Wheat wounding-responsive HD-Zip IV transcription factor GL7 is predominantly expressed in grain and activates genes encoding defensins [S]

Nataliya Kovalchuk[1], Wei Wu[1,†], Natalia Bazanova[1,‡], Nicolas Reid[1], Rohan Singh[1], Neil Shirley[1], Omid Eini[2], Alexander A. T. Johnson[3], Peter Langridge[1], Maria Hrmova[4,1,*] and Sergiy Lopato[1]

[1]School of Agriculture, Food and Wine, University of Adelaide, Glen Osmond, South Australia 5064, Australia.
[2]Department of Plant Protection, School of Agriculture, University of Zanjan, Zanjan, Iran.
[3]School of BioSciences, The University of Melbourne, Victoria 3010, Australia.
[4]School of Life Sciences, Huaiyin Normal University, China.

Current addresses:
[†]Agronomy College, Sichuan Agricultural University, Yaan 625014, China.
[‡]Commonwealth Scientific and Industrial Research Organisation, Glen Osmond 5064, South Australia, Australia.

*To whom the correspondence should be addressed; E-mail: maria.hrmova@adelaide.edu.au; Tel.: +61 8 8313 7160.

[S] This article contains Figs. 1-9, Supporting Table 1 and Supporting Fig. 1. The total word count in the manuscript is 9,323, including legend descriptions, and excluding abstract and references.

Running title: A wounding-inducible HD-Zip IV transcription factor activates defensins.

Abbreviations: AP2, APETALA2; bZIP, basic Leucine Zipper; CPL, 1-palmitoyl-2-linoleoyl-sn-glycero-3-phosphocholine; DAP, days after pollination; DLP, 1,2-dilinoleolyl-sn-glycero-3-phosphocholine; ERF, Ethylene-Responsive Element binding factor; GL2, GLABRA; GL7, GLABRA2-like clone 7; HD, homeodomain; HSF, Heat-shock transcription factor; MYB, myeloblastosis; PDB, Protein Data Bank; PDF1, PROTODERMAL FACTOR1; SEM, standard errors of mean; START, Steroidogenic acute regulatory protein-related lipid-transfer; TF(s),

transcription factor(s); UTR, 3' untranslated region; ZIP, leucine zipper; ZLZ, zipper-loop-zipper; Y1H, yeast-1-hybrid; Y2H, yeast-2-hybrid

Keywords: barley; biotechnology; molecular model; rice; structural bioinformatics; wheat; wounding; yeast-1-hybrid.

GenBank Database (https://www.ncbi.nlm.nih.gov/genbank/) accession numbers: *TaGL7* cDNA – MK583312; *TdGL7* cDNA – MK583313; *TdGL7* promoter – MK583314. Accessions will be released upon publication.

Abstract

HD-Zip class IV transcription factors constitute a family of multidomain proteins. A full-length cDNA of HD-Zip IV, designated *TaGL7* was isolated from the developing grain of bread wheat, using a specific DNA sequence as bait in the Y1H screen. 3D models of TaGL7 HD complexed with DNA *cis*-elements rationalised differences that underlined accommodations of binding and non-binding DNA, while the START-like domain model predicted binding of lipidic molecules inside a concave hydrophobic cavity. The 3'-untranslated region of *TaGL7* was used as a probe to isolate the genomic clone of *TdGL7* from a BAC library prepared from durum wheat. The spatial and temporal activity of the *TdGL7* promoter was tested in transgenic wheat, barley and rice. *TdGL7* was expressed mostly in ovary at fertilisation and its promoter was active in a liquid endosperm during cellularisation and later in the endosperm transfer cells, aleurone, and starchy endosperm. The pattern of *TdGL7* expression resembled that of genes that encode grain-specific lipid transfer proteins, particularly defensins. In addition, *GL7* expression was upregulated by mechanical wounding, similarly to defensin genes. Co-bombardment of cultured wheat cells with *TdGL7* driven by constitutive promoter and seven grain or root specific defensin promoters fused to *GUS* gene, revealed activation of four promoters. The data confirmed the previously proposed role of HD-Zip IV transcription factors in the regulation of genes that encode lipid transfer proteins involved in lipid transport and defence. The *TdGL7* promoter could be used to engineer cereal grains with enhanced resistance to insects and fungal infections.

Introduction

Transcription factors (TFs) containing homeodomain (HD) and leucine zipper (Zip) motifs constitute a large family of plant-specific proteins known as HD-Zip TFs (reviewed in Ariel et al.

2007; Chew et al. 2013; Bürglin et al. 2016). Transcription factors of this family have been grouped in four classes I to IV, based on their domain structure and DNA-binding specificity (Sessa et al. 1998). The HD-Zip class IV from *Arabidopsis*, which is also known as HD-GL2, because GLABRA2 (GL2) from *Arabidopsis* was the first identified member of this TF subfamily (Rerie et al. 1994), contains sixteen members. Classification and partial characterisation of the *Arabidopsis* HD-Zip IV subfamily revealed that five of sixteen genes of this subfamily were found expressed specifically in flowers and/or siliques Nakamura et al. (2006). Four of these genes are homologues to grain specific *OsTF1* from rice and *TaGL9* from wheat (Yang et al. 2002; Kovalchuk et al. 2012b).

The cis-element specifically activated by HD-Zip IV, TAAATG(C/T)A, was identified in the promoter of the *Arabidopsis PROTODERMAL FACTOR1* (*PDF1*) gene encoding a putative extracellular proline-rich protein, which is exclusively expressed in the L1 layer of shoot apices and the protoderm of the organ primordial (Abe et al. 2001). Several variants of the palindromic sequences including the L1-element were isolated using the PCR-assisted random oligonucleotide selection. The selection was performed in bacteria using purified HD-Zip IV proteins from sunflower and *Arabidopsis* (Tron et al. 2001; Nakamura et al. 2006). Recognition of selected oligonucleotide sequences by HD-Zip IV proteins from other plants was also demonstrated (Tron et al. 2001; Zhang et al. 2010; Kovalchuk et al. 2012b). Using a combination of yeast and plant trans-activation assays, the transcriptional activation domain of OCL1 from maize was localised at the N-terminal part of the START (Steroidogenic acute regulatory protein-related lipid-transfer) domain. A version of OCL1 with deleted activation domain was unable to *trans*-activate a synthetic reporter gene controlled by minimal promoter fused to six repeats of the L1 box (Depège-Fargeix et al. 2011). Furthermore, transcriptional activation activity of the rice HD-Zip IV TF ROC4 was also identified in the N-terminal region of the StAR-related lipid-transfer domain (Wei et al. 2016). In addition, HD-ZIP IV TFs contain CxxC motifs, which are situated downstream of or within the dimerisation leucine Zip motifs. Such combination of two motifs was designated the zipper-loop-zipper (ZLZ) motif (Ciarbelli et al. 2008; Nakamura et al. 2006). It has been suggested that the intracellular redox state could influence the activity of these TFs via cysteine residues of the ZLZ motif (Tron et al. 2002).

Expression of several HD-Zip IV factors in the upper cell layers of developing embryos and grains was demonstrated in variety of species by *in situ* hybridisation (Lu et al. 1996; Ingram et al. 2000; Ito et al. 2002; Yang et al. 2002; Ingouff et al. 2001; Ingouff et al. 2003, Guan et al. 2008) and by the analysis of transgenic plants transformed with the promoter-GUS fusion constructs (Nakamura et al. 2006; Kovalchuk et al. 2012b; Dwivedi et al. 2014).

However, not all *HD-Zip IV* genes are expressed in the epidermal cell layer, and expression of some of these genes was detected in subepidermal cell layers (Kubo et al. 1999; Ingram et al. 2000), in an outer cortical layer and the root cap of embryo and seedlings (Ingouff et al. 2003), in lateral root tips (Nakamura et al. 2006), in abscission and nutrient-transfer zones at the bases of floral organs (Dwivedi et al. 2014), in the tapetum and pollen grains (Nakamura et al. 2006), throughout developing embryo and in the whole or part of the starchy endosperm (Yang et al. 2002; Nakamura et al. 2006; Kovalchuk et al. 2012b), and in the main vascular bundle of scutellum (Kovalchuk et al. 2012b).

Several cell-wall and cuticle related genes were identified among target genes of HD-Zip IV TFs. For instance, *PDF1*, a gene encoding the epidermis-specific Pro-rich protein, is directly regulated by ATML1/PDF2 through the L1 *cis*-element (Abe et al. 2003). Genes encoding phospholipase D, AtPLDζ1, the cellulose synthase CESA5, and the xyloglucan endotransglucosylase XTH17 are directly regulated by GL2 (Ohashi et al. 2003; Tominaga-Wada et al. 2009). Fourteen downstream and one target genes were identified as regulated by the translation product of the *OCL1* gene from maize by transcriptome sequencing (Javelle et al. 2010). Most of these genes are related to lipid transport and metabolism, and to cuticle deposition and biosynthesis.

HD-Zip IV TFs are large proteins, containing several domains, and therefore it is not surprising that they can interact with several other proteins. For instance, two bHLH TFs, AtCFL1 associated protein 1 (CFLAP1) and CFLAP2, interact with HD-Zip IV TF AtCFL1 both *in vitro* and *in vivo* (Li et al. 2016). This protein complex is also involved in regulation of fatty acids, cutin and wax biosynthesis pathways, and the epicuticular crystalline wax loading. Both HDG1 and CFLAP1/2 interact with the same C-terminal C4 zinc finger domain of AtCFL1, which is essential for AtCFL1 function (Li et al. 2016).

The HD-Zip IV TF AaHD8 from *Artemisia annua*, a positive regulator that promotes the expression of another HD-Zip IV TF AaHD1, controls the glandular trichome formation and facilitates the leaf cuticle development through activation of the expression of several cuticle biosynthetic genes. It was found in a yeast-2-hybrid (Y2H) assay that AaHD8 can physically interact with MYB-type TF AaMIXTA1, which is a well-known positive regulator of cuticle-related genes. This interaction leads to the enhanced transcriptional activity of AaMIXTA1 (Yan et al. 2018).

In this work we cloned two genes of HD-Zip IV TF *GLABRA2-like clone 7* (*GL7*) from bread and durum wheat. Preferential expression of *TaGL7* in the liquid fraction of endosperm during

cellularisation was observed. Both *TaGL7* and *TdGL7* were rapidly and strongly induced by mechanical wounding in wheat leaf and grain. Wheat, barley and rice were stably transformed with the promoter-GUS fusion construct and the spatial and temporal activity of the *TdGL7* promoter was studied using histological and/or whole-mount assays. We found that the spatial and temporal patterns of the *TdGL7* promoter overlapped with earlier reported expression patterns of wheat grain-specific lipid transfer proteins (LTPs), in particular for defensins (Kovalchuck et al. 2010). The activation of three wounding-inducible defensin promoters by TaGL7 was demonstrated in a transient expression assay in wheat cells.

Materials and Methods

Gene cloning and plasmid construction

Full-length cDNA of *TaGL7* was isolated in the yeast-1-hybrid (Y1H) screen using a cDNA library prepared from the whole grain of *Triticum aestivum* collected at 0-6 days after pollination (DAP) according to the procedure described by Pyvovarenko and Lopato (2011). The four-fold repeated *cis*-element 5'-CATTAAATG-3' from *Arabidopsis* specific for HD-Zip IV TFs (Abo et al. 2001) was used as a bait. Of 48 positive clones, seven contained inserts longer than 2 kb. Two positive clones had a 2.6-kb long inserts, while five clones contained a 3.3-kb long inserts. Sequencing revealed that the differently sized inserts encoded two cDNAs containing full-length coding regions of HD-Zip IV TFs. The 2.6 kb long insert encoded the earlier reported TaGL9 TF (Kovalchuk et al. 2012b). The remaining five clones contained inserts of the same length (exact size 3,281 bp) and encoded HD-Zip IV TF of 883 residues long, designated *T. aestivum* GLABRA2-like clone 7 (TaGL7). The genomic sequence of the TaGL7 was not available at the time. Therefore, the 3' untranslated region (UTR) of *TaGL7* was used as a probe to screen a bacterial artificial chromosome (BAC) library prepared from genomic DNA of *Triticum turgidum* ssp. *durum* cv. Langdon (Cenci et al. 2003) using Southern blot hybridisation. Three BAC clones were selected for further analysis based on the strength of hybridisation signals. BAC DNA was isolated and used as a template for PCR with several primer pairs derived from the coding region of *TaGL7*. Two BAC clones produced the same predicted PCR product; one of them (#1094 M11) was used in a further work. The whole BAC clone #1094 M11 was sequenced using 454 sequencing technology (Life Sciences, Branford, CT, USA) and the full-length sequence of the gene (8,258 bp) including more than 4 kb of the promoter region of the *TaGL7* orthologue from *T. turgidum ssp durum* were assembled as a non-interrupted contig. The gene sequence was subsequently used to design forward and reverse primers for the isolation of the promoter of

TaGL7 (Supporting Table 1). The promoter fragment corresponding to 3,046 bp upstream of the translational start of *TdGL7* was amplified by PCR using AccuPrime™ Pfx DNA polymerase (Invitrogen, Carlsbad, CA, USA) from DNA of the BAC clone #1094 M11 as a template. This promoter was cloned into the pENTR-D-TOPO vector (Invitrogen); the cloned insert was verified by sequencing and subcloned into the pMDC164 vector (Curtis and Grossniklaus, 2003) using recombination cloning. The resulting binary vector, designated pTdGL7, was introduced into the *Agrobacterium tumefaciens* AGL1 strain by electroporation. For wheat transformation, the pTdGL7 construct was linearised using the unique *Pme*I site in the vector sequence. The coding region of *TdGL7* was isolated by nested RT-PCR using two sets of primers designed from the genomic sequence of *TdGL7* and a mixture of RNA samples isolated from developing grain (5-20 DAP) of *T. turgidum ssp. durum* cv. Langdon, as template. The CACC sequence required for directional cloning into pENTR-D-TOPO vector was introduced into the nested forward primer (Supporting Table 1). For transient expression assay experiments a promoter fragments were isolated using nested PCR and genomic DNA of *T. aestivum* cv Chinese Spring as a template. Primers were derived from the sequences of *TdPRPI-1* (Acc. GQ449373), *TdPRPI-5* (Acc. GQ449376), *TdPRPI-7* (Acc. GQ449377), *TdPRPI-8* (Acc. GQ449378), *TdPRPI-10* (Acc. GQ449374), *TdPRPI-11* (Acc. GQ449375) and *TdPR61* (Acc. JN400737). The CACC sequence introduced in the nested forward primer permitted directional cloning of isolated PCR products into the pENTR-D-TOPO vector; the cloned inserts were verified by sequencing and re-cloned into the pMDC164 vector. The effector construct contained the full-length coding region of *TaGL7* cloned under the polyubiquitin promoter (Ubi) of the pUbi vector. The negative controls for transient expression assays contained either the coding region of *GFP* (for 'counting GUS foci' method) or the antisense sequence of the *TaGL7* coding region (for spectrophotometric enzymatic assay) cloned into the pUbi vector under the constitutive Polyubiquitin promoter from maize.

