Sunflower germin-like proteins: evolution, gene structure and functional characterization.

Guillermo F. Ehrenbolger[1,3], Valeria Beracochea[1], Lucila Peluffo[1,3], Natalia I. Almasia[1], Cecilia Vazquez Rovere[1,3], Esteban Hopp[1,2,3], Norma Paniego[1,3] , Ruth Heinz[1,2,3] Veronica Lia[1,2,3]. vlia@cnia.inta.gov.ar.

1. Instituto de Biotecnología, CICVyA, INTA Castelar, Las Cabañas y Los Reseros, (B1712WAA) Castelar, Provincia de Buenos Aires, Argentina. 2. Facultad de Ciencias Exactas y Naturales, Universidad de Buenos Aires, Buenos Aires, Argentina. 3. CONICET

ABSTRACT

- Germin-like proteins (GLPs) are ubiquitous plant glycoproteins which belong to the cupin super family. They are encoded by a heterogeneous group of genes present in many land plants including monocots, dicots, gymnosperms and mosses. They were classified in a number of subfamilies that ranges from 6 to 12 depending on the species. Several studies have reported that GLP family members may function as complex QTL conferring broad-spectrum resistance to fungal and bacterial pathogens. Although overexpression of GLPs from different plant species has been shown to confer tolerance to *Sclerotinia sclerotiorum*, one of the main constraints of sunflower production worldwide, no data is available to date on the diversity and molecular characterization of these proteins in sunflower. Here we investigate the evolution, diversification, and function of sunflower GLPs (HaGLPs) with the aim of identifying new candidate genes for crop improvement.
- Sixteen sunflower expressed sequence tags (ESTs) containing germin motifs were retrieved from publically available databases and proprietary tissue specific EST libraries. Phylogenetic analysis was conducted using 212 protein sequences including representatives from wheat, rice, barley, Arabidopsis, soybean and wild Helianthus species, among others. Intron/exon organization was examined by comparing ESTs to the corresponding genomic sequences obtained through PCR amplification. Expression patterns were studied by Northern blot and RT-PCR. Arabidopsis transgenic plants overexpressing HaGLP1 were obtained via *Agrobacterium tumefaciens*- mediated transformation. Infection experiments with the fungal pathogens *Rhizoctonia solanii* and *S. sclerotiorum* were assayed on agar plates using 5-10 day-old seedlings.
- Phylogenetic reconstruction allowed identification of 9 putative subfamilies within HaGLPs. The HaGLP genes are divergent in terms of their primary sequence, the size of their encoded proteins and the length of the introns. Pairwise comparisons of the coding regions showed average sequence similarities of 51 and 44% at the nucleotide and amino acid levels, respectively. All HaGLPs analyzed possess N-terminal secretory signals. Expression studies showed that most HaGLPs are transcribed in major plant organs, albeit to varying degrees in different sunflower tissues (root, leaf, stem, flower, receptacle, and seed). Transgenic Arabidopsis lines overexpressing HaGLP1 showed enhanced tolerance to *Rhizoctonia solani* and 6% crude extracts (6%) of these transgenic lines inhibited the mycelial growth of *S. sclerotiorum*.
- The number of GLP subfamilies identified for sunflower is comparable to the number described for complete well-annotated angiosperm genomes, and relationships within these lineages are fairly congruent with established phylogenetic relationships among taxa. The presence of signal peptides in all HaGLPs suggests cell wall/extracellular matrix targeting. The antifungal properties of HaGLP1 will be discussed in the context of the generation of H_2O_2 produced by its enzymatic activity.
- The results from infection and mycelial inhibition assays performed on Arabidopsis transgenic plants overexpressing HaGLP1, along with the affiliations of some members of the family to confirmed defense-related germins of other species, suggest that HaGLPs are valuable candidate genes for future breeding efforts.

