTBX_MYCL_01	612	471	-	-	Dehydration-response activation

S.N. – serial number; Seq. name – respective promoters; No of modules – number of modules found in each promoter; Start pos – start position of each module; End Pos. – end position; the position of each module is given relative to the 5' end; Strand Ori – strand orientation ("+" for is the sense strand, while "–" is the for antisense strand); Chrom. No. – chromosome number; hyphenation represents the same value of the upper cell.

somes 3, 8 and 12, and were mostly related to endosperm-specific expression, dehydration and etiolation, but no such module was found on chromosome 4. An overview of the potential role of GLPs in the light of the predicted modules with respect to their chromosomal position is presented in Table 5. Most of the functional modules were found on the sense strand (41) rather than the antisense strand (36).

DISCUSSION

Phylogenetic analysis

Phylogenetic analysis of 43 *Os*GLP gene promoters revealed a narrow genetic background (0.2%), suggesting a high similarity that is smaller than the reported value of 31% [24] for seven GLP promoters from different plant species, but concurs with the previous report [25] in which 44 GLP promoters mostly from rice were considered. This could be due to the fact that in our analysis all of the promoters belong to the same species (*Oryza sativa* ssp. Japonica). Promoters located on the same chromosome shared the highest sequence similarity, which may be due to duplication and representation of the same pattern of *cis*-regulatory elements, and thus their similar roles in gene expression [31]. The phenomenon is more prominent in GLP promoters located on chromosomes 3, 8 and 12, which may be either due to recent or older duplication events, which created highly similar *cis*-regulatory elements that were selectively preserved

as such in order to enable the co-expression of these genes. These results are also supported by earlier studies in *Os*GLPs [23], *Gm*CHS7 and *Gm*CHS8 gene promoters [11] that reported the existence of common regions in these promoters. However, GLP promoters on chromosomes 1, 2, 4, 5, 9 and 11 exhibited variation that could be the result of diversification in their *cis*-regulatory elements [32], either as a result of selection pressure and changes in the environment, or because of the accumulation of mutations (due to reduced selection pressure), during which TFBSs copies were modified over time by involvement in new and diverse functional pathways, ultimately resulting in diverse expression patterns. The results are also supported by the high Tajima value (0.69), which represented the change in their *cis*-regulatory elements. All members of chromosome 8 GLPs (cluster 4), which form a separate lineage, displayed a close relationship with each other, suggesting a similar pattern in their *cis*-regulatory elements. Previously it was shown that the promoter of *OsRGLP8-10/OsRGLP2* gene was induced by salt, BAP and wounding stresses when analyzed via promoter-GUS fusion, with prominent expression in the cell wall, cell membrane, cytoplasm, vein and interveinal area [26]. Similarly, most genes of these promoters possess a strong link with the disease resistance pathotype [33], of which *OsGLP8-1-12* and *OsGLP8-14/OsGLP1* are part of the QTL that provides resistance against rice blast (*Magnaporthe oryzae*) and sheath blight (*Rhizoctonia solani*) [2,34]. The close relationship of their promoters points to their functional similarity and thus demands further study against multiple stresses. Similarly, the close relationship of *Os*RGLP1, *Os*GLP8-12 and *Os*GLP8-13 with *Os*GLP9-1, *Os*GLP1-4 and *Os*GLP3-1, respectively, points to their functional similarity. However, the distant relationship between *Os*GLP8-11, -12 and *Os*GLP8-13 can be explained by diversification in their *cis*-regulatory elements. The mechanism of defense provided by chromosome 8 in fungal pathogenicity is conserved among *Gramineae* members, such as wheat [35], rice [34] and barley [19,36], which need to be properly tested for these promoters. Similar observations were also noted for promoters on chromosomes 3 (cluster 5) and 12 (cluster 1), which points to recent duplication and the presence of similar *cis*-regulatory elements. However, none of these genes or promoters have thus far been tested against any stresses, but their close relationship with chromosome 8 promoters suggests similar regulatory mechanisms and expression patterns. Of all considered promoters, those located on chromosomes 1 and 5 have a distant relationship to all other, suggesting that they possess distinctive patterns of *cis*-regulatory elements.

