Plant Omics Journal POJ

POJ 5(6):518-531 (2012) ISSN:1836-3644

Comprehensive computational analysis of different classes of Glutathione *S*-transferases in *Triticum aestivum* L.

B. Pandey[1], P. Sharma[1*], D.M Pandey[2], J. Varshney[1], S. Sheoran[1], M. Singh[1], R. Singh[1], I. Sharma[1] and R. Chatrath[3]

[1]Plant Biotechnology, Directorate of Wheat Research, Karnal-132001, India
[2]Department of Biotechnology, BIT, Mesra, Ranchi- 835 215, India
[3]Crop Improvement, Directorate of Wheat Research, Karnal-132001, India

***Correspondence author: neprads@gmail.com**

Abstract

Glutathione *S*-transferases (GSTs) appear to be ubiquitous in plants and have defined roles in herbicide detoxification of a wide variety of xenobiotic compounds. A newly discovered plant GST subclass has been intensively studied in numerous stress responses, including those arising from oxidative stress, pathogen attack and heavy metal toxicity. In the current study, we investigated a comprehensive role of the GSTs in *Triticum aestivum* L. using *in silico* analytical approaches. The motifs predicted in GST classes were aligned to generate the position weight matrix and information content by using perl script. *cis-acting* regulatory elements present within the 5' regulatory region of the wheat GSTs, were identified using Plant CARE and PLACE databases. Prediction of folding state of wheat GSTs indicated that only PhiGST have disordered amino acid residues. Considerable degree of homology was seen in alignment of all available GST sequences in different cereal crops by ClustalW2 and Neighbour-joining method. In addition, we discovered expression profile pattern of Tau and Phi classes using the available EST information of wheat GST genes. We described the structure model for both Phi and Zeta classes using homology model, essential for defining the active site and also for designing, improving docking of small ligands to the complex target protein. This study revealed the possible role of *cis*-acting regulatory elements in the expression and regulation of GST gene families in *T. aestivum* during cellular development or environmental stress conditions.

Keywords: GSTs, *cis*-regulatory elements, Structure prediction, Expression profile.
Abbreviations: Glutathione *S*-transferases GSTs; Transcription factors (TFs); ABA responsive element (ABRE); light responsive *cis*-elements (LREs); 5' untranslated region (5' UTR); isoelectric point (pI); Multiple sequence alignment (MSA); Expressed sequence tags (ESTs); coding domain sequence (CDS); Digital differential display (DDD); Gene ontology (GO); Neighbor-joining (NJ); Position weight matrices (PWMs); Cytosolic GSTs (cGSTs).

Introduction

Bread wheat, *Triticum aestivum* L, is an allohexaploid plant. It is cultivated throughout the world's temperate regions in the Southern and Northern hemispheres and is one of the most important crops in the world in terms of planting area (Bonjean and Angus, 2001). Throughout their lives, plants respond to different kinds of biotic and abiotic stresses such as insect attack by expressing specific genes that are required to adapt defend against those stresses and attacks (Halliwell and Gutteridge, 2006). The glutathione S-transferases (GSTs, EC 2.5.1.18), a multigene family enzymes are involved in the cellular detoxification of a broad range of electrophilic xenobiotics and reactive endogenous compounds of the oxidative metabolism. The interest in plant GSTs may be attributed to their agronomic value, since it has been demonstrated that glutathione conjugation for a variety of herbicides is the major resistance and selectivity factor in plants. This group of enzymes is a diverse protein family group encoded by multi gene families (Wilce and Parker, 1994; Edwards et al., 2000; Dixon et al., 2002). GST activities have been identified in both eukaryotes and prokaryotes (Wilce and Parker, 1994). Some functions of GSTs include herbicide detoxification, signal transduction, plant protection against ozone damages and heavy metals (Esmaeili et al., 2009). The role of plant GSTs has also been proposed in the transport and metabolism of secondary compounds (Blackburn et al., 2001; Board et al., 2000). Plant GSTs also act as glutathione peroxidases (Bonjean et al., 2001; Buetler and Eaton, 1992), protect cells from oxygen toxicity and suppress apoptosis (Coelho et al., 2010) and detoxication. The most widely examined role of GSTs is their function as detoxification enzymes, where they catalyze the nucleophiles attack of glutathione (GSH) on electrophilic substrates, thereby decreasing their reactivity with cellular macromolecules (Armstrong, 1997). Based on sequence similarity and gene organization, plant GSTs can be divided into four main classes (phi, zeta, tau and theta). The phi (F) class GSTFs and the tau (U) class GSTUs are mainly plant-specific whereas, the zeta and theta classes are more phylogenetically widespread. However, the activities of different GSTs have been detected and characterized in many plants, including maize (Dixon et al., 1997), wheat (Edwards and Cole,1996), tobacco (Droog et al., 1995), soybean (Andrews et al., 1997), barley (Romano et al., 1993), chickpea (Hunatti and Ali, 1990), peanut (Lamoureux et al., 1981), sorghum (Dean et al., 1991), and sugarcane (Singhal et al., 1991). Cytosolic GSTs (cGSTs) are most abundant GST subfamily found in all aerobic organisms (Frova, 2006). cGSTC are actively involved in the detoxification of generally nonpolar compounds that contain an electrophilic carbon, nitrogen, or sulfur atom (Hayes et al., 2005). The development of different cell types from an unchanging set of genes is ruled by