Keywords: functional analysis – germin-like proteins - phylogeny – *Sclerotinia sclerotiorum* – sunflower

INTRODUCTION

Pathogen infections represent one of the main constraints on crop productivity. Sunflower head rot, caused by the fungal pathogen *Sclerotinia sclerotiorum*, is an endemic disease in the south-east area of Buenos Aires, Argentina. Outbreaks may result in yield reductions of 10-20% and, under favorable environmental conditions at flowering, infections can lead to total field loss (Pereyra and Escande 1998). Oxalic acid secretion by Sclerotinia has been reported to be an essential determinant of its pathogenicity (Godoy et al. 1990; Cessna et al. 2000). In agreement with these observations, transgenic sunflower plants overexpressing a wheat oxalate oxidase (*gf2.8*) were found to show enhanced tolerance to the pathogen (Hu et al. 2003). Wheat gf2.8is a member of the Germin family, a group of conserved and homogeneouos proteins that possess oxalate oxidase activity and are almost exclusively found among Gramineae. Germins are included withing the Cupin Superfamily, along with the more diverse Germin-like proteins (GLPs), which are present in a wide range of species and are characterized by containing two highly conserved motifs. Four GLP subfamilies were initially identified by Carter and Thornburg (1999), however, the number of lineages seems to vary among different taxa, ranging from 6 to 12 (Carter et al. 1998; Druka et al. 2002; Wu et al. 2000; Zimmerman et al. 2006; Manosalva et al. 2008).

Although the biochemical properties of many of the GLPs described to date are still unknown, superoxide dismutase activity has been reported in different species such as barley, grape, and tobacco (Zimmerman et al. 2006; Godfrey et al. 2007; Carter and Thornburg 2000). Given that most GLPs exhibit glycosylation and cell wall localization signals they have been related to cell wall strengthening and papillae formation (Wei et al. 1998). In addition, they have been proposed to play a role in reactive oxygen species detoxification and to function as signaling molecules inducing a range of defense responses in a direct or indirect manner (e.g. Lane 1994; Zhou et al. 1998).

Over the last years, evidence has accumulated regarding the involvement of GLPs in defense against pathogens (for a review see Davidson et al. 2009). Barley GLPs (HvGer) have been reported to confer resistance to *Blumeria graminis* (Zimmerman et al. 2006), whereas induction after fungal infection has also been observed in Vitis vinifera GLPs (VvGLP) (Godfrey et al. 2007). Recently, the expression of a GLP from *Beta vulgaris* **(**BvGLP-1) in *Arabidopsis* has proved to elevate H_2O_2 content and confer significant resistance to *Verticillium longisporum* and *Rhizoctonia solani* (Knecht et al. 2010).

Despite the fact that GLPs from different plant species have been shown to provide tolerance to *S. sclerotiorum* and other pathogens, no data is available to date on the diversity and molecular characterization of these proteins in sunflower. Here we investigate the evolution, diversification, and function of sunflower GLPs (HaGLPs) with the aim of identifying new candidate genes for crop improvement.

MATERIALS AND METHODS

Identification of Sunflower GLPs and phylogenetic analysis. A tBLASTn search (Altschul et al. 1990) was performed against different data bases (KEGG *Helianthus annuus* data base, DFCI Sunflower Gene Index) using as query a GLP expressed sequence tag (EST) previously identified by Fernandez et al. (2003). Using a threshold E-value of 10^{-10}, sixteen sunflower EST sequences were retrieved and subsequently used as queries to search for GLPs in other Helianthus species. Sunflower ESTs corresponding to coding regions shorter than 50 amino acids were not considered for analysis. A total of 212 GLP sequences including representatives from *Vitis vinifera*, *Hordeum vulgare*, *Beta bulgaris*, *Triticum aestivum*, *Glycine max*, *Oryza sativa* and *Lactuca sativa*, among others, were included in the phylogenetic analysis. Accession numbers of the sequences used for this study are available upon request. Nucleotide sequences were translated into the appropriate ORF and aligned using the MAFFT routine E-L-INS-i (Katoh et al. 2002). Protein alignments were used to obtain the phylogenetic relationships through Maximum Parsimony (MP) using the software TNT (Goloboff et al. 2008). Finally, sunflower orthologous and paralogous sequences were determined based on the phylogenetic relationships found for each locus.

Analysis of gene structure. DNA was extracted from lyophilized leaves of lines RHA266 and PAC2 using NucleoSpin® 96 Plant II (Macherey-Nagel, Argentina). Primers were designed for each of the Helianthus GLP families identified by phylogenetic analysis and used to amplify the corresponding regions from genomic DNA. PCR amplification was conducted as described in Fusari et al. (2008).

