*Pak. J. Bot.*, 40(4): 1627-1634, 2008.

# CLONING, SEQUENCING AND *IN SILICO* ANALYSIS OF GERMIN-LIKE PROTEIN GENE 1 PROMOTER FROM *ORYZA SATIVA* L. SSP. INDICA

**TAYYABA YASMIN[1], TARIQ MAHMOOD[2], M. ZEESHAN HYDER, SAMINA AKBAR AND S. M. SAQLAN NAQVI[*]**

*Department of Biochemistry, PMAS Arid Agriculture University, Rawalpindi*
*Present address: [1]Department of BioSciences, COMSATS, Islamabad,*
*[2]Department of Plant Sciences, Quaid-i-Azam University, Islamabad*

**Abstract**

Germin and germin-like proteins constitute a large family of plant proteins which are also considered as germination markers due to their high expression levels during germination/early growth. These proteins are known to be involved in many stress related processes as well, but their biochemical functions and physiological roles have not been fully described. In order to gain insight into the functions and regulation of a rice germin-like protein gene 1, about 1.2 kb of its upstream region was amplified, cloned, sequenced and analyzed. Analysis showed that this promoter has a very little homology with the promoters of other GLP genes within rice genome. The promoter contains putative regulatory elements of diverse functions and has distinct copy number, location and clustering pattern of regulatory elements in its sequence. This promoter being unique requires further characterization to explore its regulatory role.

## Introduction

Germin and germin-like proteins belong to a diverse family of plant proteins showing extreme resistance to heat and chemical degradation, and participate in many processes that are important for plant development and defense (Bernier & Berna, 2001). They are functionally diverse but structurally related to members of the cupin super family, which includes isomerases, cyclases, dioxygenases, sugar or auxin-binding proteins as well as monomeric or dimeric globulin seed storage proteins, such as phaseolin (Dunwell *et al*., 2000). Hence, germins and GLPs are known to play a wide variety of roles as enzymes, structural proteins and receptors (Zimmermann *et al*., 2006). As enzymes, germins have oxalate oxidase activity while some GLPs have superoxide dismutase (SOD) or phosphodiesterase activity.

The expression of germins and GLPs is known to be associated with the biotic and abiotic stresses. Over expression of wheat germin in sunflowers (*Helianthus annuus*) influences some defense-related transcripts and increases resistance to pathogens (Hu *et al*., 2003). Silencing of a germin-like gene in *Nicotiana attenuata* improves performance of native herbivores (Lou & Baldwin, 2006). A germin-like protein (McGLP) is expressed in roots of *Mesembryantemum crystallinum* in response to salt stress. Moreover, tobacco leaf apoplast proteome showed changes in response to salt stress and level of a germin-like protein increased significantly along with chitinases (Dani *et al*., 2005). Despite these and a growing number of other available reports regarding the GLPs, functions of many GLPs remain biochemically uncharacterized.

[*] Corresponding Author: saqlan@uaar.edu.pk

A functional insight into GLPs can be acquired through the analysis of corresponding promoter regions. Identification of transcriptional regulatory elements within promoter regions is of great interest for biologists since these elements govern the regulation of gene expression. Knowledge about different core elements of promoters can unveil the function/regulation of genes. The main rationale behind the present study was the cloning and sequencing of a rice germin-like protein gene 1 (*Os*RGLP1) promoter followed by computational analysis for identification/characterization of possible putative regulatory elements in it.

## Material and Methods:

**PCR amplification, cloning and sequencing:** Rice Seeds (Nonabokra) obtained from International Rice Research Institute (IRRI) were grown under green house conditions and DNA from leaf tissues was extracted by CTAB method (Richards, 1997).

The primers RGLP1P-F1:5' CCCGGGACCAACGAAAAGATTGAACA 3' and RGLP1P-R1:5' CCCGGGCATTTGTCCATGGAGAGGAT 3' were designed on chromosome 8 of *Oryza sativa* (Genbank accession AP008214) to amplify the upstream sequence of *Os*RGLP1 gene located.