TFBSs analysis

MatInspector revealed considerable diversity in TFBSs, revealing their putative roles in various plant processes. Previously, the roles of *Os*RGLP1 [22] and *Os*RGLP2 [24] were predicted by TFBS analysis with PLACE/Signal Scan. However, the present study provides a more detailed analysis. Large number of TFBSs were found on GLP promoters located on chromosomes 3, 8 and 12, which could be the result of clustering and duplication [31]. The presence of AHBP, VTBP and MYBL in all promoters suggests that their fundamental roles in regulation are conserved in the upstream aligned region from 0 to -800 bp. Conserved regions were mostly found on the promoters of chromosomes 8 and 12, suggesting their close relationship and similar expression patterns, which is in accordance with the result of the phylogenetic analysis presented in the previous section. These observations not only suggested the importance of these elements from an evolutionary point, "in which nature congregated these elements to a specific region of the GLP promoter in accordance with their demanding function", but also indicated their fundamental role in *Os*GLPs regulation. AHBP is the most abundant element reportedly involved in embryo, shoot, root patterning, shade growth control, organ fate and stem cell proliferation [37]. Most copies of this element were found in *Os*GLP2-3 (15) and *Os*GLP3-2 (13) promoters, which shows their importance. Similarly, most of the MYBL elements were found in *Os*GLP12-3 (10 copies) and *Os*GLP8-4 (11 copies), which has an important role in GA-regulated expression [38], cotton fiber development [39], endosperm development [40], organogenesis [41], gibberellin signaling [42], seed development [43], BR-induced gene expression, vascular differentiation, senescence, stress responses [44] and nitrate enhancement [39]. In the same way, VTBP is critical for promoter activity equally in plants, animals and viruses in gene-specific expression [45]. We observed that *Os*GLP12-4 has 14 copies of

this element, showing its crucial role. Interestingly, the highest number of copies of AHBP (-200 to -800 bp), MYB (0 to -400 bp) and VTBP (-600 to -800 bp) elements were congregated at upstream positions (-200 to -800 bp) in the form of clusters, possibly because of increased environmental and selection pressure [46] that led to subfunctionalization and/or neofunctionalization of genes [47]. However, these regions need to be examined further by deletion and mutational analysis to confirm their crucial role in gene regulation as has already been reported for *HvGerB, HvGerF* [14] and *AtGER3* gene promoters [48]. Aside from these common elements, other important elements include GTBX, which mostly resides in *Os*GLP3-7 (14 copies) and having role in the light of responsiveness (LRE), senescence [49], drought [50], cold, salt stress [51] and water use efficiency [39]. Similarly, the presence of 10 copies of plant-specific NAC transcription factor in *Os*GLP1-2 points to a role in homeostasis [52] and leaf senescence [53]. Likewise, Myc-like basic helix-loop-helix binding factors (MYCL) are involved in controlling light-response, tissue-specific activation of phenylpropanoid biosynthesis genes [54], fruit development [55] and auxin response [56]. Analysis showed that *Os*RGLP1 contained 10 copies of the above-mentioned element. Moreover, several unique TFBSs including EREF and IDRS were also found in *Os*GLP8-7 and *Os*GLP1-2 promoters, pointing to its role in the control of the intracellular iron status [57] and floral development [58]. Similarly, two copies of the *Arabidopsis* CDC5 homolog were found in *Os*GLP1-2 and *Os*GLP12-4 that are involved in pre-mRNA splicing [59]. The presence of unique TFBSs in rice GLPs may define their differential promoter activity which is responsible for distinct gene expression [11]. The presence of these novel elements reveals their differential expression and novel functions against various stresses.