Phylogenetic analysis of HD-Zip IV TFs

Amino acid sequences of 42 AP2 (APETALA2) HD-Zip IV TFs from *Arabidopsis* (*Arabidopsis thaliana*, *At*), rice (*Oryza sativa*, *Os*), maize (*Zea mays*, *Zm*), bread wheat (*T. aestivum*, *Ta*), durum wheat (*T. turgidum* ssp *durum*, *Td*), silver poplar (*Populus alba*, *Pa*), spreading earthmoss (*Physcomitrella patens*, *Pp*), cotton (*Gossypium hirsutum*, *Gh*), rapeseed (*Brassica napus*, *Bn*), and tomato (*Solanum lycopersicum*, Sl) were aligned with TaGL7 (Acc. MK583312) and TdGL7 (Acc. MK583313), and a phylogenetic tree, based on a crude distance measure, was generated using ProMals3D (Pei et al. 2008). The tree was visualised using TreeView (Page, 1996). The accession numbers of protein sequences used in the tree are: TaGL9 (JF332037), TdGL9H1

(JF332038), AtANL2 (Acc. NP_567183), AtHDG1 (Acc. NP_191674), ZmOCL1 (Acc. CAG38614), ZmOCL2 (Acc. CAB96422), ZmOCL3 (Acc. CAB96423), AtHDG5 (Acc. Q9FJS2), AtHDG4 (Q8L7H4), GhHOX2 (Acc. AAM97322), OsROC3 (Acc. A2ZAI7), ZmOCL4 (Acc. CAB96424), AtHDG11 (Acc. NP_ 177479), BnBBIP-1A (Acc. ABA54874), AtHDG12 (Acc. NP_564041), AtHDG10 (Acc. NP_174724), AtHDG9 (Acc. NP_197234), AtHDG8 (Acc. Q9M9P4), FWA/AtHDG6 (Acc. Q9FVI6), AtHDG2 (Acc. Q94C37), ATML1 (Acc. AL161555), ZmOCL5 (Acc. CAB96425), PpHDZ41 (Acc. DAA05775), AtHDG3 (Acc. Q9ZV65). AtGL2 (AT1G79840), OsROC9 (XM_015788973), AtREVOLUTA (AT5G60690), OsHOX9 (AY423716), OsROC7 (XM_015793053), OsROC2 (AB101645), OsROC1 (AB077993), PaHB1 (AAG43405), AtPDF2 (Q93V99), AtHDG7 (Q9LTK3), PaHB2 (AAL83725), SlCD2 (NP_001234657), OsROC4 (AB101647), OsROC5 (AB101648), OsROC6 (XM_015757135), OsROC8 (XM_015786612).

Sequence analyses, domain architecture predictions and sequence alignments of TaGL7

The primary sequence of TaGL7 (883 amino acid residues) was analysed by the BLOCKS (Henikoff and Henikoff, 1991), ProDom (Bru et al. 2005), SBASE (Vlahovicek et al. 2005) and SMART (Letunic et al. 2009) predictor tools, that delineate domain boundaries, based on the primary, secondary and tertiary structure information. The alignments of HD and START-like domain sequences of TaGL7 with selected dicot and monocot sequences of related proteins (Depège-Fargeix et al. 2011) were generated with ProMals3D (Pei et al. 2008).

Construction of 3D models of TaGL7 HD in complex with DNA cis-elements, and the START-like domain in complex with 1,2-dilinoleolyl-sn-glycero-3-phosphocholine (DLP)

The 3D model of putative HD and the START-like domain of TaGL7 was constructed with the Modeller 9v16 program (Sali and Blundell, 1993) as described (Yang et al. 2018). To identify the most suitable template for modelling of HD of TaGL7, searches were performed that unambiguously identified the mouse homeobox protein Hox-A9 (LaRonde-LeBlanc and Wolberger, 2003) (Protein Data Bank - PDB accession 1puf, chain A, 1puf:A) for HD, and the human phosphatidylcholine transfer protein in complex with DLP (Roderick et al. 2002) (PDB accession 1ln1, chain A; 1ln1:A) for the START-like domain of TaGL7. Both domains were analysed for the positions of secondary structures and the EMBOSS-Needle tool using the Needleman-Wunsch algorithm (Needleman and Wunsch, 1970) was used to calculate positional sequence identity/similarity scores between two pairs of sequences. The 1puf:A crystal structure contained a bound 20-bp DNA duplex fragment 5'-ACTCTATGATTTACGACGCT-3'. The two

9-bp fragments 5'-ATTAAATGC-3' and 5'-CAATCATTG-3' were used for modeling of HD of TaGL7 and were built based on the coordinates of a 20-bp DNA fragment, with Coot 0.8.6.1 (Emsley and Kowtan, 2004) and YASARA 17.18.15 (Krieger et al. 2009), and minimised with MMFF94 force field/charge parameters and a constant dielectric function. Aligned sequences of HDs of TaGL7 and 1puf:A, and the START-like domain of TaGL7 and a human phosphatidylcholine transfer protein were used as input parameters to build 3D models on a Linux workstation, running the Ubuntu operating system. The best scoring models of both target proteins were selected from 50 models that showed the lowest values of the 'Modeller Objective Function' and the 'Discrete Optimised Potential Energy' analysis. The stereochemical quality and overall G-factors of both templates and models of both target proteins were calculated with PROCHECK (Laskowski et al. 1993). Z-score values for combined energy profiles were evaluated by Prosa2003 (Sippl, 1993). Electrostatic potentials for HD of TaGL7 and 1puf:A were calculated with the Adaptive Poisson-Boltzmann Solver with solvent contributions (the dielectric constants of solvent and solute were 78 and 2, respectively), as implemented in PyMol (Schrődinger, USA) and mapped onto molecular surfaces that were generated with a 1.4 Å probe radius. Solvent-accessible cavity volumes of START-like domain models were estimated by CASTp, using a 1.4 Å probe radius (Tian et al. 2018). Molecular graphics was generated with PyMol.

Plant transformation and analyses

For stable transformation of wheat (*T. aestivum* L. cv. Bobwhite), the pTdGL7 construct was co-transformed together with a plant selectable marker cassette (*Ubi-hpt-nos*) into wheat using microprojectile bombardment as described in (Kovalchuk et al. 2009; Ismagul et al. 2018). The integration of promoter-GUS fusions in transgenic plants was confirmed by PCR using primers derived from promoter and *GUS* (*uidA*) sequences. After transformation, 45 transgenic T_0 wheat lines, confirmed by PCR, were analysed, and 25 of these wheat lines were selected using the GUS staining assay, from which 15 demonstrated strong GUS expression, four had middle level of expression and six showed weak expression of the reporter gene. Four lines were sterile and the analysis of these lines was not performed. Three lines, two with strong transgene expression and one with the middle level of expression were selected for further analysis. All positive lines demonstrated the same pattern of GUS expression.

The pTdGL7 construct was transformed into rice (*Oryza sativa* L. ssp. *Japonica* cv. Nipponbare) and barley (*Hordeum vulgare* cv. Golden Promise) using *Agrobacterium*-mediated transformation and the method developed by Tingay et al. (1997) and modified by Matthews et al. (2001). From

these transgenic T_0 rice lines, 16 were analysed for the GUS activity. 11 lines demonstrated a strong promoter activity and one of them had a weak GUS activity. All lines had the same pattern of gene expression. GUS expression was not detected in one line and one more line was sterile and not analysed. No phenotypic differences were found between wild-type (WT) plants and plants transformed with the control vector. From transgenic barley T_0 lines, 16 of them were analysed for the GUS activity. Eight lines showed the strong promoter activity and had the same patterns of GUS expression. WT plants and/or plants transformed with a vector containing only the selectable marker cassette were used as negative controls.

Whole-mount and histological GUS assays were performed as described by Li et al. (2008) using T_0 - T_1 transgenic plants and T_1 - T_2 seeds.

Transient expression assay

The transient promoter activation assay based on the co-bombardment of a promoter-GUS fusion constructs with the pUbi-TaGL7 construct was performed as described by Kovalchuk et al. (2012a) using the suspension cell culture of *T. monoccocum* L., which was originally developed from roots (Shimada et al. 1969). GUS activity was determined by counting the number of blue cells (foci) using a Leica DC 300F stereomicroscope (Leica, Wetzlar, Germany). For each combination of constructs three to four independent bombardments were performed. The pUbi-GFP construct was used for the control of the bombardment efficiency. Alternatively, the GUS activity was measured spectrophotometrically as described by Jefferson et al. (1987) after an independent co-bombardment experiment (in three repeats).

Quantitative PCR

Q-PCR was carried out according to Burton et al. (2004). Q-PCR analysis of the expression of the *TaGL7* gene in different tissues of WT wheat and at different stages of grain development were performed as described by Morran et al. (2011). Expression of *TaGL7* under slowly developing drought, and *TaGL7* and *TdGL7* under mechanical wounding was examined using cDNA series developed by Harris et al. (2016) and Eini et al. (2013).

Statistical analyses

These analyses were performed using the ANOVA-procedure and GenStat 9.0. The values of standard errors of mean (SEM) are shown in Figs. 4, 5 and 9 as vertical bars at $P<0.05$.

Results

Cloning of the TaGL7 and TdGL7 genes

Full-length cDNA of *TaGL7* was isolated in the yeast-1-hybrid (Y1H) screen using a cDNA library prepared from the whole grain of *T. aestivum* at 0-6 DAP. The four-fold repeated *cis*-element 5'-CATTAAATG-3' from *Arabidopsis* specific for HD-Zip IV TF (Abo et al. 2001) was used as a bait. Five independent clones contained the same 3,281 bp long inserts encoding the HD-Zip IV protein of 883 residues, designated *T. aestivum* GLABRA2-like clone 7 (TaGL7). The genomic sequence of *TaGL7* was not available, when this work was performed. Therefore, the 3' untranslated region (UTR) of *TaGL7* cDNA was used as a probe to screen the BAC library prepared from genomic DNA of *T. turgidum* ssp. *durum* cv. Langdon (Cenci et al. 2003). The selected BAC clone #1094 M11 contained the full-length sequence of the gene including more than 4 kb of the promoter region of the *TaGL7* orthologue from *T. turgidum* ssp. *durum*. The coding region of the cloned gene, designated *TdGL7*, contained nine introns. Coding region of the *TdGL7* cDNA was isolated using nested PCR. The protein sequence of TaGL7 was highly homologous to TdGL7, and the alignment revealed 98% identity between the two sequences.

Protein structure and phylogenetic relation to the other members of the HD-Zip IV subfamily

Search through databases using the TaGL7 protein sequence identified this protein to be a member of the HD-Zip class IV family of TFs. The phylogenetic relationships based on amino acid sequences of TaGL7, TdGL7 and the sequences of the HD-Zip class IV proteins from other plant species (Fig. 1) revealed that the closest homologues of the TaGL7 protein is ZmOCL4 (Ingram et al. 2000; Vernoud et al. 2009), and not yet characterised OsROC3 (Q336P2) from rice, and AtHDG4 (Nakamura et al. 2006), AtHDG5 (Nakamura et al. 2006; Kamata et al. 2013) and GhHOX2 (GenBank accession AAM97322) from dicotyledonous *Arabidopsis* and cotton.

Domain boundaries and the identification of multiple domains in TaGL7

The TaGL7 sequence could be subdivided to at least seven domains (Fig. 2A; Supporting Fig. 1), amongst them HD, also known as the homeobox-like domain, and other six domains (exocyst complex component Sec6-like domain, chorismate synthase-like domain, TraG-like N-terminal-like domain, lipid-binding START-like domain, protein kinase-like domain, and uroporphyrin-III C/tetrapyrrole methyltransferase-like domain).

The sequence of the 2nd HD domain (Fig. 2B) spanned 76 residues between 136 and 211 residues of the full-length TaGL7, although, the precise domain boundaries slightly varied amongst predictors. The superposition (Pei et al. 2008) of HDs of TaGL7 and other closely related plant proteins, amongst them the closest structural template, mouse HD (PDB accession1puf, domainA) showed that there was high conservation of residues regardless the origin of sequences; the conservation of residues on the 9-5 scale is shown at the top of Fig. 2B. These absolutely conserved (green boxes in Fig. 2B) and highly conservative (yellow boxes in Fig. 2B) regions were spread throughout HD sequences, and presumably reflected folding patterns of HDs and identified residues that mediated contacts with DNA (Fig. 2B, 2C). The 5th domain represented the lipid-binding START-like domain, the sequence of which spanned 248 residues between 353 and 600 residues of the full-length TaGL7 (Fig. 2A; Fig. 3A). The superposition of the START-like domain sequence and those from related monocot species (Depège-Fargeix et al. 2011) showed some degree of conservation of residues amongst the monocot entries (the conservation of residues on the 9-5 scale is shown at the top of Fig. 3A). These absolutely conserved residues (green boxes in Fig. 3A) were not many but were spread throughout the sequences, obviously reflecting the folding patterns of proteins.

3D model of HD of TaGL7 in complex with DNA cis-elements

The most appropriate structural template for HD of TaGL7 was the mouse homeobox protein Hox-A9 (PDB 1puf:A;) in complex with 20-bp DNA duplex fragment 5'-ACTCTATGATTTACGACGCT-3' (LaRonde-LeBlanc and Wolberger, 2003). The positional sequence identity and similarity between the 1puf:A and the HD of TaGL7 sequences (65 residues) were 41% and 78%, respectively. The stereochemical parameters of 1puf:A and the modeled HD of TaGL7 calculated by PROCHECK (Ramachandran et al. 1963) showed that 100% of all residues were located in the most favoured, additionally allowed or generously allowed regions; this indicated that the stereochemistry of the HD domain of TaGL7 was correct. The overall average G factor values of 1puf:A and HD of TaGL7, as measures of normality of main chain bond lengths and bond angles (Laskowski et al. 1993), were 0.39, 0.16 (for the HD model of TaGL7 with 5'-ATTAAATGC-3') and 0.15 (for the HD of TaGL7 with 5'-CAATCATTG-3'), respectively. The Z-score values for combined energy profiles were -4.91, -6.15 and -5.80 for 1puf:A and HD of TaGL7 HD with (5'-ATTAAATGC-3') and (5'-CAATCATTG-3'), respectively. These scores reflected the complexity of modelling (Sanchez and Sali, 1998; Sippl and Wiederstein, 2008) and the validity of HD models.

The secondary structure analyses indicated that the mouse homeobox protein Hox-A9 and HD of TaGL7 folded into α-helical orthogonal bundle structures (Fig. 2C-D). The analysis of sequences showed that the positions of positively charged Arg and Lys residues would be involved in binding of DNA. These residues were highly conserved in both template and target sequences (Fig. 2B). Molecular folds of the model of TaGL7 HD and the template structure in complex with the cognate DNA duplexes (Fig. 2C-D) denoted that the two structures matched closely with a root-mean-square-deviation value of 1.04 Å for Cα backbone positions in 65 superposed positions. The spatial positions of the 9-bp DNA 5'-ATTAAATGC-3' (binding DNA) and 5'-CAATGATTG-3' (non-binding DNA) fragments in HD of the TaGL7 model, based on the 1puf:A structure, were also determined (Fig. 2C-D). The data in Fig. 2C showed that HD of TaGL7 accommodated DNA well through its major groove, while the NH_2-terminal loop was localised to the minor groove of DNA (Fig. 2C-D). Although structural models could explain positions of binding and non-binding DNA fragments in TaGL7 HD, their validity needs to be tested in future experiments. One option would be to introduce variations in key binding and non-binding DNA positions.

3D model of the START-like domain of TaGL7 in complex with 1,2-dilinoleolyl-sn-glycero-3-phosphocholine (DLP)

The most suitable template for the START-like domain was found to be the human phosphatidylcholine transfer protein in complex with DLP [Protein Data Bank (PDB) accession 1ln1:A) (Roderick et al. 2002]. The positional sequence identity and similarity between the 1ln1:A and the START-like domain sequences (203 residues) were 21% and 32%, respectively. The stereochemical parameters of 1ln1:A and the modeled START-like domain showed that 99.4% and 99.1% of all residues were located in the most favoured, additionally allowed or generously allowed regions; this indicated that the stereochemistry of the model was correct. The overall average G factor values of 1ln1:A and the START-like domain were 0.07, and -0.42, respectively, and the Z-score values for combined energy profiles were -7.52 and -3.41. These scores reflected a very high complexity of modelling in the so-called near 'twilight zone' (Sanchez and Sali, 1998; Sippl and Wiederstein, 2008). However, the evaluation criteria indicated that the START-like domain model was reliable.