The PCR mixture contained 50 pM of each primer, 1X PCR buffer, 15 mM $MgCl_2$, 2 mM each of the dNTPs and 1.5 units of *Taq* polymerase (Fermentas). PCR was performed with a pre-amplification denaturation at 94 $^oC$ for 5 minutes followed by thirty five cycles each of 94 $^oC$ for 30 seconds, 55 $^oC$ for 30 seconds and 72 $^oC$ for 1 minute, and final extension at 72$^oC$ for 10 minutes. PCR products were ligated into TA cloning vector pTZ57R (Fermentas) according to the manufacturer instructions and electroporated into DH5α cells at medium voltage (2.5 KV/0.15 cm) in *E. coli* Porator (Life Technologies GBCO BRL). The cells were selected for resistance to ampicillin and screened for the inactivation of β-glycosidase gene according to Sambrook & Russel (2001).

The plasmid was isolated by Eppendorf Fastplasmid$^{TM}$ Mini kit and sequencing was carried out using CEQ™ Dye Terminator Cycle sequencing kit (Beckman Coulter Inc. USA) according to the manufacturers' instructions with universal M13 and gene specific primers and run on the CEQ 8000 DNA Analyzer (Beckman Coulter Inc. USA).

**Sequence analysis:** The promoter sequence was aligned with other GLP promoters using CLUSTAL W (Thompson *et al*., 1994) available at www.ebi.ac.uk/clustalw to find out homology. Presence of open reading frame was searched through GENSCAN (Burge & Karlin, 1997) available at http://genes.mit.edu/GENSCAN.html. To find out the regulatory elements, their types and positions, whole promoter sequence was analyzed using PLACE/Signal Scan (Higo *et al*. 1999) available at www.dna.affrc.go.jp. Many regulatory elements were found by Signal Scan on both the strands of promoter; however only cis elements without any ambiguous nucleotide were considered and further characterized.

## Results and Discussion

**Promoter sequence analysis:** Sequencing of the cloned fragment revealed a stretch of 1228 bp. The sequence of this promoter region has been submitted to Genbank

(Accession no. EU742684). Alignment of this sequence with other rice GLP gene promoters by CLUSTAL W showed no significant homology. GENSCAN predicted a 40 bp promoter region, extending two nucleotides into the transcriptional region of *Os*RGLP1 gene.

**Mapping of the regulatory elements on *Os*RGLP1:** Many putative regulatory elements of variable lengths were identified by PLACE/Signal Scan (Table 1) and mapped on the promoter (Figure 1).The putative elements were distributed evenly in two main clusters (-40 to -650 and -800 to -1120) on whole promoter sequence except from -1120 to -1228 bp region where density of elements was very much low, suggesting that the promoter lies within -1120 region of the cloned sequence and that whole of it has been sequenced.

**Reported functions of putative elements:** The search of regulatory elements by the Signal Scan resulted in number of noteworthy elements with diverse functions.

**Tissue and developmental stage specific elements:** The expression of germin like proteins is reported to be tissue and developmental stage specific (Schlesier *et al*., 2004), this necessitates the presence of regulatory elements which can confer tissue and developmental stage specificity in the upstream regions of GLP genes. Regulatory element AGAAA is reported to be responsible for pollen specific activation of tomato *lat52* gene, also found in the promoter of tomato endo-beta-mannanase gene which is involved in hemicellulose degradation (Filichkin *et al*., 2004). Four copies of this element were found in *Os*RGLP1 promoter region. Seven copies of ATATT element are located at different positions on *Os*RGLP1 promoter region, which is reported to be expressed specifically in root elongation zone and in vascular tissue (Elmayan & Tepfer, 1995).

GTGA element was first reported by Rogers *et al* (2001) as *cis*-regulatory element within the promoter of the tobacco late pollen gene *g10*. Six copies of this element were present on *Os*RGLP1 promoter. Six copies of CAAT were also observed in the promoter region at different locations (table 1) which is known for tissue specific activity of a pea legumin gene in tobacco (Shirsat *et al.*, 1989). This observation may indicate the common regulatory mechanism of *Os*RGLP2 and *Os*RGLP1 promoters.

**Elements related to inducible expression:** Presence of five copies of CAAT box have been reported in *Os*RGLP2 promoter which is located upstream of a germin like protein gene in rice (Mahmood *et al*., 2007). The involvement of CAAT box in inducible gene expression is well established. Upstream regions of light-regulated genes in both monocots and dicots were reported to contain GATAA conserved sequence, also designated as I-box (Terzaghi & Cashmore, 1995). There are three copies of this element located at various positions in the promoter region of *Os*RGLP2 and similar number has been reported for *Os*RGLP2 promoter as well. Light regulated gene expression in plants involves another important element, GATA which was located on *Os*RGLP2 promoter at seven positions (Mahmood *et al*., 2007). Interestingly the same number of GATA element has been observed during the analysis of *Os*RGLP1 promoter region (Mahmood *et al*., 2007).

