Molecular characterization revealed the role of thaumatin-like proteins in stress response in bread wheat

Alok Sharma[1], Himanshu Sharma[1, 2], Ruchika Rajput[3], Ashutosh Pandey[3] and Santosh Kumar Upadhyay[*1]

1. Department of Botany, Panjab University, Chandigarh, India-160014

2. I.K. Gujral Punjab Technical University, Kapurthala, Jalandhar, India-144603

3. National Institute of Plant Genome Research, Aruna Asaf Ali Marg, New Delhi, India-110067

*Corresponding Author

***Corresponding Author: Dr. Santosh Kumar Upadhyay**, Department of Botany, Panjab University, Chandigarh, India-160014. Email: skupadhyay@pu.ac.in, Tel: +91-172-2534001, Fax: +91-172-2779510

ABSTRACT

Thaumatin-like proteins (TLPs) are related to the pathogenesis-related-5 (PR-5) family and involved in stress response. Herein, a total of 93 *TLP* genes were identified in the genome of *Triticum aestivum.* Further, we identified 26, 27, 39 and 37 *TLP* genes in the *Brachypodium distachyon*, *Oryza sativa*, *Sorghum bicolor* and *Zea mays* genomes for comparative characterization, respectively. They could be grouped into small and long TLPs with conserved thaumatin signature motif. Tightly clustered genes exhibited conserved gene and protein structure. The physicochemical analyses suggested significant differences between small and long TLPs. Evolutionary analyses suggested the role of duplication events and purifying selection in the expansion of the TLP gene family. Expression analyses revealed the possible roles of *TLPs* in plant development and abiotic and fungal stress response. Recombinant expression of *TaTLP2-B* in *Saccharomyces cerevisiae* provided significant tolerance against cold, heat, drought and salt stresses. The results depicted the importance of TLPs in cereal crops that would be highly useful in future crop improvement programs.

Keywords: Abiotic stress, Biotic stress, Cereals, Thaumatin-like proteins, *TaTLP2-B*, *Triticum aestivum*

1. Introduction

The growth and development of plants have always been affected by various abiotic and biotic stress conditions. In response to these stress conditions, plants produce numerous molecules including pathogenesis-related (PR) proteins [1,2]. PR proteins belong to a superfamily of defence-related proteins consisting of PR1-PR17 protein families, that have been classified based on amino acid sequences, serological reaction, enzymatic activity etc. [3]

Thaumatin-like proteins (TLPs) are related to the PR5 family of the PR superfamily [3]. These are called TLPs due to their significant similarity with a ~23 kDa sweet-tasting protein known as thaumatin [4]. Thaumatin protein was first isolated from an African shrub known as *Thaumatococcus danielli* [5]. Over the period, TLPs have been reported in a diverse group of organisms including insects, nematodes, fungi and plants [6,7]. In plants, TLPs are reported from algae to angiosperms [6-9].

Based on the molecular weight (MW), TLP proteins are grouped as long (L-type) and small (S-type) TLPs, having the MW ~21-26 kDa and ~16-17 kDa, respectively. Long TLPs have been identified from each group of plants, while small TLPs are confined to only gymnospermous and monocot plants [6,9]. A total of 16 and 10 conserved cysteine residues, forming eight and five disulfide linkages, are known to be present in long and small TLPs, respectively. These disulfide bonds provide resistance against extreme pH, heat and protease degradation [6,10]. Each TLP consists of a conserved REDDD motif and a thaumatin signature motif G-X-[GF]-X-C-X-T-[GA]-D-C-X- (1,2)-G-X-(2,3)-C. The REDDD motif is involved in receptor binding for their antifungal action [6]. Numerous studies reported that the overexpression of TLPs provides significant resistance against various fungi in both dicot and monocot plants [11-13]. The mechanism of action of TLPs in fungal resistance is ambiguous; however, it has been presumed

that they work by degradation and permeabilization of the fungal cell walls [6,14]. Besides, TLPs are also known to be involved in antifreeze action and abiotic stress resistance in plants [15]. Further, they are also reported to be involved in flowering, fruit ripening and seed germination [16-18].

Despite utmost importance, a detailed characterization of TLPs is still limited in various cereal crops. Therefore, the current study aims at an inclusive characterization of TLPs in bread wheat, an important cereal crop. Besides, we have also performed a comparative analysis of various features of TLPs in other cereals including *B. distachyon*, *O. sativa*, *S. bicolor* and *Z. mays*. Chromosomal distribution, phylogeny, physicochemical properties, gene duplication events, *cis*-regulatory elements, and gene and protein structural analyses were performed. To decipher their roles in plant development and under abiotic and biotic stresses, their expression analysis was performed using high throughput RNA-seq data. The expression of eight selected *TaTLP* genes was validated using qRT-PCR. An abiotic stress-responsive gene, *TaTLP2-B*, of *T. aestivum* was cloned and used for functional characterization in *Saccharomyces cerevisiae* (yeast) cells. Recombinant expression of TaTLP2-B provided significant tolerance to the yeast in spot assay against various abiotic stresses. The study will increase the knowledge about numerous TLP proteins and provide the base for the functional characterization of identified genes in future studies.

2. Material and methods

2.1. Plant materials, growth conditions and stress treatments

Surface sterilization of bread wheat seeds (cv. Chinese Spring) was done with 1.2% sodium hypochlorite in 10% ethanol followed by the washing with double autoclaved water. The

sterilized seeds were kept overnight on moist Whatman filter papers at 4°C for stratification and were further kept at room temperature for germination. The seedlings were then transferred to fresh phytaboxes and were allowed to grow in plant growth chambers under 60% relative humidity, 22°C temperature and a 16 and 8 h light and dark period, respectively [19]. Seven-days old seedlings were kept under heat (40°C), drought (20% PEG) and combined heat and drought stress treatments for 6, 12, 24 and 48 hours (h). For salinity stress, plants were subjected to 150 mM NaCl for 6, 12, 24 and 48 h. The plant samples including roots and shoots were collected and stored at -80°C till further experiments.

2.2. Identification and nomenclature of the TLP genes

Extensive BLAST searches were used to identify the TLP proteins in five different cereals including *B. distachyon*, *O. sativa*, *S. bicolor*, *T. aestivum* and *Z. mays*. Arabidopsis TLP sequences were used as a query against the protein model sequences of each crop downloaded from the Ensembl Plants [9,20]. The presence of the thaumatin (PF00314) domain was analyzed using the Hidden Markov Model (HMM) and Pfam BLAST searches at e-value 10^{-10}. Identified TLP sequences were further subjected to the NCBI Conserved Domain Database (CDD) BLAST to further confirm the presence of the thaumatin domain. The *TLP* genes, having a complete thaumatin family signature, were selected for further analysis. Identified *TLPs* were named as per their sequence of occurrence at various chromosomes in each crop, except *T. aestivum*. The international rules for gene symbolization of *T. aestivum* (http://wheat.pw.usda.gov/ggpages/wgc/98/Intro.htm) were followed for the nomenclature of *TaTLPs*.

2.3. Chromosomal localization and duplication events

Plant Ensemble (http://plants.ensembl.org/) was used for gathering the chromosomal and sub-genomic location of *TLPs* of all five crops. The homeologous *TLP* genes in *T. aestivum* were identified based on ≥90% sequence similarity and their occurrence at the related chromosomes. The MapInspect software was used for the graphical representation of *TLP* genes on their respective chromosomes (http://www.plantbreeding.wur .nl/uk/software_mapinspect.html.2012). Duplication events (DEs) were identified using a bidirectional blast hit approach with sequence similarity of ≥80%, while tandem and segmental DEs were segregated based on their distance and occurrence at respective chromosomes as per the previous studies [21].

2.4. Multiple sequence alignments, phylogeny and structural analysis

The Muscle and Multalin tools were used for the multiple sequence alignments of all the TLP*s* with a known thaumatin protein (P02883.2|THM1_THADA) to find the conserved residues [22,23]. The phylogenetic tree was constructed by the maximum likelihood method with 1000 bootstraps using the MEGA X software [24].

2.5. Ka/Ks and Tajima's relativity test

The alignment of the protein and nucleotide sequences of the paralogous gene pairs was done using the ClustalOmega server (https://www.ebi.ac.uk/Tools/msa/clustalo). The synonymous substitution per synonymous site (Ks), the non-synonymous substitution per non-synonymous site (Ka), and their ratio (Ka/Ks) were calculated using the PAL2NAL programme [25,26]. The calculation of the divergence time of each pair of duplicated genes was done using the formula T=Ks/2r, where T represents the divergence time and r represents the divergence rate. The divergence rate was assumed to be 6.5×10^{-9} for cereals [27]. The Tajima's relativity test was performed to find out the evolutionary rate between paralogous genes [28].

2.6. Gene structure organization

The genomic and coding DNA sequence (CDS) of identified *TLP* genes were used for the analysis of gene structure in terms of exon-intron organization and intron phases using the GSDS 2.0 server as done in earlier studies [29,30].

2.7. Physicochemical properties of the TLPs

Various physicochemical characteristics such as peptide length, molecular weight (MW) and isoelectric point (pI) were analyzed using the ExPasy tool, which were further confirmed from the Ensemble plants and Sequence Manipulation Suite [31-33]. Tools including CELLO v.2.5, ngLOC, ProtComp9 and WoLF PSORT were used for the prediction of subcellular localization of TLP proteins [34-37]. Tools such as Phobius and DAS-Tmfilter were used for the prediction of transmembrane regions [38,39]. The signal peptides were detected using the tools Phobius and SignalP [38,40]. SMART server was used for the domain analysis, whereas, motifs were analyzed using MEME v.4.11.4 [41,42].

2.8. Cis-regulatory element analysis

For *cis*-regulatory elements analysis, 1.5 kb upstream genomic sequences from the initiation codon were retrieved for each *TLP* gene. These promoter sequences were analyzed using the PLACE software [43]. Identified *cis*-regulatory elements were categorized based on their functions.

2.9. Expression profiling using RNA-seq data

Expression analysis in various tissues of each crop was done using the high throughput RNA-seq data retrieved from the URGI database (wheaturgi.versailles.inra.fr/files/RNASeqWheat/) and Expression ATLAS [44-46]. In *T. aestivum*, the expression data generated in replicates for various tissue developmental stages, under biotic (fungal pathogen) and abiotic (heat, drought

and salt) conditions were used [47-49]. Data generated for three developmental stages of root, stem, leaf, spike and grain was used to analyze the tissue-specific expression in wheat. Further, the RNA-seq data (PRJNA243835) generated in triplicates after the 24, 48 and 72 h of inoculation of *Blumeria graminis* f. sp. *tritici* (Bgt) and *Puccinia striiformis* f. sp. *tritici* (Pst) in 7-d-old seedlings were used for the *TaTLP* expression analysis under biotic stress [47]. Under abiotic stress conditions, duplicate RNA-seq data (SRP045409) for 1 and 6 h of treatments of heat (40º C), drought (20% PEG 6000) and a combination of both heat and drought stresses were used to study the *TaTLP* expression [48]. Root RNA-seq data (SRP062745) available in triplicates for the treatment of 150 mM NaCl at 6, 12, 24 and 48 h treatment were used to analyze the expression of *TaTLPs* under salt stress [49]. Trinity package was used to calculate the expression value in terms of FPKM value [50]. Hierarchical Clustering Explorer 3.5 was used to generate heat maps of differentially expressed genes, which were clustered using the Euclidean distance method [51].

2.11. RNA isolation, cDNA synthesis and qRT-PCR

The total RNA of each sample was isolated using the Spectrum™ Plant Total RNA kit (Sigma, USA). Seedlings (shoot and root) grown under normal conditions were used as a control. TURBO DNA-free™ Kit (Invitrogen, USA) was used to remove the genomic DNA contamination. The qualitative and quantitative analysis of RNA was done using agarose gel electrophoresis and nanodrop quantification. The Superscript III First-Strand Synthesis Super-mix (Invitrogen, USA) was used to synthesize the cDNA from one microgram of total RNA. A real-time qRT-PCR was performed with the gene-specific primers of selected genes using SYBR Green at the 7900 HT Fast Real-Time PCR System (Applied Biosystems) following the method established in our laboratory [21]. The fold expression change was calculated using the delta-

delta CT method ($2^{-\Delta\Delta CT}$) using the expression of ADP-ribosylation factor (*TaARF*) as an internal control [52]. All the experiments were carried out in triplicates and expressed as mean ± SD. A significant difference between the control and treatments was examined by using the two-tailed student's t-test.

2.12. Cloning and functional characterization

The full-length open reading frame (ORF) of the *TaTLP2-B* gene was amplified from cDNA using the *TaTLP2-B* forward (5' GTAATGGCTCTTCTTCCTCCTCTGCTTCTG 3') and *TaTLP2-B* reverse (5' AATCTGGGCCACACGATCGCCCC 3') primers. The amplified DNA was cloned into the pJET1.2 cloning vector and sequenced. *TaTLP2-B* gene was re-amplified from the confirmed clone using the same primers and ligated into the pYES2.1/V5-His-Topo vector (Invitrogen, USA). The recombinant plasmid (pYES2.1-*TaTLP2-B*) was transformed into *S. cerevisiae* (W303) (yeast) cells for further characterization. For control, *lacz* gene containing pYES2.1 vector (pYES2.1/V5-His/lacZ) was used during the studies. Spot assays using recombinant yeast cells were performed under various abiotic stress conditions. Recombinant yeast cells were grown overnight in SD/−ura (with 2% dextrose) medium at 30°C and 200 rpm as primary culture and further inoculated for secondary culture (1:100 dilutions). As soon as OD_{600} reached 0.4, 2% galactose was added to induce the expression of recombinant protein in yeast and kept at 30°C and 200 rpm for 6 h. The OD_{600} was adjusted to 0.4 and an equal volume (500 µl) of each induced culture was further diluted in 10 ml medium. The diluted cultures were treated with heat (37°C), cold (4°C), drought (30% PEG), combined heat- drought and salt (1 M NaCl) stresses for 24 h, separately. After the treatments, serial dilutions (10^0, 10^{-2}, 10^{-4} and 10^{-6}) were prepared, 5 µl of each dilution was spotted on SD/-ura agar plates, and incubated at 30°C

for 2–3 d [21]. All the experiments were performed in triplicates and results were compared visually.

2.13. Subcellular localization of the TaTLP2-B

Subcellular localization of TaTLP2-B was analyzed using CaMV35S-driven C-terminal YFP fusion construct generated by Gateway LR-recombination into binary vector PEG101. Plasmids were then transformed into *Agrobacterium tumefaciens* GV3101 strain. The cells were resuspended in a freshly prepared infiltration medium (10 mM $MgCl^2$, 10 mM MES/KOH, pH 5.7 and 150 μM acetosyringone). The resulting constructs were infiltrated onto the abaxial surface of *Nicotiana benthamiana* leaf and kept at 22°C for 48 h. The YFP fluorescence was observed using a Leica TCS SP8 (Leica Microsystems, Wetzlar, Germany) laser-scanning confocal microscope at 514-527 nm.

3. Results

3.1. Identification and chromosomal localization of the TLP genes

Numerous properties of TLPs ascribed to defence and development pathways and lack of such studies in numerous cereals lead us to perform a comprehensive analysis of the TLP family in five major cereals. An extensive BLAST search identified 26, 27, 39, 93 and 37 *TLP* genes in *B. distachyon, O. sativa, S. bicolor, T. aestivum* and *Z. mays*, respectively. The genes lacking or having incomplete thaumatin signature motifs were excluded from the study. Four *TLPs* were identified as small *TLPs* (sTLPs) in each *B. distachyon, O. sativa* and *Z. mays*, while 10 and 20 *sTLPs* were detected in *S. bicolor* and *T. aestivum*, respectively, (**supplementary file S1**). In the case of *T. aestivum*, identified *TLP* genes from the A, B and D sub-genomes formed 32

homeologous groups based on their sequence homology (≥ 90%). All of the identified proteins exhibited a complete thaumatin signature motif (**supplementary fig. S1)**.