physiological and biochemical processes that control gene activity. The selective expression of this set of genes directs plant development, cellular differentiation and responses to environmental stimuli (Franklin and Cande, 1999). Transcription in gene expression is the most essential step and has been extensively studied in cell and molecular biology (Wyeth and Albin, 2004). Gene specific regulation of transcription is of great significance for all phases of cellular functions. Transcriptional gene regulation is important for the function and development of all organisms. DNA binding transcription factors (TFs) are one of the important components of this network. Transcription is created by the interaction between TFs that usually bind to the *cis*-acting regulatory elements in the gene with some additional co-factors to activate or repress gene transcription in response to change in the environment, as well as during development. A single TF rarely controls gene expression, but precise combinations of many TFs are essential for the differential gene expression in higher organisms (Lloyd et al., 2001). This usually occurs on the promoters of each gene close to the transcription start site (Qui, 2003). Therefore, understanding of the *cis*-acting regulatory region bound by TFs that control gene expression will offer the essential information to build transcriptional regulatory systems (Qui, 2003). As well acknowledged, abiotic and biotic stress causes major losses in crop productivity worldwide. In plants, multiple signaling pathways regulate stress responses, and these often overlap with one another in response to different stresses. As an example, fungal infestation on plant gene expression profiles was found to overlap with those obtained for wounding (Durrant et al., 2000).Stress induced genes occur mostly at transcription levels, and their temporal and spatial expression regulatory patterns are vital part of the plant stress response system (Rushton and Somssich,1998). Usually strongly related TFs have the possibility to activate or suppress the transcription of genes by interacting with the *cis*-acting regulatory sequences that respond to a particular stress. Plants have been attested to devote a large portion of their genome capacity to transcription, with the *Arabidopsis* genome coding in excess of 1500 transcription factors (Riechmann et al., 2000). In this study, the establishment of an extensive picture of GST gene family in wheat was attempted. To date, no work has been published on the *cis*-acting regulatory elements that regulate the expression of GST genes. Therefore, the aim of this study was the prediction of available computational tools to depict a comprehensive account on *cis*-acting regulatory elements that are present within the 5'regulatory region of the DNA sequences of GST gene families in wheat as a source of potentially useful information for the prediction of expression and regulation of these genes. To understand the sequence and structure correlation, we have carried out bioinformatics analysis to deduce theoretical pI, conserved domain sequences, subcellular locations, recombinant proteins solubility and folding states. In order to see the evolutionary relationship among the wheat GST proteins, we performed a phylogenetic analysis based on their protein sequence and developed 3-D models for the Tau, Phi and zeta GSTs for structural comparison. Using the available EST information as a source of expression data, GST genes from wheat were detected in as many as ten wheat tissues.

Results

Conserved motif scanning, Position weight matrices (PWMs), Information content and analysis

The curated nucleotide sequence data of Tau (TaGSTU), Phi (TaGSTF) and Zeta (TaGATZ) GSTs were retrieved in FASTA format (Table 1) and subjected to MEME tool (http://meme.sdsc.edu/meme/meme-intro.html)fortheelucidation of motifs (Supplementary Table S1) in wheat GST classes. Single motif was predicted in the N'-terminal of the TaGSTU ranged between 492-515bp. PWM had four rows (one for each of A, C, G and T) and the number of columns was equal to the length of the motif. The PWMs for these motifs were derived by the perl script (http://www.perl.org/) and we found that start position (1-5) and end position (12-14) were highly conserved in the entire TaGSTU motif whereas positions 6, 7, 8, 9 and 15 has a strong preference for nucleotide T (Fig 1A).We measured the positions of high information content in the motif and plotted graph by using information theory. Information content of Tau motif exhibits that there was a higher degree of information content from position 1 to 5 and 12 to14 with 0 values whereas position 6 to 11 and 15 had lower information content value ranging from 1 to 1.2 (Fig 1C; Supplementary Table S2a). The consensus sequence for a PWM was constructed by choosing the nucleotide with the highest frequency of occurrence at each position. Finally CGAGTCCC TCATCAT was the strongly conserved sequence with bigger log odd score among all predicted motif sequence for Tau class (Supplementary Table S3a). A single Phi motif ranges from 587-634bp, in which 10-15 position were highly conserve for nucleotide T and G whereas 4 and 5 having 100% probability for nucleotide A. Rest of the position had maximum probability for nucleotide G (Fig 1B). The information content for the Phi motif was maximum at position 3 (1), 6 (0.6) and 9 (1.6) leading to low degree of conservation (Fig 1D; Supplementary Table S2b). GTGAAGGCCTGG TGG was the strongly conserved sequence for Phi GST motif (Supplementary Table S3b).

cis-acting regulatory elements analysis

The complete nucleotide sequence and the coding domain sequence (CDS) of the GST gene family in wheat (Table 1) were computationally examined to identify the *cis*- regulatory region. In TaGSTU gene family, the core elements: AACGG, CAAT, CAAACAC and ACGT motifs which are closest to the translational start site (ATG) were found to be very close to one another. However, RAV1AAT was found at the distal part of the 5' regulatory region in all of the genes. Conversely in TaGSTF gene family; AACCAA, CAAT, CCTTTTG, TTTCTTCTCT, CTCTT and CCAACC motifs were closest to the translational site and found to be very close to one another in several of the genes (Table 2). MYB family motifs were found in the regulatory regions of TaGST1A, TaGSTU2, TaGSTF5, TaGSTF3 and TaGSTZ gene families whereas Erd1 was present in TaGSTU3 gene and CAAACAC were present in TaGSTU1C. ABA responsive (ABRE) element found in TaGSTF3 has been functionally identified in the 5'regulatory region of several genes that are induced or regulated by ABA. The W-box motif was only present TaGSTZ gene family; whereas WRKY was present TaGSTF6b and TaGSTZ. The GARE and P-box motif was found within the regulatory regions of TaGSTF2, whereas TaGSTF1 gene contained 5'UTR Py-rich stretch and TCA regulatory elements, involve d in high transcription levels and salicylic acid responsiveness. TCA, a *cis*-acting regulatory element involved in salicylic responsiveness. GATA-box and Sp1motif were located at TaGSTU3 gene, known as light responsive *cis*-elements (LREs). SORLIP5AT, REALPHALGLHCB21, CACTFTPPCA1, CURE, RHERPATEXPA7 were present in TaGST6b. WRKY71OS was present in TaGSTZ and TGST6b, while CTCTT and ACACGA motifs in TaGSTF5 and TaGSTF4, respectively.