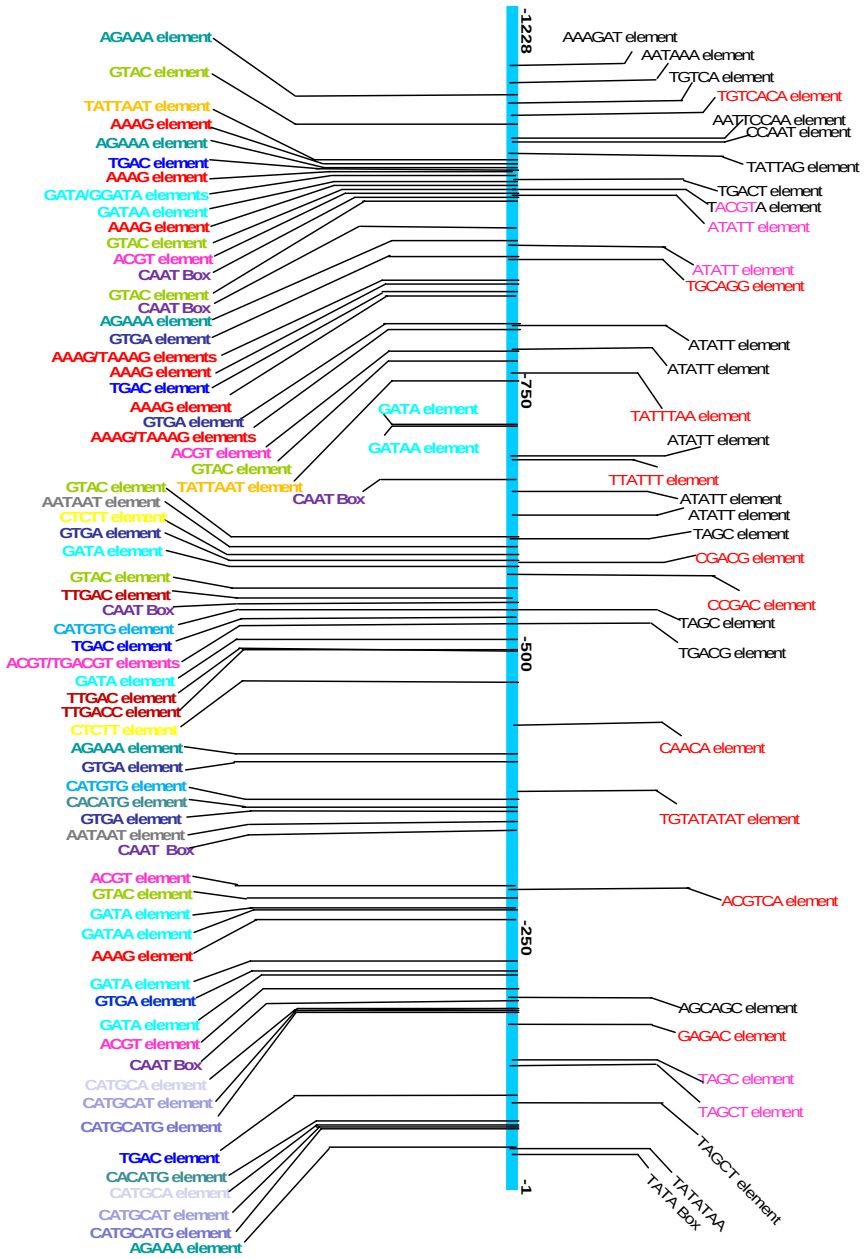


**Figure 1: Description of regulatory elements found in full length cloned region.**

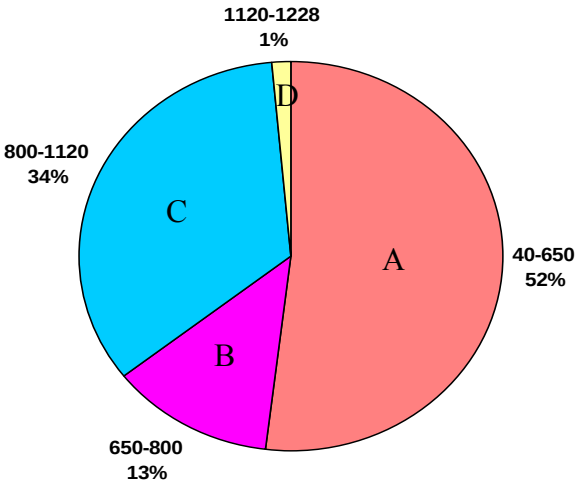


**Figure 2:** Distribution/density of regulatory elements on *Os*RGLP1 promoter.
A-C=High density of regulatory elements
D= Lowest density of regulatory elements

Five copies of ACGT have been detected on promoter sequence of *Os*RGLP1 while only one copy of ACGT was observed on *Os*RGLP2 promoter region (Mahmood *et al*., 2007). ACGT is reportedly involved in drought and senescence responses (Simpson *et al*., 2003). Seven copies of another element GTAC are present in *Os*RGLP1 promoter. GTAC has been identified as core of a copper-response element (CuRE) found in promoters of Cyc6 and Cpx1 genes in Chlamydomonas and is also involved in oxygen-response of these genes (Quinn *et al*., 2002).

**Biotic or abiotic stress:** Plants, like other living organisms are exposed to various biotic and abiotic stresses and they have to cope with such situations. The expression of various proteins involved in stress responses is regulated through the interaction of transcription factors with upstream regions of genes. A number of stress related regulatory elements have been identified which are specifically involved in expression of their cognate genes. AAAG is a core site required for binding of Dof proteins, which are DNA binding proteins involved in expression of photosynthetic genes, seed storage protein genes and genes responsive to various stresses (Yanagisawa and Schmidt, 1999; Kisu *et al*., 1998; Mena *et al*., 1998). It is quite interesting to note that *Os*RGLP1 promoter contains nine copies of this element which are located at different positions. Strikingly, three copies of AAAG have been clustered in a very small region of 20 bp only (from -1022 to -1041) on *Os*RGLP1 promoter. Presence of ten copies of this element reported in *Os*RGLP2 promoter (Mahmood *et al*., 2007) is remarkable and may highlight the involvement of similar regulatory behavior of both GLP promoters.

**Table 1: Regulatory elements found in promoter of *Os*RGLP1**

| Regulatory element | Number of copies | Location | Functions | Reference |