The chromosomal localization suggested the scattered distribution of the *TLP* genes at the majority of chromosomes in each crop (**Fig. 1**). In *B. distachyon,* chromosome 4 consisted of a maximum of 10 *TLP* (both long and small) genes, while in the case of *O. sativa, S. bicolor, T. aestivum* and *Z. mays*, the majority of genes (6, 11, 18 and 11) were localized onchromosomes 12, 8, 5A and 1, respectively.

3.2. Multiple sequence alignments

TLP proteins consist of various characteristic and conserved residues, which are important for their functional activities. Multiple sequence alignments of TLP proteins with a known thaumatin protein (P02883.2|THM1_THADA) revealed the identification of 16 and 10 conserved cysteine residues in the majority of long and small TLP proteins of the studied cereal crops, respectively. Moreover, 7-15 cysteine residues had also been observed in a few TLPs of *O. sativa, S. bicolor, T. aestivum* and *Z. mays*. The REDDD motif was also found conserved in the majority of TLP proteins. However, in certain TLP proteins, a few amino acid residues (AAs) were replaced by other AAs having either similar or different properties. For instance, Arginine (R) was replaced by Lysine (K) or Asparagine (N). This could be responsible for the differential evolvement of various characteristic features of TLPs during evolution. The thaumatin signature motif, which is a characteristic feature of TLPs [6], was found to be highly conserved in all the identified TLP proteins. Some other regions like FF hydrophobic motif and bottom of acidic cleft forming amino acids were also found well conserved (**Fig. 2**, **supplementary fig. S1**).

3.3. Evolutionary analyses

3.3.1. Phylogeny

To decipher the evolutionary relationship, a phylogenetic tree was built using the full-length TLP protein sequences of the five cereals and *A. thaliana*. These were clustered into 11 different clades based on their phylogenetic relatedness, named as groups I to XI (**Fig. 3**). The highest number of TLPs were found in group XI, followed by group II and group X, while group IV was the smallest with only five genes. All the sTLPs were tightly clustered into group XI, which could be due to their smaller size. Further, the majority of groups consisted of TLPs from all the five cereals, except groups III, V and VI that lacked members from one or more plant species. Besides, group IV comprised only three TaTLP and two AtTLP proteins. Besides, the homeologous TaTLPs of *T. aestivum* were tightly clustered in proximity.

3.3.2. Duplication events investigation

Duplication event (DE) analysis was carried out to understand their roles in the evolution and expansion of the TLP gene family in the studied crop species. A total of one, four, three, six and eight DEs were predicted in *B. distachyon, O. sativa, S. bicolor, T. aestivum* and *Z mays*, respectively (**Fig. 1**, **supplementary file S2**). Only tandem duplication events (TDEs) were found in the case of paralogous genes of *B. distachyon* (*BdTLP1-BdTLP2*), *O. sativa* (*OsTLP5-OsTLP7, OsTLP22-OsTLP24, etc.*) and *S. bicolor* (*SbTLP28-SbTLP29, SbTLP31-SbTLP34, etc.*). However, in the case of *T. aestivum*, five DEs (*TaTLP1-A-TaTLP24-D, TaTLP1-B-TaTLP24-B, etc.*) were segmental and one was TDE (*TaTLP16-A-TaTLP17-A*). In *Z. mays*, seven segmental (*ZmTLP1-ZmTLP35, ZmTLP2-ZmTLP34* etc.) and one TDE (*ZmTLP6-*

ZmTLP7) were found. All the duplicated gene pairs of each cereal crop were tightly clustered in proximity in the phylogenetic tree.

3.3.3. Ka/Ks and Tajima's relative rate test

Over the course of evolution, various evolutionary forces and natural pressures affected the duplicated genes [53]. To understand the evolutionary divergence between the paralogous gene pairs, the Ka/Ks analysis was carried out. The Ka/Ks ratio of more than one suggests the positive (non-purifying) and less than one indicates negative (purifying) selection pressure. All the paralogous genes showed the Ka/Ks value lesser than one, which suggested the negative or purifying selection on duplicated *TLP* genes (**Table 1**). However, Ka/Ks analysis for *ZmTLP6* and *ZmTLP7* was not performed due to their 100% similarity. Additionally, the divergence time of DEs was also calculated using the Ks value and previously described methods [54,55]. The divergence time of duplicated genes was calculated as 83 million years ago (MYA) in *BdTLPs*, while the range varied from 52-108, 6-137, 63-99 and 1-63 MYA in *O. sativa, S. bicolor, T. aestivum* and *Z. mays*, respectively (**Table 1**). Tajima's relative rate test found the insignificant $\chi 2$ value for all the duplicated gene pairs at $P > 0.05$ (**Table 2**). The P-value of more than 0.05 depicted their acceptance of the molecular evolutionary clock hypothesis [28].

3.4. Gene architecture, domain and motif analyses

The number of exons varied from one to three in *O. sativa,* while one to four in *B. distachyon, S. bicolor* and *T. aestivum.* In the case of *Z. mays,* most of the *TLPs* consisted of one to three exons, while *ZmTLP16, ZmTLP20, ZmTLP25* and *ZmTLP33* exhibited eleven, seven, eight and ten exons, respectively. Moreover, a total of 44% (98/222) identified *TLPs* were intronless.

Intriguingly, all the *sTLPs* were intronless, except *TaTLP6-A* and *TaTLP20-A1*. The intron phase analysis revealed that the occurrence of a maximum number of introns in phase 1, followed by phase 2, while the least number of introns were in phase 0 (**supplementary fig. S2**).

The functional nature of a protein depends upon the occurrence of domain composition. All of the identified cereals' TLPs consisted of a thaumatin domain (PF00314), which confirmed that they are thaumatin-like proteins. The size of the thaumatin domain ranged from 202-217 AAs to134-154 AAs in the long and small TLPs, respectively. In addition to the thaumatin domain, ZmTLP16, ZmTLP20, ZmTLP25 and ZmTLP33 also consisted of a nuclear protein 96 (NUP96) domain of ~211 AAs at the C-terminus of these proteins (**supplementary file S3**).

Motif investigation revealed the occurrence of 10 highly conserved motifs in the TLP proteins. Motifs 1-9 were parts of the thaumatin domain, while motif 10 was unknown. Motifs 5, 6 and 8 were the most conserved motifs in TLP proteins, in which the thaumatin signature sequence was present in motif 8 (**supplementary fig. S2**). The occurrence of conserved motifs across the cereal species suggested the conserved nature of TLP proteins in related plant species.

3.5. Physicochemical properties

To understand the various important features, several physicochemical properties of identified TLPs were studied. The TLPs were analyzed for molecular weight (MW), peptide length, isoelectric point (pI), subcellular localization, transmembrane (TM) helix and signal peptide (**supplementary file S4**). The average length of the long and small TLPs ranged from284-334 AAs to 175-185 AAs in the studied cereal crops, respectively. Similarly, the average MWs ranged from ~29-35 kDa to ~17-19 kDa, respectively. However, the average pI ranged from 5.89

to 6.95 and from 4.90 to 6.82 for the long and small TLPs, respectively (**Table 3**, **supplementary file S4**).

The majority of TLP proteins lacked TM helices, which suggested their soluble/cytoplasmic nature. However, five TLPs of *B. distachyon* and *O. sativa*, seven of *T. aestivum* and *Z. mays*, and 10 TLPs of *S. bicolor* consisted of a single TM region, and OsTLP18, TaTLP4-D and TaTLP5-D comprised of two TM helices, which suggested their membrane-bound nature (**supplementary file S4**).

A total of 23, 21, 36, 86 and 33 TLPs of *B. distachyon*, *O. sativa*, *S. bicolor*, *T. aestivum* and *Z. mays*, consisted of the N-terminal signal peptide, respectively. Further, the majority of TLP proteins were predicted to be localized in the extracellular region, while three TLPs from *O. sativa* (OsTLP10, OsTLP15 and OsTLP16) showed nuclear, and BdTLP16, TaTLP21-A, ZmTLP20, ZmTLP25 and ZmTLP33 showed plasma membrane localization (**supplementary file S4**). To validate the subcellular localization, a *TLP* gene (*TaTLP2-B*) was cloned and expressed with YFP as a translational fusion protein. Analysis of YFP fluorescence of fusion protein confirmed their extracellular localization (**Fig. 4**). Moreover, we could also see some fluorescence in the cytoplasmic region. The results suggested that the majority of TaTLP2-B was localized in the extracellular region, where they could be involved in various defence related functions (such as anti-fungal etc.). The cytoplasmic localization could be due to their translation in the cytoplasm, or it might also function inside the cytoplasm.

3.6. Promoter region analyses

Promoter elements are necessary for the regulation of gene expression under different conditions. Therefore, we performed the *cis*-regulatory analysis of *TLPs* to foresee their regulatory mechanisms. Based on their putative functions, the identified *cis*-regulatory elements were

segregated into four groups; growth and development, light, hormone and stress-responsive elements (**supplementary table S1**). The promoter regions of *TLP* genes comprised of several growth-related elements including TE2F2NTPCNA, EBOX, RYREPEAT4, RAV1AAT, POLLEN1LEAAT52 etc. In the case of light-responsive elements, GT1CONSENSUS, GATA box, GBOXLERBCS, BOXII, IBOXCORE were some common *cis*-regulatory elements. However, under the hormonal responsive category, DPBFCOREDCD3, ARFAT, ARR1, ERELEE4 and GARE10OSREP1 were abscisic acid, auxin, cytokinin, ethylene and gibberellic acid-responsive elements, respectively.

3.7. Expression profiling of the TLP genes under different tissues and developmental stages

Expression analysis of genes is a significant way to understand their involvement in various developmental and physiological processes. Previously, TLPs are reported to be involved in plant growth and developmental processes [56,57]. Therefore, we performed the expression analysis of TLPs under various tissues and their developmental stages. A total of 23, 27, 37, 84 and 29 *TLP* genes showed expression in one or more tissue developmental stages in *B. distachyon, O. sativa, S. bicolor, T. aestivum* and *Z. mays*, respectively (**Fig. 5A-E**). Based on expression profile, these genes were clustered into 3-5 groups in various crop species. *BdTLPs* of group 1 were highly expressed in early and emerging inflorescence stages, while *BdTLP9* and *BdTLP12*, and *BdTLP20* and *BdTLP23* were upregulated in endosperm and embryo, respectively. Group 2 genes were highly expressed in the leaf, pistil, embryo and endosperm tissues. However, group 3 genes were highly expressed at 10 days after pollination (DAP) of seeds; moreover, *BdTLP8* and *BdTLP18* were upregulated in the pistil, as well (**Fig. 5A**). In *O. sativa*, the majority of group 1 genes were highly expressed in seed, group 2 in callus, group 3

genes in the shoot, group 4 in inflorescence and group 5 genes in emerging inflorescence and anther (**Fig. 5B**). In *S. bicolor*, group 1 and group 4 genes showed higher expression in root and leaf tissues, respectively. However, groups 2 and 3 *TLP* genes were highly expressed in various reproductive tissues like anther, pistil, endosperm, embryo and flower (**Fig. 5C**).

In *T. aestivum*, the majority of group 1 genes exhibited higher expression in root tissue followed by the shoot. However, most of the groups 2, 3 and 4 *TLP* genes were highly expressed in various developmental stages of spike and grain tissues (**Fig. 5D)**.

In the case of *Z. mays*, group 1 genes showed significant expression in unpollinated silk, pericarp and aleurone layer. While, the majority of group 2 genes were specifically upregulated in the root tissues, and group 4 genes in the leaf and internode tissues. Group 3 genes were specifically upregulated in the embryo (**Fig. 5E)**. These results suggested the role of *TLP* genes in both vegetative and reproductive tissues development in all the studied cereal crops. However, the higher expression of a few genes in a specific tissue or developmental stages suggested their precise role in particular tissue development.

3.8. Expression profiling of the TaTLP genes under biotic and abiotic stress conditions

Since TLPs belong to the defence-related protein family and are well known for their stress-responsive behaviour. The expression analysis of *TaTLP* genes was also performed under various biotic and abiotic stress conditions to reveal their roles in stress resistance [47-49]. For biotic stress, the available RNA-seq data generated at 24, 48 and 72 h of infestation of two important fungal pathogens, namely, *Blumeria graminis* f. sp. *tritici* (Bgt) and *Puccinia striiformis* f. sp. *tritici* (Pst) was used for expression analysis [47]. In *T. aestivum*, a total of 50 *TaTLP* genes exhibited differential expression (≥2 fold) in these stress conditions, which could

be clustered into four groups (**Fig. 6A**). The majority of group 1 *TaTLP* genes were upregulated at 24 and 48 h of Bgt infestation. However, all the group 3 and 4 *TaTLP* genes were significantly upregulated at the late (72 h) and early (24 h) periods of Pst infestations, respectively. *TaTLP3-A2*, *TaTLP12-A*, and *TaTLP32-A* were the most upregulated genes after Bgt infection with 37, 36 and 35 folds, respectively. However, in the case of Pst infestation, *TaTLP10-A* (51 fold up) and *TaTLP27-D* (17 fold up) were highly upregulated.

Abiotic stresses are another major threat to plants, which affect various physiological and biological processes. To understand the putative role of *TaTLPs* under abiotic stress conditions, expression analysis was carried out under drought (DS), heat (HS) and combined heat and drought (HD) stress using online available RNA-seq data [48]. A total of 52 *TaTLPs* exhibited significant differential expression in these stresses, which formed four clusters in the heatmap (**Fig. 6B**). Almost all the genes of group 1 were downregulated in the DS, HS and HD stresses, except *TaTLP12-A* that was slightly upregulated in DS treatment. In the case of group 2, the majority of genes were upregulated at 6 h of DS, HS and HD treatments, while either unaffected or downregulated at 1 h of treatment. These might be the late responsive *TaTLP* genes. Most of the group 3 and group 4 *TaTLP* genes were found to be DS and HS responsive, respectively.

Expression analysis of the *TaTLP* genes was also performed at 6, 12, 24 and 48 h of salt stress using available RNA-seq data [49]. Out of 93 genes, 63 *TaTLP* genes were found to be differentially expressed under salt stress, which were clustered into 4 groups, based on their expression patterns (**Fig. 6C**). All the group 1 genes were downregulated at all the stages of salt stress treatment. In contrast, all the group 3 genes were found significantly upregulated at 48 h of treatment. However, group 2 genes were upregulated at the early stage (6 h) of treatment, while

they get normalized at later stages. The results suggested that the group 2 and group 3 genes are early and late responsive *TaTLP* genes, respectively.

To validate the RNA-seq expression data, qRT-PCR analysis of eight *TaTLP* genes was performed using gene-specific primers at various hours of heat, drought and salt treatments (**Fig. 7A-P, supplementary file S5**). Overall the result was in agreement with the expression observed using RNA-seq data, with a few exceptions. In the case of heat stress, *TaTLP2-B*, *TaTLP7-D*, *TaTLP14-B1* and *TaTLP25-B* were more upregulated at 1 h, while other genes were highly upregulated at 6 h of treatment. In the case of drought stress, *TaTLP-7D* was exclusively more upregulated at 1 h, whereas other genes were upregulated at 6 h. However, *TaTLP2-B* and *TaTLP10-D* were downregulated at DS 1 h. The majority of genes were highly upregulated at 1 h HD treatment. In the case of salt stress, seven genes were highly upregulated at 12 h of salt stress, except *TaTLP14-B1* that showed maximum expression at 48 h. The results further confirmed that a few *TaTLP* genes are early while others are late responsive.