Analysis of module

The presence of three novel *cis*-regulatory modules (AHBP_VTBP, MYBS_MYBS and VTBP_VTBP) in all promoters further confirmed the crucial co-regulated role of VTBP, MYB and AHBP in *Os*GLP genes expression. The co-occurrence of these elements in such a regular pattern in all promoters points to their fundamental role in *Os*GLPs regulation. Previously, a similar module (MYCL_MYBL_01) was found to be active during brassinosteroid (BR)-targeted gene expression [44]. Similarly, the combined role of MYCS_P1BS and GAMYB_DOF in the regulation of mycorrhiza-activated phosphate transporters [60], seed development and germination was previously validated [43]. Likewise, the synergetic effect of the GC-rich region and TATA box was found to be critical for adam8 promoter activation [45], and the role of DOF and HD-Zip transcription factors was observed to be important in the regulation of cell-specific expression of the *Atsuc2* gene (*Arabidopsis thaliana* sucrose transporter-2 gene) [61]. However, a detailed study is needed to further clarify the role of these novel modules in rice GLP genes expression. Further, *in silico* functional analysis revealed the presence of various functional modules, the highest being DOFF_OPAQ_03 and GTBX_MYCL_01, which have roles in endosperm-specific expression [62], dehydration, etiolation, tuberization and cotyledon-specific expression [63]. The observed result is in close agreement with the observed function of GLPs as germination markers [1]. *Os*GLP3-5 and *Os*GLP3-6 possess the highest number of these modules, confirming their role in endosperm-specific expression, dehydration and etiolation. Three unique modules, including NACF_LEGB_01, OCSE_DOFF_01 and OPAQ_DOFF_01, that cause iron deficiency-, glutathione S-transferase (GST)- and seed storage-specific expression were identified in *Os*GLP8-12, *Os*GLP8-7 and *Os*GLP3-5 respectively. Other unique modules include AREF_MYCL_01 (found in *Os*GLP3-4 and *Os*GLP8-1) and GARP_GARP_01 (found in LOC_*Os*05g10830, *Os*GLP3-3) which play a role in BR- and cytokinin-induced expression. Similarly, *Os*GLP3-1, *Os*GLP8-1, -6, *Os*GLP12-1 and *Os*GLP12-2 each contained 4 modules, revealing their roles in dehydration, endosperm-specific expression and transcription control. The presence of novel modules revealed the functional diversity of rice GLP promoters. The highest number of modules were found in promoters situated on chromosome 8 (30), 3 (23) and 12 (12), which revealed their functional importance related to tuberization-, cotyledon- and endosperm-specific expression, while the presence of several new modules related to hormonal stress and light responsiveness pointed to its diverse role. Most promoters on chromosomes 8 and 12 have nearly the same patterns of *cis*-regulatory element, revealing their co-regulated role in response to environmental stresses.

This finding is in close agreement with the previous works of different authors that together validate the importance of this region (5185878- 7994721) located on chromosome 8 [2,5,23,26,34] in responses to different stresses. This has not only been established in rice but also in other members of *Gramineae* i.e. *Hordeum vulgare*, *Triticum aestivum* and *Brachypodium distachyon* [34]. Overall, the analysis showed that most genes on chromosomes 2, 3 and 12 are regulated by endosperm-specific activity, while those of chromosome 8 exhibited more diverse roles, in dehydration, tuberization and in signaling pathways (brassinosteroid, cytokinin, gibberellin, etc.), possibly due to the congregation of specific TFBSs in their promoters. No functional module was found on the promoters of chromosome 4 ((LOC_Os04g52720), and some other promoters (*Os*GLP1-1, -2, -3, *Os*GLP2-3, -4, *Os*GLP3-2, *Os*GLP8-2 and *Os*-GLP9-3)), which may be due to the accumulation of mutations due to reduced selection pressure. A more detailed analysis is presented in Table 5 which provides information about the possible regulatory role of these promoters in different parts of the plant in response to various stresses.

CONCLUSIONS

Recognition of regulatory *cis*-acting elements is an important step towards an improved understanding of gene expression and its regulatory mechanisms. The presented data show that *Os*GLP gene promoters are under considerable environmental pressure which has resulted variations in their *cis*-regulatory elements and phylogenetic relationship. Certain regions (-200 to -800 bp) of these promoters harbor a large number of specific *cis*-regulatory elements (AHBP, MYBL and TBP) whose interaction appears to be decisive for their regulation. The presence of these elements in the form of functional modules provides evidence for their significant involvement in various fundamental biological processes in response to various stresses. Using the above data, the functioning and expression patterns of these genes and promoters can be predicted to a very high level of certainty, which will pave the way for their use in crop biotechnology. Certain promoters, particularly those located on chromosomes 3, 8 and 12, are of considerable importance and can be used in the development of resistant cultivars against various stresses.

Acknowledgments: We are grateful to the Higher Education Commission, Islamabad, Pakistan for providing financial assistance.

Authors' contribution: MI is the main author who designed the research, collected the data, applied various software and wrote the manuscript. TM and SMSN helped with the interpretation of the results and manuscript writing by reviewing several drafts of the manuscript. All authors read and approved the final manuscript.

Conflict of interest disclosure: The authors declare that there are no conflicts of interest.