The analyses of sequences in template and target sequences showed (Fig. 3A) showed that the positions of several positively (arginine) and negatively (glutamic and aspartic acid) charged residues were conserved, in addition to aromatic residues (tryptophan). The structural analyses indicated that the START-like domain of TaGL7 and the human phosphatidylcholine transfer

protein folded into 'α/β-helix grip' structures (Fig. 3B) (Roderick et al. 2002; Thorsell et al. 2011), where a curved antiparallel β-sheet packed by several α-helices, formed a concave hydrophobic cavity that faced the antiparallel β-sheet. This cavity could accommodate lipidic molecules and several crystal structures showed that these lipids could be DLP, 1-palmitoyl-2-linoleoyl-sn-glycero-3-phosphocholine (CPL) or other hydrophobic peptides and molecules (Thorsell et al. 2011). It was notable that non-polar (hydrophobic) residues constituted 53% of the START-like domain protein sequence.

Molecular folds of the model of the START-like domain of TaGL7 and the human phosphatidylcholine transfer protein (template) structure in complex with DLP (Fig. 3B) indicated that the two 3D structures matched. Here, three Arg383, Arg593 and Gln487 residues were involved in binding of DLP in the START-like domain of TaGL7, while Arg78, Tyr72 and Glc157 mediated contacts of in the human phosphatidylcholine transfer protein with DLP. Although these residues did not sterically matched, it is known that binding of lipids in these cavities varies, depending on the lipid and protein structural features (Thorsell et al. 2011). Importantly, the DLP molecule in the START-like domain of the TaGL7 model (Fig. 3B, left panel) was buried in a largely hydrophobic cavity. The residues delineating these hydrophobic cavities matched spatially in both proteins, whose shape could change based on the structure of a bound lipid indicating its structural flexibility (Thorsell et al. 2011). However, because the cavity with lipids is enclosed in both proteins, binding of lipids, before they are enclosed in cavities, would require substantial structural re-arrangement of both proteins. Calculations of the volumes of solvent-accessible cavities (1,360 $Å^3$ and 949 $Å^3$ for the START-like domain of TaGL7 and the human phosphatidylcholine transfer protein, respectively) by CASTp (Tian et al. 2018) indicated that the volume of the more elongated START-like domain of TaGL7 was by around 30% larger than that of the human phosphatidylcholine transfer protein.

Expression of GL7 genes in different wheat tissues in the absence of stress, under drought and in response to mechanical wounding

The expression of the *TaGL7* gene in different wheat tissues and grain at various stages of plant development was studied by quantitative RT-PCR (Fig. 4). Weak expression of the gene was detected in all tested tissues (Fig. 4A). Strong expression of the gene was found in the liquid fraction of the syncytial endosperm at 5 DAP, although the level of *TaGL7* expression in pericarp of grains collected at 5 DAP was low. Further, the level of expression of the *TaGL7* gene was

higher in shoots of seedlings, in flowers during meiosis, and in the grain fraction enriched for embryos.

Expression of *TaGL7* in leaves declined gradually under slowly developing drought and under strong drought, and the level of *TaGL7* decreased two-fold, but after re-watering the expression level was quickly returned to the normal level of unstressed plants (Fig. 4B). In contrast, under mechanical wounding, the number of *TaGL7* transcripts rapidly increased (within one hour) but returned to the original number three hours after wounding occurred (Fig. 4C).

The activation of the *GL7* gene expression by wounding was studied in a more detail in leaves and developing grain (2-6 DAP) of durum wheat (Fig. 5). The changes in expression of *TdGL7* were detected as early as 15 minutes after wounding, gene expression reached maximum at 1.5-2 hours and declined to original levels three to four hours after wounding. In wounded leaves the number of transcripts increased about nine-fold, compared to their number in control unwounded leaves (Fig. 5A). In grain we observed only three- to four-fold increase in expression because of the higher basal levels of the *TdGL7* expression in unstressed developing grain compared to those in unstressed leaves (Fig. 5B).

Spatial and temporal activity of the* TdGL7 *promoter in wheat, barley and rice

Analysis of the temporal and spatial activity of the *TdGL7* promoter was performed using the *GUS* reporter system in transgenic wheat, barley and rice stably transformed with the same construct. In both barley and wheat, the activity of *TdGL7* promoter was initially detected in the embryo sac during fertilisation, and later it became strong in syncytium and the cellularised endosperm (Figs. 6 and 7). This data correlate with the Q-PCR data obtained for wheat grain fractions: the strongest expression of *TaGL7* was detected in the liquid fraction of endosperm (Figs. 6B and 7F-G). On the fifth day of grain development, a strong *GUS* expression occurred in aleurone, endosperm transfer cells (ETC) and embryo (Fig. 6H-J and 7I,-J). At the end of cellularisation, *GUS* expression declined in the endosperm of both plants, but later, approximately at 10 DAP in wheat and at 15 DAP in barley, the strong activity of the promoter was detected in endosperm again (Figs. 6K, M and 7K-P). GUS staining was detected in both endosperm and embryo at least until 30 DAP. However, at 30 DAP, GUS expression in the embryo declined and could be seen only in some regions (Fig. 7O-P). Although the low level of *TaGL7* expression was demonstrated by Q-PCR in most wheat tissues, we could not detect the GUS expression driven by the *TdGL7* promoter in both unstressed or wounded leaves, and stems of wheat and barley.

GUS expression driven by the *TdGL7* promoter in rice grains was overall stronger than in wheat and barley. This staining was observed not only in grain, but also in the upper cell layers of lemma and palea (Fig. 8A). Starting from the point of rice endosperm cellularisation and until grain maturity, GUS activity was detected in aleurone and endosperm (Fig. 8B-H), but not in the upper cell layer of grain (Fig. 8K-N). However, until 10 DAP, GUS staining was stronger in aleurone than in endosperm (Fig. 8B-C), but after 10 DAP, the activity of GUS started to continuously increase in the starchy endosperm (Fig. 8D-H) and it became strong at 56 DAP. In contrast to endosperm, the continuous increase of the embryo-localised GUS activity during grain development was not observed. At some stages of grain development, the embryo either showed no GUS expression, or expression was seen only in some parts of grain (Fig. 8E-H). At other stages of grain development, the activity of the *TdGL7* promoter in embryo had the same or similar strength as in endosperm (Fig. 8D, F, G). The reasons of such behaviour of the promoter during the embryo development remain unclear.

Mechanical wounding induced a very faint GUS coloration in vascular bundles of the rice stem, and a stronger promoter response on stress was detected in veins of leaves (Fig. 8I-J). In contrast, induction of the GUS activity in response to wounding in rice grains was strong at all tested stages of grain development (Fig. 8K-N).

In transgenic rice, but not in wheat or barley, the GUS activity was observed during grain germination (Fig. 8O-S). Initially it was seen in the surrounding embryo endosperm and emerging coleoptile (Fig. 8O). Later the GUS expression was observed mainly in vascular system of roots and some parts of coleoptile, it became faint 15 days after germination (DAG) and later became undetectable (Fig. 8P-S).

Activation of grain specific defensin promoters by transient expression of TaGL7 in wheat cell culture

Transient expression assay in cultured wheat cells was used to reveal if promoters of wounding-inducible LTP genes, and particularly earlier described genes of grain-specific defensins (Kovalchuk et al. 2010) could be activated by TdGL7. The same effector construct was used in co-bombardment with each promoter-GUS fusion construct. It contained *TdGL7* driven by the strong constitutive Ubi promoter from maize. Reporter constructs contained promoters of different LTP genes fused to the *GUS* reporter gene. The durum wheat defensin promoters, active

predominantly in grain (*TdPRPI-1*, *TdPRPI-7*, *TdPRPI-8*, *TdPRPI-10*, and *TdPRPI -11*) and the root specific defensin promoter *TdPRPI-5* were tested by co-bombardment with the effector construct. Driven by the constitutive pUbi promoter *GFP* and cloned in an antisense orientation, *TaGL7* were used as negative controls. A strong activation was observed for the *TdPRPI-1* promoter, relatively strong activation was detected when the *TdPRPI-11* promoter was used, and a weak activation was found for the *TdPRPI-10* promoter (Fig. 9A). These three defensin promoters were previously reported as wounding inducible (Kovalchuk et al. 2010). In addition, the weak activation of the *TdPR61* promoter by TaGL7 was also detected in a transient assay. It was previously shown that this grain specific promoter supports GUS expression in ETC, adjacent to ETC parts of aleurone, the embryo surrounding region (ESR) and embryo (Kovalchuk et al. 2012a). No statistically significant activation by TdGL7 was detected in the case of three other promoters, *TdPRPI-5*, *TdPRPI-7* and *TdPRPI-8*.

Based on our previous experience, the activity of most promoters usually remains in the range of 10 to 600 GUS foci, which is a suitable range for quantitative evaluations (Eini et al. 2013; Bi et al. 2016; Yang et al. 2018). Because of the unusually strong activation of the *TdPRPI-1* promoter (Fig. 9A), the evaluation of the number of GUS foci was uneasy and hence the resultant data could be unprecise. Therefore, the transient expression assay was repeated, and a spectrophotometric analysis was used to compare GUS activities of three defensin promoters, which were active in the first experiment, with activity of the maize Ubi promoter (Jefferson et al. 1987). We usually avoid using this method because it is a labour and time consuming, provides the high background values due to the presence of wheat-specific coloured substances in supernatants, and has a lower sensitivity than the method of quantitative evaluations by counting GUS foci under the microscope. Nevertheless, the results of the spectrophotometric analysis confirmed the activation of all three promoters. The activity of the *TdPRPI-1* promoter was the strongest one, which was 1.5-fold stronger than the activity of the Ubi promoter (positive control) that is usually used as a strong promoter for the constitutive overexpression of genes in transgenic wheat plants.

Discussion

While searching for the early grain specific wheat genes, we isolated cDNAs of two TFs using a Y1H screen of a grain cDNA library with the four-fold repeated palindromic sequence CATTAAATG, which is specific for *Arabidopsis* HD-Zip IV TFs (Tron et al. 2001). The bait sequence also contained the three-fold repeated L1-box sequence TAAATGCA, which is an

asymmetric *cis*-element previously identified in the promoters of several target genes from *Arabidopsis* (Abe et al. 2001; Ohashi et al. 2003). Initially, Nakamura et al. (2006) demonstrated that in *Arabidopsis* several HD-Zip IV proteins are specifically expressed in flowers and developing seeds, later Javelle et al. (2011) has shown that many members of the HD-Zip IV family from monocotyledonous plants are preferably or specifically expressed in flowers and developing grain. However, neither wheat genes nor full-length contigs of low-abundant wheat HD-Zip IV cDNA sequences were yet available, when we commenced this work; thus the use of PCR amplification was impossible. However, we had prepared several good quality cDNA libraries from wheat developing grain and grain components for the Y1H/Y2H cloning, therefore the Y1H system was selected as a cloning method. As DNA binding domains in HD-Zip IV TFs are situated near the proximity of N-termini of proteins, the chance of isolating cDNAs containing full-length coding regions by the Y1H screen was high. Indeed, two cDNAs encoding different full-length HD-Zip IV TFs were identified in the YIH screen. It was previously shown that expression of one of the identified genes, *TaGL9*, was specific for parts of grain and was not detected in other plant tissues. The promoter of the *TaGL9* homologue from durum wheat, *TdGL9H1*, was initially active in the epidermal cell layer of the embryo and in surrounding endosperm during endosperm cellularisation, and later expression was detected in walls of the main vascular bundle of scutellum (Kovalchuk et al. 2012b).

In this work we characterise the second HD-Zip IV gene, *TaGL7*, isolated from the grain cDNA library and its homologue *TdGL7* (there is 98% protein sequence identity between TaGL7 and TaGL9 proteins) isolated by PCR from durum wheat grain cDNA. Although TaGL7 and earlier reported TaGL9 bind the same *cis*-element, the TaGL7 protein is significantly longer than TaGL9 TF and belongs to a different clade in the phylogenetic tree (Fig. 1). The closest homologues of GL7 proteins were OCL4 from maize (Ingram et al. 2000; Vernoud et al. 2009), ROC3 from rice, HDG4 and HDG5 from *Arabidopsis* (Nakamura et al. 2006; Kamata et al. 2013) and GhHOX2 from cotton. The conserved protein domains and motifs of the TaGL7 and TdGL7 proteins are underlined and specified in the protein alignment of TaGL7 to its closest homologues from monocotyledonous plants (Supporting Fig. 1).

Molecular modelling is an *in-silico* approach that complements functional understanding of proteins; this sheds light on structural and functional properties of proteins. As we currently observe in our on-going structural proteomics era, new 3D structures and reliable molecular models are being generated daily, and these structure/function analyses are becoming the integral components of biological research.

To this end, molecular models of the wheat HD of TaGL7 protein were built, using the homeobox protein Hox-A9 from mouse (LaRonde-LeBlanc and Wolberger, 2003; PDB accession 1puf:A) in complex with two DNA fragments that can (5'-ATTAAATGC-3') or cannot (5'-CAATCATTG-3') be bound by HD of TaGL7. The template was identified by a multitude of prediction tools, as described previously (Yang et al. 2018). This 65-residue protein folds into a canonical 'all alpha protein' class that is categorised in the SCOP protein classification system (Andreeva et al. 2008). More precisely, 1puf:A belongs to the 'DNA/RNA-binding 3-helical bundle' fold and 'homeodomain-like' superfamily of proteins, where other RNA/DNA-binding proteins are also classified.

For construction of models of TaGL7 with *cis*-elements, we used 9-bp DNA fragments 5'-ATTAAATGC-3' and 5'-CAATCATTG-3'. While the first fragment is known to bind to HD of TaGL7, the second fragment does not (current work). The two models of HDs of TaGL7 with bound 5'-ATTAAATGC-3' and 5'-CAATCATTG-3' DNA fragments were compared with HD of HoaxA9 that contained bound 5'-ATTTACGAC-3' (Fig. 2C). The detailed analysis of HD of TaGL7 and that of HoaxA9 in complex with the DNA fragments indicated differences in binding of DNA duplexes. Predominantly positively charged lysine and arginine residues, stemming from an α-helix 3, located perpendicularly to the viewer (Fig. 2C) contributed to binding in both HDs, although a series of polar asparagine and glutamine residues also participated. The NH_2-terminal part of the protein bound in the minor groove of DNAs. The bound nucleotides interacting with HDs indicated that only certain parts of HDs are making close contacts with DNAs. During binding of 5'-ATTAAATGC-3'/(3'-GCATTTAAT-5') predominantly the TAA (in coding strand) and GC (in complementary strand) motifs formed interactions with HD of TaGL7 (Fig. 2C). On the contrary, the complex built with 5'-CAATCATTG-3'/(3'-CAATCATTG-5') (this fragment doesn't bind HD of TaGL7) indicated that the AT (in coding strand) and TG (in complementary strand) motifs precluded efficient DNA binding to HD of TaGL7. Further, the key residues that were involved in binding of 5'-ATTAAATGC-3'/3'-GCATTTAAT-5' (such as Q180 in Fig. 2C), were not engaged in close contacts with a non-binding DNA duplex. The other difference that was observed with HD in complex with both DNA fragments was, that the HD complex with non-binding DNA showed smaller negative charge in the major binding cleft of HD (Fig. 2D, left and middle panels). This could mean that the protein dynamics of HDs could play a key role during the formation of TF-DNA complexes and decide on the strength of these contacts.