3.10. Comparative expression profiling of the TaTLP paralogous genes

To understand the function of duplicated genes, a comparative expression profiling of each paralogous pair was performed. Generally, based on expression pattern, duplicated genes could categorize into the retention of function, pseudo-functionalization and neo-functionalization. Out of six paralogous gene pairs, three pairs (*TaTLP1-A-TaTLP24-D, TaTLP1-B-TaTLP24-B, TaTLP15-A-TaTLP18-B*) showed a similar trend of expression, suggested retention of function in these duplicated genes (**Fig. 8A-C**). Two duplicated pairs (*TaTLP12-D-TaTLP13-D* and *TaTLP16-A-TaTLP17-A*) showed insignificant expression of one of the gene suggested pseudo-functionalization (**Fig. 8D, E**). Retention of function in the majority of duplicated genes and the

absence of neo-functionalization suggested functional conservation in *TLP* genes during evolution.

3.11. Cloning and functional characterization of the TaTLP2-B in Saccharomyces cerevisiae

In our study, since the *TaTLP2-B* exhibited significant differential expression under abiotic stress conditions, it was selected for functional characterization. *TaTLP2-B* open reading frame was amplified using end primers and cloned into pYES2.1 yeast expression vector. The *LacZ* gene-containing vector was used as control (**Fig. 9A-F**). In the spot assay, we observed similar growth of both control (LacZ) and TaTLP2-B expressing yeast cells in the controlled conditions (**Fig. 9A**). However, in the case of cold, drought, heat, combined heat drought and salt stress conditions, TaTLP2-B recombinant cells exhibited higher growth than the control cells (**Fig. 9B-F).** The results revealed that the overexpression of the TaTLP2-B provided abiotic stress tolerance to the recombinant yeast cells. This gene can be used for the development of abiotic stress-resistant transgenic crops in future studies.

4. Discussion

TLPs are important proteins involved in the tissue development and stress resistance pathways of plants [6]. Therefore, we have studied the TLPs in five major cereals including *B. distachyon*, *O. sativa*, *S. bicolor*, *T. aestivum* and *Z. mays*. In the present study, we have identified a total of 26, 27, 39, 93 and 37 *TLPs* in the genome of *B. distachyon*, *O. sativa*, *S. bicolor*, *T. aestivum* and *Z. mays*, respectively. Variable numbers of the *TLP* genes in different plants have been reported in

earlier studies. For instance, 24, 33 and 49 *TLP* genes have been reported in diploid *A. thaliana*, *Vitis vinifera* and *Populus trichocarpa*, respectively [7,58]. In the case of *O. sativa* and *Z. mays*, 49 *TLP* genes in each are reported in the earlier study [7], because probably they had not removed the genes with incomplete thaumatin signature motifs, as done in the present study. The occurrence of the highest number of *TLP* genes in *T. aestivum* could be attributed to their complex allohexaploid genome configuration (AABBDD) [20]. Further, differential expansion of the *TLP* genes in various plant species might be attributed to the number of duplication events.

The duplication events are responsible for the expansion of gene families and these are the major forces to attain neo-functionality by causing genetic variability [59]. In DEs analysis, our results suggested that tandem DEs are a major factor in *B. distachyon*, *O. sativa* and *S. bicolor* and segmental DEs in *T. aestivum* and *Z. mays* behind the gene expansion of TLP family. The role of tandem and segmental duplications in the evolution of TLP gene family has been earlier reported in various plant species including *A. thaliana*, *O. sativa*, and *Z. mays* [7]. The higher number of paralogous genes in *T. aestivum* and *Z. mays* might be attributed to their larger genome size and the presence of more transposable elements [53,60].

Furthermore, the divergence time of monocots from eudicots was estimated as ~170-235 Mya, which were further diverged into grasses around ~77 Mya [60,61]. Therefore, the paralogous *TLP* genes in *B. distachyon* might have evolved before the divergence of grasses. The main duplication incidence in *S. bicolor* and *O. sativa* genome was estimated at around ~70 Mya [62]. However, our results indicated that the majority of paralogous *TLP* genes of *O. sativa* and *S. bicolor* are evolved after the main DE occurrence. In *Z. mays*, paralogous *TLP* genes were probably formed close to the divergence time of maize and sorghum, i.e. ~11-28 Mya [62]. However, in the case of *T. aestivum*, paralogous *TLP* genes were probably evolved earlier than

the hybridization event of A, B and D subgenomes [63]. However, two duplicated pairs, having NUP 96 domain showed their recent incidence of DEs. Duplicated genes face various environmental forces and natural pressures during the process of evolution [53]. Our results suggested the purifying selection of *TLP* genes as a major force of selection, as reported in the various other plant species [7,64]. Moreover, positive selection has also been reported for *TLP* genes in poplar, which suggested varied natural selection procedures during TLPs' evolution [8]. Since the TLPs are paraphyletic in origin, they are supposed to be derived from multiple ancestral genes [65]. Clustering of identified TLPs in multiple clades is in agreement with the paraphyletic nature of origin. Further, a variable number of TLPs in each clade is due to the differential duplication of *TLP* genes. Similarly, the uneven distribution and paraphyletic nature of TLPs have also been reported by previous studies [7,64]. The tight clustering of homeologous and paralogous genes could be due to the high sequence homology among the respective sequences.

The organization of introns and exons of *TLPs* have been reported to be in the range of one to ten exons [6,7,9,53]. On similar patterns, our results suggest comparable findings with one to four exons in most of the TLPs. However, most of the monocots' TLPs have intron-less nature, which was also found in our analysis [7]. Additionally, our results with respect to the protein lengths, molecular weights and isoelectric points of long and small TLP proteins, found in accordance with previously studied TLPs [6,7,9]. In our analysis, the TLP2-B was found to be localized in extracellular region. Similarly, the extracellular localization of a wheat TLP (TaLr19TLP1) is also reported in an earlier study [66].

In addition, the presence of thaumatin-like domain in all the TLP proteins, make our studies similar to previously reported findings. However, four ZmTLPs (ZmTLP16, ZmTLP20,

ZmTLP25, and ZmTLP33) have an additional NUP 96 domain. The NUP 96 domain is reported to be involved in plant development and stress resistance against pathogens [67]. Various abiotic and biotic stress-responsive elements were also found in most of the studied *TLPs*, for instance, DRE2COREZMRAB17, ASF1, BIHD1, CBFHV, CGG-box, GCCCORE, LTRECOREATCOR15, MYB, MYC, WRKY1, W-box, T/G-box etc. Various plant hormone-related and stress-related *cis*-regulatory elements have also been reported in four cotton species [56]. Moreover, in previous studies, the elements like ERE, W-box, WRKY, were found to be involved in plant growth and stress resistance [68-70]. The occurrence of numerous *cis*-regulatory elements suggested diverse functions of *TLP* genes in plants. Further, the expression profiling of *TLPs* under various tissues and developmental stages also advocated the putative involvement of these genes in development of vegetative and reproductive tissues. Similar expression trend in various plant tissues and their role in reproductive tissues like flowers and seeds has also been reported in earlier studies, for instance, *TLP* genes in *Gossypium hirsutum*, also showed their varied expression pattern in vegetative as well as in reproductive organs [18, 56, 71]. The role of *TLP* genes in fungal resistance has been reported in earlier studies against numerous pathogens including *Fusarium* sp. [72], *Microdochium nivale* [73], *Rhizoctonia solani* [74] and *Verticillium dahlia* [56] in wheat and other plant species. The differential expression of *TaTLP* genes in response to Pst and Bgt infestation further established their roles in fungal stress responses. Moreover, similar to our finding's differential expression and upregulation of *TLP* genes at various hours of drought, salt and cold stress treatment has also been reported in other plants like *Brassica rapa* and *Gossypium* sp. [56,75]. Further, the over-expression of *TLP* genes provided increased tolerance against various abiotic stresses in tobacco, cotton, Arabidopsis etc. [11,56,76]. In a recently conducted study, the important role of a cotton *TLP* (*GhTLP19*) gene

was studied in combating drought stress [56]. Moreover, the GbTLP1 and ObTLP1 of *G. barbadense* and *Ocimum basilicum*, respectively, were found effective against drought and salt stresses [57,76]. Collectively, current findings depict that the TLP proteins could be utilized to develop the stress resistance transgenic cereal crops, which will be a boost for the agriculture.

5. Conclusions

Thaumatin-like proteins (TLPs), a part of the PR5 protein family, are known to be involved in various biotic and abiotic stresses. The curiosity to unravel the various possibilities led us to a detailed characterization of a total of 222 *TLP* genes in five cereal crops. Phylogenetic analysis revealed the paraphyletic origin of TLPs in cereals, while the occurrence of duplication events suggested the role of paralogous genes in the expansion of the TLP gene family. Gene expression analysis using RNA-seq-data and qRT-PCR suggested the important roles of *TLP* genes in plant growth and development and abiotic and biotic stress responses. Significant tolerance in the *TaTLP2-B* expressing yeast cells further established their role in stress response. Overall the study revealed that the *TLP* genes can be used for stress-resistant transgenic crop development. Since the identified genes are directly derived from the food crops, they would be easily de-regulated from the regulatory authority of the transgenic crops. Further, the current study will facilitate the detailed functional characterization of *TLP* genes of cereal crops in future studies.

Competing interests

The authors declare that they have no competing interests

Author's contribution

SKU conceived the idea. AS and SKU designed the experiments. AS, HS and RR performed the experiments. AS, HS and AP analyzed the data. AS, HS and SKU wrote the manuscript. All the authors have read and finalized the manuscript.

Acknowledgements

The authors are grateful to the Panjab University, Chandigarh, and National Institute of Plant Genome Research, New Delhi, India for research facilities, Ensembl Plants, URGI and NCBI for data availability. AS and HS is grateful to CSIR for the senior research fellowship. HS is also thankful to the IKGPTU, Jalandhar for the Ph.D registration. SKU is grateful to the Council of Scientific and Industrial Research (CSIR) and the Department of Science and Technology (DST), Government of India for a research grant (No. 38(1489)/19/EMR-II) and partial financial support under the Promotion of University Research and Scientific Excellence (PURSE) grant scheme, respectively.

References

1. E. Kombrink, I.E. Somssich, Pathogenesis-Related Proteins and Plant Defense, in: Plant Relationships, Springer Berlin Heidelberg, Berlin, Heidelberg, 1997: pp. 107–128. https://doi.org/10.1007/978-3-662-10370-8_7.
2. P. Chadha, R.H. Das, A pathogenesis-related protein, AhPR10 from peanut: an insight of its mode of antifungal activity, Planta. 225 (2006) 213–222. https://doi.org/10.1007/s00425-006-0344-7.
3. L.C. van Loon, M. Rep, C.M.J. Pieterse, Significance of inducible defense-related proteins in infected plants, Annu. Rev. Phytopathol. 44 (2006) 135–162. https://doi.org/10.1146/annurev.phyto.44.070505.143425.
4. R. Velazhahan, S.K. Datta, S. Muthukrishnan, The PR-5 family: Thaumatin-like proteins, in: Pathog. Proteins Plants, CRC Press, 1999: pp. 107–129. https://doi.org/10.1201/9781420049299.
5. H. van der Wel, K. Loeve, Isolation and characterization of thaumatin I and II, the sweet-tasting proteins from *Thaumatococcus daniellii* Benth., Eur. J. Biochem. 31 (1972) 221–5. https://doi.org/10.1111/j.1432-1033.1972.tb02522.x.
6. J.-J. Liu, R. Sturrock, A.K.M. Ekramoddoullah, The superfamily of thaumatin-like proteins: its origin, evolution, and expression towards biological function, Plant Cell Rep. 29 (2010) 419–436. https://doi.org/10.1007/s00299-010-0826-8.
7. J. Cao, Y. Lv, Z. Hou, X. Li, L. Ding, Expansion and evolution of thaumatin-like protein (TLP) gene family in six plants, Plant Growth Regul. 79 (2016) 299–307. https://doi.org/10.1007/s10725-015-0134-y.

8. J.-J. Liu, A. Zamani, A.K.M. Ekramoddoullah, Expression profiling of a complex thaumatin-like protein family in western white pine, Planta. 231 (2010) 637–651. https://doi.org/10.1007/s00425-009-1068-2.
9. B. Petre, I. Major, N. Rouhier, S. Duplessis, Genome-wide analysis of eukaryote thaumatin-like proteins (TLPs) with an emphasis on poplar, BMC Plant Biol. 11 (2011) 33. https://doi.org/10.1186/1471-2229-11-33.
10. R. Ghosh, C. Chakrabarti, Crystal structure analysis of NP24-I: a thaumatin-like protein, Planta. 228 (2008) 883–890. https://doi.org/10.1007/s00425-008-0790-5.
11. M. V. Rajam, N. Chandola, P. Saiprasad Goud, D. Singh, V. Kashyap, M.L. Choudhary, D. Sihachakr, Thaumatin gene confers resistance to fungal pathogens as well as tolerance to abiotic stresses in transgenic tobacco plants, Biol. Plant. 51 (2007) 135–141. https://doi.org/10.1007/s10535-007-0026-8.
12. K. Datta, R. Velazhahan, N. Oliva, I. Ona, T. Mew, G.S. Khush, S. Muthukrishnan, S.K. Datta, Over-expression of the cloned rice thaumatin-like protein (*PR-5*) gene in transgenic rice plants enhances environmental friendly resistance to *Rhizoctonia solani* causing sheath blight disease, Theor. Appl. Genet. 98 (1999) 1138–1145. https://doi.org/10.1007/s001220051178.
13. X. Wang, C. Tang, L. Deng, G. Cai, X. Liu, B. Liu, Q. Han, H. Buchenauer, G. Wei, D. Han, L. Huang, Z. Kang, Characterization of a pathogenesis-related thaumatin-like protein gene *TaPR5* from wheat induced by stripe rust fungus, Physiol. Plant. 139 (2010) 27–38. https://doi.org/10.1111/j.1399-3054.2009.01338.x.
14. L.R. Abad, M.P. D'Urzo, D. Liu, M.L. Narasimhan, M. Reuveni, J.K. Zhu, X. Niu, N.K. Singh, P.M. Hasegawa, R.A. Bressan, Antifungal activity of tobacco osmotin has

specificity and involves plasma membrane permeabilization, Plant Sci. 118 (1996) 11–23. https://doi.org/10.1016/0168-9452(96)04420-2.