Table 1. The complete nucleotide sequence and the coding domain sequence (CDS) of the GST gene family in wheat.

Gene	Accession number	Number of base pairs	CDS region	5' UTR before ATG
TaGSTU1A gene	AJ414697.1	1085	46- 714	1-45
TaGSTU1B gene	AJ414698.1	1051	42-710	1-41
TaGSTU1C gene	AJ414699.1	1008	54-722	1-53
TaGSTU2 gene	AJ414700.1	926	66-76	1-65
TaGSTU3 gene	AJ414701.1	1043	39-770	1-38
TaGSTF6b gene	AJ441055.1	904	69-725	1-68
TaGSTF4 gene	AJ440793.1	721	21-689	1-20
TaGSTF2 gene	AJ440791.1	865	54-728	1-53
TaGSTF1 gene	AJ440796.1	927	72-710	1-71
TaGSTF5 gene	AJ440794.1	866	45-686	1-44
TaGSTF3 gene	AJ440792.1	930	60-728	1-59
TaGSTZ gene	AF109714.1	2947	17-98,603-647,1111-1196,1298-1369,1558-1611, 1719-1794,2036-2097,2314-2373,2477-2581	1-16

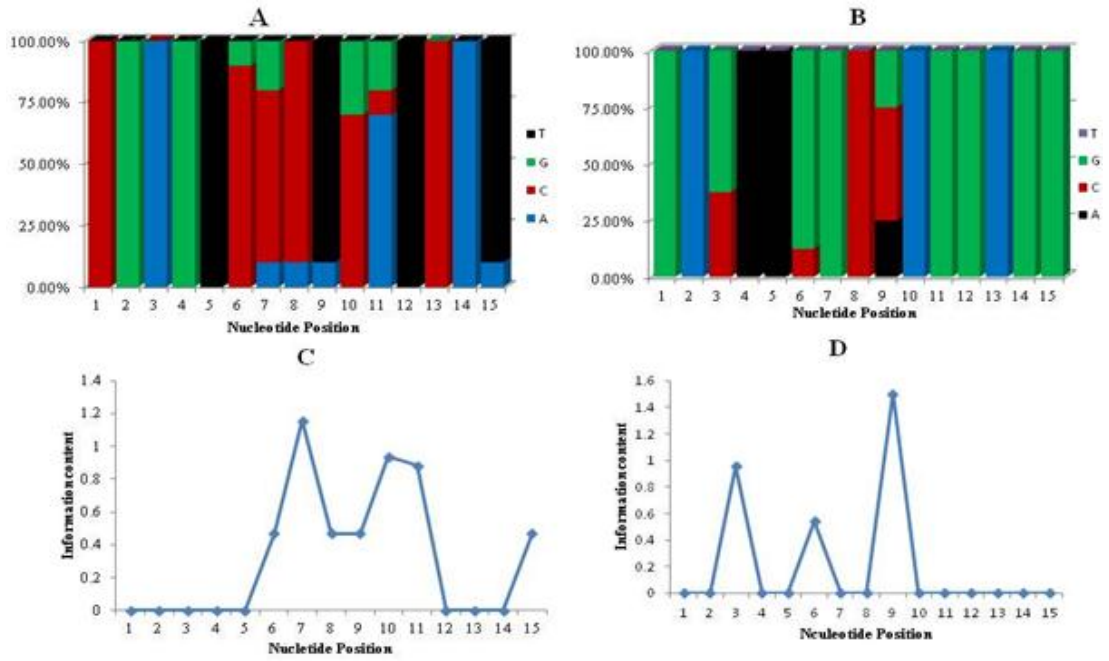

Fig 1. PWMs for (A) TaGSTU and (B) TaGSTU from wheat. We used color-coded vertical bars stacked one over the other to represent the percentage of each nucleotide, Green (A), Blue (C), Black (G) and Red (T). The bars are ordered by their percentages. Information content graph for (C) TaGSTU and (D) TaGSTU showed high degree of conservation at the start and end position of the in the motif sites.

Functional characterization of GST protein

The folding state of wheat GST was predicted by the FoldIndex program (Prilusky et al., 2005). The TaGSTF class showed small unfolding. The highest percentage of disordered residue was achieved 14.28% in TaGSTF2 followed by 5.40 % in TaGSTF4 (Fig 2). However, other two classes of GST were completely stable. Sub-cellular location result showed mitochondrial targeting peptide in TaGSTU3 and chloroplast-targeting peptide in case of TaGSTZ, whereas, Phi family members lack such peptide signal .The statistical model used for the prediction of protein solubility assuming the protein is being over-expressed in *Escherichia coli*. The result showed that the TaGSTU and TaGSTZ has less than 60%, chance of solubility when overexpressed in *E. coli* while Phi has more than 60% chance. The values of theoretical pI for Tau class was same for TaGSTU1A, TAGSTU1B and TaGSTU2 (pI= 6.34) whereas TaGSTU1C and TaGSTu2 had pI value less than 6. All TaGSTF had theoretical pI less than 6 except for TaGSTF6B (pI = 6.5). Zeta class with single sequence had 6.5 theoretical pI values.