Exon-intron structure was established by comparison of genomic vs. EST sequence data. Prediction of N-terminal signal peptides was conducted by using the SignalP 4.0 server (http://www.cbs.dtu.dk/services/SignalP) (Petersen et al. 2011)

Analysis of gene expression. RNA was isolated from different tissues and developmental stages (leaves, stems, roots, seeds, R3 florets, R3 receptacles, R5.2 florets, R5.2 receptacles, R6 florets, and R6 receptacles) of inbred line 5393-E (Sunflower Breeding Program INTA Manfredi). The RNA was extracted with RNAqueous® kit (Ambion®, USA), according to manufacturer's instructions. To remove polysaccharides and polyphenolics the tissue was homogenized in a mixture of RNA Isolation Aid (Ambion®, USA) and RNAqueous® Lysis/Binding solution. One µg of DNAse-treated RNA was used to perform RT-PCR with the SuperScript III Reverse Transcriptase (Invitrogen, USA) using random primers according to the kit instructions. The resulting cDNA was used as template for PCR amplification using primer sequences unique to each of the HaGLP genes.

Arabidopsis transformation. *HaGLP1* was cloned into the Gateway ™ pK2 binary vector under the transcriptional control of the 35S promoter. The recombinant binary vector was transformed into *Agrobacterium tumefaciens* GV3101. *Arabidopsis thaliana* (ecotype Columbia-0) was transformed by using the floral dip method (Clough and Bent 1998). Transgenic plants were propagated under selective condition to the T3 generation. The homozygous phenotype was selected by kanamycin resistance. The presence of the transgene was confirmed by transgene-specific PCR amplification and Northern Blot assays.

Effect of leaf extract from transgenic plants on fungal mycelial growth *in vitro*. Leaf tissue (15 g) from transgenic control plants was ground in sodium phosphate buffer (pH 7.0) and incubated at room temperature for 30 min. The leaf extract was centrifuged at 1 700 *g* and the supernatant was filter-sterilized. Different aliquots of the extract (1.5, 3, and 6%) were added to minimal medium after cooling to about 46 °C. Mycelial plugs of *R. solani* and *S. sclerotiorum* were inoculated in the centre of the leaf extract-amended medium and the radial growth of the fungi was recorded (Almasia 2009).

Infection assay of Arabidopsis plants with *R. solani* and *S. sclerotiorum* on agar plates. Seedling infection assays were conducted as described in Knecht et al. (2010) with minor modifications. Briefly, seedlings were incubated with 4 mm diameter mycelium plugs from *R. solani* or *S. sclerotiorum* cultures. At 3, 6, 9, and 14 days after inoculation, the plant survival ratio was scored.

RESULTS AND DISCUSSION

Sixteen *H. annuus* ESTs containing germin motifs were retrieved from public databases and proprietary tissue specific EST libraries. After quality checking and assembly, nine of them were subsequently used for further analysis (HaGLP1, HaGLP2, HaGLP3, HaGLP4, HaGLP5, HaGLP6, HaGLP7, HaGLP8 and HaGLP16). Phylogenetic analysis of the complete data matrix (212 terminal nodes – 204 characters) revealed that each of these genes can be assigned to a different subfamily. A reduced cladogram (81 terminal nodes) depicting the more relevant affiliations is presented in Fig. 1. Five main groups can be distinguished. Groups II, III, IV and V are in general agreement with the major subfamilies reported to date (Carter and Thornburg 2000, Davidson et al. 2009), whereas group I has a basal position on the tree and consists exclusively of Helianthus representatives, including HaGLP7. HaGLP5, HaGLP6 and HaGLP16 are placed within clade II, which also contains HvGer5a and HvGer6. HaGLP1 and HaGLP2 cluster within group III, along with HvGER4 and VvGLP3, both of which have demonstrated SOD activities (Christensen et al. 2004, Godfrey et al. 2007). HaGLP3 and HaGLP4 belong to the same group as BvGLP-1, HvGER2a and VvGLP6.

The HaGLP genes are divergent in terms of their primary sequence, the size of their encoded proteins and the length of the introns. The deduced open reading frames range in length from 209 to 224 amino acid residues. A single exon contains the complete coding region of HaGLP7, HaGLP3 and HaGLP4. In contrast, the remaining HaGLPs are organized into two exons. A short intron (ca. 100 bp) was found in HaGLP1 and HaGLP2, whereas HaGLP5, HaGLP6, HaGLP8 and HaGLP16 exhibited larger intervening sequences (1.5-2 Kb). The three germin motifs described by Godfrey et al. (2007) were identified in all HaGLPs, although with varying degrees of conservation. Histidine and glutamate residues that have been reported to bind manganese at the active site of the protein (Requena and Bornemann 1999) were found in the nine HaGLPs examined.