15. Y. Zhang, D.S. Shih, Isolation of an osmotin-like protein gene from strawberry and analysis of the response of this gene to abiotic stresses, J. Plant Physiol. 164 (2007) 68–77. https://doi.org/10.1016/j.jplph.2006.02.002.
16. A.D. Neale, J.A. Wahleithner, M. Lund, H.T. Bonnett, A. Kelly, D.R. Meeks-Wagner, W.J. Peacock, E.S. Dennis, chitinase, [beta]-1,3-glucanase, osmotin, and extensin are expressed in tobacco explants during flower formation, PLANT CELL ONLINE. 2 (1990) 673–684. https://doi.org/10.1105/tpc.2.7.673.
17. R.A. Salzman, I. Tikhonova, B.P. Bordelon, P.M. Hasegawa, R.A. Bressan, Coordinate accumulation of antifungal proteins and hexoses constitutes a developmentally controlled defense response during fruit ripening in grape., Plant Physiol. 117 (1998) 465–72. http://www.ncbi.nlm.nih.gov/pubmed/9625699.
18. P.J. Seo, A.-K. Lee, F. Xiang, C.-M. Park, Molecular and functional profiling of Arabidopsis pathogenesis-related genes: Insights into their roles in salt response of seed germination, Plant Cell Physiol. 49 (2008) 334–344. https://doi.org/10.1093/pcp/pcn011.
19. P. Lei, X. Wei, R. Gao, F. Huo, X. Nie, W. Tong, W. Song, Genome-wide identification of PYL gene family in wheat: Evolution, expression and 3D structure analysis, Genomics. 113 (2021) 854–866. https://doi.org/10.1016/j.ygeno.2020.12.017.
20. A. Sharma, Shumayla, S. Tyagi, A. Alok, K. Singh, S.K. Upadhyay, Thaumatin-like protein kinases: Molecular characterization and transcriptional profiling in five cereal crops, Plant Sci. 290 (2020). https://doi.org/10.1016/j.plantsci.2019.110317.

21. Shumayla, S. Tyagi, A. Sharma, K. Singh, S.K. Upadhyay, Genomic dissection and transcriptional profiling of Cysteine-rich receptor-like kinases in five cereals and functional characterization of *TaCRK68-A*, Int. J. Biol. Macromol. 134 (2019) 316–329. https://doi.org/10.1016/j.ijbiomac.2019.05.016.
22. R.C. Edgar, MUSCLE: multiple sequence alignment with high accuracy and high throughput, Nucleic Acids Res. 32 (2004) 1792–1797. https://doi.org/10.1093/nar/gkh340.
23. F. Corpet, Multiple sequence alignment with hierarchical clustering., Nucleic Acids Res. 16 (1988) 10881–90. http://www.ncbi.nlm.nih.gov/pubmed/2849754 (accessed May 27, 2017).
24. S. Kumar, G. Stecher, M. Li, C. Knyaz, K. Tamura, MEGA X: Molecular Evolutionary Genetics Analysis across Computing Platforms, (n.d.). https://doi.org/10.1093/molbev/msy096.
25. M. Suyama, D. Torrents, P. Bork, PAL2NAL: Robust conversion of protein sequence alignments into the corresponding codon alignments, Nucleic Acids Res. 34 (2006). https://doi.org/10.1093/nar/gkl315.
26. M. Nagaraju, P.S. Reddy, S.A. Kumar, A. Kumar, G. Rajasheker, D.M. Rao, P.B. Kavi Kishor, Genome-wide identification and transcriptional profiling of small heat shock protein gene family under diverse abiotic stress conditions in *Sorghum bicolor* (L.), Int. J. Biol. Macromol. 142 (2020) 822–834. https://doi.org/10.1016/j.ijbiomac.2019.10.023.
27. B.S. Gaut, B.R. Morton, B.C. McCaig, M.T. Clegg, Substitution rate comparisons between grasses and palms: synonymous rate differences at the nuclear gene Adh parallel

rate differences at the plastid gene rbcL., Proc. Natl. Acad. Sci. U. S. A. 93 (1996) 10274. https://doi.org/10.1073/PNAS.93.19.10274.

28. F. Tajima, Simple methods for testing the molecular evolutionary clock hypothesis., Genetics. 135 (1993).
29. B. Hu, J. Jin, A.-Y. Guo, H. Zhang, J. Luo, G. Gao, GSDS 2.0: an upgraded gene feature visualization server., Oxford University Press, 2015. https://doi.org/10.1093/bioinformatics/btu817.
30. H. Sharma, M. Taneja, S.K. Upadhyay, Identification, characterization and expression profiling of cation-proton antiporter superfamily in *Triticum aestivum* L. and functional analysis of *TaNHX4-B*, Genomics. 112 (2020) 356–370. https://doi.org/10.1016/j.ygeno.2019.02.015.
31. E. Gasteiger, C. Hoogland, A. Gattiker, S. Duvaud, M.R. Wilkins, R.D. Appel, A. Bairoch, Protein Identification and Analysis Tools on the ExPASy Server, in: Proteomics Protoc. Handb., Humana Press, Totowa, NJ, 2005: pp. 571–607. https://doi.org/10.1385/1-59259-890-0:571.
32. T. Nussbaumer, M.M. Martis, S.K. Roessner, M. Pfeifer, K.C. Bader, S. Sharma, H. Gundlach, M. Spannagl, MIPS PlantsDB: a database framework for comparative plant genome research., Nucleic Acids Res. 41 (2013) D1144-51. https://doi.org/10.1093/nar/gks1153.
33. P. Stothard, The sequence manipulation suite: JavaScript programs for analyzing and formatting protein and DNA sequences., Biotechniques. 28 (2000) 1102, 1104. https://doi.org/10.2144/00286ir01.

34. C.-S. Yu, Y.-C. Chen, C.-H. Lu, J.-K. Hwang, Prediction of protein subcellular localization, Proteins Struct. Funct. Bioinforma. 64 (2006) 643–651. https://doi.org/10.1002/prot.21018.
35. B.R. King, C. Guda, ngLOC: an n-gram-based Bayesian method for estimating the subcellular proteomes of eukaryotes, Genome Biol. 8 (2007) R68. https://doi.org/10.1186/gb-2007-8-5-r68.
36. O. Emanuelsson, H. Nielsen, S. Brunak, G. Von Heijne, Predicting subcellular localization of proteins based on their N-terminal amino acid sequence, J. Mol. Biol. 300 (2000) 1005–1016. https://doi.org/10.1006/jmbi.2000.3903.
37. P. Horton, K.-J. Park, T. Obayashi, N. Fujita, H. Harada, C.J. Adams-Collier, K. Nakai, WoLF PSORT: protein localization predictor, Nucleic Acids Res. 35 (2007) W585–W587. https://doi.org/10.1093/nar/gkm259.
38. L. Käll, A. Krogh, E.L.. Sonnhammer, A combined transmembrane topology and signal peptide prediction method, J. Mol. Biol. 338 (2004) 1027–1036. https://doi.org/10.1016/j.jmb.2004.03.016.
39. M. Cserzo, F. Eisenhaber, B. Eisenhaber, I. Simon, TM or not TM: transmembrane protein prediction with low false positive rate using DAS-TMfilter, Bioinformatics. 20 (2004) 136–137. https://doi.org/10.1093/bioinformatics/btg394.
40. T.N. Petersen, S. Brunak, G. von Heijne, H. Nielsen, SignalP 4.0: discriminating signal peptides from transmembrane regions, Nat. Methods. 8 (2011) 785–786. https://doi.org/10.1038/nmeth.1701.
41. I. Letunic, T. Doerks, P. Bork, SMART: recent updates, new developments and status in 2015., Nucleic Acids Res. 43 (2015) D257-60. https://doi.org/10.1093/nar/gku949.

42. T.L. Bailey, M. Boden, F.A. Buske, M. Frith, C.E. Grant, L. Clementi, J. Ren, W.W. Li, W.S. Noble, MEME SUITE: tools for motif discovery and searching, 2009. https://doi.org/10.1093/nar/gkp335.
43. K. Higo, Y. Ugawa, M. Iwamoto, H. Higo, PLACE: a database of plant cis -acting regulatory DNA elements, Nucleic Acids Res. 26 (1998) 358–359.
44. F. Choulet, A. Alberti, S. Theil, N.M. Glover, V. Barbe, J. Daron, L. Pingault, P. Sourdille, A. Couloux, E. Paux, Structural and Functional Partitioning of Bread Wheat Chromosome 3B, (n.d.). https://doi.org/10.1126/science.1249721ï.
45. L. Pingault, F. Choulet, A. Alberti, N. Glover, P. Wincker, C. Feuillet, E. Paux, Deep transcriptome sequencing provides new insights into the structural and functional organization of the wheat genome, Genome Biol. 16 (2015) 29. https://doi.org/10.1186/s13059-015-0601-9.
46. I. Papatheodorou, N.A. Fonseca, M. Keays, Y.A. Tang, E. Barrera, W. Bazant, M. Burke, A. Füllgrabe, A.M.P. Fuentes, N. George, L. Huerta, S. Koskinen, S. Mohammed, M. Geniza, J. Preece, P. Jaiswal, A.F. Jarnuczak, W. Huber, O. Stegle, J.A. Vizcaino, A. Brazma, R. Petryszak, Expression Atlas: Gene and protein expression across multiple studies and organisms, Nucleic Acids Res. 46 (2018) D246–D251. https://doi.org/10.1093/nar/gkx1158.
47. H. Zhang, Y. Yang, C. Wang, M. Liu, H. Li, Y. Fu, Y. Wang, Y. Nie, X. Liu, W. Ji, Large-scale transcriptome comparison reveals distinct gene activations in wheat responding to stripe rust and powdery mildew, BMC Genomics. 15 (2014) 898. https://doi.org/10.1186/1471-2164-15-898.

48. Z. Liu, M. Xin, J. Qin, H. Peng, Z. Ni, Y. Yao, Q. Sun, Temporal transcriptome profiling reveals expression partitioning of homeologous genes contributing to heat and drought acclimation in wheat (*Triticum aestivum* L.), BMC Plant Biol. 15 (2015) 152. https://doi.org/10.1186/s12870-015-0511-8.

49. Y. Zhang, Z. Liu, A.A. Khan, Q. Lin, Y. Han, P. Mu, Y. Liu, H. Zhang, L. Li, X. Meng, Z. Ni, M. Xin, Expression partitioning of homeologs and tandem duplications contribute to salt tolerance in wheat (*Triticum aestivum* L.)., Sci. Rep. 6 (2016) 21476. https://doi.org/10.1038/srep21476.

50. B.J. Haas, A. Papanicolaou, M. Yassour, M. Grabherr, P.D. Blood, J. Bowden, M.B. Couger, D. Eccles, B. Li, M. Lieber, M.D. MacManes, M. Ott, J. Orvis, N. Pochet, F. Strozzi, N. Weeks, R. Westerman, T. William, C.N. Dewey, R. Henschel, R.D. LeDuc, N. Friedman, A. Regev, De novo transcript sequence reconstruction from RNA-seq using the Trinity platform for reference generation and analysis, Nat. Protoc. 8 (2013) 1494–1512. https://doi.org/10.1038/nprot.2013.084.

51. J. Seo, H. Gordish-Dressman, E.P. Hoffman, An interactive power analysis tool for microarray hypothesis testing and generation, Bioinformatics. 22 (2006) 808–814. https://doi.org/10.1093/bioinformatics/btk052.

52. K.J. Livak, T.D. Schmittgen, Analysis of relative gene expression data using Real-Time quantitative PCR and the 2−ΔΔCT method, Methods. 25 (2001) 402–408. https://doi.org/10.1006/meth.2001.1262.

53. H. Ellegren, Comparative genomics and the study of evolution by natural selection, Mol. Ecol. 17 (2008) 4586–4596. https://doi.org/10.1111/j.1365-294X.2008.03954.x.

54. S. Huang, A. Sirikhachornkit, X. Su, J. Faris, B. Gill, R. Haselkorn, P. Gornicki, Genes encoding plastid acetyl-CoA carboxylase and 3-phosphoglycerate kinase of the Triticum/Aegilops complex and the evolutionary history of polyploid wheat, Proc. Natl. Acad. Sci. U. S. A. 99 (2002) 8133–8138. https://doi.org/10.1073/pnas.072223799.

55. M. Sharma, A. Singh, A. Shankar, A. Pandey, V. Baranwal, S. Kapoor, A.K. Tyagi, G.K. Pandey, Comprehensive expression analysis of rice armadillo gene family during abiotic stress and development, DNA Res. 21 (2014) 267–283. https://doi.org/10.1093/dnares/dst056.

56. Z. Li, X. Wang, Y. Cui, K. Qiao, L. Zhu, S. Fan, Q. Ma, Comprehensive genome-wide analysis of thaumatin-like gene family in four cotton species and functional identification of ghtlp19 involved in regulating tolerance to verticillium dahlia and drought, Front. Plant Sci. 11 (2020) 1597. https://doi.org/10.3389/fpls.2020.575015.

57. M.F.H. Munis, L. Tu, F. Deng, J. Tan, L. Xu, S. Xu, L. Long, X. Zhang, A thaumatin-like protein gene involved in cotton fiber secondary cell wall development enhances resistance against *Verticillium dahliae* and other stresses in transgenic tobacco, Biochem. Biophys. Res. Commun. 393 (2010) 38–44. https://doi.org/10.1016/j.bbrc.2010.01.069.

58. X. Yan, H. Qiao, X. Zhang, C. Guo, M. Wang, Y. Wang, X. Wang, Analysis of the grape (*Vitis vinifera* L.) thaumatin-like protein (*TLP*) gene family and demonstration that *TLP29* contributes to disease resistance, Sci. Rep. 7 (2017) 4269. https://doi.org/10.1038/s41598-017-04105-w.

59. R. Appels, K. Eversole, N. Stein, C. Feuillet, B. Keller, J. Rogers, et al., Shifting the limits in wheat research and breeding using a fully annotated reference genome, Science (80-.). 361 (2018) eaar7191. https://doi.org/10.1126/science.aar7191.

60. B.S. Gaut, Evolutionary dynamics of grass genomes, New Phytol. 154 (2002) 15–28. https://doi.org/10.1046/j.1469-8137.2002.00352.x.

61. Y.W. Yang, K.N. Lai, P.Y. Tai, W.H. Li, Rates of nucleotide substitution in angiosperm mitochondrial DNA sequences and dates of divergence between Brassica and other angiosperm lineages, J. Mol. Evol. 48 (1999) 597–604. https://doi.org/10.1007/PL00006502.

62. A.H. Paterson, J.E. Bowers, B.A. Chapman, Ancient polyploidization predating divergence of the cereals, and its consequences for comparative genomics, Proc. Natl. Acad. Sci. U. S. A. 101 (2004) 9903–9908. https://doi.org/10.1073/pnas.0307901101.

63. T. Marcussen, S.R. Sandve, L. Heier, M. Spannagl, M. Pfeifer, K.S. Jakobsen, et al., Ancient hybridizations among the ancestral genomes of bread wheat, Science (80-.). 345 (2014) 1250092–1250092. https://doi.org/10.1126/science.1250092.

64. Y. Liu, J. Cui, X. Zhou, Y. Luan, F. Luan, Genome-wide identification, characterization and expression analysis of the TLP gene family in melon (*Cucumis melo* L.), Genomics. 112 (2020) 2499–2509. https://doi.org/10.1016/j.ygeno.2020.02.001.

65. R.G. Shatters, L.M. Boykin, S.L. Lapointe, W.B. Hunter, A.A. Weathersbee, Phylogenetic and structural relationships of the *PR5* Gene family reveal an ancient multigene family conserved in plants and select animal taxa, J. Mol. Evol. 63 (2006) 12–29. https://doi.org/10.1007/s00239-005-0053-z.

66. J. Zhang, F. Wang, F. Liang, Y. Zhang, L. Ma, H. Wang, D. Liu, Functional analysis of a pathogenesis-related thaumatin-like protein gene *TaLr35PR5* from wheat induced by leaf rust fungus, BMC Plant Biol. 18 (2018) 1–11. https://doi.org/10.1186/s12870-018-1297-2.

67. Y. Zhang, X. Li, A putative nucleoporin 96 is required for both basal defense and constitutive resistance responses mediated by suppressor of npr1-1, constitutive 1, Plant Cell. 17 (2005) 1306–1316. https://doi.org/10.1105/tpc.104.029926.