Analysis of the available information about close homologues of wheat *GL7* genes revealed that most of these genes are not seed specific but are expressed in several plant tissues. The ZmOCL4 protein/gene is better characterised than other four genes/proteins, which are grouped in the same clade as GL7. Transcripts of the *ZmOCL4* gene were found in vegetative tissues, inflorescence, floral apices and adaxial faces of maize embryos, and therefore expression of *ZmOCL4* cannot be considered to be grain-specific (Ingram et al. 2000). It was shown that *AtHDG4* and *AtHDG5* products are involved in the development of a flower, rather than the trichome (Nakamura et al. 2006; Kamata et al. 2013). One of them, *AtHDG5*, is weakly expressed in all tested tissues except roots, but strongly expressed in flowers and siliques. Another one, *AtHDG4*, was weekly expressed only in flowers. We were surprised to find out that both *TaGL7* and *TdGL7* were expressed mostly in wheat grain. In the absence of stress, weak expression of *TaGL7* was detected in other tested tissues of bread wheat by Q-PCR. However, *TdGL7* expression was not detected in the wheat and barley vegetative tissues and flowers by using the GUS reporter gene fused to the *TdGL7* promoter.

To our knowledge, another not yet reported feature of the HD-Zip IV genes was the strong response of both *GL7* genes to mechanical wounding (Figs. 4C and 5). Earlier induction by wounding was described for members of HD-Zip I (Manavella et al. 2008, Ré et al. 2011), HD-Zip II (Dezar et al. 2011), HD-Zip III (Baima et al. 2001) subfamilies.

Both *TaGL7* and *TdGL7* genes were rapidly induced during the first hour after wounding with a metal brush; the response on stress was transient and after three hours *GL7* genes returned to the original levels of expression observed in unstressed leaves and grain. The activation of the *TdGL7* promoter was analysed in a stably transformed wheat plants using the *GUS* reporter gene. In this case the activity of the promoter in the absence of stress was detected only in wheat grain. This discrepancy of data obtained by Q-PCR and by using the promoter-GUS activity assay can be explained by potential differences between sequences of bread and durum wheat promoters.

While the *TdGL7* promoter was grain-specific and wounding-inducible, it can be potentially applied in plant biotechnology as a tissue-specific promoter responsive to grain damage by insects and fungal infections, and hence be useful for the expression of defence-related genes in grain. Therefore, we explored the possibility to use this promoter in two other agriculturally important crop cereals, in transgenic barley and rice plants. Practically no significant differences in spatial and temporal patterns of the *TdGL7* promoter expression were found in these phylogenetically close species, such as wheat and barley. However, the activity of the promoter in transgenic rice

638 was overall stronger than that in wheat and barley and the GUS activity was detected in some
639 flower tissues, around embryo during grain germination, and in coleoptiles and roots of seedlings.
640 Notably, the lower tissue specificity of grain-specific wheat promoters in transgenic rice than in
641 transgenic wheat plants has already been observed (Kovalchuk et al. 2009; Kovalchuk et al. 2012a).
642
643 We have found that the pattern of expression of the HD-Zip IV TF from durum wheat, *TdGL9H1*
644 (Kovalchuk et al. 2012b), which is expressed in the main vascular bundle of the scutellum during
645 grain development, is very similar to those of several other genes encoding non-specific LTPs
646 from wheat: TaLtp7.2a, TaLtp9.1a, and TaLtp9.3e (Boutrot et al. 2007). Unfortunately, our
647 attempt to reveal, if TdGL9H1 is a regulator of the *TaLtp9.1a* promoter activity using transient
648 expression assay was unsuccessful (Kovalchuk et al. 2012b). The reason for this negative result
649 could be either in the insufficient length of the used promoter fragment, or in the absence of a
650 promoter activation co-factor in cultured root cells (Shimada et al. 1969), which could also be
651 specific only for the surface cell layer of the scutellum vascular bundle.
652
653 Spatial patterns of *GUS* expression driven by the *TdGL7* promoter resembled those of several grain
654 specific LTPs, including wounding inducible defensins (Kovalchuck et al. 2012a; Kovalchuck et
655 al. 2010). At the same time the spatial pattern of expression of *TdGL7* and defensins in grain were
656 less specific than those of *TdGL9*. In addition, low levels of *TdGL9* expression detectable only by
657 Q-PCR could exist in vegetative wheat tissues. These facts and the availability of a small collection
658 of grain specific promoters at the University of Adelaide encouraged us to test if some of these
659 promoters could be activated by co-bombardment with the effector cassette pUbi-TdGL7. Indeed,
660 in the case of *TdGL7*, we succeeded compared to our previous attempts with *TdGL9*, since we
661 found four promoters that could be activated by *TdGL7*. It is noteworthy, that among these four
662 promoters are three defensin promoters, which classified in the wounding-responsive group of
663 defensin genes. This may suggest that *TdGL7* is positioned at the end of the wounding-inducible
664 pathway upstream of the grain specific genes, products of which are involved in defence and/or in
665 the delivery of lipids for the repair of damaged cell walls and/or cuticle.
666
667 Large losses of grain yields in cereal plants often occur as a result of insect and fungal attacks
668 during vulnerable stages of grain development and germination. It was demonstrated that
669 overexpression of a single plant defensin gene in transgenic wheat (Li et al. 2011; Kaur et al. 2016;
670 Sasaki et al. 2016), rice (Kanzaki et al. 2002; Jha and Chattoo, 2010) and other plants (reviewed
671 by Carvalho and Gomes, 2009; Stotz et al. 2009; Parisi et al. 2018) strongly increased the

resistance of plants to damage produced by insects and fungal infections. The gamma-thionin family of plant defensins comprises a group of a low molecular mass cysteine-rich proteins (Stotz et al. 2009; Tang et al. 2018). Members of the ERF (Ethylene-Responsive Element binding factor), bZIP (basic Leucine Zipper), MYB (myeloblastosis), HSF (Heat-shock transcription factor) and WRKY families of proteins are among already known activators of defensin genes (Zarei et al. 2011; Carvalho and Gomes 2009; Brown et al. 2003). It would be elegant to use these upstream TFs, which regulate expression of defensin genes in wheat, and particularly in the wheat grain, for the simultaneous up-regulation of a variety of wheat defensins, and thus to enhance the whole plant or grain protection from a broad assortment of fungal and bacterial infections.

In summary, the *TaGL7* gene directed by an appropriate promoter (*e.g.* constitutive or tissue specific, alike to a strong defensin promoter) could potentially become a useful tool in biotechnological attempts to improve grain protection. However, the possible influence of the *TaGL7* overexpression on phenotypes of transgenic or gene-edited cereal plants should be thoroughly examined before the conclusion could be made.

Acknowledgements

We acknowledge the contributions of Ainur Ismagul and Serik Eliby in plant transformation. We thank Margaret Pallotta for technical assistance with BAC library screening and promoter cloning. Ursula Langridge, Lorraine Carruthers and Alex Kovalchuk are thanked for their assistance with growing of plants in the glasshouse. This work was supported by the Australian Research Council (LP120100201 to M.H. and S.L.), the Grains Research and Development Corporation and the Government of South Australia.

Author contributions statement

Conceived, designed experiments and analysed data: NK and SL. Cloning and Y1H: NB and WW. Transient expression assays: NR and OE. Q-PCR experiments: NS. Plant transformation and analysis: RS, AATJ, NK. 3D molecular modelling and bioinformatics: MH. Discussed the data: SL, NK, PL and MH. Writing of the manuscript: SL. Contributed to writing: MH.

Compliance with ethical standards

Authors declare that they have no conflict of interest.

Supporting Information

Additional Supporting Information may be found in the online version of this article.

References

Abe M, Katsumata H, Komeda Y, Takahashi T (2003) Regulation of shoot epidermal cell differentiation by a pair of homeodomain proteins in *Arabidopsis*. Development. 130: 635-643.

Abe M, Takahashi T, Komeda Y (2001) Identification of a cis-regulatory element for L1 layer-specific gene expression, which is targeted by an L1-specific homeodomain protein. Plant J 26: 487-494.

Andreeva A, Howorth D, Chandonia J-M, Brenner SE, Hubbard TJP, Chothia C, Murzin AG (2008) Data growth and its impact on the SCOP database: new developments. Nucleic Acids Res 36: D419-D425.

Ariel FD, Manavella PA, Dezar CA, Chan RL (2007) The true story of the HD-Zip family. Trends Plant Sci 12: 419-426.

Bi H, Luang S, Li Y, Bazanova N, Morran S, Song Z, Perera MA, Hrmova M, Borisjuk N, Lopato S (2016) Identification and characterization of wheat drought-responsive MYB transcription factors involved in the regulation of cuticle biosynthesis. J Exp Bot 67: 5363-5380.

Baima S, Possenti M, Matteucci A, Wisman E, Altamura MM, Ruberti I, Morelli G (2001) The *Arabidopsis* ATHB-8 HD-Zip protein acts as a differentiation-promoting transcription factor of the vascular meristems. Plant Physiol 126: 643-655.

Boutrot F, Meynard D, Guiderdoni E, Joudrier P, Gautier MF (2007) The *Triticum aestivum* non-specific lipid transfer protein (*TaLtp*) gene family: comparative promoter activity of six *TaLtp* genes in transgenic rice. Planta 225: 843-862.

Brown RL, Kazan K, McGrath KC, Maclean DJ, Manners JM (2003) A role for the GCC-box in jasmonate-mediated activation of the *PDF1.2* gene of *Arabidopsis*. Plant Physiol 132, 1020-1032.

Bru C, Courcelle E, Carrère S, Beausse Y, Dalmar S, Kahn D (2005) The ProDom database of protein domain families: more emphasis on 3D. Nucleic Acids Res 33: D212-D215.

Bürglin TR, Affolter M (2016) Homeodomain proteins: an update. Chromosoma 125: 497-521.

Burton RA, Shirley NJ, King BJ, Harvey AJ, Fincher GB (2004) The *CesA* gene family of barley. Quantitative analysis of transcripts reveals two groups of co-expressed genes. Plant Physiol 134: 224-236.

736 Carvalho Ade O, Gomes VM (2009) Plant defensins - prospects for the biological functions and
737 biotechnological properties. Peptides 30: 1007-1020.
738 Cenci A, Chantret N, Kong X, Gu Y, Anderson OD, Fahima T, Distelfeld A, Dubcovsky J (2003)
739 Construction and characterization of a half million clone BAC library of durum wheat (*Triticum*
740 *turgidum* ssp. durum). Theor Appl Genetics 107: 931-939.
741 Chew W, Hrmova M, Lopato S (2013) Role of Homeodomain leucine zipper (HD-Zip) IV
742 transcription factors in plant development and plant protection from deleterious environmental
743 factors. Intern J Mol Sci 14: 8122-8147.
744 Ciarbelli AR, Ciolfi A, Salvucci S, Ruzza V, Possenti M, Carabelli M, Fruscalzo A, Sessa G,
745 Morelli G, Ruberti I (2008) The *Arabidopsis* homeodomain-leucine zipper II gene family:
746 diversity and redundancy. Plant Mol Biol 68: 465-478.
747 Curtis MD, Grossniklaus U (2003) A gateway cloning vector set for high-throughput functional
748 analysis of genes in planta. Plant Physiol 133: 462-469.
749 Depège-Fargeix N, Javelle M, Chambrier P, Frangne N, Gerentes D, Perez P, Rogowsky PM,
750 Vernoud V (2011) Functional characterization of the HD-ZIP IV transcription factor OCL1 from
751 maize. J Exp Bot 62: 293-305.
752 Dezar CA, Giacomelli JI, Manavella PA, Ré DA, Alves-Ferreira M, Baldwin IT, Bonaventure G,
753 Chan RL (2011) HAHB10, a sunflower HD-Zip II transcription factor, participates in the induction
754 of flowering and in the control of phytohormone-mediated responses to biotic stress. J Exp Bot 62:
755 1061-1076.
756 Dwivedi KK, Roche DJ, Clemente TE, Ge Z, Carman JG (2014) The OCL3 promoter from
757 *Sorghum bicolor* directs gene expression to abscission and nutrient-transfer zones at the bases of
758 floral organs. Annals Bot 114: 489-498.
759 Eini O, Yang N, Pyvovarenko T, Pillman K, Bazanova N, Tikhomirov N, Eliby S, Shirley N,
760 Sivasankar S, Tingey S, Langridge P, Hrmova M, Lopato S (2013) Complex regulation by
761 Apetala2 domain-containing transcription factors revealed through analysis of the stress-
762 responsive *TdCor410b* promoter from durum wheat. PLoS One 8: e58713.
763 Emsley P, Cowtan K (2004) Coot: model-building tools for molecular graphics. Acta Cryst D 60:
764 2126-2132.
765 Guan XY, Li QJ, Shan CM, Wang S, Mao YB, Wang LJ, Chen XY (2008) The HD-Zip IV gene
766 *GaHOX1* from cotton is a functional homologue of the Arabidopsis *GLABRA2*. Physiol Plantarum
767 134: 174-182.

Harris JC, Sornaraj P, Taylor M, Bazanova N, Baumann U, Lovell B, Langridge P, Lopato S, Hrmova M (2016) Molecular interactions of the γ-clade homeodomain-leucine zipper class I transcription factors during the wheat response to water deficit. Plant Mol Biol 90: 435-452.

Henikoff S, Henikoff JG (1991) Automated assembly of protein blocks for database searching. Nucleic Acids Res 19: 6565-6572.

Ingouff M, Farbos I, Lagercrantz U, Arnold S (2001) *PaHB1* is an evolutionary conserved HD-GL2 homeobox gene expressed in the protoderm during Norway spruce embryo development. Genesis 30, 220-230.

Ingouff M, Farbos I, Wiweger M, von Arnold S (2003) The molecular characterization of *PaHB2*, a homeobox gene of the HD-GL2 family expressed during embryo development in *Norway spruce*. J Exp Bot 54: 1343-1350.

Ingram GC, Boisnard-Lorig C, Dumas C and Rogowsky PM (2000) Expression patterns of genes encoding HD-ZipIV homeo domain proteins define specific domains in maize embryos and meristems. Plant J 22: 401-414.

Ismagul A, Yang N, Maltseva E, Iskakova G1,3, Mazonka I1, Skiba Y, Bi H, Eliby S, Jatayev S, Shavrukov Y, Borisjuk N, Langridge P (2018) A biolistic method for high-throughput production of transgenic wheat plants with single gene insertions. BMC Plant Biol 18: 135.

Ito M, Sentoku N, Nishimura A, Hong SK, Sato Y, Matsuoka M (2002) Position dependent expression of GL2-type homeobox gene, *Roc1*: significance for protoderm differentiation and radial pattern formation in early rice embryogenesis. Plant J 29: 497-507.

Javelle M, Klein-Cosson C, Vernoud V, Boltz V, Maher C, Timmermans M, Depège-Fargeix N, Rogowsky PM (2011) Genome-wide characterization of the HD-ZIP IV transcription factor family in maize: preferential expression in the epidermis. Plant Physiol 157: 790-803.

Javelle M, Vernoud V, Depège-Fargeix N, Arnould C, Oursel D, Domergue F, Sarda X, Rogowsky PM (2010) Overexpression of the epidermis-specific homeodomain-leucine zipper IV transcription factor Outer Cell Layer1 in maize identifies target genes involved in lipid metabolism and cuticle biosynthesis. Plant Physiol 154: 273-286.

Jefferson RA, Kavanagh TA and Bevan MW (1987) GUS fusions: β-glucuronidase as a sensitive and versatile gene fusion marker in higher plants. EMBO J 6: 3901-3907.

Jha S, Chattoo BB (2010) Expression of a plant defensin in rice confers resistance to fungal phytopathogens. Transgenic Res 19: 373-384.