REFERENCES

1. Thompson E, Lane B. Relation of protein synthesis in imbibing wheat embryos to the cell-free translational capacities of bulk mRNA from dry and imbibing embryos. J Biol Chem. 1980;255(12):5965-70.
2. Banerjee J, Maiti MK. Functional role of rice germin-like protein1 in regulation of plant height and disease resistance. Biochem Bioph Res Co. 2010;394(1):178-83.
3. Barman AR, Banerjee J. Versatility of germin-like proteins in their sequences, expressions, and functions. Funct Integr Genomic. 2015;15(5):533-48.
4. Segarra CI, Casalongue CA, Pinedo ML, Ronchi VP, Conde RD. A germin-like protein of wheat leaf apoplast inhibits serine proteases. J Exp Bot. 2003;54(386):1335-41.
5. Davidson RM, Reeves PA, Manosalva PM, Leach JE. Germins: A diverse protein family important for crop improvement. Plant Sci. 2009;177(6):499-510.
6. Dunwell JM, Gibbings JG, Mahmood T, Saqlan Naqvi S. Germin and germin-like proteins: evolution, structure, and function. Crit Rev Plant Sci. 2008;27(5):342-75.
7. Park SH, Yi N, Kim YS, Jeong MH, Bang SW, Choi YD, Kim JK. Analysis of five novel putative constitutive gene promoters in transgenic rice plants. J Exp Bot. 2010;61(9):2459-67.
8. Cartharius K, Frech K, Grote K, Klocke B, Haltmeier M, Klingenhoff A, Frisch M, Bayerlein M, Werner T. MatInspector and beyond: Promoter analysis based on transcription factor binding sites. Bioinformatics. 2005;21(13):2933-42.
9. Zhao J, Wang J, An L, Doerge R, Chen ZJ, Grau CR, Meng J, Osborn TC. Analysis of gene expression profiles in response to Sclerotinia sclerotiorum in Brassica napus. Planta. 2007;227(1):13-24.
10. Chawade A, Bräutigam M, Lindlöf A, Olsson O, Olsson B. Putative cold acclimation pathways in Arabidopsis thaliana identified by a combined analysis of mRNA co-expression patterns, promoter motifs and transcription factors. BMC Genomics. 2007;8(1):304.
11. Yi J, Derynck MR, Chen L, Dhaubhadel S. Differential expression of CHS7 and CHS8 genes in soybean. Planta. 2010;231(3):741-53.
12. Ibraheem O, Botha CEJ, Bradley G. In silico analysis of cis-acting regulatory elements in 5′ regulatory regions of sucrose transporter gene families in rice (Oryza sativa Japonica) and Arabidopsis thaliana. Comput Biol Chem. 2010;34(5–6):268-83.
13. Yadav DK, Shukla D, Tuteja N. Rice heterotrimeric G-protein alpha subunit (RGA1): In silico analysis of the gene and pro-