|---|---|---|---|---|
| **GTAC [core of a CuRE (copper-response element)]** | 7 | -1114,-1018,-1011,-776,-648,-581,-280 | Involved in oxygen-response of Cyc6 and Cpx1 genes | Quinn *et al*., 2002 |
| **AAAG** | 9 | -1213,-1038,-1034,-1022,-937,-926,-897,-866,-259 | Binding site for Dof proteins | Yanagisawa and Schmidt, 1999 |
| **TGAC (Core sequence in W box)** | 7 | -1087,-946,-568,-555,-501,-111, -95 | Pathogen inducibility, gibberellin signaling pathway | Eulgem *et al*., 2000 Wang *et al.*, 1998 |
| **AGAAA** | 4 | -1035,-961,-427,-47 | Pollen specific expression | Filichkin *et al.* 2004 |
| **GATA** | 7 | -211,-239,-272,-525,-622,-729,-1030 | light regulated and tissue specific gene expression | Lam and Chua, 1989 |
| **ACGT** | 5 | -1016,-796,-553,-295, -192 | Dark and senescence-induced expression | Simpson *et al*., 2003 |
| **GTGA** | 6 | -947,-874,-624,-415,-361, -220 | Found in promoter of tobacco late pollen gene g10 | Rogers *et al.*, 2001 |
| **GATAA** | 3 | -1029,-726,-271 | Function in expression of light-regulated genes | Terzaghi, and Cashmore, 1995 |
| **TATTAAT** | 2 | -1055,-750 | Transcription initiation | |
| **TTGAC/TTGACC** | 3 | -568,-501,-95 | Pathogen inducibility, gibberellin signaling pathway | Eulgem *et al*., 2000 Wang *et al.*, 1998 |
| **CATGTG** | 2 | -563,-362 | Involved in water stress | Simpson *et al*., 2003 |
| **CACATG** | 2 | -364,-73 | Involved in water stress | Abe *et al*., 1997 |
| **CAAT** | 6 | -146, -330,-545,-688,-957, -1072 | tissue specific promoter activity | Shirsat *et al.,* 1989 |
| **ATATT** | 7 | -661, -683, -695, -763, -921,- 1059, -1069 | Root specific expression | Elmayan *et al*., (1995) |