Kamata N, Okada H, Komeda Y, Takahashi T (2013) Mutations in epidermis-specific HD-ZIP IV genes affect floral organ identity in *Arabidopsis thaliana*. Plant Journal 75: 430-40.

Kanzaki H, Nirasawa S, Saitoh H, Ito M, Nishihara M, Terauchi R, Nakamura I (2002) Overexpression of the wasabi defensin gene confers enhanced resistance to blast fungus (*Magnaporthe grisea*) in transgenic rice. Theor Appl Genetics 105: 809-814.

Kaur J, John F, Adholeya A, Velivelli S, Kaoutar El-Mounadi, Natalya N, Thomas C, Shah D. (2016). Expression of apoplast-targeted plant defensin MtDef4.2 confers resistance to leaf rust pathogen *Puccinia triticina* but does not affect mycorrhizal symbiosis in transgenic wheat. Transgenic Res 26: 37-49.

Kovalchuk N, Li M, Wittek F, Reid N, Singh R, Shirley N, Ismagul A, Eliby S, Johnson A, Milligan AS, Hrmova M, Langridge P, Lopato S (2010) Defensin promoters as potential tools for engineering disease resistance in cereal grains. Plant Biotechnol J 8: 47-64.

Kovalchuk N, Smith J, Bazanova N, Pyvovarenko T, Singh R, Shirley N, Ismagul A, Johnson A, Milligan AS, Hrmova M, Langridge P, Lopato S (2012a) Characterization of the wheat gene encoding a grain-specific lipid transfer protein TdPR61, and promoter activity in wheat, barley and rice. J Exp Bot 63: 2025-2040.

Kovalchuk N, Smith J, Pallotta M, Singh R, Ismagul A, Eliby S, Bazanova N, Milligan AS, Hrmova M, Langridge P, Lopato S (2009) Characterization of the wheat endosperm transfer cell-specific protein TaPR60. Plant Mol Biol 71: 81-98.

Kovalchuk N, Wu W, Eini O, Bazanova N, Pallotta M, Shirley N, Singh R, Ismagul A, Eliby S, Johnson A, Langridge P, Lopato S (2012b) The scutellar vascular bundle-specific promoter of the wheat HD-Zip IV transcription factor shows similar spatial and temporal activity in transgenic wheat, barley and rice. Plant Biotechnol J 10: 43-53.

Krieger E, Joo K, Lee J, Lee J, Raman S, Thompson J, Tyka M, Baker D, Karplus K (2009) Improving physical realism, stereochemistry, and side-chain accuracy in homology modeling: Four approaches that performed well in CASP8. Proteins 77: 114-122.

Kubo H, Peeters AJ, Aarts MG, Pereira A, Koornneef M (1999) *ANTHOCYANINLESS2*, a homeobox gene affecting anthocyanin distribution and root development in *Arabidopsis*. Plant Cell 11: 1217-1226.

LaRonde-LeBlanc NA, Wolberger C (2003) Structure of HoxA9 and Pbx1 bound to DNA: Hox hexapeptide and DNA recognition anterior to posterior. Gen Dev 17: 2060-2072.

Laskowski RA, MacArthur MW, Moss DS, Thornton JM (1993) PROCHECK: a program to check the stereochemical quality of protein structures. J Appl Cryst 26: 283-291.

Letunic I, Doerks T, Bork P (2009) SMART 6: recent updates and new developments. Nucleic Acids Res 37: D229-D232.

834 Li M, Singh R, Bazanova N, Milligan AS, Shirley N, Langridge P, Lopato S (2008) Spatial and
835 temporal expression of endosperm transfer cell-specific promoters in transgenic rice and barley.
836 Plant Biotechnol J 6: 465-476.
837 Li S, Wang X, He S, Li J, Huang Q, Imaizumi T, Qu L, Qin G, Qu LJ, Gu H (2016) CFLAP1 and
838 CFLAP2 are two bHLH transcription factors participating in synergistic regulation of AtCFL1-
839 mediated cuticle development in *Arabidopsis*. PLoS Genetics 12: e1005744.
840 Li Z, Zhou M, Zhang Z, Ren L, Du L, Zhang B, Xu H, Xin Z (2011) Expression of a radish defensin
841 in transgenic wheat confers increased resistance to *Fusarium graminearum* and *Rhizoctonia*
842 *cerealis*. Funct Integr Genomics 11: 63-70.
843 Lu P, Porat R, Nadeau JA, O'Neill SD (1996) Identification of a meristem L1 layer-specific gene
844 in Arabidopsis that is expressed during embryonic pattern formation and defines a new class of
845 homeobox genes. Plant Cell 8: 2155-2168.
846 Manavella PA, Dezar CA, Bonaventure G, Baldwin IT, Chan RL (2008) HAHB4, a sunflower
847 HD-Zip protein, integrates signals from the jasmonic acid and ethylene pathways during wounding
848 and biotic stress responses. Plant J 56: 376-388.
849 Matthews PR, Wang MB, Waterhouse PM, Thornton S, Fieg SJ, Gubler F, Jacobsen JV (2001)
850 Marker gene elimination from transgenic barley, using co-transformation with adjacent 'twin T-
851 DNAs' on a standard *Agrobacterium* transformation vector. Mol Breed 7: 195-202.
852 Morran S, Eini O, Pyvovarenko T, Parent B, Singh R, Ismagul A, Eliby S, Shirley N, Langridge
853 P, Lopato S (2011) Improvement of stress tolerance of wheat and barley by modulation of
854 expression of DREB/CBF factors. Plant Biotechnol J 9: 230-249.
855 Nakamura M, Katsumata H, Abe M, Yabe N, Komeda Y, Yamamoto KT, Takahashi T (2006)
856 Characterization of the class IV Homeodomain-Leucine Zipper gene family in *Arabidopsis*. Plant
857 Physiol 141: 1363-1375.
858 Needleman SB, Wunsch CD (1970) A general method applicable to the search for similarities in
859 the amino acid sequence of two proteins. J Mol Biol 48: 443-453.
860 Ohashi Y, Oka A, Rodrigues-Pousada R, Possenti M, Ruberti I, Morelli G, Aoyama T (2003)
861 Modulation of phospholipid signaling by GLABRA2 in root-hair pattern formation. Science 300:
862 1427-1430.
863 Page RDM (1996) TreeView: an application to display phylogenetic trees on personal computers.
864 Comp Appl Biosci 12: 357-358.
865 Parisi K, Shafee TMA, Quimbar P, van der Weerden NL, Bleackley MR, Anderson MA (2019)
866 The evolution, function and mechanisms of action for plant defensins. Seminar Cell Dev Biol
867 https://doi.org/0.1016/j.semcdb.2018.02.004

Pei J, Kim B-H, Grishin NV (2008) PROMALS3D: a tool for multiple sequence and structure alignment. Nucleic Acids Res 36: 2295-2300.

Pyvovarenko T, Lopato S (2011) Isolation of plant transcription factors using a yeast one-hybrid system. Meth Mol Biol 754: 45-66.

Ramachandran GN, Ramakrishnan C, Sasisekharan V (1963) Stereochemistry of polypeptide chain configurations. J Mol Biol 7: 95-99.

Rerie WG, Feldmann KA, Marks MD (1994) The GLABRA2 gene encodes a homeo domain protein required for normal trichome development in Arabidopsis. Genes Dev 8: 1388-1399.

Ré DA, Dezar CA, Chan RL, Baldwin IT, Bonaventure G (2011) *Nicotiana attenuata* NaHD20 plays a role in leaf ABA accumulation during water stress, benzylacetone emission from flowers, and the timing of bolting and flower transitions. J Exp Bot 62: 155-166.

Roderick SL, Chan WW, Agate DS, Olsen LR, Vetting MW, Rajashankar KR, Cohen DE (2002) Structure of human phosphatidylcholine transfer protein in complex with its ligand. Nature Struct Biol 9: 507-511.

Sali A, Blundell TL (1993) Comparative protein modeling by satisfaction of spatial restrains. J Mol Biol 234: 779-815.

Sanchez R, Sali A (1998) Large-scale protein structure modeling of the *Saccharomyces cerevisiae* genome. Proc Natl Acad Sci USA 95: 13597-13602.

Sasaki K, Kuwabara C, Umeki N, Fujioka M, Saburi W, Matsui H, Abe F, Imai R (2016) The cold-induced defensin TAD1 confers resistance against snow mold and *Fusarium* head blight in transgenic wheat. J Biotechnol 228: 3-7.

Sessa G, Steindler C, Morelli G, Ruberti I (1998) The *Arabidopsis Athb-8, -9, -14* genes are members of a small gene family coding for highly related HD-ZIP proteins. Plant Mol Biol 38: 609-622.

Shimada TT, Sasakuma T, Tsunewaki K (1969) In vitro culture of wheat tissues. I. Callus formation, organ redifferentiation and single cell culture. Canadian J Genetics Cytol 11: 294-304.

Sippl MJ (1993) Recognition of errors in three-dimensional structures of proteins Proteins 17: 355-362.

Stotz HU, Thomson JG, Wang Y (2009) Plant defensins: defense, development and application. Plant Signal Behav 4: 1010-1012.

Tang SS, Prodhan ZH, Biswas SK, Le CF, Sekaran SD (2018) Antimicrobial peptides from different plant sources: Isolation, characterisation, and purification. Phytochemistry 154: 94-105.

900 Tominaga-Wada R, Iwata M, Sugiyama J, Kotake T, Ishida T, Yokoyama R, Nishitani K, Okada
901 K, Wada T (2009) The GLABRA2 homeodomain protein directly regulates *CESA5* and *XTH17*
902 gene expression in *Arabidopsis* roots. Plant J 60: 564-574.
903 Thorsell AG, Lee WH, Persson C, Siponen M, Nilsson M, Busam RD, Kotenyova T, Schüler H,
904 Lehtiö L (2011) Comparative structural analysis of lipid binding START domains. PLoS One 6:
905 e19521.
906 Tian W, Chen C, Lei X, Zhao J, Liang J (2018) CASTp 3.0: computed atlas of surface topography
907 of proteins. Nucleic Acids Res 46: W363-W367.
908 Tingay S, McElroy D, Kalla R, Fieg S, Wang MB, Thornton S, Brettell R (1997) *Agrobacterium*
909 *tumefaciens*-mediated barley transformation. Plant J 11: 1369-1376.
910 Tron AE, Bertoncini CW, Palena CM, Chan RL, Gonzalez DH (2001) Combinatorial interactions
911 of two amino acids with a single base pair define target site specificity in plant dimeric
912 homeodomain proteins. Nucleic Acids Res: 29, 4866-4872.
913 Tron AE, Bertoncini CW, Chan RL, Gonzalez DH (2002) Redox regulation of plant homeodomain
914 transcription factors. J Biol Chem 277: 34800-34807.
915 Vernoud V, Laigle G, Rozier F, Meeley RB, Perez P, Rogowsky PM (2009) The HD-ZIP IV
916 transcription factor OCL4 is necessary for trichome patterning and anther development in maize.
917 Plant J 59: 883-894.
918 Vlahovicek K, Kajan L, Agoston V, Pongor S (2005) The SBASE domain sequence resource,
919 release 12: prediction of protein domain-architecture using support vector machines. Nucleic
920 Acids Res 33: D223-D225.
921 Wei J, Choi H, Jin P, Wu Y, Yoon J, Lee YS, Quan T, An G (2016) GL2-type homeobox gene
922 *Roc4* in rice promotes flowering time preferentially under long days by repressing Ghd7. Plant Sci
923 252: 133-143.
924 Yang JY, Chung MC, Tu CY, Leu WM (2002) *OSTF1*: a HD-GL2 family homeobox gene is
925 developmentally regulated during embryogenesis in rice. Plant Cell Physiol 43: 628-638.
926 Yang Y, Luang S, Harris J, Riboni M, Li Y, Bazanova N, Hrmova M, Haefele S, Kovalchuk N,
927 Lopato S (2018) Overexpression of the class I homeodomain transcription factor TaHDZipI-5
928 increases drought and frost tolerance in transgenic wheat. Plant Biotechnol J 16: 1227-1240.
929 Yan T, Li L, Xie L, Chen M, Shen Q, Pan Q, Fu X, Shi P, Tang Y, Huang H, Huang Y, Huang Y,
930 Tang K (2018) A novel HD-ZIP IV/MIXTA complex promotes glandular trichome initiation and
931 cuticle development in *Artemisia annua*. New Phytol 218: 567-578.

Zarei A, Körbes AP, Younessi P, Montiel G, Champion A, Memelink J (2011) Two GCC boxes and AP2/ERF-domain transcription factor ORA59 in jasmonate/ethylene-mediated activation of the *PDF1.2* promoter in *Arabidopsis*. Plant Mol Biol 75: 321-331.

Zhang F, Zuo K, Zhang J, Liu X, Zhang L, Sun X, Tang K (2010) An L1 box binding protein, GbML1, interacts with GbMYB25 to control cotton fibre development. J Exp Bot 61: 3599-3613.

Figure legends

Fig. 1. Amino acid sequences of 42 HD-Zip IV TFs from *Arabidopsis* (*Arabidopsis thaliana*, *At*), rice (*Oryza sativa*, *Os*), maize (*Zea mays*, *Zm*), bread wheat (*T. aestivum*, *Ta*), durum wheat (*T. turgidum* ssp *durum*, *Td*), silver poplar (*Populus alba*, *Pa*), spreading earthmoss (*Physcomitrella patens*, *Pp*), cotton (*Gossypium hirsutum*, *Gh*), rapeseed (*Brassica napus*, *Bn*), were aligned with TaGL7 and TdGL7. The phylogenetic tree (accession numbers specified in Materials and methods) was generated by ProMals3D and visualised by TreeView.

Fig. 2. Molecular analysis of TaGL7 and modeling of HD of TaGL7 in complex with a 9-bp DNA fragment. (A) A TaGL7 sequence has at least seven domains as predicted by SBASE (Vlahovicek et al. 2005). The 2nd HD or homeobox-like domain binds DNA and the 5th START-like domain binds lipid molecules. Respective regular and bold types indicate numbering of residues and domain numbering. (B) Multiple sequence alignment of selected HD domains using ProMals3D (Pei et al. 2008). Predicted secondary structures of α-helices are shown in magenta. The black box indicates the boundaries of HD domains. The conservation of residues on the scale 9-5 (9 - most conserved in brown) is stated on the top of sequences. The absolutely conserved and similar residues are highlighted in green and yellow boxes, respectively. (C) Molecular folds of HDs of TAGL7 (left) and HoxA9 used as the template (right) in complex with respective DNA fragments 5'-ATTAAATGC-3'/3'-GCATTTAAT-5' and 5'-ATTTACGAC-3'/3'-GTCGTAAAT-5'. Ribbon representations show the disposition of secondary structure elements, where predominantly α-helix 3 (perpendicular to the viewer) carries residues that mediate contacts between HD and DNA *cis*-elements. The ribbons are coloured in cyan and yellow for TAGL7 and HoxA9 HD domains, respectively, and duplex *cis*-elements are shown in cpk-magenta (coding strand) and cpk-green (complementary strand). The residues interacting with nucleotides are shown in sticks and atomic colours. Top and lower black arrows point to NH_2- and COOH-termini of HDs. The separations of ≥ 3.4 Å between residues (1-letter codes) and *cis*-elements are indicated by black dotted lines. The interplay of interacting residues within HDs suggests that the structural

rigidity and/or flexibility could impact upon the selectivity of DNA binding. (D) Molecular surfaces with projected electrostatic surfaces of HDs of TAGL7 with 5'-ATTAAATGC-3'/3'-GCATTTAAT-5' (binding DNA) (left) and 5'-CAATGATTG-3'/3'-CAATCATTG-5' (non-binding DNA) (middle), and HoxA9 with 5'-ATTTACGAC-3'/3'-GTCGTAAAT-5' (right). Black arrows point to NH_2-termini of HDs. Blue and red patches indicate electropositive and electronegative regions, respectively, contoured at ± 5 $k.Te^{-1}$, calculated by the Adaptive Poisson-Boltzman Solver in PyMol.