68. K. Hiroyuki and R. Terauchi, Regulation of expression of rice thaumatin-like protein: inducibility by elicitor requires promoter W-box elements, Plant Cell Rep. 27 (2008) 1521–1528. doi: 10.1007/s00299-008-0536-7

69. P.J. Rushton, I.E. Somssich, P. Ringler, Q.J. Shen, WRKY transcription factors, Trends Plant Sci. 15 (2010) 247–258. https://doi.org/10.1016/j.tplants.2010.02.006.

70. J.Y. Gou, L.J. Wang, S.P. Chen, W.L. Hu, X.Y. Chen, Gene expression and metabolite profiles of cotton fiber during cell elongation and secondary cell wall synthesis, Cell Res. 17 (2007) 422–434. https://doi.org/10.1038/sj.cr.7310150.

71. B.R. Fils-Lycaon, P.A. Wiersma, K.C. Eastwell, P. Sautiere, A cherry protein and its gene, abundantly expressed in ripening fruit, have been identified as thaumatin-like., Plant Physiol. 111 (1996) 269–73. http://www.ncbi.nlm.nih.gov/pubmed/8685266.

72. C.A. Mackintosh, J. Lewis, L.E. Radmer, S. Shin, S.J. Heinen, L.A. Smith, M.N. Wyckoff, R. Dill-Macky, C.K. Evans, S. Kravchenko, G.D. Baldridge, R.J. Zeyen, G.J. Muehlbauer, Overexpression of defense response genes in transgenic wheat enhances resistance to Fusarium head blight, Plant Cell Rep. 26 (2007) 479–488. https://doi.org/10.1007/s00299-006-0265-8.

73. C. Kuwabara, D. Takezawa, T. Shimada, T. Hamada, S. Fujikawa, K. Arakawa, Abscisic acid- and cold-induced thaumatin-like protein in winter wheat has an antifungal activity against snow mould, *Microdochium nivale*, Physiol. Plant. 115 (2002) 101–110. https://doi.org/10.1034/j.1399-3054.2002.1150112.x.

74. G. Naseri, M.M. Sohani, A. Pourmassalehgou, S. Allahi, In planta transformation of rice (Oryza sativa) using thaumatin-like protein gene for enhancing resistance to sheath blight, African J. Biotechnol. 11 (2012) 7885–7893. https://doi.org/10.5897/AJB11.3331.

75. N.U. Ahmed, J.-I. Park, H.-J. Jung, K.-K. Kang, Y.-P. Lim, Y. Hur, I.-S. Nou, Molecular characterization of thaumatin family genes related to stresses in *Brassica rapa*, Sci. Hortic. (Amsterdam). 152 (2013) 26–34. https://doi.org/10.1016/j.scienta.2013.01.007.

76. R.C. Misra, Sandeep, M. Kamthan, S. Kumar, S. Ghosh, A thaumatin-like protein of *Ocimum basilicum* confers tolerance to fungal pathogen and abiotic stress in transgenic Arabidopsis, Sci. Rep. 6 (2016) 25340. https://doi.org/10.1038/srep25340.

Figure legends

Fig. 1 Chromosomal distribution and paralogous gene analysis of *TLPs* in five cereal crops. Image depicts the scattered localization of *TLP* genes of **(A)** *B. distachyon*, **(B)** *O. sativa*, **(C)** *S. bicolor*, **(D)** *T. aestivum* and **(E)** *Z. mays* at their respective chromosomes. Red lines and blue lines denote the tandem and segmental duplication events, respectively.

Fig. 2 Multiple sequence alignment of selected long and small TLPs of *B. distachyon, O. sativa, S. bicolor, T. aestivum* and *Z. mays* with a known thaumatin protein sequence. Figure shows the alignment of long (OsTLP5, SbTLP2, BdTLP13, ZmTLP13 and TaTLP1-A) and small (BdTLP10, OsTLP25, ZmTLP6, SbTLP29 and TaTLP6-B) TLPs with a known thaumatin (P02883.2|THM1_THADA) protein sequence. The characteristic thaumatin signature motif is shown in the sky blue box. All the conserved cysteine residues were marked with an asterisk. The REDDD motif was marked with a red circle. The FF hydrophobic motif and bottom of acidic cleft forming regions are marked with a green triangle and brown box, respectively. The conserved regions of the thaumatin domain are marked with a black line.

Fig. 3 Phylogenetic tree analysis of *A. thaliana, B. distachyon, O. sativa. S, bicolor, T. aestivum and Z. mays* using full-length peptide of TLPs. The figure shows the phylogenetic tree of TLPs, built by the maximum likelihood approach with 1000 bootstrap values using MEGAX software. The phylogenetic tree displays the clustering of TLPs of six plant species in I-XI different groups, each group is coloured differently. The TLPs from each plant are marked with a different colour.

Fig. 4 Subcellular localization analysis of TaTLP2-B by YFP fusion. The YFP fluorescence depicts the localization of TaTLP2-B protein majorly in the extracellular region.

Fig. 5 Expression analysis of *TLP* genes of *B. distachyon* (*BdTLP*), *O. sativa* (*OsTLP*). *S. bicolor* (*SbTLP*), *T. aestivum* (*TaTLP*) and *Z. mays* (*ZmTLP*) under various tissue and developmental stages. Heatmaps **(A-E)** shows the clustering and expression profiling of *TLP* genes in (**A**) *B. distachyon*, **(B)** *O. sativa*, **(C)** *S. bicolor*, **(D)** *T. aestivum* and **(E)** *Z. mays*. The different clusters are formed based on expression values of *TLP* genes in the respective crop which are represented with different colours. The colour bar depicts the upregulated and downregulated expression with red and green colours, respectively.

Fig. 6 Expression profiling of *TLP* genes of *T. aestivum* under biotic and abiotic stress conditions. Heat maps **(A-C**) shows the clustering and expression profiling of *TaTLP* genes under **(A)** *Blumeria graminis* (Bgt) and *Puccinia striiformis* (Pst), **(B)** heat (HS), drought (DS) and combined heat-drought (HD) and **(C)** salt stress conditions. The clustering of *TaTLP* genes in each heat map is based on their expression pattern in respective stress conditions. The colour bar shows high and low expressions with red and green colours, respectively.

Fig. 7 qRT-PCR analysis of eight *TaTLP* genes. Expression patterns in the form of fold changes have been shown in the bar graphs. Figures **7 A, C, E, G, I, K, M, and O** show expression of *TaTLP2-B, TaTLP5-D, TaTLP7-D, TaTLP9-D, TaTLP10-D, TaTLP14-B1, TaTLP25-B and TaTLP29-A* genes, under HS, DS and HD stress conditions, while figure **7 B, D, F, H, J, L, N, and P** shows the expression pattern of these genes under salt stress, respectively. Significance between the control and treated conditions is carried using a two-tailed student's t-test. The ns, *, ** and *** markings represent the significance at p-value >0.05, <=0.05, <=0.01, <=0.001, respectively.

Fig. 8 Comparative expression analysis of duplicated gene pairs in *T. aestivum*. Graphs **(A-E)** represents the comparative expression pattern of paralogous *TaTLPs* under various tissue and developmental stages. Based on their expression trend, *TaTLPs* are categorized into the retention of function **(A-C)** and pseudo functionalization types **(D, E)**.

Fig. 9. Spot assay of recombinant *Saccharomyces cerevisiae* (yeast) cells under various stress conditions. Figure **(A)** shows similar growth of both TaTLP2-B (pYES2.1-TaTLP2-B) and control vector (pYES2.1/V5-His/lacZ) containing yeast cells under control conditions. Figure **(B-F)** shows the spot assay analysis of *TaTLP2-B* and control vector containing yeast cells in (**B**) cold (4°C), (**C**) drought (30% PEG), (**D**) heat (37°C), (**E**) combined heat and drought (37°C and 30% PEG), and (**F**) salt (1M NaCl) stress conditions, respectively.

Supplementary files

Figure S1. Multiple sequence alignment of 222 TLPs of *B. distachyon, O. sativa, S. bicolor, T. aestivum* and *Z. mays*. Conserved cysteine residues are marked with an asterisk. The red dot denotes the REDDD motif. The FF hydrophobic motif is highlighted with a green triangle. The thaumatin signature motif, conserved domain, amino acids forming the bottom of the acidic cleft are marked with sky blue, red and brown lines, respectively.

Figure S2. Intron/exon organization and conserved motifs of TLPs of *B. distachyon*, *O. sativa*, *S. bicolor, T. aestivum* and *Z. mays.*

File S1. List of identified *TLP* genes with their proposed nomenclature in *B. distachyon, O. sativa S, bicolor, T. aestivum* and *Z. mays*. The Red colour represents the *sTLPs* of each crop.

File S2. List of paralogous *TLP* genes in *B. distachyon*, *O. sativa*, *S. bicolor*, *T. aestivum* and *Z. mays*.

File S3. Domain organization of TLP proteins of *B. distachyon*, *O. sativa*, *S. bicolor*, *T. aestivum* and *Z. mays*.

File S4. The characterization table of *B. distachyon, O. sativa*, *S. bicolor*, *T. aestivum* and *Z. mays* TLPs.

File S5. List of qRT-PCR primers.

Table S1. List of predicted *cis*-regulatory elements in the promoter region of *TLP* genes of *B. distachyon, O. sativa, S. bicolor, T. aestivum* and *Z, mays*.

Table 1 The Ka/Ks ratio and divergence time of duplicated *TLP* gene pairs.

Gene A	Gene B	Ka	Ks	Ka/Ks	Selection pressure	T = Ks/2r
BdTLPK1	*BdTLPK2*	0.1102	1.0810	0.1020	Purifying	83.2
OsTLP5	*OsTLP7*	0.1014	0.6788	0.1494	Purifying	52.2
OsTLP22	*OsTLP24*	0.0697	0.4134	0.3190	Purifying	27.7
OsTLP23	*OsTLP25*	0.1149	0.3603	0.31646	Purifying	29.4
OsTLP26	*OsTLP27*	0.1569	1.3980	0.1122	Purifying	107.6
SbTLP28	*SbTLP29*	0.0911	0.6808	0.1338	Purifying	52.4
SbTLP31	*SbTLP34*	0.1388	1.7816	0.0779	Purifying	137
SbTLP32	*SbTLP33*	0.0055	0.0716	0.0771	Purifying	5.5
TaTLP1-A	*TaTLP24-D*	0.1224	0.8739	0.1401	Purifying	67.2
TaTLP1-B	*TaTLP24-B*	0.0937	0.8163	0.1148	Purifying	62.8
TaTLP12-B	*TaTLP13-D*	0.1193	1.1836	0.1008	Purifying	91
TaTLP15-A	*TaTLP18-B*	0.1368	0.9879	0.1384	Purifying	76
TaTLP15-B2	*TaTLP18-A*	0.1423	0.9582	0.1486	Purifying	73.7
TaTLP16-A	*TaTLP17-A*	0.1456	1.2825	0.1135	Purifying	98.7
ZmTLP1	*ZmTLP35*	0.0955	0.3929	0.2431	Purifying	30.2
ZmTLP2	*ZmTLP34*	0.1044	0.8149	0.1281	Purifying	62.7
ZmTLP3	*ZmTLP22*	0.0495	0.3036	0.1630	Purifying	23.4
ZmTLP9	*ZmTLP12*	0.0767	0.3783	0.2027	Purifying	29.1

ZmTLP13	*ZmTLP37*	0.0523	0.3200	0.1634	Purifying	24.7
ZmTLP16	*ZmTLP33*	0.0133	0.0141	0.9483	Purifying	1.1
ZmTLP20	*ZmTLP25*	0.0797	0.0889	0.8965	Purifying	6.9

Ka- non-synonymous substitutions per non-synonymous site, **Ks**- synonymous substitutions per synonymous site, **T**- Divergence time

Table 2 Tajima's relative rate test of paralogous genes.

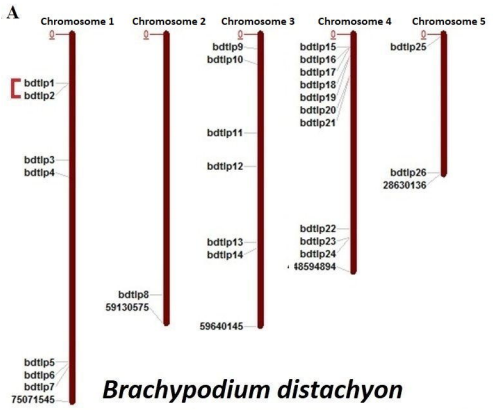
Group A	Group B	Outgroup	Nt	Na	Nb	χ^2	P
BdTLP1	*BdTLP2*	*BdTLP11*	403	47	42	0.28	0.59611
OsTLP5	*OsTLP7*	*OsTLP18*	401	28	29	0.02	0.89463
OsTLP22	*OsTLP24*	*OsTLP18*	333	15	17	0.13	0.72367
OsTLP23	*OsTLP25*	*OsTLP18*	313	20	27	1.04	0.30723
OsTLP26	*OsTLP27*	*OsTLP18*	411	34	34	0	1
SbTLP28	*SbTLP29*	*SbTLP5*	316	19	30	2.47	0.11608
SbTLP31	*SbTLP34*	*SbTLP5*	312	27	32	0.42	0.51508
SbTLP32	*SbTLP33*	*SbTLP5*	347	2	2	0	1
TaTLP1-A	*TaTLP24-D*	*TaTLP4-D*	483	37	30	0.73	0.39245
TaTLP1-B	*TaTLP24-B*	*TaTLP4-D*	486	36	31	0.37	0.5413
TaTLP12-B	*TaTLP13-D*	*TaTLP4-D*	321	24	33	1.42	0.23323
TaTLP15-A	*TaTLP18-B*	*TaTLP4-D*	391	36	38	0.05	0.81615
TaTLP15-B2	*TaTLP18-A*	*TaTLP4-D*	260	28	27	0.02	0.89274
TaTLP16-A	*TaTLP17-A*	*TaTLP4-D*	291	36	31	0.37	0.5413
ZmTLP1	*ZmTLP35*	*ZmTLP28*	551	36	33	0.13	0.71798
ZmTLP2	*ZmTLP34*	*ZmTLP28*	541	34	25	1.37	0.24132
ZmTLP3	*ZmTLP22*	*ZmTLP28*	560	11	14	0.36	0.54851
ZmTLP9	*ZmTLP12*	*ZmTLP28*	372	6	12	2	0.1573
ZmTLP13	*ZmTLP37*	*ZmTLP28*	464	19	22	0.22	0.63941

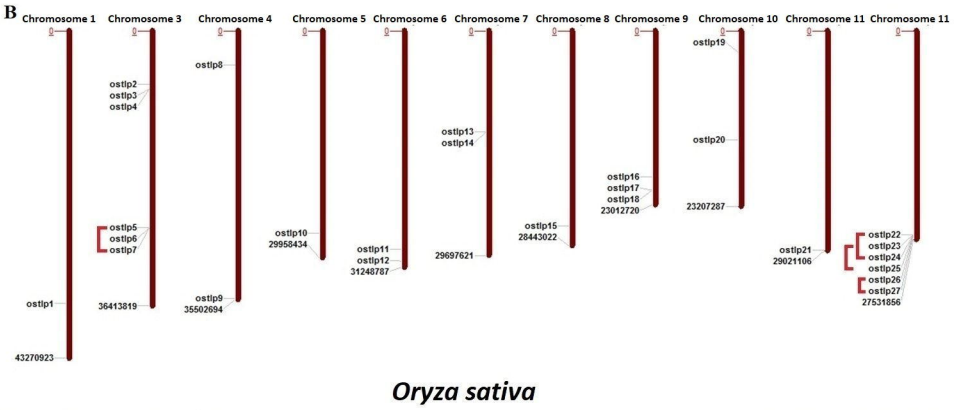
ZmTLP16	*ZmTLP33*	*ZmTLP28*	551	5	7	0.33	0.5637
ZmTLP20	*ZmTLP25*	*ZmTLP28*	525	17	23	0.9	0.34278