moter and its upregulation under abiotic stress. Plant Physiol Biochem. 2013;63:262-71.
14. Federico ML, Iñiguez-Luy FL, Skadsen RW, Kaeppler HF. Spatial and temporal divergence of expression in duplicated barley germin-like protein-encoding genes. Genetics. 2006;174(1):179-90.
15. Fan Z, Gu H, Chen X, Song H, Wang Q, Liu M, Qu LJ, Chen Z. Cloning and expression analysis of Zmglp1, a new germin-like protein gene in maize. Biochem Biophys Res Commun. 2005;331(4):1257-63.
16. Sassaki FT, Bravo JP, González ER, Maia IG. Expression pattern and promoter analysis of a Eucalyptus grandis germin-like gene. Plant Mol Biol Rep. 2014;33(1):12-21.
17. Mathieu M, Neutelings G, Hawkins S, Grenier E, David H. Cloning of a pine germin-like protein (GLP) gene promoter and analysis of its activity in transgenic tobacco Bright Yellow-2 cells. Physiol Plant. 2003;117(3):425-34.
18. Chen J-Y, Wen F-X, Chen X-J, Cheng X-Y, Xu H-X. Nucleotide Sequence of Promoter of Triticum aestivum Germin-like Protein 3. Mol Plant Breeding. 2005;3(6):901-2.
19. Himmelbach A, Liu L, Zierold U, Altschmied L, Maucher H, Beier F, Müller D, Hensel G, Heise A, Schützendübel A, Kumlehn J, Schweizer P. Promoters of the barley germin-like GER4 gene cluster enable strong transgene expression in response to pathogen attack. Plant Cell. 2010;22(3):937-52.
20. Yang L, Li T, Zhang S, Gao G, Yang C. Characterization of the GLP13 gene promoter in Arabidopsis thaliana. Biol Plantarum. 2013;57(2):231-7.
21. Ruiz-López N, Haslam RP, Venegas-Calerón M, Li T, Bauer J, Napier JA, Sayanova O. Enhancing the accumulation of omega-3 long chain polyunsaturated fatty acids in transgenic Arabidopsis thaliana via iterative metabolic engineering and genetic crossing. Transgenic Res. 2012;21(6):1233-43.
22. Yasmin T, Mahmood T, Hyder MZ, Akbar S, Naqvi SS. Cloning, sequencing and in silico analysis of germin-like protein gene 1 promoter from Oryza sativa L. ssp. indica. Pakistan J Bot. 2008;40(4):1627-34.
23. Mahmood T, Zeeshan Hyder M, Saqlan Naqvi S. Cloning and sequence analysis of germin-like protein gene 2 promoter from Oryza sativa L. ssp. indica. Mitochondrial DNA. 2007;18(1):26-32.
24. Mahmood T, Nazar N, Yasmin T, Abbasi BH, Ahmad M, Naqvi SMS. Comparative analysis of regulatory elements in different germin-like protein gene promoters. Afri J Biotechnol. 2010;9(13):1871-80.
25. Ali I, Mahmood T. Identification and analysis of regulatory elements in the germin and germin-like proteins family promoters in rice. Turk J Bot. 2015;39:389-400.
26. Mahmood T, Yasmin T, Haque MI, Naqvi SMS. Characterization of a rice germin-like protein gene promoter. Genet Mol Res. 2013;12(1):360-9.
27. Munir F, Hayashi S, Batley J, Naqvi SMS, Mahmood T. Germin-like protein 2 gene promoter from rice is responsive to fungal pathogens in transgenic potato plants. Funct Integr Genomics. 2016:16(1):19-27.
28. Yasmin T, Mahmood T, Hyder MZ, Akbar S, Naqvi SS. Cloning, sequencing and in silico analysis of germin-like protein gene-promoter from Oryza sativa L. ssp. indica. Pakistan J Bot. 2008;40(4):1627-34.