Fig. 3. Molecular analysis of the lipid-binding START-like domain and its 3D modeling in complex with 1,2-dilinoleolyl-sn-glycero-3-phosphocholine (DLP). (A) An activation sub-domain of the START-like domain of TaGL7 was aligned with monocot sequences using ProMals3D (Pei et al. 2008). The black box indicates the boundaries of predicted activation sub-domains of TaGL7 START-like domains and other entries. Consensus residues are indicated in bold. Predicted secondary structures of α-helices and β-sheets are shown in magenta and blue. Green colouring highlights the 100-residue NH_2-terminal activation motif shown in Fig. 3A. The conservation of residues on the scale 9-5 (9 - most conserved in brown) is stated on the top of sequences. (B) Molecular folds of START-like domains of TaGL7 (left) and the human phosphatidylcholine transfer protein used as the template (right), both in complex with DLP. Ribbon representations showing the disposition of secondary structure elements are coloured in cyan and yellow for START-like domains of TaGL7 and a human phosphatidylcholine transfer protein, respectively; DLP molecules are shown in sticks and atomic colours. Residues positioned outside of the activation domain that interact with DLP molecules are shown in sticks and atomic colours. Bottom and top black arrows point to NH_2- and COOH-termini of domains. The separations of $\geqslant$ 3.5 Å between residues (1-letter codes) and DLPs are indicated by black dotted lines. The interplay of interacting residues within START-like domains suggests that the structural flexibility of residues within the cavity of domains impacts the selectivity of lipid- binding.

Fig. 4. Q-PCR analysis of *TaGL7* expression in bread wheat. (A) *TaGL7* expression in different wheat tissues, at different stages of grain development and in several grain fractions. (B) Influence of slowly developing drought on *TaGL7* expression. (C) Induction of *TaGL7* expression by mechanical wounding. SEM values at $P<0.05$ are shown.

Fig. 5. Q-PCR analysis of the *TdGL7* expression induced by mechanical wounding in leaves (A) and developing grains (B) of durum wheat. SEM values at $P<0.05$ are shown.

Fig. 6. Spatial and temporal GUS expression in wheat directed by the *TdGL7* promoter. (A, B) Strong GUS expression in T_1 grain from the transgenic wheat Line 2 at 2 DAP: (A) the intact grain, crease down, (B) longitudinal section with the cavity containing liquid endosperm. GUS expression is observed mainly in the upper half of pericarp and in syncytium. (C-E) Histochemical GUS assay counterstained with Safranin O in the 10 μm thick cross section in the middle of grain (D) and longitudinal sections (C, E) of grain at 2 DAP. (F, G) Wheat grain at 5 DAP: (F) the intact grain, crease down, (G) histochemical GUS assay of the cross section in the middle of grain. (H-J, L) Wheat grain at 10 DAP: (I) cross section in the middle of grain, the strongest GUS staining can be seen in aleurone and ETC. (H, J, L) histochemical GUS assay of magnified cross sections: (H) aleurone, (J) ETC and (L) the pigment strand. (K, M). Histochemical assay of the longitudinal section with strongly stained in starchy endosperm at 11 DAP. (N) Histochemical assay of longitudinal section of grain at 14 DAP. Control WT grains of the samestages of development are shown in the left parts of A, B, F, and I panels. Bars = 200 μm.

Fig. 7. Spatial and temporal GUS expression in barley directed by the *TdGL7* promoter. (A-C) GUS staining in pericarp: (A) before anthesis, (B, C) at anthesis. (D-P) GUS expression in the barley grain at different stages of development. (D, E) GUS expression was observed in ovary, no staining was detected in other flower tissues at 1 DAP. Strong GUS expression in partially cellularised endosperm at 5 DAP is shown in longitudinal (F) and cross (G) sections. Preferable GUS expression in embryo, ETC, and aleurone in longitudinal (I) and cross section (J) of barley grain at 10 DAP. (K, L) Strong GUS staining in starchy endosperm, the embryo surrounding region and parts of the embryo are shown on longitudinal section at 15 DAP. (M-P) At later stages of grain development staining was observed in endosperm and aleurone, but not in embryo.

Fig. 8. GUS expression in rice directed by the *TdGL7* promoter. (A-H) spatial and temporal GUS expression: (A) the promoter is active in ovary and upper layers of bracts at 2 DAP. (B-H) Longitudinal sections of control WT and transgenic rice grains at different stages of development. (I-N) Induction of the *TdGL7* promoter by mechanical wounding: (I, J) induction of GUS expression in vascular tissues of stem (I) and leaf (J) of rice plants. (K-N) Induction of GUS in

wounded areas of rice grains of different age. (O-S) GUS expression in germinating grain (O) and roots (P, R) and coleoptiles (Q, S) of seedlings. DAG – days after germination.

Fig. 9. Transient expression assay in the cultured cells of *Triticum monococcum* L. (A) co-bombardment of the effector (pUbi-TaGL7) with various reporter (promoter-GUS) constructs to wheat cells with the quantification of the promoter activation using either (A) counting of GUS foci or (B) the spectrophotometric enzymatic assay. pUbi-GFP and pUBI-TaGL7 (antisense strand) were used as negative controls and pUbi-GUS in the same concentration (as defensin promoters) was used as the positive control. Panel (A) shows two graphs with different scales, because of large differences in activities of the *TdPRPI-1* and other tested promoters. SEM values at $P<0.05$ are shown.

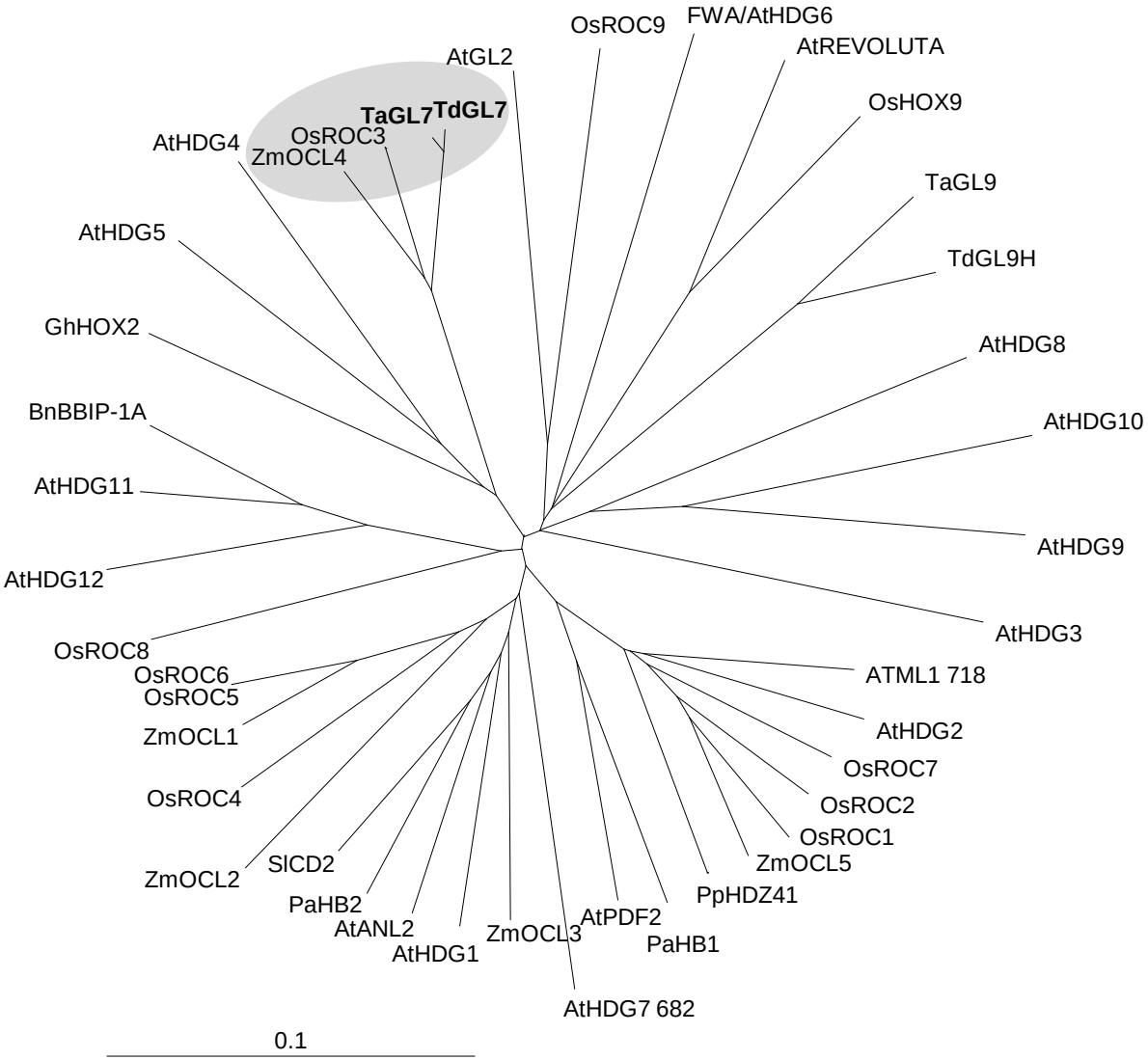
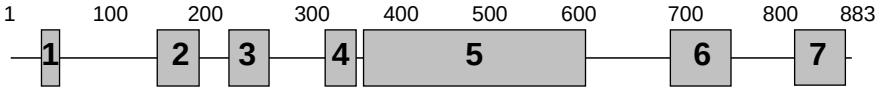

OsROC9
FWA/AtHDG6
AtREVOLUTA
AtGL2
OsHOX9
TaGL7
TdGL7
AtHDG4
OsROC3
ZmOCL4
TaGL9
AtHDG5
TdGL9H
GhHOX2
AtHDG8
BnBBIP-1A
AtHDG10
AtHDG11
AtHDG9
AtHDG12
AtHDG3
OsROC8
ATML1 718
OsROC6
OsROC5
AtHDG2
ZmOCL1
OsROC7
OsROC4
OsROC2
OsROC1
SlCD2
ZmOCL5
ZmOCL2
PpHDZ41
PaHB2
AtPDF2
AtANL2
ZmOCL3
PaHB1
AtHDG1
AtHDG7 682
0.1

Figure 1

(A)

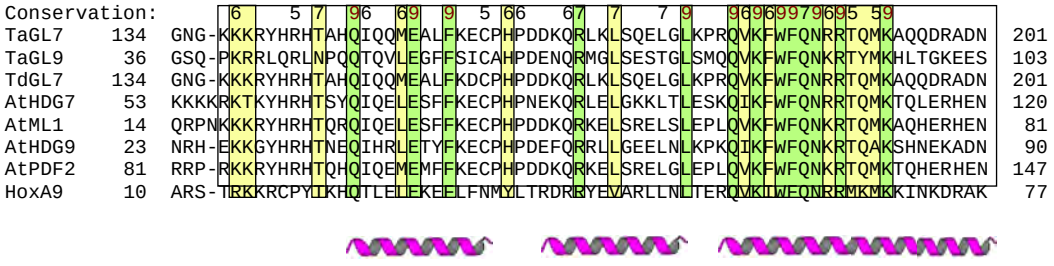
Domain architecture prediction of TaGL7



DOMAIN 1	25 - 40	Exocyst complex component Sec6-like domain
DOMAIN 2	141 - 199	Homeodomain (homeobox-like domain)
DOMAIN 3	219 - 261	Chorismate synthase-like domain
DOMAIN 4	318 - 353	TraG-like N-terminal-like domain
DOMAIN 5	353 - 600	Lipid-binding START-like domain
DOMAIN 6	682 - 756	Protein kinase-like domain
DOMAIN 7	819 - 870	Uroporphyrin-III C/tetrapyrrole methyltransferase-like domain

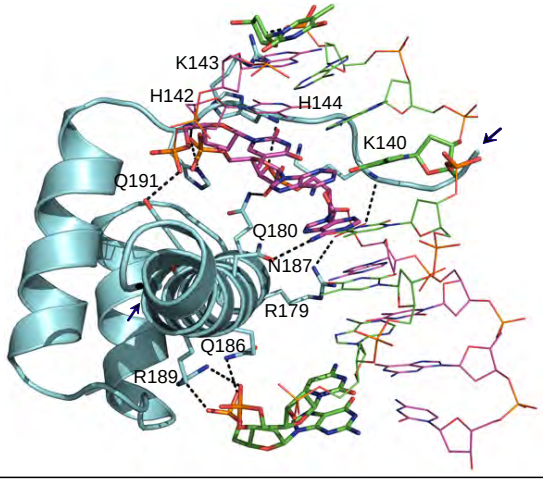
(B)

Conservation of HD of TaGL7 and related entries

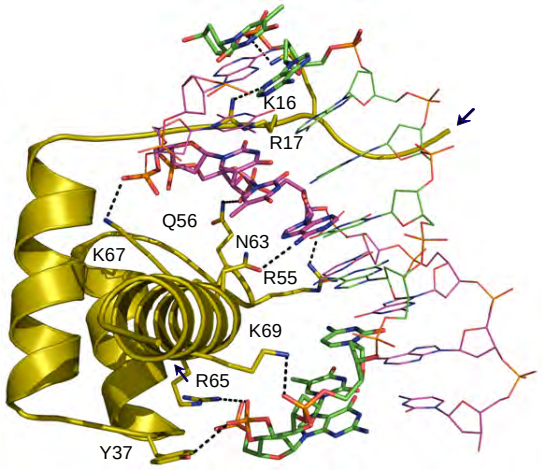


(C)

HD of TaGL7/
5’-ATTAAATGC-3’/3’-GCATTTAAT-5’

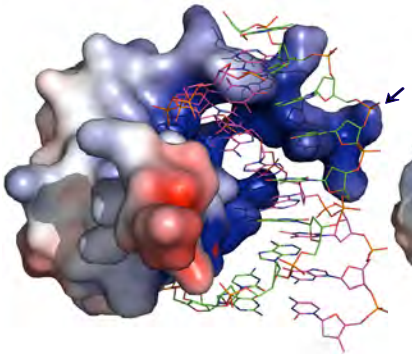


HD of HoxA9/
5’-ATTTACGAC-3’/3’-GTCGTAAAT-5’

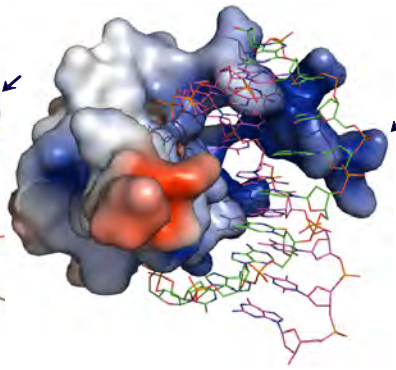


(D)

HD of TaGL7/
5’-ATTAAATGC-3’/3’-GCATTTAAT-5’



HD of TaGL7/
5’-CAATGATTG-3’/3’-CAATCATTG-5’



HD of HoxA9/
5’-ATTTACGAC-3’)/(3’-GTCGTAAAT-5’