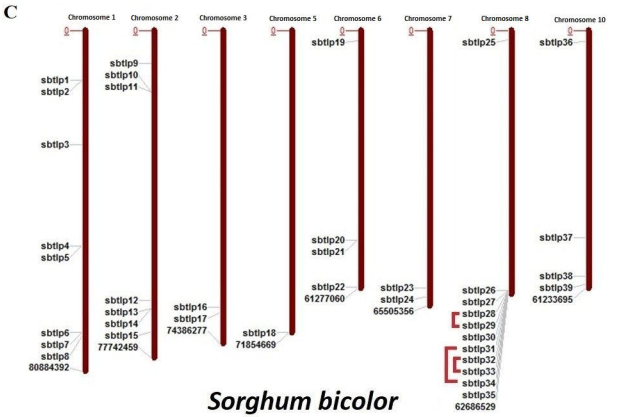
Nt- Identical sites in all three sequences**, Na**-Unique differences in Sequence A, **Nb-** Unique differences in Sequence B

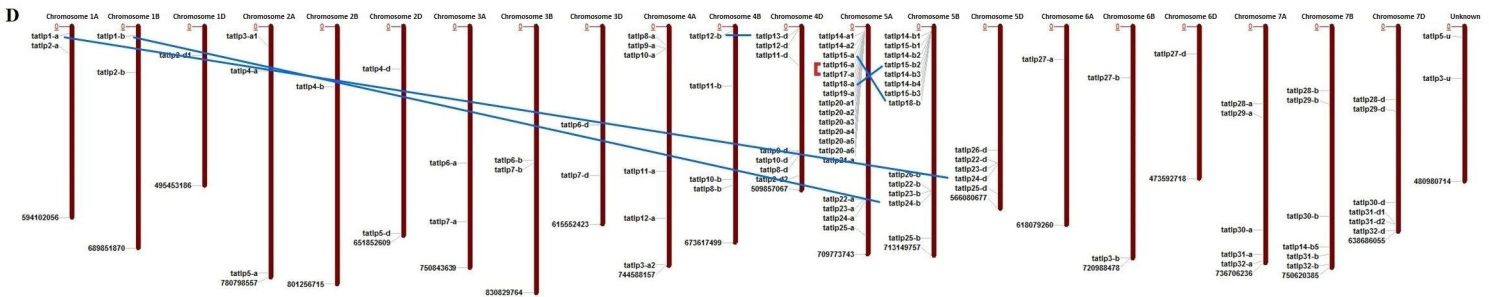
Table 3. Physicochemical properties of long and small TLP proteins.

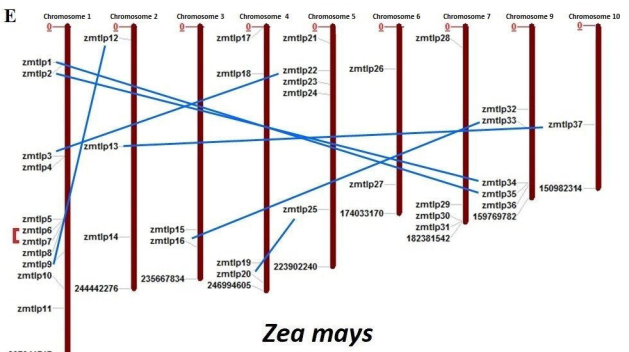
Property	Type	*Brachypodium distachyon*	*Oryza sativa*	*Sorghum bicolor*	*Triticum aestivum*	*Zea mays*
Average Protein length	Long TLPs	299	289	294	284	334
	Small TLPs	185	178	178	177	175
Average Molecular weight (kDa)	Long TLPs	30.8	29.8	30.0	29.1	35.0
	Small TLPs	19.5	18.1	18.3	18.22	17.7
Average pI	Long TLPs	6.18	6.95	6.05	6.18	5.89
	Small TLPs	6.82	6.06	5.74	6.17	4.90

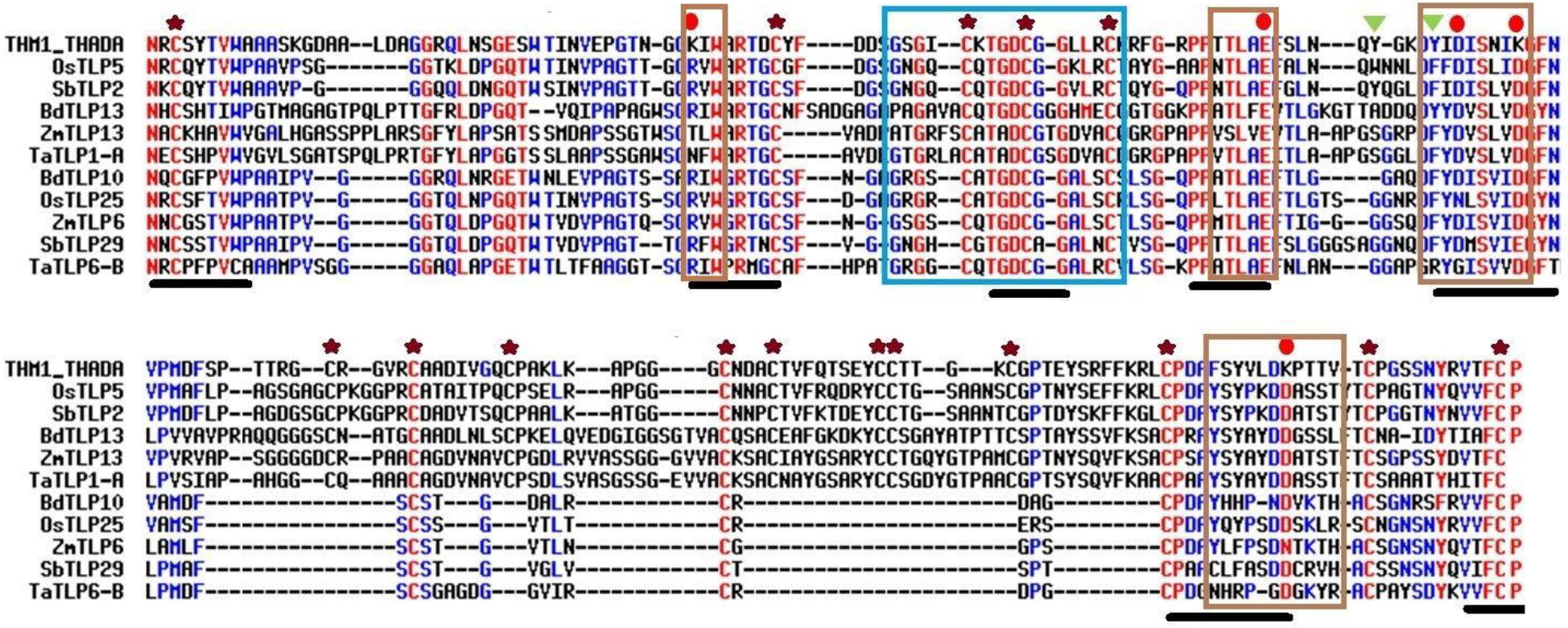

A
Chromosome 1
Chromosome 2
Chromosome 3
Chromosome 4
Chromosome 5
bdtip1
bdtip2
bdtip3
bdtip4
bdtip5
bdtip6
bdtip7
75071545
bdtip8
59130575
bdtip9
bdtip10
bdtip11
bdtip12
bdtip13
bdtip14
59640145
bdtip15
bdtip16
bdtip17
bdtip18
bdtip19
bdtip20
bdtip21
bdtip22
bdtip23
bdtip24
48594894
bdtip25
bdtip26
28630136
Brachypodium distachyon


B
Chromosome 1
Chromosome 3
Chromosome 4
Chromosome 5
Chromosome 6
Chromosome 7
Chromosome 8
Chromosome 9
Chromosome 10
Chromosome 11
Chromosome 11
ostip1
43270923
ostip2
ostip3
ostip4
ostip5
ostip6
ostip7
36413819
ostip8
ostip9
35502694
ostip10
29958434
ostip11
ostip12
31248787
ostip13
ostip14
29697621
ostip15
28443022
ostip16
ostip17
ostip18
23012720
ostip19
ostip20
23207287
ostip21
29021106
ostip22
ostip23
ostip24
ostip25
ostip26
ostip27
27531856
Oryza sativa


C
Chromosome 1
Chromosome 2
Chromosome 3
Chromosome 5
Chromosome 6
Chromosome 7
Chromosome 8
Chromosome 10
sbtip1
sbtip2
sbtip3
sbtip4
sbtip5
sbtip6
sbtip7
sbtip8
80884392
sbtip9
sbtip10
sbtip11
sbtip12
sbtip13
sbtip14
sbtip15
77742459
sbtip16
sbtip17
74386277
sbtip18
71854669
sbtip19
sbtip20
sbtip21
sbtip22
61277060
sbtip23
sbtip24
65505356
sbtip25
sbtip26
sbtip27
sbtip28
sbtip29
sbtip30
sbtip31
sbtip32
sbtip33
sbtip34
sbtip35
62686529
sbtip36
sbtip37
sbtip38
sbtip39
61233695
Sorghum bicolor


D
Chromosome 1A
Chromosome 1B
Chromosome 1D
Chromosome 2A
Chromosome 2B
Chromosome 2D
Chromosome 3A
Chromosome 3B
Chromosome 3D
Chromosome 4A
Chromosome 4B
Chromosome 4D
Chromosome 5A
Chromosome 5B
Chromosome 5D
Chromosome 6A
Chromosome 6B
Chromosome 6D
Chromosome 7A
Chromosome 7B
Chromosome 7D
Unknown
Triticum aestivum


E
Chromosome 1
Chromosome 2
Chromosome 3
Chromosome 4
Chromosome 5
Chromosome 6
Chromosome 7
Chromosome 9
Chromosome 10
zmtip1
zmtip2
zmtip3
zmtip4
zmtip5
zmtip6
zmtip7
zmtip8
zmtip9
zmtip10
zmtip11
307041717
zmtip12
zmtip13
zmtip14
244442276
zmtip15
zmtip16
235667834
zmtip17
zmtip18
zmtip19
zmtip20
246994605
zmtip21
zmtip22
zmtip23
zmtip24
zmtip25
223902240
zmtip26
zmtip27
174033170
zmtip28
zmtip29
zmtip30
zmtip31
182381542
zmtip32
zmtip33
zmtip34
zmtip35
zmtip36
159769782
zmtip37
150982314
Zea mays