29. Tamura K, Stecher G, Peterson D, Filipski A, Kumar S. MEGA6: molecular evolutionary genetics analysis version 6.0. Mol Biol Evol. 2013;30(12):2725-9.
30. Frech K, Danescu-Mayer J, Werner T. A novel method to develop highly specific models for regulatory units detects a new LTR in GenBank which contains a functional promoter. J Mol Biol. 1997;270(5):674-87.
31. Kafri R, Levy M, Pilpel Y. The regulatory utilization of genetic redundancy through responsive backup circuits. National Acad Sciences. 2006;103(31):11653-8.
32. Ober D. Seeing double: Gene duplication and diversification in plant secondary metabolism. Trends Plant Sci. 2005;10(9):444-9.
33. Druka A, Kudrna D, Kannangara CG, von Wettstein D, Kleinhofs A. Physical and genetic mapping of barley (Hordeum vulgare) germin-like cDNAs. National Acad Sciences. 2002;99(2):850-5.
34. Manosalva PM, Davidson RM, Liu B, Zhu X, Hulbert SH, Leung H, Leach JE. A germin-like protein gene family functions as a complex quantitative trait locus conferring broad-spectrum disease resistance in rice. Plant Physiol. 2009;149(1):286-96.
35. Hsam S, Lapochkina I, Zeller F. Chromosomal location of genes for resistance to powdery mildew in common wheat (Triticum aestivum L. em Thell.). 8. Gene Pm32 in a wheat-Aegilops speltoides translocation line. Euphytica. 2003;133(3):367-70.
36. Zimmermann G, Bäumlein H, Mock H-P, Himmelbach A, Schweizer P. The multigene family encoding germin-like proteins of barley. Regulation and function in basal host resistance. Plant Physiol. 2006;142(1):181-92.
37. Lohmann JU, Hong RL, Hobe M, Busch MA, Parcy F, Simon R, Weigel D. A molecular link between stem cell regulation and floral patterning in Arabidopsis. Cell. 2001;105(6):793-803.
38. Gubler F, Jacobsen JV. Gibberellin-responsive elements in the promoter of a barley high-pI alpha-amylase gene. Plant Cell. 1992;4(11):1435-41.
39. Wang R, Guan P, Chen M, Xing X, Zhang Y, Crawford NM. Multiple regulatory elements in the Arabidopsis NIA1 promoter act synergistically to form a nitrate enhancer. Plant Physiol. 2010;154(1):423-32.
40. Diaz I, Vicente-Carbajosa J, Abraham Z, Martinez M, Isabel-La Moneda I, Carbonero P. The GAMYB protein from barley interacts with the DOF transcription factor BPBF and activates endosperm-specific genes during seed development. Plant J. 2002;29(4):453-64.
41. Guo M, Thomas J, Collins G, Timmermans MC. Direct repression of KNOX loci by the Asymmetric Leaves1 complex of Arabidopsis. Plant Cell. 2008;20(1):48-58.
42. Zhu Y, Cai XL, Wang ZY, Hong MM. An interaction between a MYC protein and an EREBP protein is involved in transcriptional regulation of the rice Wx gene. J Biol Chem. 2003;278(48):47803-11.
43. Rubio-Somoza I, Martinez M, Diaz I, Carbonero P. HvMCB1,a R1MYB transcription factor from barley with antagonistic regulatory functions during seed development and germination. Plant J. 2006;45(1):17-30.