Gene ontology

Gene ontology (GO) annotation was performed using the AmiGO annotation tool and GST sequences were classified into three categories, namely, biological process, cellular component and molecular function. Molecular function assignment revealed members of broad range of categories (Table 2) including TaGSTZ genes with putative catalytic activity, protein binding, oxidoreductase activity, transferase activity, isomerase activity, glutathione transferase activity, and peroxidase activity whereas, TaGSTU and TaGSTF were found to be involved in glutathione transferase activity and peroxidase activity. The differentially regulated TaGSTZ gene was member of variety cellular components with a large

Table 2. Cis regulatory element in 5' regulatory region of the GST class in wheat.

Gene	Accession number	Factor name	Position		Function
TaGSTU1A gene	AJ414697.1	2SSEEDPROTBANAPA	1	CAAACAC	Important for high activity of the napA promoter
		ARR1AT	24	NGATT	Response regulators operate as transcriptional activators
		MYBCORE	34	CNGTTR	Induced by dehydration stress and its gene product binds to the conserved MYB recognition sequence
		DPBFCOREDCDC3	4	ACACNNG	Encodes a bZIP transcription
TaGSTU1B gene	AJ414698.1	ARE	10	TGGTTT	Cis-acting regulatory element essential for the anaerobic induction
		CAAT	14	CAAT	Common cis-acting element in promoter and enhancer regions
TaGSTU1C gene	AJ414699.1	2SSEEDPROTBANAPA	9	CAAACAC	Important for high activity of the napA promoter
		RAV1AAT	4	CAACA	Protein contain AP2-like and B3 like domains
TaGSTU2 gene	AJ414700.1	ACGTATERD1	15	ACGT	required for etiolation-induced expression of erd1
		MYBCOREATCYCB1	31	AACGG	able to activate reporter gene without leading to M-phase-specific expression
		GTGANTG10	28	GTGA	cis-regulatory elements within the promoter
TaGSTU3 gene	AJ414701.1	Sp1	16	CC(G/A)CCC	light responsive element
		GATABOX	11	GATA	Molecular light switches for plant genes
		Erd1	33	ACGT	required for etiolation-induced DE expression of erd1
		CACTFTPPCA1	9	YACT	mesophyll-specific gene expression inthe C4 plant
TaGSTF6b gene	AJ441055.1	WRKY71OS	5	TGAC	encodes a transcriptional repressor of the gibberellin signaling pathway in aleurone cells
		SORLIP5AT	32	GAGTGAG	involved in the network of phytochrome A-regulated gene
		REALPHALGLHCB21	55	AACCAA	closely defined phytochrome regulatory elements
		CACTFTPPCA1	34	YACT	key component of Mem1
		CURE	47	GTAC	copper-sensing signal transduction
TaGSTF4 gene	AJ440793.1	RHERPATEXPA7	5	KCACGW	Root Hair Cell-Specific cis-Element
TaGSTF2 gene	AJ440791.1	CAAT-box	12	CAAT	common cis-acting element in promoter and enhancer regions
		GARE-motif	35	AAACAGA	gibberellin-responsive element
		P-box	43	CCTTTTG	gibberellin-responsive element
TaGSTF1 gene	AJ440796.1	5UTR Py-rich stretch	26	TTTCTTCTCT	cis-acting element conferring high transcription levels
		TCA-element	17	GAGAAGAATA	cis-acting element involved in salicylic acid responsiveness
TaGSTF5 gene	AJ440794.1	MYBCORE	9	CNGTTR	flavonoid biosynthesis; KW leaf; shoot;
		OSE2ROOTNODULE	32	CTCTT	activated in infected cells of root nodules
TaGSTF3 gene	AJ440792.1	ABRE	1	MACGYGB	as Ca2+-Responsive
		CACTFTPPCA1	30	YACT	key component of Mem1
		MYB1AT	43	WAACCA	promoters of the dehydration-responsive gene
		MYBPZM	54	CCWACC	controls pigmentation
TaGSTZ	AF109714.1	WBOXHVISO1	9	TGACT	participates in sugar signaling
		WRKY71OS	10	TGAC	a transcriptional repressor of the gibberellin signaling pathway

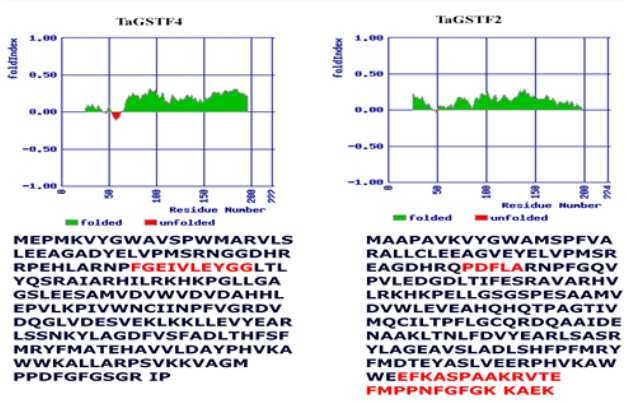

Fig 2. FoldIndex plotted with window size 51, for three GST classes, the plot showed that only TaGSTF2 and TaGSTF2 contain unstructured regions in its native state. Positive and negative numbers represent ordered and non-ordered protein, respectively. Amino acids suggested as being ordered is shown in green and unordered in red respectively.

representation from the cytosolasm. In biological process, GO assignment illustrated that TaGSTU genes involved in oxidation-reduction process, toxin, toxin catabolic process, aromatic amino acid familyand metabolic process. However, no information was available for TaGSTF and TaGSTZ for both cellular and biological GO (Table 3).