The W box [(T) TGAC(C/T)] is the binding site for members of the WRKY family of transcription factors. There is increasing evidence that W boxes are a major class of *cis* -acting elements responsible for the pathogen inducibility of many plant genes (Wang *et al*., 1998) and, are also involved in senescence and trichome development (Eulgem *et al*., 2000). W box was also found to be involved as an enhancer in the induction of salicylic acid in response to wounding and bacterial infection in *Brassica oleracea*. The TGAC element was first reported in rice (Eulgem *et al*., 1999) and found to be involved in gibberellin signaling pathway. The TGAC motif is found as an invariant core sequence in W box which serves as a binding site for a rice transcription factor, WRKY71. Seven copies of TGAC are present in the promoter region of *Os*RGLP1. The TTGAC/ TTGACC element is also part of W box and reportedly involved in pathogen inducibility in Parsley (Rushton *et al.*, 1996). Three copies of this element are also detected in the promoter region under study.

Two elements, CATGTG and CACATG reported to be involved in water stress (Simpson *et al*., 2003) response were also found on the promoter region with two copies of each, indicating that this promoter might be involved in water stress responses.

Analysis further showed that the promoter region has regulatory elements related to gibberellin signaling pathway and defense. Various regulatory elements were found during Signal Scan analysis of promoter region, at different locations. The regulatory elements varied in density and ranged from 2 to 9 in the sequenced region. The elucidation of significance of promoter in regulation however needs further confirmation by mutations/truncations followed by *in vivo* analysis of expression driven by the promoter.

**Conclusion:** The *Os*RGLP1 promoter was sequenced and cloned in *E.coli*. Sequence analysis did not show any significant homology with other GLP gene promoters in rice. Many putative regulatory elements were identified in *Os*RGLP1 promoter sequence. The elements were distinct with respect to copy number, position on the promoter and clustering pattern. The regulatory elements seem to be associated with diverse functions.

**Acknowledgement:** This work was supported by the grant #01-03-01-19 from Agricultural Linkages Program of the Pakistan Agricultural Research Council (PARC) to S.M. Saqlan Naqvi.

**References**

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

Burge, C. and S. Karlin. 1997. Prediction of complete gene structures in human genomic DNA. *J. Mol. Biol*., 268(1):78-94.

Dunwell, J.M., S. Khuri and P.J. Gane. 2000. Microbial relatives of the seed storage proteins of higher plants: conservation of structure and diversification of function during evolution of the cupin superfamily. *Microbiol. Mol. Biol. Rev*., 64: 153-179.

Eulgem, T., P.J. Rushton, E. Schmelzer, K. Hahlbrock and I.E. Somssich.1999. Early nuclear events in plant defense signaling: rapid gene activation by WRKY transcription factors *EMBO J*., 18: 4689-4699.

Eulgem, T., P.J. Rushton, S. Robatzek and I.E. Somssich. 2000. The WRKY superfamily of plant transcription factors. *Trends Plant Sci*., 5:199-206.

Elmayan, T. and M. Tepfer .1995. Evaluation in tobacco of the organ specificity and strength of the rol D promoter, domain A of the 35S promoter and the 35S 2 promoter. *Transgenic Res*., 4: 388-396.

Filichkin, S.A., J.M. Leonard, A. Monteros, P.P. Liu and H. Nonogaki. 2004. A novel endo-beta-mannanase gene in tomato LeMAN5 is associated with anther and pollen development. *Plant Physiol*., 134: 1080-1087.

Higo, K., Y. Ugawa, M. Iwamoto and T. Korenaga. 1999. Plant cis-acting regulatory DNA elements (PLACE) database. *Nucleic Acids Res*., 27(1):297-300.

Hu, X., D.L. Bidney, N. Yalpani, J.P. Duvick, O. Crasta, O. Folkerts and G. Lu. 2003. Over expression of a gene encoding hydrogen peroxide-generating oxalate oxidase evokes defense responses in sunflower. *Plant Physiol*., 133: 170-181.