In agreement with previous studies suggesting cell wall and extracellular matrix targeting (Davidson et al. 2009 and references therein), N-terminal peptides between 18 and 25 residues long were

also predicted for all the sequences analysed here. Pairwise comparisons of the coding regions showed average sequence similarities of 51 and 44% at the nucleotide and amino acid levels, respectively.

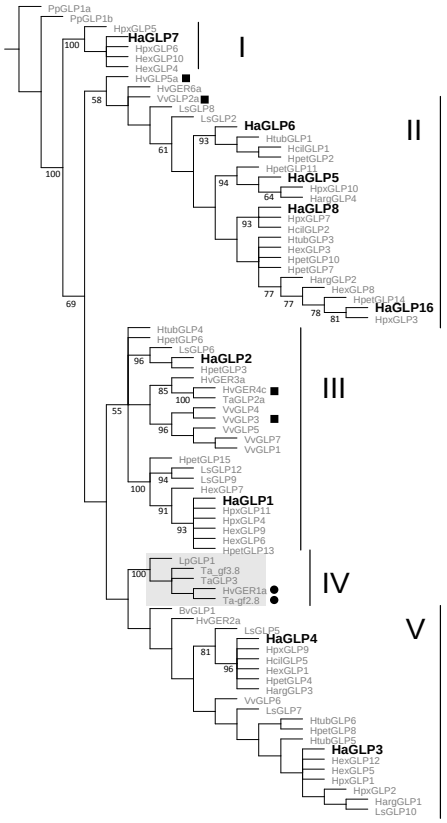

Fig. 1. Phylogenetic Analysis of Sunflower Germin-like proteins (HaGLPs). Strict consensus of 300 maximum parsimony trees. Jacknife support values are given below branches. Genes with experimentally confirmed Oxalate oxidase activity are indicated with circles, genes with experimentally confirmed Superoxide dismutase activity are indicated with squares. The Germin clade is shaded in grey. Bv: *Beta vulgaris*; Ha: *Helianthus annuus*; Harg: *Helianthus argophyllus*; Hex: *Helianthus exilis*; Hcil: *Helianthus ciliaris*; Hpx: *Helianthus paradoxus*; Hpet: *Helianthus petiolaris*; Htub: *Helianthus tuberosus*; Hv: *Hordeum vulgare*; Ls: *Lactuca sativa;* Lp: *Lolium perenne*; Pp: *Physcomitrella patens*; Ta: *Triticum aestivum*; Vv: *Vitis vinifera.*

Semiquantitative RT-PCR experiments showed that HaGLPs are transcribed in all major plant organs, albeit to varying degrees in different sunflower tissues (root, leaf, stem, flower, receptacle, and seed). HaGLP7 and HaGLP4 were found to be the most ubiquitous, also showing the highest expression levels. HaGLP3 exhibited the lowest abundance and the most restricted expression pattern. Interestingly, HaGLP3 transcripts were the only ones detected in mature seeds. The results presented here are in agreement with the recent review by Davidson et al. (2009), in which the expression patterns of 89 germin-like protein families from different species were examined. In general, germin-like genes are expressed in all tissue types and are induced by biotic and/or abiotic stresses. Most of them (74%) are expressed in leaf tissue including cotyledons, shoots, young and mature leaves, and somewhat fewer are expressed in root tissue (61%) including young and mature roots. Finally, a limited number of germins are expressed in flowers (29%) and seeds (45%).

Five transgenic lines of *A. thaliana* overexpressing HaGLP1 were obtained for functional characterization of the gene. Total leaf protein extracts of the transgenic line 2B were able to reduce

mycelial growth of *S. sclerotiorum* when applied at 6% v/v (Fig. 2A). No effect was observed either for the other lines and concentrations tested or for the pathogen *R. solani*.

To further investigate the role of HaGLP-1 in plant-phatogen interactions, transgenic plants were inoculated with *S. sclerotiorum* and *R. solani* on agar plates. No significant differences were observed among controls and transgenic lines after infection with *S. sclerotiorum*. Interestingly, fourteen days after infection with *R. solani*, most of the transgenic plants showed normal growth, with the survival rate being always above 85% while non-transgenic Col-0 plants showed strong disease symptoms, and only 58% of them sur