Cysteine residue

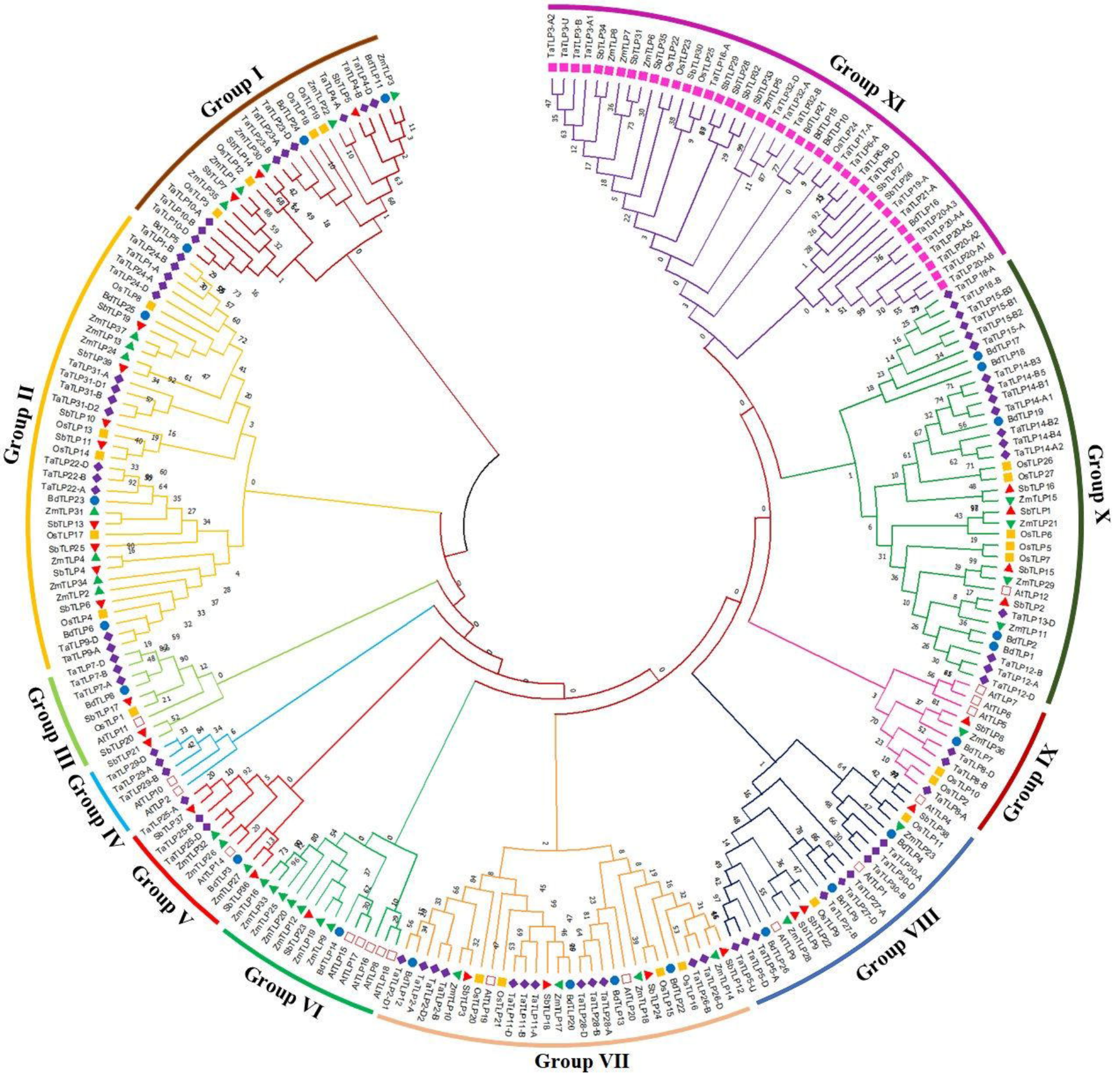
FF hydrophobic motif

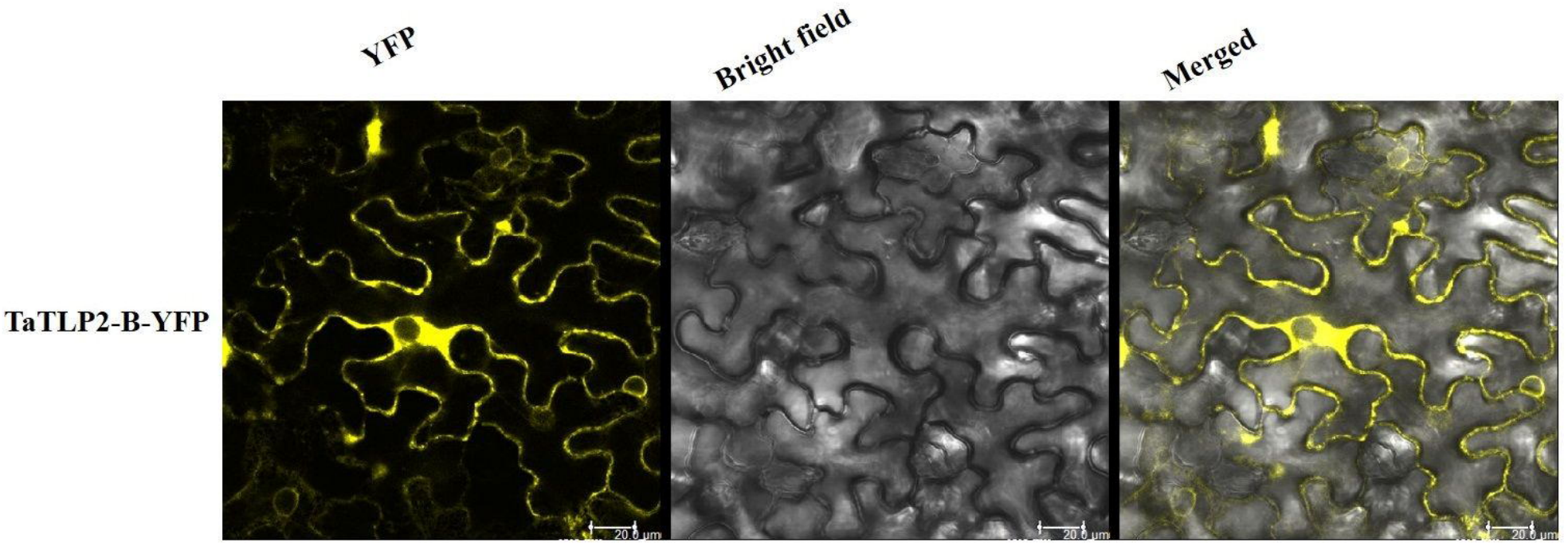
Signature motif

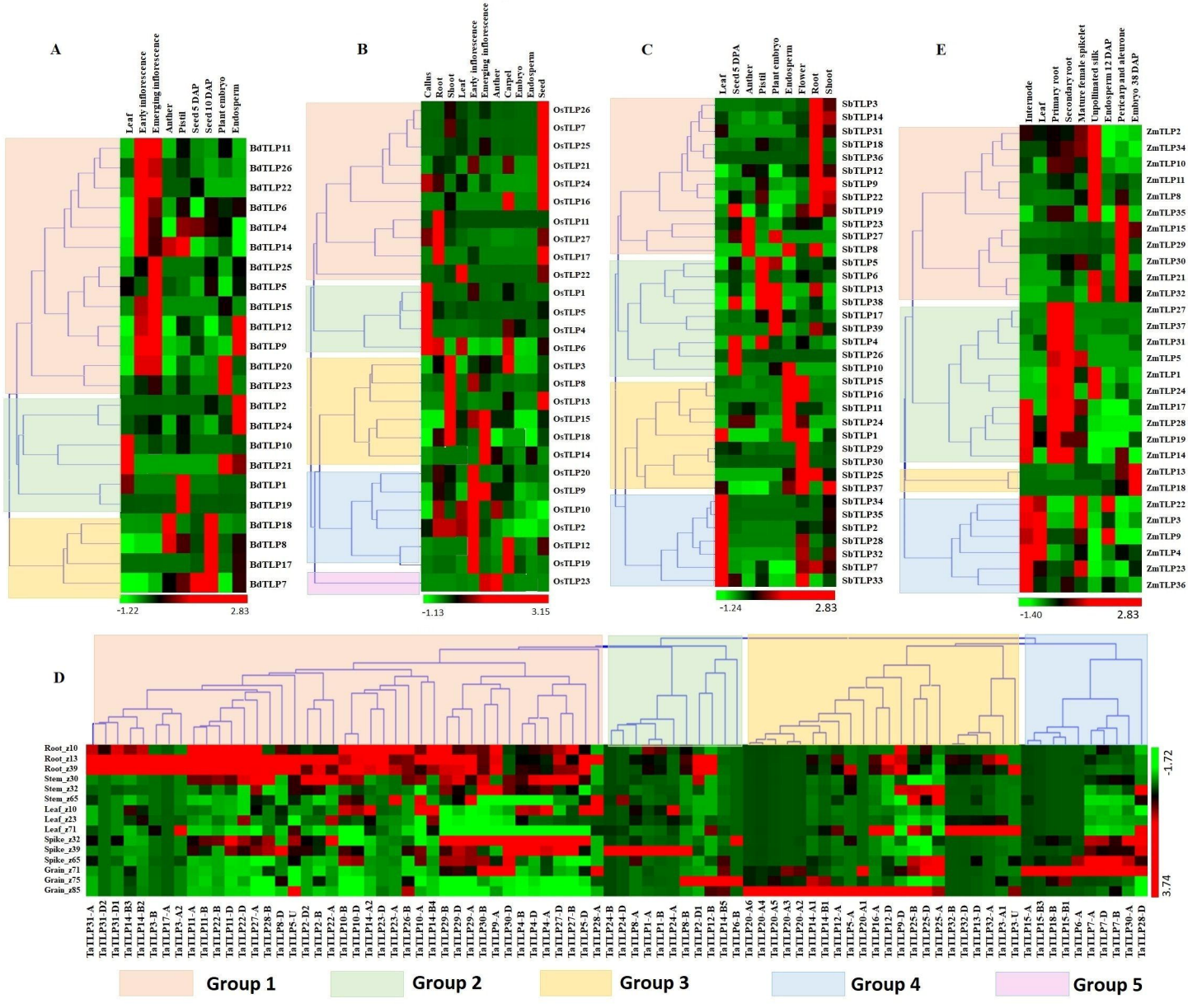
REDDD motif

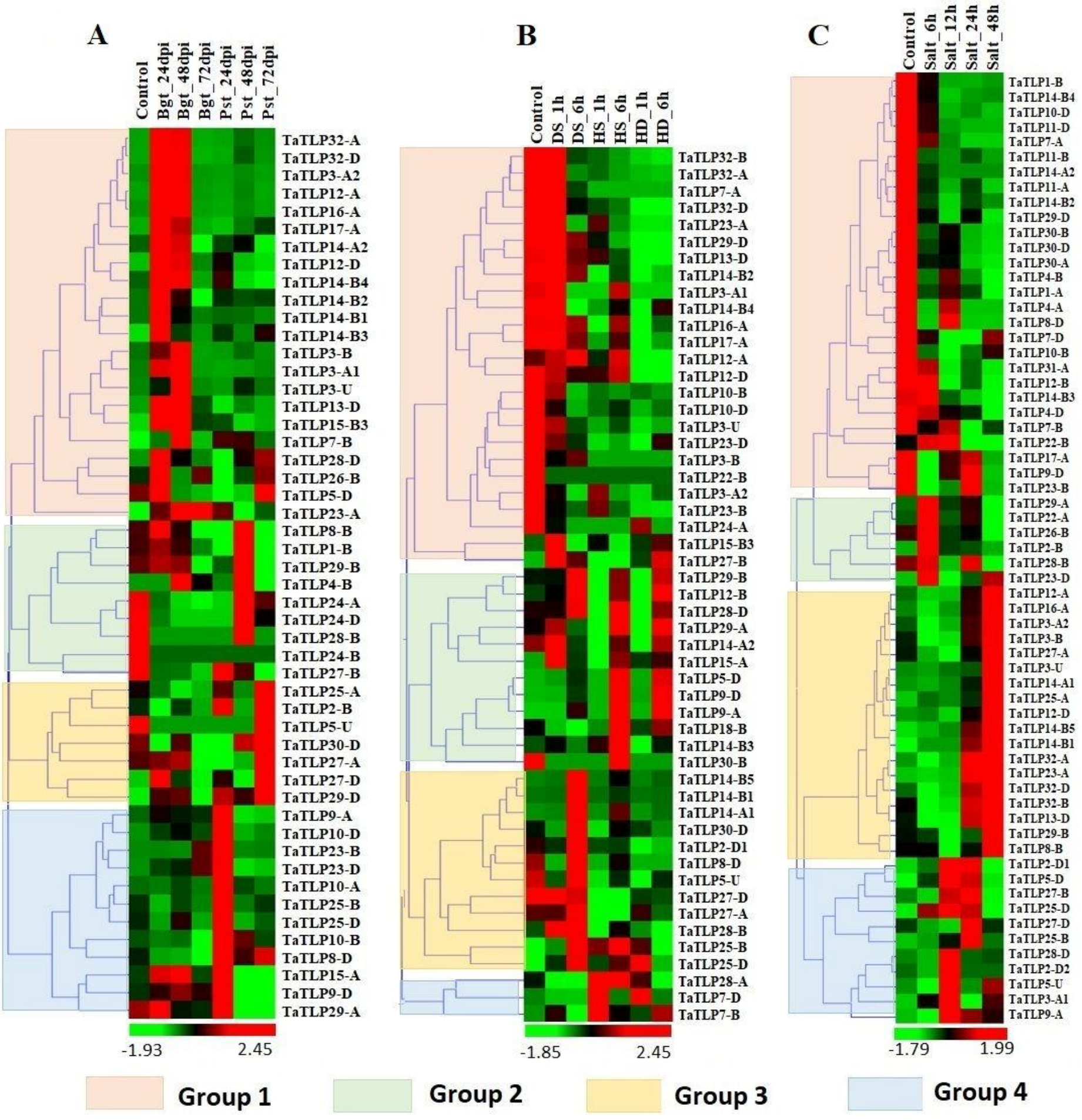
Bottom of acidic cleft forming amino acids

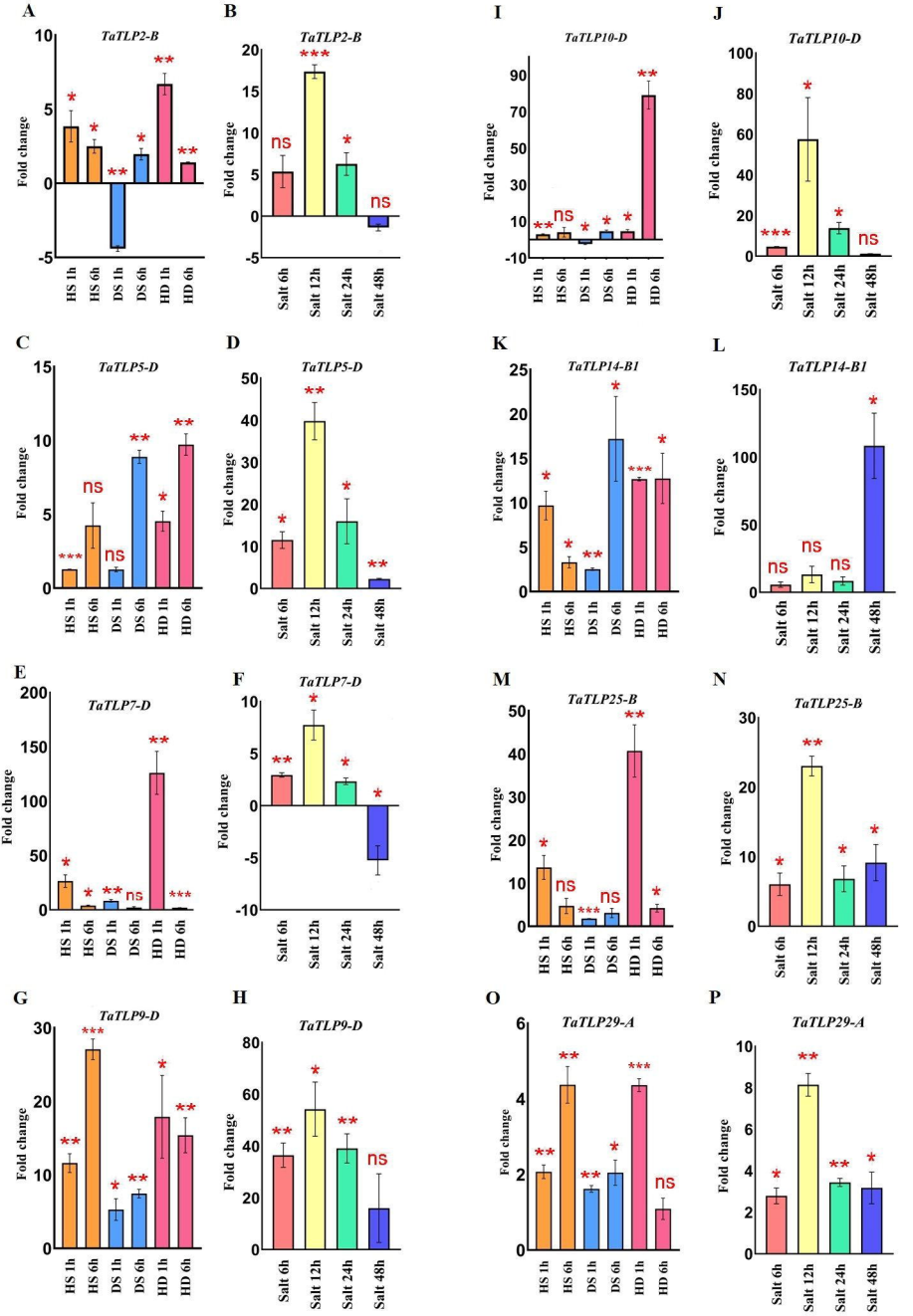
Thaumatin conserverd domain

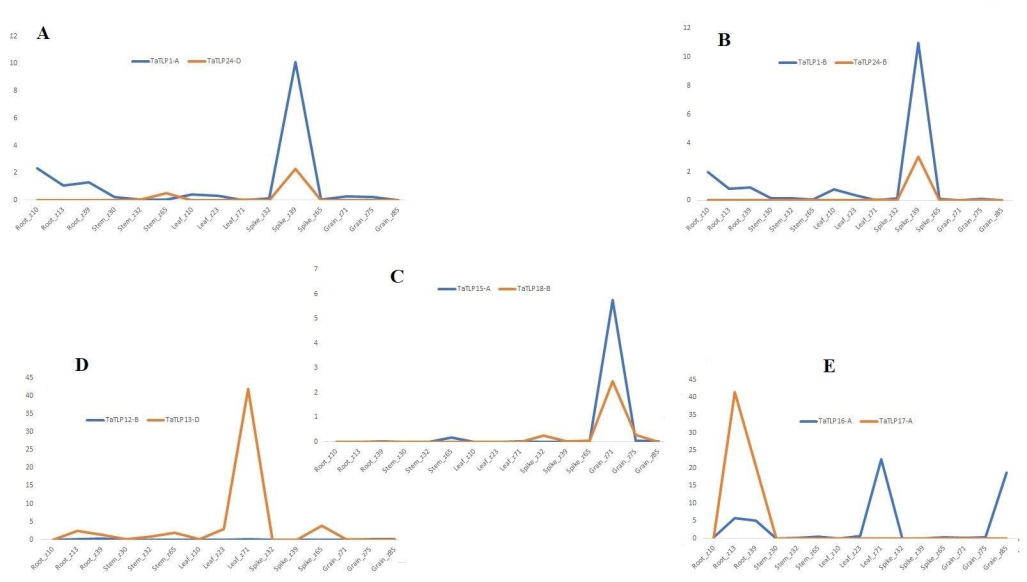

Group I
Group II
Group III
Group IV
Group V
Group VI
Group VII
Group VIII
Group IX
Group X
Group XI