Kisu Y., T. Ono, N. Shimofurutani, M. Suzuki and M. Esaka. 1998. Characterization and expression of a new class of zinc finger protein that binds to silencer region of ascorbate oxidase gene. *Plant Cell Physiol*., 39:1054-1064.

Lam E. and N.H. Chua. 1989. ASF-2: A factor that binds to the cauliflower mosaic virus 35S promoter and a conserved GATA motif in cab promoters. *Plant Cell*, 1:1147-1156.

Lou Y. and I.T. Baldwin. 2006. Silencing of a Germin-Like Gene in *Nicotiana attenuata* Improves Performance of Native Herbivores. *Plant Physiol*., 140:1126-1136.

Mahmood T., M.Z. Hyder and S.M. Saqlan Naqvi. 2007. Cloning and sequence analysis of germin-like protein gene 2 promoter from *Oryza sativa* L. ssp. Indica. *DNA Sequence*, 18 (1): 26 - 32.

Mena M., J.Vicente-Carbajisa, R.J. Schmidt and P. Carbonero. 1998. An endosperm-specific Dof protein from barley, highly conserved in wheat, binds to and activates transcription from the prolamin box of a native b-hordein promoter in barley endosperm. *Plant J*., 16:53-62.

Quinn, J.M., M. Eriksson, J.L. Moseley and S. Merchant. 2002. Oxygen deficiency responsive gene expression in Chlamydomonas reinhardtii through a copper-sensing signal transduction pathway. *Plant Physiol*., 128:463-471.

Richards, E.J. 1997. Preparation of plant DNA using CTAB. In: *Short protocols in molecular biology*. Pp. 2.10-2.11. (Eds.): F. Ausubel, R. Brent, R.E. Kingston, D.D. Moore, J.G. Seidman, J.A. Smith and K. Struhl. John Wiley, New York.

Rogers, H.J., N. Bate, J. Combe, J. Sullivan, J. Sweetman, C. Swan, D.M. Lonsdale and D. Twell. 2001. Functional analysis of cis-regulatory elements within the promoter of the tobacco late pollen gene g10. *Plant Mol. Biol*., 45: 577-585.

Rushton, P.J., J.T. Torres, M. Parniske, P. Wernert, K. Hahlbrock and I.E. Somssich. 1996. Interaction of elicitor-induced DNA-binding proteins with elicitor response elements in the promoters of parsley PR1 genes. *EMBO J*., 15:5690-5700.

Sambrook, J. and D.W. Russell. 2001. *Molecular Cloning*: A Laboratory Manual. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, New York.

Schlesier, B., A. Berna, F. Bernier and H-P. Mock. 2004. Proteome analysis differentiates between two highly homologous germin-like proteins in *Arabidopsis thaliana* ecotypes Col-0 and Ws-2. *Phytochemistry*, 65 (11):1565-1574.

Shirsat, A., N. Wilford, R. Croy and D. Boulter.1989. Sequences responsible for the tissue specific promoter activity of a pea legumin gene in tobacco. *Mol. Genet*., 215:326-331.

Simpson, S.D., K. Nakashima, Y. Narusaka, M. Seki, K. Shinozaki and K.R.A. Yamaguchi-Shinozaki. 2003. 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*., 33: 259-270.

Terzaghi, W.B. and A.R. Cashmore.1995. Light-regulated transcription. *Annu Rev Plant Physiol Plant Mol. Biol*., 46:445-474.

Thompson, J.D., D.G. Higgins and T.J. Gibson. 1994. CLUSTALW: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. *Nucleic Acids Res*., 22(22):4673-80.

Wang, Z., P. Yang, B. Fan and Z. Chen. 1998. An oligo selection procedure for identification of sequence-specific DNA-binding activities associated with plant defense. *Plant J*., 16: 515-522.

Yanagisawa, S., and R.J. Schmidt. 1999. Diversity and similarity among recognition sequences of Dof transcription factors. *Plant J*., 17:209-214.

Zimmermann, G., H. Baumlein, H.P. Mock, A. Himmelbach and P. Schweizer. 2006. The Multigene Family Encoding Germin-Like Proteins of Barley; Regulation and Function in Basal Host Resistance. *Plant Physiol*., 142:181-192.

(Received for publication 21 April, 2008)