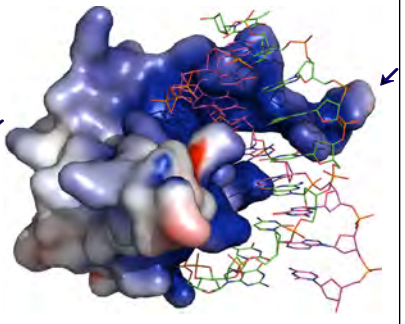


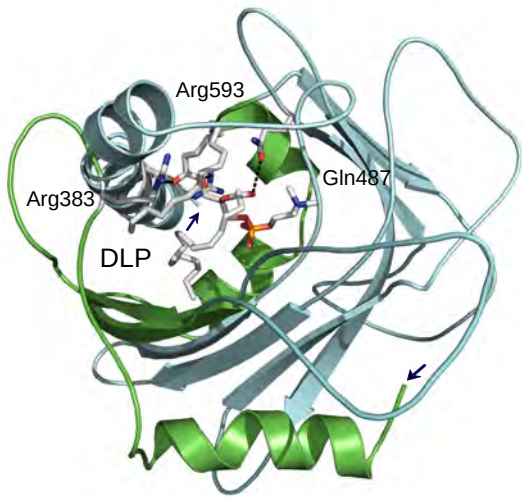
Figure 2

(A)

```
Conservation:              6 57  5      69              7   67   6                   9569    7       7 5 579   5   65
PpHDZ41          1  --------MEELVRMAQTEEPLWV-NMGEGK----EQLNYEEYMRQFPRI-------G-MCPPGLKTEATRETALVMMN-GVNLVETLMDATQWMDMFPCI  79
PpHDZ44          1  --------MEELVRMVQAEEPLWV-RAGGGR----EELNYEEYVRQFPRM-------G-MKPAGLRTEASRETALVMMN-GVNLVETLLDASQWAEMFPCV  79
PpHDZ42          1  --------MEELVRMAQSEEPLWV-RTPEGR----EQLNYDEYLRQSPRI-------G-MKPPGLKTEVTRETAMVMMN-GVNLVETLMDATQWIDMYPCM  79
PpHDZ43          1  --------MEELVRMAQAEEPLWL--SMDGK----AQLNYDEYMRQFPRI-------G-MRPSGLKPEATRETALVMMN-GVNLVETLMDATQWMDMFPCM  78
ZmOCL3           1  --------MEELMKVAQMDEPLWL-RSPDG-SGL-EALNLDEYHRAFARVF------G-PSPAGYVTEATREAGVAITS-SVDLVDSLMDAARWSEMFPCI  82
OsROC6           1  --------MDELVKVAQMDEPLWL-PSLDGGF---ETLNYDEYHRAFARVV------G-QCPAGYVSEATRESGIAIIS-SVDLVDSLMDAPRWSEMFPCV  81
OsROC5           1  --------MDELVKMAQMDDPLWV-PALPGSPSK-EVLNFEEYLHSFLPCI------G-MKPAGYVSEASRESGLVIIDNSLALVETLMDERRWSDMFSCM  84
ZmOCL2           1  --------MNELVKMARMNEPLWI-PTIPGSPIM-ETLNFKEYLKAFSPCV------G-VKPTGFVSEASRESGIVTIDSSAALMEVFMDERRWSDIFYCI  84
ZmOCL1           1  --------MDELVKLAQVDEPLWL-PSLSGSPDK-KLLNFEEYAHSFSPSV------GAVKPVGYVSEASRESGLVIIDNSLALVETLMDVRRWSDMFSCM  85
OsROC4           1  --------MDELVKMAQMGDPLWI-PGASSSPAK-ESLNFEEYLNTFPPCI------G-VKPEGYVSEASRESGIVIIDDGAALVETLMDERRWSDMFSCM  84
ZmOCL8           1  --------MEELVRMAQLDEPLWN-APAGS-AEE-ETLNEEEYARLFPL--------G-PKPYGLNSEASRDSAVVIMT-HANLVEILMDVNQYAAVFSSI  80
ZmOCL6           1  --------MEELVRMAQLDEPLWN-APAGSA----ETLNEEEYARMFVPL-------GLKKQYGFKSEASRDSSVVIMT-HASLVEILMDVNQYATVFSSI  80
OsROC2           1  --------MEELVRMAQLDEPLWSVAPPTA-AAM-ETLSEEEYARMFPL--------G-PKQYGLRSEASRDSAVVIMT-HANLVEILMDANQYAAVFSNI  81
OsROC1           1  --------MDELVQMAQLDEPLWSSSSEPAA----ALLDEEEYARMFPL--------G-PKQYGLKSEASRHGAVVIMT-HSNLVEILMDVNQFATVFSSI  79
ZmOCL5           1  --------MDELLRMARPDALLWG-GGASA-GAQ-QQLDEEEYVRTFPL--------G-PRQYGLRPEASRDSAVVIMT-CDSLIEILMDANRFAAVFSSI  80
ZmOCL7           1  --------MEELVRMALLGEPLWI-PALATT----ETLNEEEYARGFFPV-------G-PKLPELRSEASREAVVVIMN-HVSLIEMLMDVNQWSTLFSSI  79
OsROC7           1  --------MEELVRMAQLGEPLWA-PALG------EALGEEEYARTFPRGL------G-PKSPELRSEASREATVVIMN-HVSLVEMLMDVGQWTALFSSI  78
OsROC8           1  --------MDELIRLAQAGDHIWS-KSPGGGDAR-ETLNVDTYDSIFSKPG------GSYRAPSINVEGSRESGLVLMS-AVALADVFMDTNKWMEFFPSI  84
ZmOCL4           1  --------ADTLAKMCRAGESLWL-RRRGS-----EVMVADEHARMFSWPQ------DSSAGAAARTEGSRDGTVVIMN-SITLVDAFLDANKWMELFPSI  80
TaGL7            1  -MELASTAADHLIKMCRTSEPLWA-RRRGTGPSSGEVMDAEEHARMLSWPIDSGKQQQSDVAHSARTEGTRDSAVVIMN-SITLVDAFLDANKWMELFPSI  98
TdGL7            1  -MELASTAADHLIKMCRTSEQLWA-RRRGAGPSSGEVMDAEEHARMLSWPIDSGKQQQADVAHAARTEGTRDSAVVIMN-SITLVDAFLDANKWMELFPSI  98
OsROC3           1  --------ADQLARMCRAGEPLWV-RQRG------EVMAVEEHARMFSWPQ--------GDGAVARAEGTRDNAVVIMN-SINLVDAFLDANKWMELFPSI  77
OsROC9           1  --------LDELVGMCSSGEPVWV-RGVEGR----DILNYDEYVRLFRRDH-----GG-SGMAGWTVEASRECGLVYLD-TMHLVHTFMDVDKWKDLFPTM  81
TdGL9            1  IAELAQSAMHEFVVLSESGGPLWM-PVPGGSL---DVLNKMAYAQTFGAGS------S-ANAIGFMTEATRADGMVMMD-AKQIVDYIMDSECYTSFCPGL  89
TaGL9            1  IAELVQNAMHEFVVLSESGGPLWM-PVPGGSL---DLLNKVAYAQTFGARS------S-ANAIGFRVEATRADGMVMME-AKQIVDYIMDSECYTSFCPGT  89
Consensus aa:         ........h-ELl.htp.s-PLW...s.st.....-hLs.-EYhp.Fs.........t.....th.sEtR-stlVhhs.s.pLV-hhMDspp@.phFsth
```

(B)

SLD of TaGL7 in complex with DLP



SLD of Phosphatidylcholine transfer protein in complex with DLP

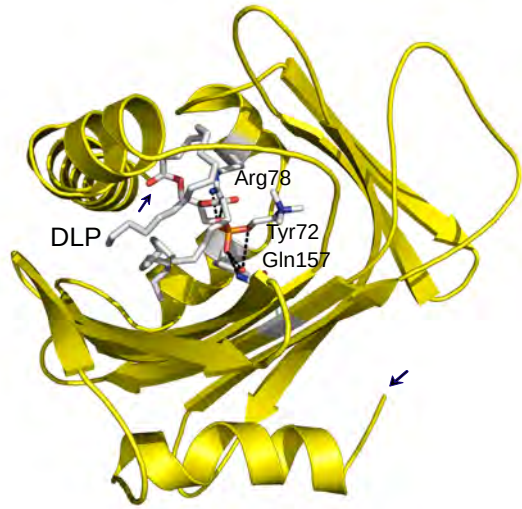


Figure 3

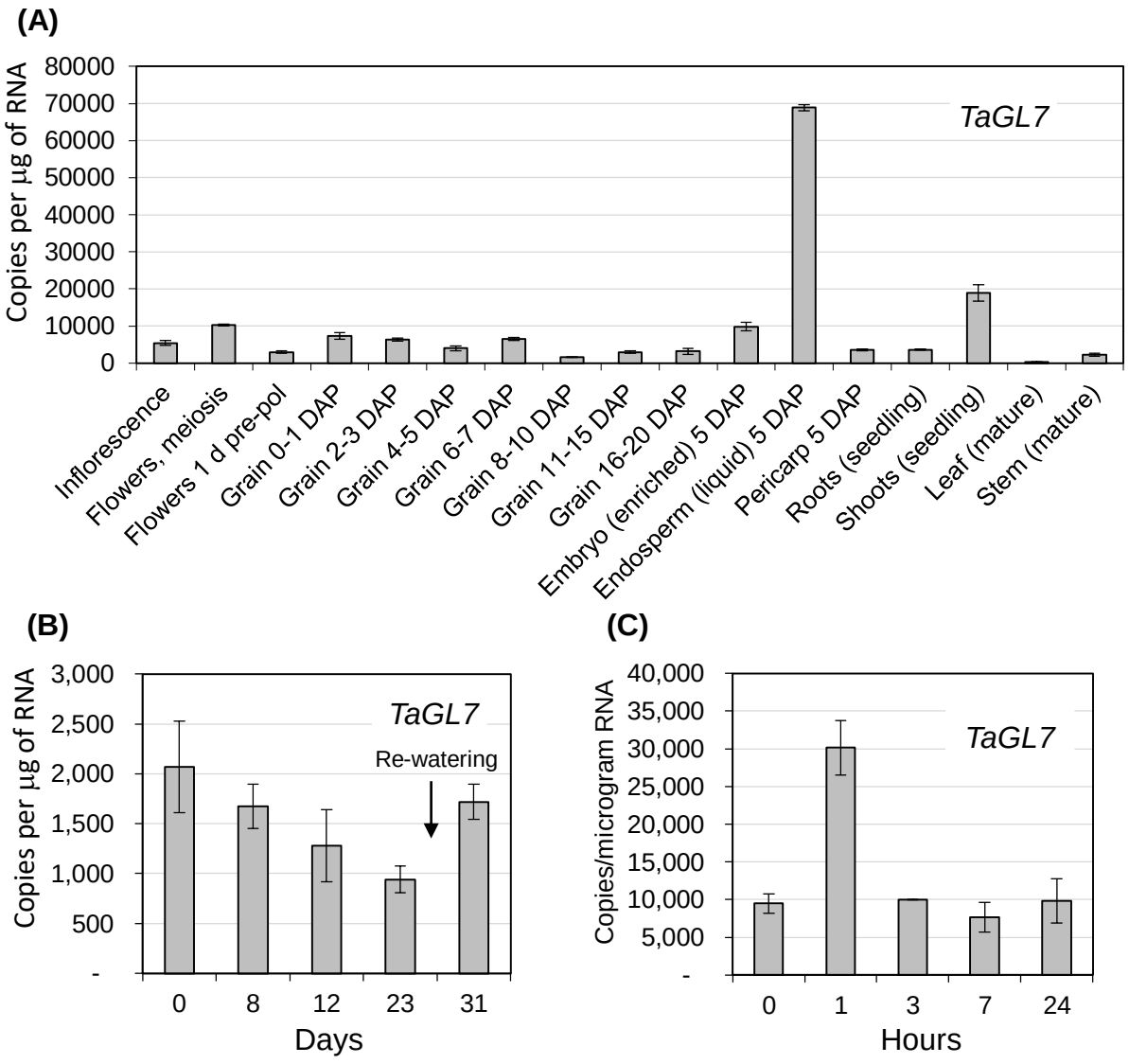
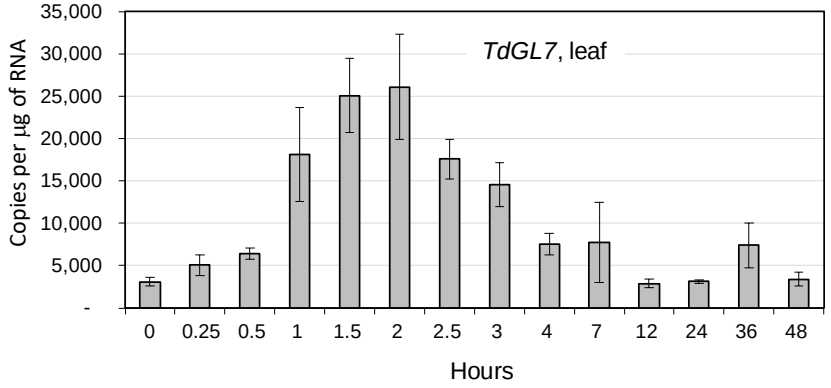

(A)
Copies per µg of RNA
80000
70000
60000
50000
40000
30000
20000
10000
0
TaGL7
Inflorescence
Flowers, meiosis
Flowers 1 d pre-pol
Grain 0-1 DAP
Grain 2-3 DAP
Grain 4-5 DAP
Grain 6-7 DAP
Grain 8-10 DAP
Grain 11-15 DAP
Grain 16-20 DAP
Embryo (enriched) 5 DAP
Endosperm (liquid) 5 DAP
Pericarp 5 DAP
Roots (seedling)
Shoots (seedling)
Leaf (mature)
Stem (mature)
(B)
Copies per µg of RNA
3,000
2,500
2,000
1,500
1,000
500
-
TaGL7
Re-watering
0
8
12
23
31
Days
(C)
Copies/microgram RNA
40,000
35,000
30,000
25,000
20,000
15,000
10,000
5,000
-
TaGL7
0
1
3
7
24
Hours

Figure 4

(A)


35,000
30,000
25,000
20,000
15,000
10,000
5,000
-
Copies per µg of RNA
TdGL7, leaf
0
0.25
0.5
1
1.5
2
2.5
3
4
7
12
24
36
48
Hours

(B)

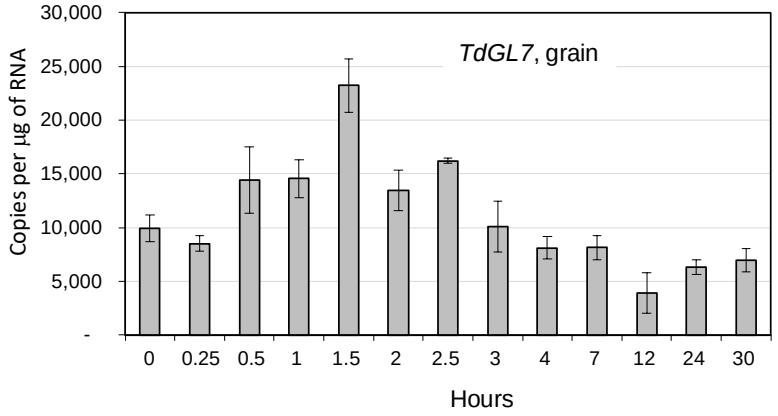

30,000
25,000
20,000
15,000
10,000
5,000
-
Copies per µg of RNA
TdGL7, grain
0
0.25
0.5
1
1.5
2
2.5
3
4
7
12
24
30
Hours