44. Li L, Yu X, Thompson A, Guo M, Yoshida S, Asami T, Chory J, Yin Y. Arabidopsis MYB30 is a direct target of BES1 and cooperates with BES1 to regulate brassinosteroid-induced gene expression. Plant J. 2009;58(2):275-86.
45. Sriraman V, Eichenlaub-Ritter U, Bartsch JW, Rittger A, Mulders SM, Richards JS. Regulated expression of ADAM8 (a disintegrin and metalloprotease domain 8) in the mouse ovary: evidence for a regulatory role of luteinizing hormone, progesterone receptor, and epidermal growth factor-like growth factors. Bio Rep. 2008;78(6):1038-48.
46. Papp B, Pál C, Hurst LD. Evolution of cis-regulatory elements in duplicated genes of yeast. Trends in Genet. 2003;19(8):417-22.
47. Davidson R, Manosalva P, Snelling J, Bruce M, Leung H, Leach J. Rice germin-Like proteins: Allelic diversity and relationships to early stress responses. Rice. 2010;3(1):43-55.
48. Staiger D, Apel K, Trepp G. The Atger3 promoter confers circadian clock-regulated transcription with peak expression at the beginning of the night. Plant Mol Biol. 1999;40(5):873-82.
49. Gao MJ, Lydiate DJ, Li X, Lui H, Gjetvaj B, Hegedus DD, Rozwadowski K. Repression of seed maturation genes by a trihelix transcriptional repressor in Arabidopsis seedlings. Plant Cell. 2009;21(1):54-71.
50. Yoo CY, Pence HE, Jin JB, Miura K, Gosney MJ, Hasegawa PM, Mickelbart MV. The Arabidopsis GTL1 transcription factor regulates water use efficiency and drought tolerance by modulating stomatal density via transrepression of SDD1. Plant Cell. 2010;22(12):4128-41.
51. Xi J, Qiu Y, Du L, Poovaiah BW. Plant-specific trihelix transcription factor AtGT2L interacts with calcium/calmodulin and responds to cold and salt stresses. Plant Sci. 2012;186:274-80.
52. Ogo Y, Kobayashi T, Nakanishi Itai R, Nakanishi H, Kakei Y, Takahashi M, Toki S, Mori S, Nishizawa NK. A novel NAC transcription factor, IDEF2, that recognizes the iron deficiency-responsive element 2 regulates the genes involved in iron homeostasis in plants. J Biol Chem. 2008;283(19):13407-17.
53. Balazadeh S, Kwasniewski M, Caldana C, Mehrnia M, Zanor MI, Xue GP, Mueller-Roeber B. ORS1, an H2O2-responsive NAC transcription factor, controls senescence in Arabidopsis thaliana. Moleculer Plant. 2011;4(2):346-60.
54. Hartmann U, Sagasser M, Mehrtens F, Stracke R, Weisshaar B. 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 Mol Biol. 2005;57(2):155-71.
55. Girin T, Paicu T, Stephenson P, Fuentes S, Korner E, O'Brien M, Sorefan K, Wood TA, Balanzá V, Ferrándiz C, Smyth DR, Østergaard L. Indehiscent and Spatula interact to specify carpel and valve margin tissue and thus promote seed dispersal in Arabidopsis. Plant Cell. 2011;23(10):3641-53.
56. Walcher CL, Nemhauser JL. Bipartite promoter element required for auxin response. Plant Physiol. 2012;158(1):273-82.
57. Petit JM, van Wuytswinkel O, Briat JF, Lobreaux S. Characterization of an iron-dependent regulatory sequence involved in the transcriptional control of AtFer1 and ZmFer1 plant ferritin genes by iron. J Biol Chem. 2001;276(8):5584-90.
58. Nole-Wilson S, Krizek BA. DNA binding properties of the Arabidopsis floral development protein Aintegumenta. Nucleic Acids Res. 2000;28(21):4076-82.
59. McDonald WH, Ohi R, Smelkova N, Frendewey D, Gould KL. Myb-related fission yeast cdc5p is a component of a 40S snRNP-containing complex and is essential for pre-mRNA splicing. Mol Cell Biol. 1999;19(8):5352-62.
60. Chen A, Gu M, Sun S, Zhu L, Hong S, Xu G. Identification of two conserved cis-acting elements, MYCS and P1BS, involved in the regulation of mycorrhiza-activated phosphate transporters in eudicot species. New Phytol. 2011;189(4):1157-69.
61. Schneidereit A, Imlau A, Sauer N. Conserved cis-regulatory elements for DNA-binding-with-one-finger and homeodomain-leucine-zipper transcription factors regulate companion cell-specific expression of the Arabidopsis thaliana Sucrose transporter 2 gene. Planta. 2008;228(4):651-62.
62. Kawakatsu T, Yamamoto MP, Hirose S, Yano M, Takaiwa F. Characterization of a new rice glutelin gene GluD-1 expressed in the starchy endosperm. J Exp Bot. 2008;59(15):4233-45.
63. Simpson SD, Nakashima K, Narusaka Y, Seki M, Shinozaki K, Yamaguchi-Shinozaki K. Two different novel cis-acting elements of erd1, a clpA homologous Arabidopsis gene function in induction by dehydration stress and dark-induced senescence. Plant J. 2003;33(2):259-70.
64. Zou X, Neuman D, Shen QJ. Interactions of two transcriptional repressors and two transcriptional activators in modulating gibberellin signaling in aleurone cells. Plant Physiol. 2008;148(1):176-86.
65. Ramireddy E, Brenner WG, Pfeifer A, Heyl A, Schmulling T. In planta analysis of a cis-regulatory cytokinin response motif in Arabidopsis and identification of a novel enhancer sequence. Plant Cell Physiol. 2013;54(7):1079-92.
66. Kawagoe Y, Campbell BR, Murai N. Synergism between CACGTG (G-box) and CACCTG cis-elements is required for activation of the bean seed storage protein beta-phaseolin gene. Plant J. 1994;5(6):885-90.
67. Gilmartin PM, Chua NH. Spacing between GT-1 binding sites within a light-responsive element is critical for transcriptional activity. Plant Cell. 1990;2(5):447-55.
68. Abe H, Urao T, Ito T, Seki M, Shinozaki K, Yamaguchi-Shinozaki K. Arabidopsis AtMYC2 (bHLH) and AtMYB2 (MYB) function as transcriptional activators in abscisic acid signaling. Plant Cell. 2003;15(1):63-78.
69. Chen W, Chao G, Singh KB. The promoter of a H2O2-inducible, Arabidopsis glutathione S-transferase gene contains closely linked OBF- and OBP1-binding sites. Plant J. 1996;10(6):955-66.
70. Yamamoto MP, Onodera Y, Touno SM, Takaiwa F. Synergism between RPBF Dof and RISBZ1 bZIP activators in the regulation of rice seed expression genes. Plant Physiol. 2006;141(4):1694-707.
71. Lin T, Sharma P, Gonzalez DH, Viola IL, Hannapel DJ. The impact of the long-distance transport of a BEL1-like messenger RNA on development. Plant Physiol. 2013;161(2):760-72.