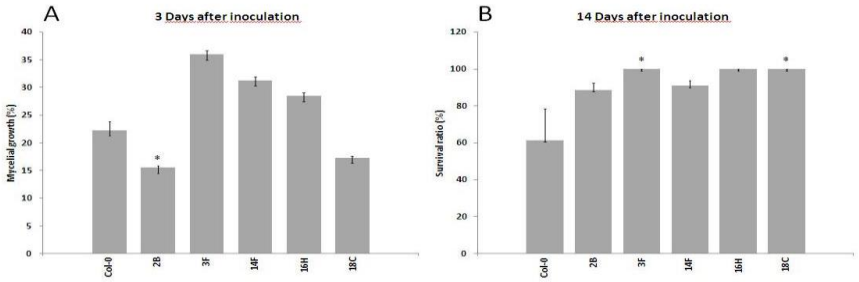

Fig. 2. Characterization of transgenic Arabidopsis plants overexpressing HaGLP1. **A.** Effect of leaf extracts on the mycelial growth of S. sclerotiorum. **B.** Infection assays with R. solani on agar plates. * Student t-test, $p < 0.05$.

The data presented here suggest that the expression of HaGLP1 in *A.thaliana* interferes with *R. solani* infection. The role of this gene in controlling *S. sclerotiorum* infection still needs further clarification, since only one of the lines exhibited inhibitory activity, and no evidences of resistance were observed in the infection assays. The participation of GLP genes in resistance to *R. solani* has also been reported for other species. Rice GLP genes, including a tandemly duplicated cluster of 12 members, have been shown to contribute collectively to resistance against the fungal pathogens *Magnaporthe oryzae* and *R. solani* (Manosalva et al. 2008). In addition, overexpreesion of BvGLP1 in Arabidopsis yielded enhanced tolerance to this pathogen and to *Verticillium longisporum* (Knecht et al. 2010).

Although the molecular mechanisms by which HaGLP1 may aid in pathogen resistance remain largely unknown, its phylogenetic relatedness to GLP proteins of proven SOD activity may allow hypothesizing about its biochemical function. Superoxide dismutases are proposed to be involved in basal defense responses, specifically through H_2O_2 generation (Christensen et al., 2004; Zimmermann et al., 2006). H_2O_2 in the extracellular matrix can act as a signaling molecule and also can participate in several cell wall remodeling processes, including cell wall polymer crosslinking, peroxidase-dependent wall lignification, formation of wall appositions (papillae), and cell expansion and elongation (Davidson et al. 2009). Considering the diversity of sunflower GLPs, future work will be focused on establishing the biochemical properties of the different germin-like families and their role in different aspects of basal defense responses.

AKNOWLEDGEMENTS

Several grants from CONICET, INTA and the ANPCyT are gratefully acknowledged.

REFERENCES

Petersen, T.N., Brunak, S., von Heijne G., and H. Nielsen. 2011. SignalP 4.0: discriminating signal peptides from transmembrane regions. *Nature Methods*, 8:785-786.

Almasia, N.I. 2009. Estudio Genómico y Funcional del péptido antimicrobiano SNAKIN-1. Ph.D diss.,Univ. of Buenos Aires, Buenos Aires, Argentina.

Altschul, S.F, Gish, W., Miller, W., Myers E.W., and D.J. Lipman. 1990. Basic local alignment search tool. J Mol Biol 1990, 215:403-410.

Carter, C., Graham R.A., and R.W. Thornburg. 1998. *Arabidopsis thaliana* contains a large family of germin-like proteins: characterization of cDNA and genomic sequences encoding 12 unique family members. Plant Mol Biol 38: 929–943