Multiple sequence alignment, structural analysis and phylogenetic tree construction

To clarify the relationship among Tau, Phi and Zeta classes of wheat GST, the multiple sequence alignment (MSA) of *T. aestivum*, *O. sativa*, *H. vulgare*, *Z. mays* and *G. max*, (Fig 3A &B) was carried out, MSA revealed a large number of conserved residues in the N-terminal domain of the sequences. The GST fold contains an N'-terminal TRX-fold domain and α-helical domain at C'-terminal. The predicted N'-terminal domain (approximately residues 1 to 80) was classified as part of the thioredoxin superfamily fold. This domain revealed to consists of α/β structure with the folding topology βαβαββα. The core domain was consisting of three layers with the β-sheet sandwiched between α-helices (α/β/α). Furthermore sequences alignment analysis showed that residue at N'-terminal were more conserved as compared to the C'-terminal (86-220 residues) and the three GSTclasses share somewhat same secondary structure. Analysis of the 5' regulatory regions revealed that TaGSTF genes were more divergent (16.90% - 44.2% identity) than TaGSTU (23.08–71.70% identity; Supplementary Table S4a). The coding regions of the TaGSTF gene appeared to be poorly conserved (46.64–77% identity), whereas the coding region for the TaGSTU genes appear to be more conserved (47.93–93%identity; supplementary Table S4b). The analysis of the protein sequences of wheat GST classes revealed higher conservation in the sequences. A 41–99% identity was obtained for the TaGSTU protein sequences (Supplementary Table S5) while 38–79% identity was obtained for TaGSTF (Supplementary Table S6). To evaluate the evolutionary relationships of the wheat GST gene family and GSTs homologous from different plant species, phylogentic study was constructed. Phylogentic analysis of GST has been done earlier with different cereal crops (Mohsenzadeh et al., 2011). In this analysis, we included newly sequenced monocot model plant *Brachypodium distachyon*s – which revealed close evolutionary relationship and better conservation of micro-colinearity between wheat and *Brachypodium* orthologous regions than between wheat and rice. This is in agreement with an earlier study (Bossolini et al., 2007). Phylogenetic analysis of Tau, Phi and Zeta wheat GST formed two distinct clusters. Cluster I comprised of sequences from TauGST and cluster II included sequences from Phi class and both clusters were supported by high bootstrap value (Fig 4A). The Zeta class formed an out group. Phylogenetic analysis of amino acid sequence of GSTs homologous from different plant species revealed four GST families, clustered separately in the phylogenetic analysis. Cluster I include TauGST sequences, Cluster II comprised of sequences from Zeta and Theta-GST and cluster III comprised of PhiGST sequences and both clusters were supported by high bootstrap value (Fig 4B). *B. distachyon* was phylogenetically most closely TaGSTU3 whereas TaGSTU2 was closely related to *O. sativa* (supported by 56% bootstrap value) and shared a common evolutionary path. Comparison of the deduced amino acid sequences of the three GST genes showed that the TaGSTU1A, TaGSTU1B and TaGSTU1C proteins were 96% identical and formed a clade with *H. vulgare*, *Z. mays* and *B. distachyon* supported with 52% bootstrap value. TaGST6B showed relatedness with *H. vulagre* with high bootstrap value of 87% and TaGSTF4 illustrate relatedness with *O. sativa and H. vulagre* Phi whereas TaGSTF2 and TaGSTF1 showed relationship with *Z. mays* Phi. TaGSTF5 showed closeness with *O. sativa* as well as with *Z. mays*, Phi supported by 60% bootstrap value. TaGSTZ was closely related to *O. sativa* which and share a common evolutionary relationship.

Analysis of 3D-structure of Tau, Phi, and Zeta GSTs

The 3D-structure of Tau, Phi and Zeta was predicted by homology modeling. TaGSTU showed 68% identity with PDB code 1OYJ (Crystal Structure Solution of Rice Gst1). Similarly, TaGSTF showed 48% identity with 1BYE (glutathione-S-transferase I from mains in complex with atrazine glutathione conjugate) and Zeta revealed 52% identity with 1E6B (structure of a zeta-class glutathione S-transferase from *A. thaliana*). Sequence alignment between query and template indicated that the catalytic residue, metal binding site and ligand binding site are well conserved in Tau- and Phi-classes as shown in figure 5. Specifically, Ser17 (when aligned with template, the position shifted to 12) of the plant specific Phi and Tau have a crucial role in the catalytic activation of GSH (Fig 6). By using site-directed mutagenesis, the Ser residues have proven catalytically essential in GST catalysis in different organisms (Oztetik, 2008) was well conserved in both template and target. Ramachandran plot showed 94.2%, 91.8% and 90.2%, residues in most favoured regions in Tau, Phi and Zeta respectively. Only in Zeta-class, 1.6 residues were in disallowed regions (Table 4).