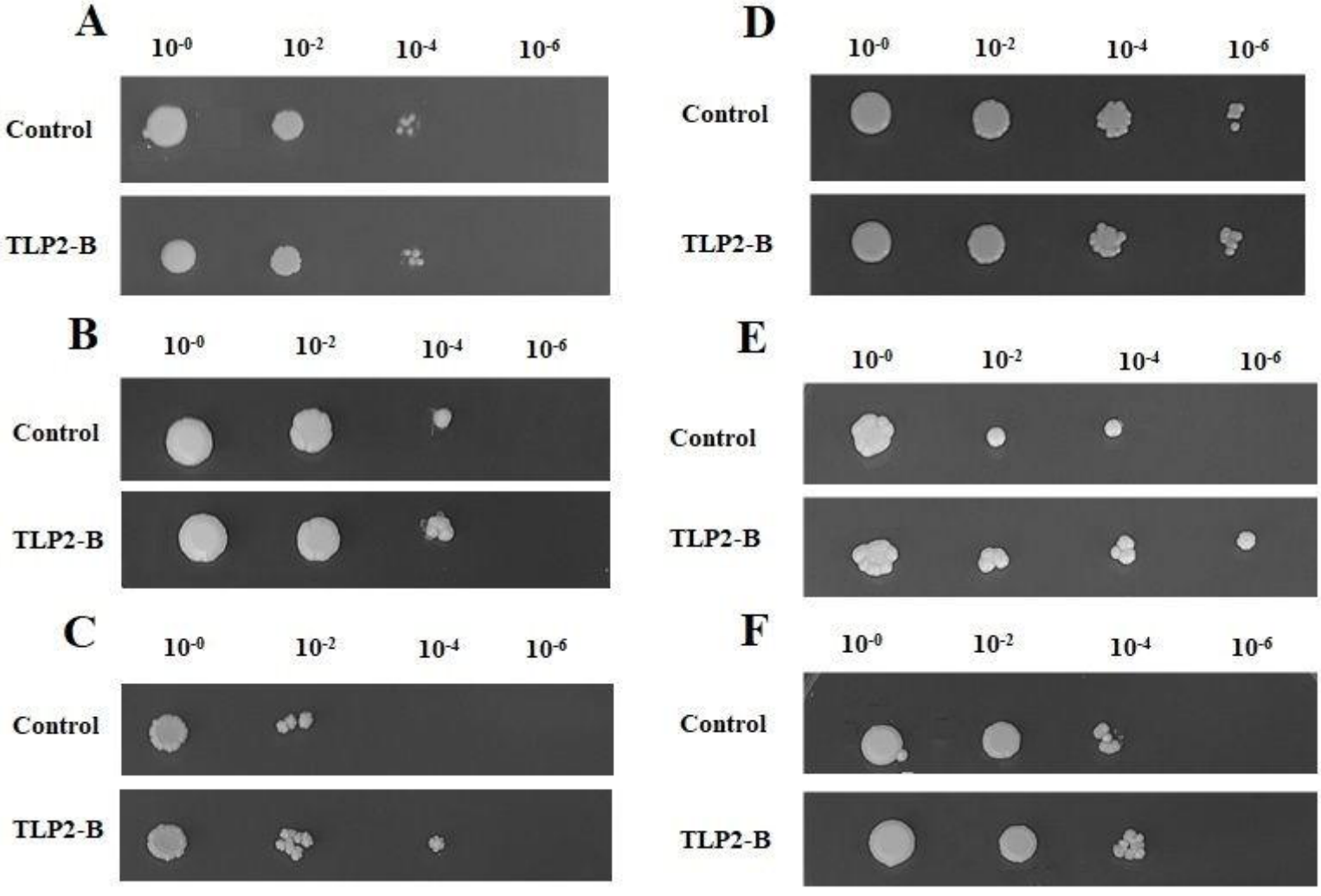

YFP
Bright field
Merged
TaTLP2-B-YFP
20.0 µm
20.0 µm
20.0 µm


A
Leaf
Early inflorescence
Emerging inflorescence
Anther
Pistil
Seed 5 DAP
Seed 10 DAP
Plant embryo
Endosperm
BdTLP11
BdTLP26
BdTLP22
BdTLP6
BdTLP4
BdTLP14
BdTLP25
BdTLP5
BdTLP15
BdTLP12
BdTLP9
BdTLP20
BdTLP23
BdTLP2
BdTLP24
BdTLP10
BdTLP21
BdTLP1
BdTLP19
BdTLP18
BdTLP8
BdTLP17
BdTLP7
-1.22
2.83
B
Callus
Root
Shoot
Leaf
Early inflorescence
Emerging inflorescence
Anther
Carpel
Embryo
Endosperm
Seed
OsTLP26
OsTLP7
OsTLP25
OsTLP21
OsTLP24
OsTLP16
OsTLP11
OsTLP27
OsTLP17
OsTLP22
OsTLP1
OsTLP5
OsTLP4
OsTLP6
OsTLP3
OsTLP8
OsTLP13
OsTLP15
OsTLP18
OsTLP14
OsTLP20
OsTLP9
OsTLP10
OsTLP2
OsTLP12
OsTLP19
OsTLP23
-1.13
3.15
C
Leaf
Seed 5 DPA
Anther
Pistil
Plant embryo
Endosperm
Flower
Root
Shoot
SbTLP3
SbTLP14
SbTLP31
SbTLP18
SbTLP36
SbTLP12
SbTLP9
SbTLP22
SbTLP19
SbTLP23
SbTLP27
SbTLP8
SbTLP5
SbTLP6
SbTLP13
SbTLP38
SbTLP17
SbTLP39
SbTLP4
SbTLP26
SbTLP10
SbTLP15
SbTLP16
SbTLP11
SbTLP24
SbTLP1
SbTLP29
SbTLP30
SbTLP25
SbTLP37
SbTLP34
SbTLP35
SbTLP2
SbTLP28
SbTLP32
SbTLP7
SbTLP33
-1.24
2.83
E
Internode
Leaf
Primary root
Secondary root
Mature female spikelet
Unpollinated silk
Endosperm 12 DAP
Pericarp and aleurone
Embryo 38 DAP
ZmTLP2
ZmTLP34
ZmTLP10
ZmTLP11
ZmTLP8
ZmTLP35
ZmTLP15
ZmTLP29
ZmTLP30
ZmTLP21
ZmTLP32
ZmTLP27
ZmTLP37
ZmTLP31
ZmTLP5
ZmTLP1
ZmTLP24
ZmTLP17
ZmTLP28
ZmTLP19
ZmTLP14
ZmTLP13
ZmTLP18
ZmTLP22
ZmTLP3
ZmTLP9
ZmTLP4
ZmTLP23
ZmTLP36
-1.40
2.83
D
Root_z10
Root_z13
Root_z39
Stem_z30
Stem_z32
Stem_z65
Leaf_z10
Leaf_z23
Leaf_z71
Spike_z32
Spike_z39
Spike_z65
Grain_z71
Grain_z75
Grain_z85
-1.72
3.74
TaTLP31-A
TaTLP31-D1
TaTLP31-D2
TaTLP14-B3
TaTLP14-B2
TaTLP7-A
TaTLP3-B
TaTLP3-A2
TaTLP11-A
TaTLP11-B
TaTLP22-B
TaTLP11-D
TaTLP22-D
TaTLP27-A
TaTLP28-B
TaTLP8-D
TaTLP5-U
TaTLP2-D2
TaTLP2-B
TaTLP22-A
TaTLP10-B
TaTLP10-D
TaTLP14-A2
TaTLP23-D
TaTLP23-A
TaTLP26-B
TaTLP10-A
TaTLP14-B4
TaTLP29-B
TaTLP29-D
TaTLP29-A
TaTLP30-B
TaTLP9-A
TaTLP30-D
TaTLP4-B
TaTLP4-D
TaTLP4-A
TaTLP27-D
TaTLP27-B
TaTLP5-D
TaTLP28-A
TaTLP24-B
TaTLP24-D
TaTLP8-A
TaTLP1-A
TaTLP1-B
TaTLP24-A
TaTLP8-B
TaTLP2-D1
TaTLP12-B
TaTLP14-B5
TaTLP6-B
TaTLP20-A6
TaTLP20-A4
TaTLP20-A5
TaTLP20-A3
TaTLP20-A2
TaTLP14-A1
TaTLP14-B1
TaTLP12-A
TaTLP5-A
TaTLP20-A1
TaTLP16-A
TaTLP12-D
TaTLP9-D
TaTLP25-B
TaTLP25-D
TaTLP25-A
TaTLP32-B
TaTLP32-D
TaTLP13-D
TaTLP32-A
TaTLP3-A1
TaTLP3-U
TaTLP15-A
TaTLP15-B3
TaTLP18-B
TaTLP15-B1
TaTLP6-A
TaTLP7-A
TaTLP7-D
TaTLP7-B
TaTLP30-A
TaTLP28-D
Group 1
Group 2
Group 3
Group 4
Group 5


A
Control
Bgt_24dpi
Bgt_48dpi
Bgt_72dpi
Pst_24dpi
Pst_48dpi
Pst_72dpi
TaTLP32-A
TaTLP32-D
TaTLP3-A2
TaTLP12-A
TaTLP16-A
TaTLP17-A
TaTLP14-A2
TaTLP12-D
TaTLP14-B4
TaTLP14-B2
TaTLP14-B1
TaTLP14-B3
TaTLP3-B
TaTLP3-A1
TaTLP3-U
TaTLP13-D
TaTLP15-B3
TaTLP7-B
TaTLP28-D
TaTLP26-B
TaTLP5-D
TaTLP23-A
TaTLP8-B
TaTLP1-B
TaTLP29-B
TaTLP4-B
TaTLP24-A
TaTLP24-D
TaTLP28-B
TaTLP24-B
TaTLP27-B
TaTLP25-A
TaTLP2-B
TaTLP5-U
TaTLP30-D
TaTLP27-A
TaTLP27-D
TaTLP29-D
TaTLP9-A
TaTLP10-D
TaTLP23-B
TaTLP23-D
TaTLP10-A
TaTLP25-B
TaTLP25-D
TaTLP10-B
TaTLP8-D
TaTLP15-A
TaTLP9-D
TaTLP29-A
-1.93
2.45
B
Control
DS_1h
DS_6h
HS_1h
HS_6h
HD_1h
HD_6h
TaTLP32-B
TaTLP32-A
TaTLP7-A
TaTLP32-D
TaTLP23-A
TaTLP29-D
TaTLP13-D
TaTLP14-B2
TaTLP3-A1
TaTLP14-B4
TaTLP16-A
TaTLP17-A
TaTLP12-A
TaTLP12-D
TaTLP10-B
TaTLP10-D
TaTLP3-U
TaTLP23-D
TaTLP3-B
TaTLP22-B
TaTLP3-A2
TaTLP23-B
TaTLP24-A
TaTLP15-B3
TaTLP27-B
TaTLP29-B
TaTLP12-B
TaTLP28-D
TaTLP29-A
TaTLP14-A2
TaTLP15-A
TaTLP5-D
TaTLP9-D
TaTLP9-A
TaTLP18-B
TaTLP14-B3
TaTLP30-B
TaTLP14-B5
TaTLP14-B1
TaTLP14-A1
TaTLP30-D
TaTLP2-D1
TaTLP8-D
TaTLP5-U
TaTLP27-D
TaTLP27-A
TaTLP28-B
TaTLP25-B
TaTLP25-D
TaTLP28-A
TaTLP7-D
TaTLP7-B
-1.85
2.45
C
Control
Salt_6h
Salt_12h
Salt_24h
Salt_48h
TaTLP1-B
TaTLP14-B4
TaTLP10-D
TaTLP11-D
TaTLP7-A
TaTLP11-B
TaTLP14-A2
TaTLP11-A
TaTLP14-B2
TaTLP29-D
TaTLP30-B
TaTLP30-D
TaTLP30-A
TaTLP4-B
TaTLP1-A
TaTLP4-A
TaTLP8-D
TaTLP7-D
TaTLP10-B
TaTLP31-A
TaTLP12-B
TaTLP14-B3
TaTLP4-D
TaTLP7-B
TaTLP22-B
TaTLP17-A
TaTLP9-D
TaTLP23-B
TaTLP29-A
TaTLP22-A
TaTLP26-B
TaTLP2-B
TaTLP28-B
TaTLP23-D
TaTLP12-A
TaTLP16-A
TaTLP3-A2
TaTLP3-B
TaTLP27-A
TaTLP3-U
TaTLP14-A1
TaTLP25-A
TaTLP12-D
TaTLP14-B5
TaTLP14-B1
TaTLP32-A
TaTLP23-A
TaTLP32-D
TaTLP32-B
TaTLP13-D
TaTLP29-B
TaTLP8-B
TaTLP2-D1
TaTLP5-D
TaTLP27-B
TaTLP25-D
TaTLP27-D
TaTLP25-B
TaTLP28-D
TaTLP2-D2
TaTLP5-U
TaTLP3-A1
TaTLP9-A
-1.79
1.99
Group 1
Group 2
Group 3
Group 4


A
TaTLP2-B
Fold change
HS 1h
HS 6h
DS 1h
DS 6h
HD 1h
HD 6h
B
TaTLP2-B
Fold change
Salt 6h
Salt 12h
Salt 24h
Salt 48h
I
TaTLP10-D
Fold change
HS 1h
HS 6h
DS 1h
DS 6h
HD 1h
HD 6h
J
TaTLP10-D
Fold change
Salt 6h
Salt 12h
Salt 24h
Salt 48h
C
TaTLP5-D
Fold change
HS 1h
HS 6h
DS 1h
DS 6h
HD 1h
HD 6h
D
TaTLP5-D
Fold change
Salt 6h
Salt 12h
Salt 24h
Salt 48h
K
TaTLP14-B1
Fold change
HS 1h
HS 6h
DS 1h
DS 6h
HD 1h
HD 6h
L
TaTLP14-B1
Fold change
Salt 6h
Salt 12h
Salt 24h
Salt 48h
E
TaTLP7-D
Fold change
HS 1h
HS 6h
DS 1h
DS 6h
HD 1h
HD 6h
F
TaTLP7-D
Fold change
Salt 6h
Salt 12h
Salt 24h
Salt 48h
M
TaTLP25-B
Fold change
HS 1h
HS 6h
DS 1h
DS 6h
HD 1h
HD 6h
N
TaTLP25-B
Fold change
Salt 6h
Salt 12h
Salt 24h
Salt 48h
G
TaTLP9-D
Fold change
HS 1h
HS 6h
DS 1h
DS 6h
HD 1h
HD 6h
H
TaTLP9-D
Fold change
Salt 6h
Salt 12h
Salt 24h
Salt 48h
O
TaTLP29-A
Fold change
HS 1h
HS 6h
DS 1h
DS 6h
HD 1h
HD 6h
P
TaTLP29-A
Fold change
Salt 6h
Salt 12h
Salt 24h
Salt 48h


A
TaTLP1-A
TaTLP24-D
B
TaTLP1-B
TaTLP24-B
C
TaTLP15-A
TaTLP18-B
D
TaTLP12-B
TaTLP13-D
E
TaTLP16-A
TaTLP17-A
Root_210
Root_213
Root_259
Stem_230
Stem_232
Stem_265
Leaf_210
Leaf_223
Leaf_271
Spike_232
Spike_259
Spike_265
Grain_271
Grain_275
Grain_285


A
10-0
10-2
10-4
10-6
Control
TLP2-B
B
10-0
10-2
10-4
10-6
Control
TLP2-B
C
10-0
10-2
10-4
10-6
Control
TLP2-B
D
10-0
10-2
10-4
10-6
Control
TLP2-B
E
10-0
10-2
10-4
10-6
Control
TLP2-B
F
10-0
10-2
10-4
10-6
Control
TLP2-B