Carter, C., and R.W. Thornburg. 1999. Germin-like proteins: structure, phylogeny and function. J. Plant Biol. 42: 97–108.
Carter, C., and R.W. Thornburg. 2000. Tobacco Nectarin I. Purification and characterization as a germin-like manganese superoxide dismutase implicated in the defense of floral reproductive tissues. J. Biol. Chem. 275: 36726-36733.
Cessna, S.G., Sears, V.E., Dickman, M.B., and P.S. Low. 2000. Oxalic acid, a pathogenicity factor for Sclerotinia sclerotiorum, suppresses the oxidative burst of the host plant. Plant Cell 12: 2191-2200.
Christensen, A.B., Thordal-Christensen, H., Zimmermann, G., Gjetting, T., Lyngkjaer, M.F., Dudler, R., and P. Schweizer. 2004. The germinlike protein GLP4 exhibits superoxide dismutase activity and is an important component of quantitative resistance in wheat and barley. Mol Plant Microbe Interact 17: 109–117.
Clough, S.J., and A.F. Bent. 1998. Floral dip: a simplified method for Agrobacterium-mediated transformation of Arabidopsis thaliana. Plant J. 16(6):735-43.
Davidson, R.M., Reeves, P., Manosalva, P.M., and J.E. Leach. Germins: a diverse protein family important for crop improvement. Plant Sci. 2009;177:499–510.
Druka, A., Kudrna, D., Kannangara, C.G., von Wettstein, D., and A. Kleinhofs. 2002. Physical and genetic mapping of barley (Hordeum vulgare) germin-like cDNAs. Proc Natl Acad Sci USA 99: 850–855
Fusari, C.M., Lia, V.V., Hopp, H.E., Heinz, R.A., Paniego N.B. 2008. Identification of single nucleotide polymorphisms and analysis of linkage disequilibrium in sunflower elite inbred lines using the candidate gene approach. BMC Plant Biology 8:7.
Fernandez, P., Paniego, N., Lew, S., Hopp, H.E., and R.A. Heinz. 2003. Differential representation of sunflower ESTs in enriched organ-specific cDNA libraries in a small scale sequencing project. BMC Genomics 4:40.
Godfrey, D., Able, A.J., and I.B. Dry. 2007. Induction of a Grapevine Germin-Like Protein (VvGLP3) Gene Is Closely Linked to the Site of Erysiphe necator Infection: A Possible Role in Defense? Molecular Plant-Microbe Interactions 20: 1112-1125.
Goloboff, P.A., Farris, J.S., and K.C. Nixon. 2008. TNT, a free program for phylogenetic analysis. Cladistics 24:774-786.
Hu, X., Bidney, D.L., Yalpani, N., Duvick, J.P., Crasta, O., Folkerts, O., Lu, G. 2003. Overexpression of a gene encoding hydrogen peroxide-generating oxalate oxidase evokes defense responses in sunflower. Plant Physiol. 133:170-81.
Katoh, K., Misawa, K., Kuma, K., Miyata, T. 2002. MAFFT: a novel method for rapid multiple sequence alignment based on fast Fourier transform. Nucl Acids Res 2002, 30:3059-3066.
Knecht, K., Seyffarth, M., Desel, C. , Thurau, T., Sherameti, I., Lou, B., Oelmüller, R., and D. Cai. 2010. Expression of *BvGLP-1* Encoding a Germin-Like Protein from Sugar Beet in *Arabidopsis thaliana* Leads to Resistance Against Phytopathogenic Fungi. Mol. Plant. Microbe Interact. 23: 446-57.
Lane, B.G. 1994. Oxalate, germin, and the extracellular matrix of higher plants. FASEB J 8: 294–301
Lane, B.G.. 2002. Oxalate, Germins, and Higher-Plant Pathogens. IUBMB Life 53: 67-75.
Pereyra V, Escande A. 1994. Enfermedades del girasol en la Argentina. Manual de reconocimiento. INTA, Buenos Aires, Argentina.
Manosalva, P.M., Davidson, R.M., Liu, B., Zhu, X., Hulbert, S.H., Leung, H., and J.E. Leach, A germin-like protein gene family functions as a complex quantitative trait locus conferring broad-spectrum disease resistance in rice, Plant Physiol. 149 (2009) 286–296.
Requena, L., and Bornemann, S. 1999. Barley (*Hordeum vulgare*) oxalate oxidase is a manganese-containing enzyme. Biochem. J. 343:185-190.
Wei, Y.D., Zhang, Z.G., Andersen, C.H., Schmelzer, E., Gregersen, P.L., Collinge, D.B., Smedegaard-Petersen, V., and H.Thordal-Christensen. 1998. An epidermis/papilla-specific oxalate oxidase-like protein in the defence response of barley attacked by the powdery mildew fungus. Plant Mol Biol 36: 101–112.
Wu, S.P., Druka, A., Horvath, H., Kleinhofs, A., Kannangara, C.G., von Wettstein, D. 2000. Functional characterization of seed coat-specific members of the barley germin gene family. Plant Physiol. Biochem. 38:685-698.
Zimmermann, G., Baumlein, H., Mock, H-P, Himmelbach, A., and P. Schweizer P. 2006. The multigene family encoding germin-like proteins of barley. Regulation and function in Basal host resistance. Plant Physiol.142:181-92.
Zhou, F.S., Zhang, Z.G., Gregersen, P.L., Mikkelsen, J.D., de Neergaard, E., Collinge, D.B., and H. Thordal-Christensen. 1998. Molecular characterization of the oxalate oxidase involved in the response of barley to the powdery mildew fungus. Plant Physiol 117: 33–41