Expression model with wheat GST genes

There are approximately one million wheat EST deposits in the GenBank. These partial length genes encoding sequence provided useful bioinformatics source. To demonstrate the patterns of expression of these GST genes from wheat, we performed gene expression analysis in ten wheat tissues, callus, cell culture, crown, inflorescence, leaf, flower, root, stem, sheath and seed (Fig 5). Our results showed that TaGSTU2C and TaGSTU3 genes were detected in most of the tissues. It is noteworthy that a relatively very high expression level of TaGSTU2C was found in stem tissues and TaGSTU3 in root tissues (Fig 6A). The lowest level of expression was shown in the majority of tissues examined in case of in TaGSTU1A and TaGSTU2. All genes expressed in each one or two tissues were

Table 3. Gene Ontolgy of three classes of TaGST.

	Cellular component	Biological process	Biochemical function
TaGSTU	cytoplasm	response to stress	catalytic activity
	cytosol	oxidation-reduction process	protein binding
		response to toxin	oxidoreductase activity
		toxin catabolic process	transferase activity
		aromatic amino acid family metabolic process	isomerase activity
		response to stress	peroxidase activity
		oxidation-reduction process	glutathione transferase activity
		response to toxin	catalytic activity
		toxin catabolic process	protein binding
TAGSTF			glutathione transferase activity
			transferase activity
TaGSTZ			glutathione transferase activity
			transferase activity

Fig 3. Multiple sequence alignment using amino acid sequence of (A) TaGSTU and (B) TaGSTF and (C) TaGSTZ from different cereal crop. Conserved amino acids are shown in black boxes. Protein structural features are indicated above the alignment. Alpha helix and beta strands are elements of secondary structure represented as rods and arrow. The GST sequence comprised of N'-terminal TRX-fold domain and C'-terminal α-helical domain.

showing much higher levels than in other tissues with the exception that TaGSTU3 preferentially expressed in callus, stem and root. In phi-GST, the expression level was only in three or four tissues for TaGSTF2 and TaGSTF3. The TaGSTF5 and TaGSTF4 showed expression in nine tissues while TaGSTF4 showed equal expression in sheath and root (Fig 6B). However, higher expression of TaGSTF6B was detected in 10 tissues with highest expression in sheath, root, cell culture and flower.

Discussion

Plants have to respond to biotic and abiotic stresses and environmental stimuli during their growth and developmental periods. Precise gene regulations play essential roles in these processes. Plant gene expressions display accurate location, when and where the exact expression occurs is well regulated. GSTs are multifunctional proteins involved in diverse intracellular events such as primary and secondary metabolisms, stress metabolism, herbicide detoxification and plant protection against ozone damages and heavy metals. To date, the crystal structures of over 200 soluble GSTs, present in the main classes of plants, animals and bacteria have been resolved. At present, structural information about plant GSTs is available for phi GST from *Arabidopsis* (Reinemer et al., 1996), a tau class from wheat (Thom et al., 2002), rice GST1 (OsGSTU1) (Dixon et al., 2003) and maize (Neuefeind et al., 1997), and for a zeta-class GST from *Arabidopsis* (Thom et al., 2001). In this study we have extended the comprehensive role of phi, theta and zeta GSTs in wheat especially. Motif discovery operation was performed separately through motif discovery tool MEME on GST classes. The N'-terminal domain was less conserved than the C'-terminal domain (Offen et al., 2006; Shao et al., 2005). Thus, C-terminal motif was selected for constructing position weight matrix. A stretch of nucleotide, CGAGTCCCTCATCAT and GTGAAGGCCTGGTGG was the strongly conserved between TaGSTU and TaGSTF sequences with bigger log odd score. The information content is a measure of the degree of conservation in motif sequences, with lower information content corresponding to a higher degree of conservation, and vice versa (Hertz and Stormo 1999; Schneider and Stephens, 1990). Using position weight matrix and information content of the Tau and Phi motifs, we identified a higher degree of conservation at the start and end position of the in the motif sites (value from 0-0.6), which is important as it might be the splice site junction between the exon and intron with conserve start and end sites (Lim and Burge , 2001) . Finally the consensus sequences obtained for TaGSTU and TaGSTF on the basis of the log odd score further used for determining other regulatory motifs such as transcription factor binding site, splice site and signal recognition site. *cis*-acting regulatory elements analysis was performed to identify the presence of several *cis*-acting regulatory elements, which are known to be involved in transcription regulation of plant genes and which could possibly be controlling the expression of the GST genes in wheat .This region was used in present study as no introns were detected upstream of the translational start site of the studied genes, and thus possible regulatory information within the 5' untranslated region (5' UTR) be identified. Several functional significant *cis*-acting regulatory elements that are associated with plant development, hormonal regulation and stress response were identified in the upstream region of the GST gene families. The names of the identified *cis*-acting elements and their predicted functions are summarized in Table 2. CAAT boxes as well as other potentially important *cis*-acting regulatory elements that are required for minimal promoter GST activity in plants. MYB and Erd1 family motifs *cis*-elements have been reported to be required for the binding of transcriptional factors that are required for drought inducible gene expression. MYB-proteins play crucial roles in cell proliferation and differentiation (Luscher and Eisenman, 1990). W-boxes have been found to interact with zinc-fingered WRKY proteins; a family of TFs that are involved in response to pathogen infection and other plant stresses (Ross et al., 2007). The CAAACAC motif is conserved in many storage protein gene promoters and important for high activity of the napA promoter (Stalberg et al. 1996). ABRE is also an example of an extended *cis*-element that contain the G-box core sequence [(C/A) ACG (T/C)G(T/C/G)] (Busk and Pages, 1998). Garciarrubio et al. (1997) found that ABA affects the germination of mature seeds and growth of young seedlings of Arabidopsis by limiting the availability of nutrients. It might be possible that ABA may regulate the expression of GST gene, required for cellular and metabolic activities. Gibberellin responsive element (GARE) sequence plays a fundamental function in GA action because changes within this sequence caused a huge reduction in GA-driven gene expression (Gubler and Jacobsen, 1992). GARE have been found in most genes responsible for synthesis, transport and signaling of other hormones, but the role of GARE is yet to be clarified, although uncharacterized crosstalk between GA and other hormones have been proposed (Ogawa et al., 2003). Plant genes which are induced by one or more forms of stress (Goldsbrough et al., 1993). TCA, a *cis*-acting regulatory element involved in salicylic acid responsiveness is known to be present in the non-translated regions of many monocot and dicot plant genes which are induced by one or more forms of stress (Goldsbrough et al., 1993). GATA-box and Sp1motif is known as light responsive cis-element (LREs) are found in the regulatory region of light-regulated genes, apparently essential for light-controlled transcriptional activity (Lamm and Chua, 1989; Gilmartin et al., 1990). Deletion or mutation of these LRE sequences compromises the ability of the regulatory region respond to light and ultraviolet radiation. Presence of these sequence elements in wheat GSTU3 gene regulatory regions demonstrate that GSTU3 gene expression may be well regulated by light, whereas 5'-UTR Py-rich stretch, *cis*-acting element conferring high transcription levels. CTCTT, one of the consensus sequence motifs of organ-specific elements (OSE) is characteristic of the promoters activated in infected cells of root nodules and in the arbuscule-containing cells of mycorrhizal roots (Vieweg et al., 2004) RHERPATEXPA7 was the root-hair-specific *cis*-element (Murray et al., 2007) in angiosperms, whereas SORLIP5AT and REALPHALGLHCB21 were *cis*-regulatory elements involved in the network of phytochrome A-regulated gene (Degenhardt and Tobin, 1996). WRKY71OS encodes a transcriptional repressor of the gibberellin signaling pathway in aleurone cells (Gomez-Cadenas et al, 2001). Family members of WRKY transcription factors appear to be involved in the regulation of various physiological programs that are unique to plants, including pathogen defense, senescence, trichome development plant growth and development. The rice WRKY gene superfamily has also been implicated in the regulation of abscisic acid signaling in aleurone cells (Xie et al., 2005). DPBFCOREDCDC3 encodes a bZIP transcription and ARE cis-acting regulatory element essential for the anaerobic induction. The CuRE motif has been associated with genes that require copper for their expression, through copper-response elements associated with them (Quinn et al., 2000). These include photosynthetic genes such as: plastocyanin; an electron transfer protein in the thylakoid lumen, plastid CuZn-superoxide dismutase, polyphenol oxidase; also in the thylakoid lumen, mitochondrial cytochrome oxidase, extracellular