Figure 5

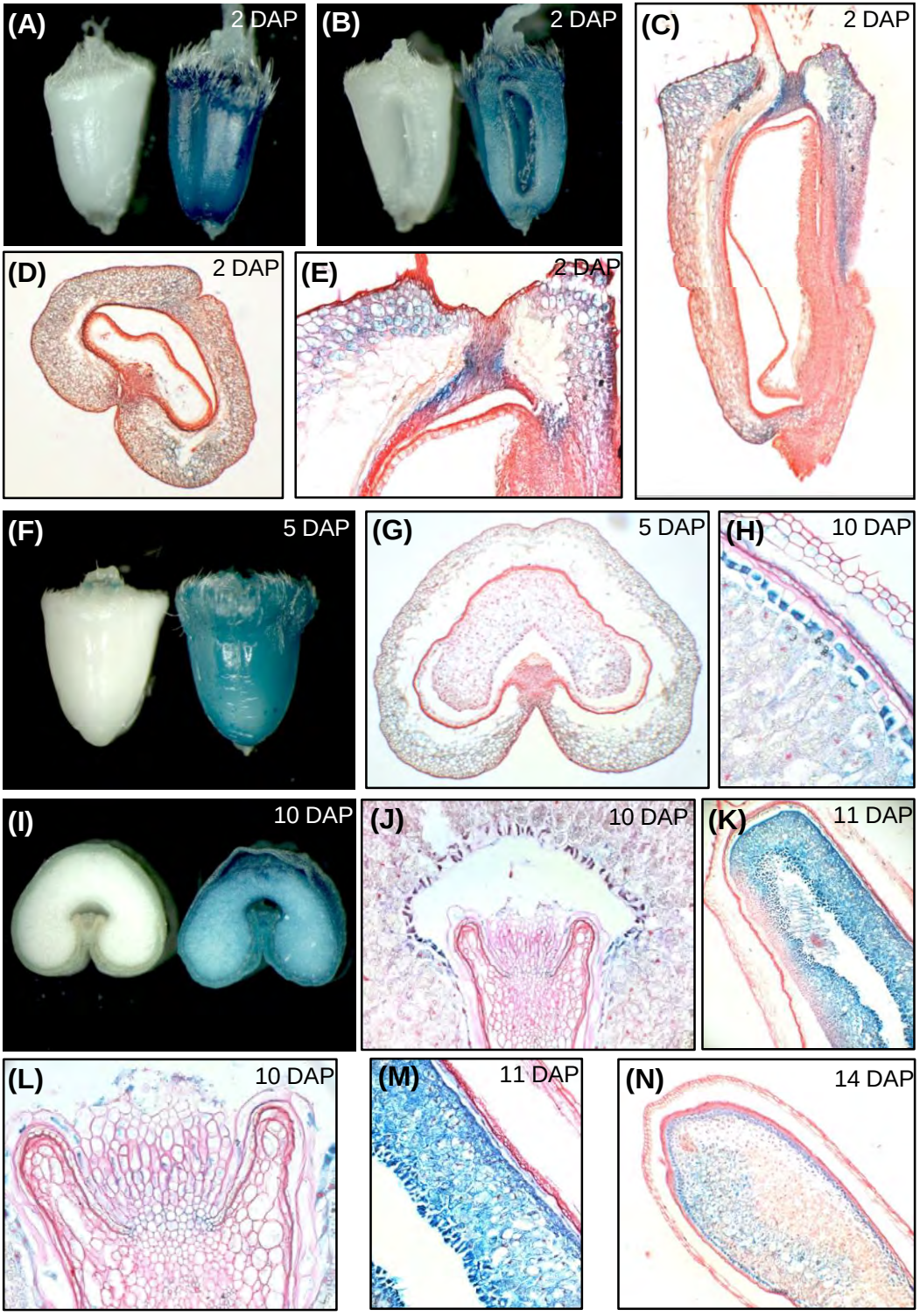

(A)
2 DAP
(B)
2 DAP
(C)
2 DAP
(D)
2 DAP
(E)
2 DAP
(F)
5 DAP
(G)
5 DAP
(H)
10 DAP
(I)
10 DAP
(J)
10 DAP
(K)
11 DAP
(L)
10 DAP
(M)
11 DAP
(N)
14 DAP

Figure 6

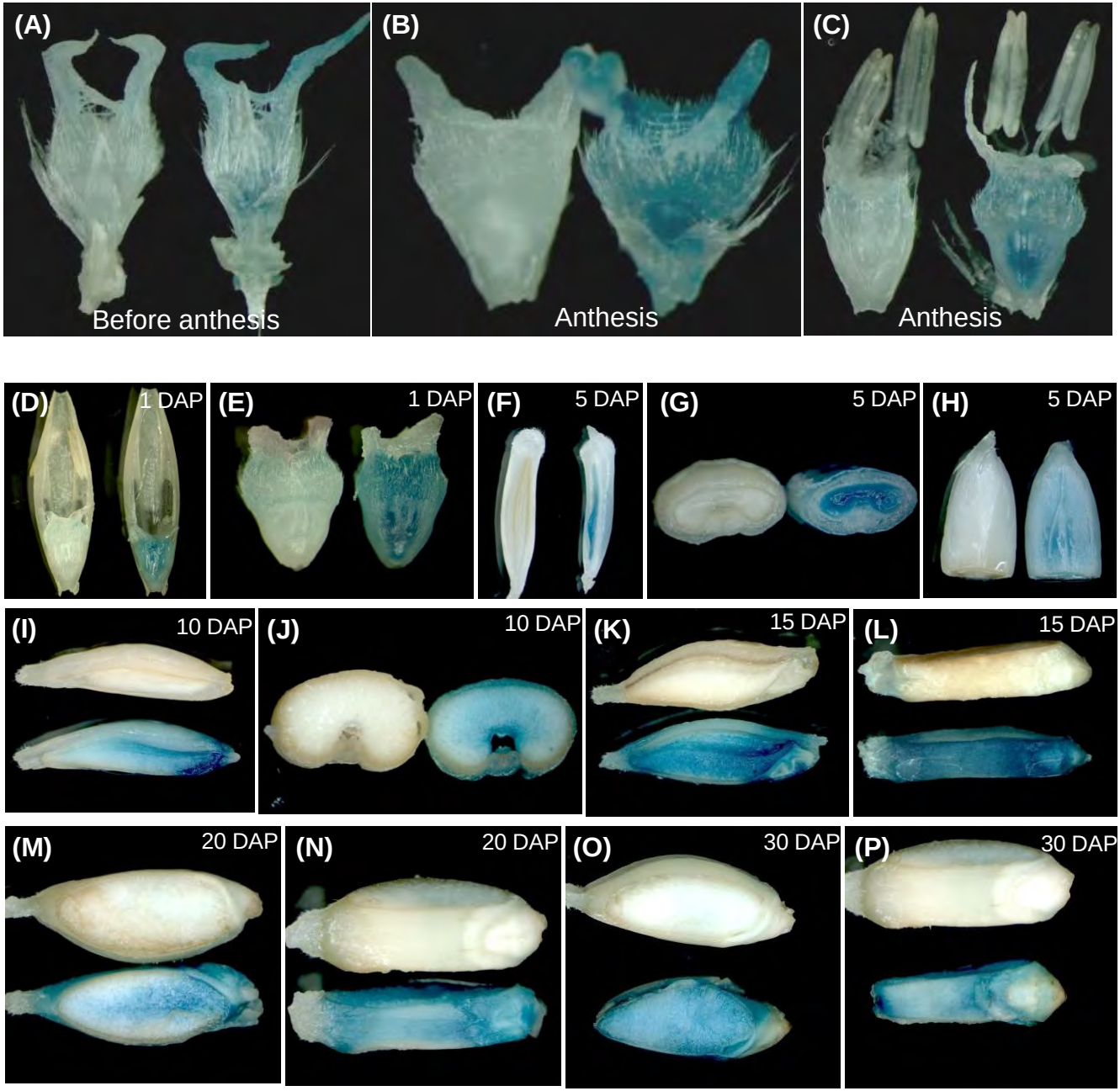

(A)
Before anthesis
(B)
Anthesis
(C)
Anthesis
(D)
1 DAP
(E)
1 DAP
(F)
5 DAP
(G)
5 DAP
(H)
5 DAP
(I)
10 DAP
(J)
10 DAP
(K)
15 DAP
(L)
15 DAP
(M)
20 DAP
(N)
20 DAP
(O)
30 DAP
(P)
30 DAP

Figure 7

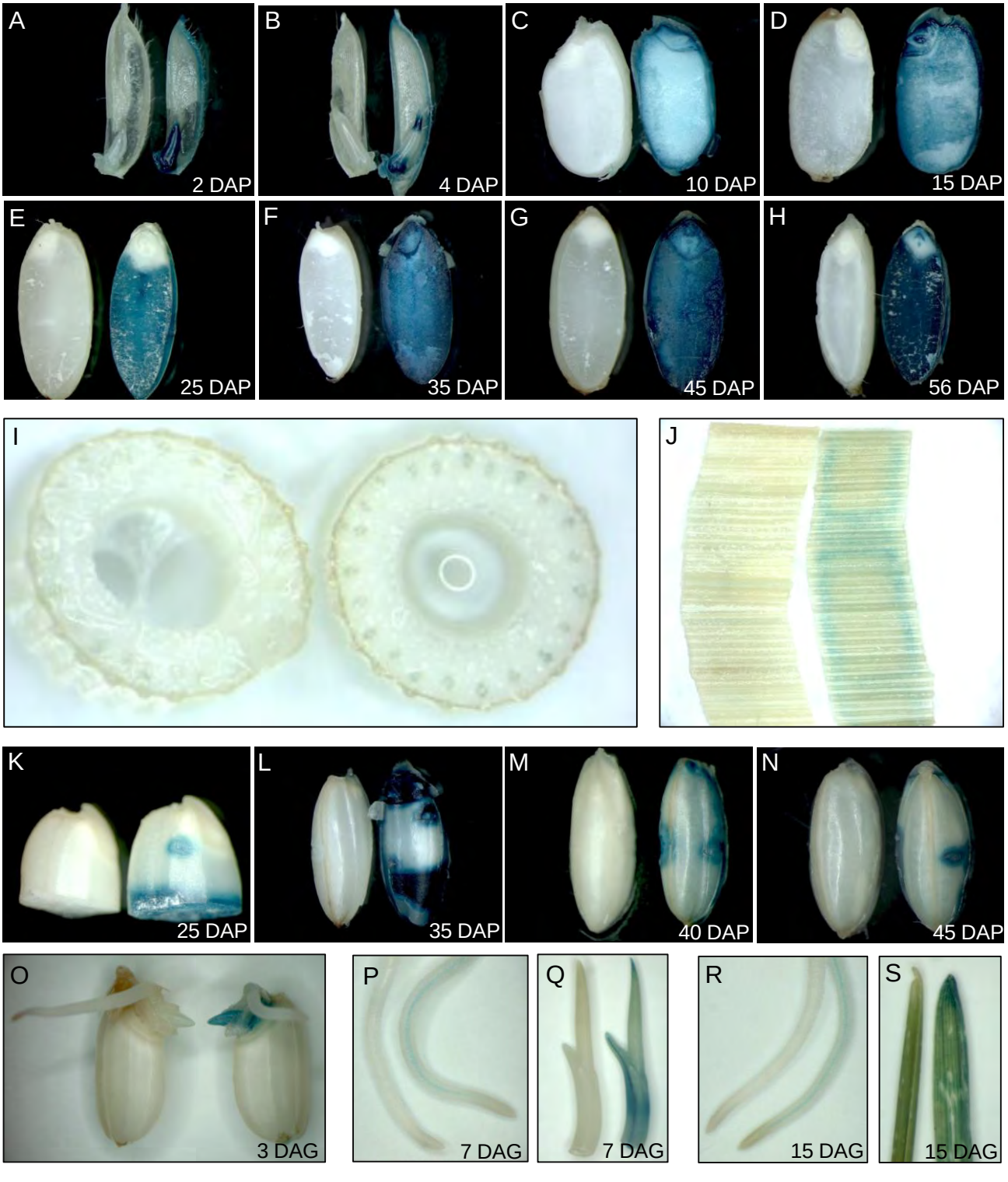
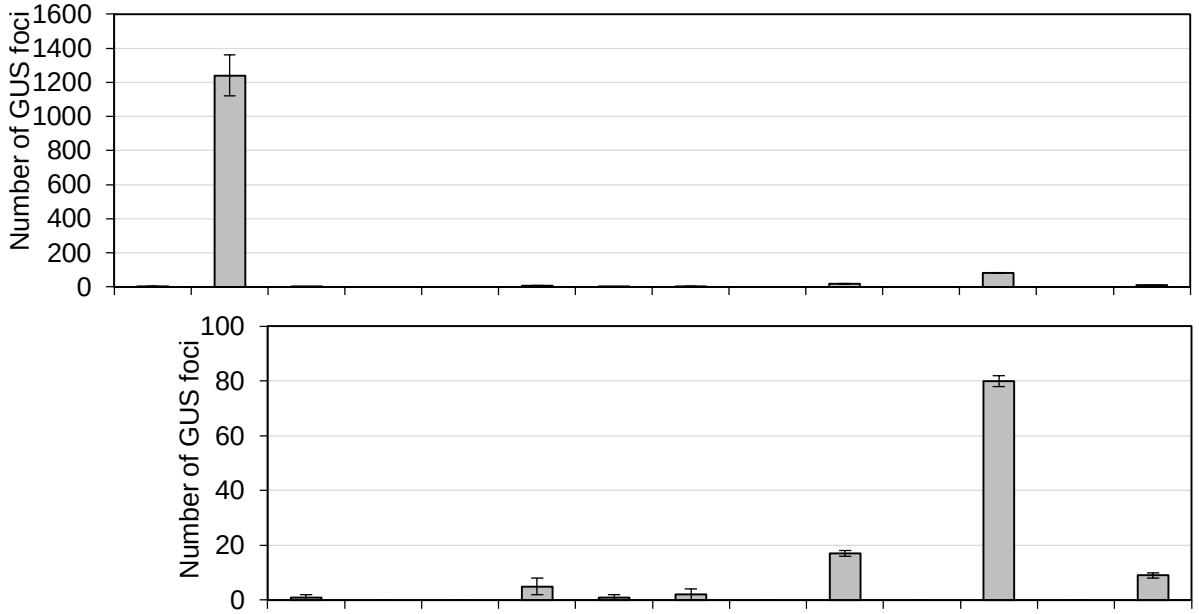

A
2 DAP
B
4 DAP
C
10 DAP
D
15 DAP
E
25 DAP
F
35 DAP
G
45 DAP
H
56 DAP
I
J
K
25 DAP
L
35 DAP
M
40 DAP
N
45 DAP
O
3 DAG
P
7 DAG
Q
7 DAG
R
15 DAG
S
15 DAG

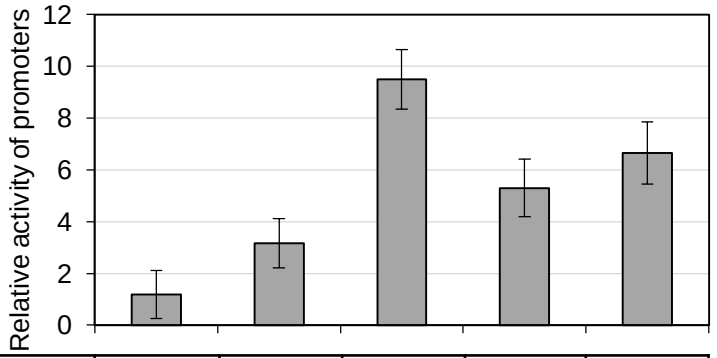
Figure 8

(A)



Promoter	PI-1	PI-1	PI-5	PI-5	PI-7	PI-7	PI-8	PI-8	PI-10	PI-10	PI-11	PI-11	PR 61	PR 61
Effector	C	GL7	C	GL7	C	GL7	C	GL	C	GL7	C	GL7	C	GL7

(B)



Promoter	PI-1	PI-10	PI-1	PI-11	pUbi
Effector	GL7 antisense	GL7	GL7	GL7	GL7 antisense

Figure 9