Table 4. Ramachandran's plot statistics of models generated by MODELLER after energy minimization.

Server	Ramachandran's plot statistics	Glutathione Transferases		
		Tau	Phi	Zeta
Modeller9v8	Residues in most favored region	94.2%	91.8%	90.2%
	Residues in additionally allowed region	3.7%	6.7%	7.1%
	Residues in generously allowed region	1.6%	1.5%	1.1%
	Residues in disallowed region	0.0	0.0.%	1.6%

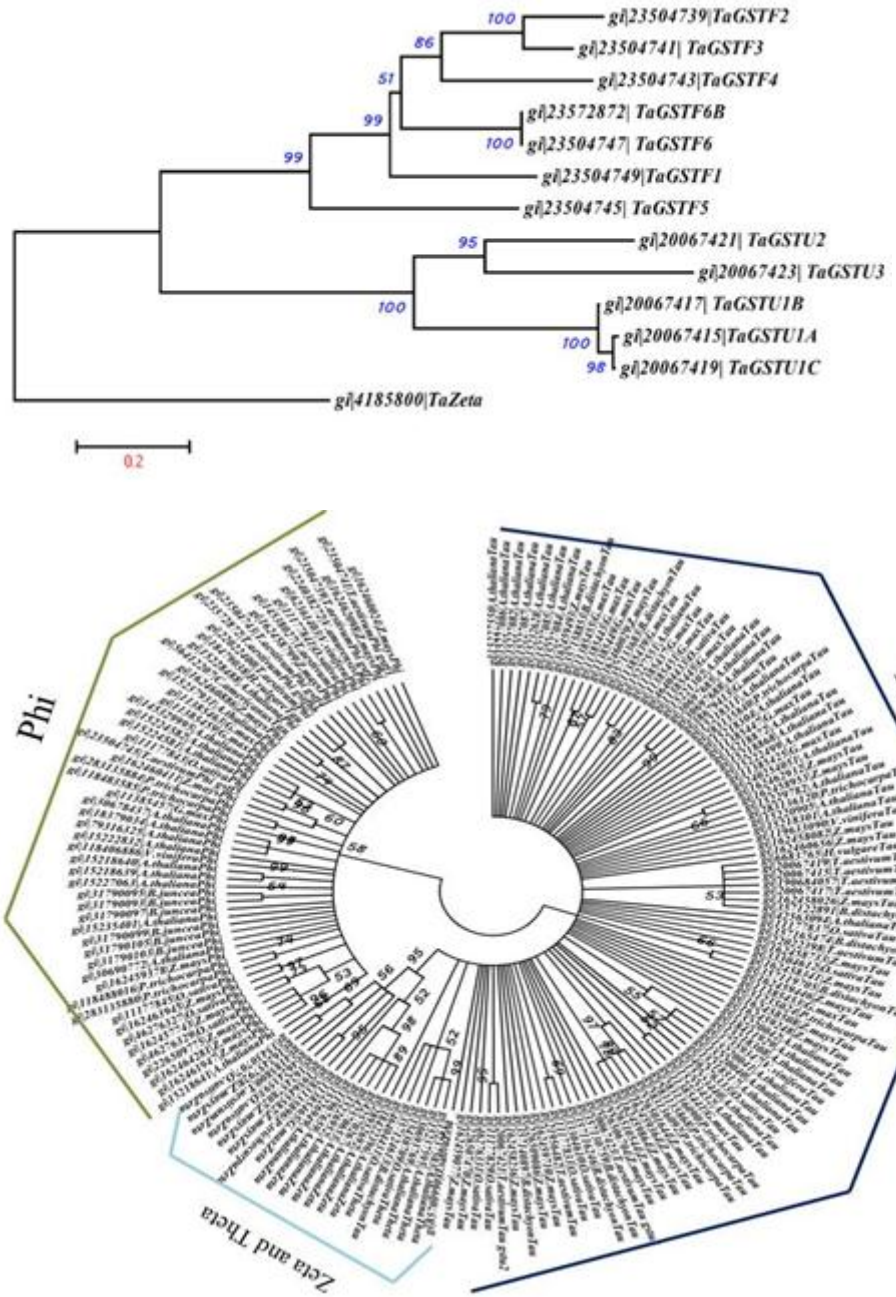

Fig 4. Phylogenetic trees of GST genes studied in this work. (A) Phylogenetic analysis among the Tau, Phi and Zeta of wheat GST showed two large clusters. Cluster I containing of sequences from TauGST (GSTU) and cluster II included sequences from Phi class (GSTF) and Zeta formed an out group. Bootstrap values are indicated against each branch; (B) Phylogenetic analysis of Tau, Phi and Zeta class of GST homologues from different plant species constructed by the Neighbor-joining method using the MEGA 4 program with bootstrap values are indicated against each branch.

laccase; involved in lignifications, and several oxidases (Hansch and Mendel, 2009). High numbers of stress associated *cis*-elements were found in all GST gene families in wheat. In present investigation, stress associated *cis*-elements were found highest in TaGSTF6b and TaGSTU3 and lowest in TaGSTU1B and TaGSTF5; (Table 4). These observations infer that the expression of GST might be well linked to and regulated by environmental stress conditions and could potentially represent good candidates for further systematic experimental evaluations. Though these sequence patterns require experimental validation nevertheless, our current findings may open new avenues for studying the regulation of gene expression in cereal crops. The folded or unfolded structure of wheat GST protein has not been studied clearly till now. It is now possible to predict extended disordered regions in gene sequences with a high degree of accuracy (Prilusky et al., 2005). In this study, we noticed highest percentage of disordered residue was noticed in TaGSTF2 followed by TaGSTF4 (Fig 1). However, other two classes of GST were completely stable. The results obtained in this investigation could be used to estimate the local and general probability for the provided protein sequences, under specific conditions to fold.These differences of folding state amongst GST classes are our future research targets. Estimating the amount of protein distributed in each subcellular location is essential for quantitative understanding and modeling of protein dynamics and to study their effect on cell behaviours (Coelho et al., 2010). Sub-cellular localization result showed mitochondrial targeting peptide in TaGSTU3 and chloroplast-targeting peptide in case of Zeta-GST, whereas, Phi family GST members lack such kind of peptide signal , though differential mRNA splicing leads to the majority of the derived transcripts lacking this signal, giving cytosolic protein (Thatcher et al., 2007). Alternatively, it is also possible that these GSTs associate with different sub-cellular fractions as a consequence of their association with other proteins or substrates and as such an understanding of their sub-cellular localization could be very important in unraveling their endogenous function . Protein solubility is an important pre-requisite for structural and biophysical studies. This will help in selection of proteins with high solubility based on a statistical solubility model of in *E. coli,* has been a successful approach to expressing a target insoluble protein in soluble form as part of a fusion protein (Harrison, 1999) we showed that TaGSTU and TaGSTZ has less than 60%, chance of solubility when over-expressed in *E. coli* ,while Phi has more than 60% chance.

The calculated isoelectric point (pI) for these three classes of wheat GST will be useful in purification and separation, because proteins are separated according to their pI using isoelectric focusing method. The proteins are insoluble at pI and will be soluble both below and above this point help in predicting the pH of the elution buffer system in purification process (Bjellqvist et al., 1982). Under the GO Annotation system at the highest level, there are three major gene properties: cellular component, molecular function, and biological process. Our result showed that the TaGSTZ was cellular, molecular and biologicallywell annotated as compared to TaGSTU and TaGSTF . Analysis of non-coding and coding region of the GST classes revealed TaGSTF genes were highly divergent (16.90% - 44.2% identify) than TaGSTU (23.08–71.70% identity,) in non-coding region whereas coding regions of the TaGSTF gene appeared to be poorly conserved (46.64–77%identity) with respect to TaGSTU genes (47.93–93%identity). The high divergence identity witnessed in the 5' regulatory regions of the GST gene families in wheat could be as a result of loss of common *cis*-elements from the ancestral GST gene. This is in agreement with the degenerative complementation model, which proposes that after gene duplication each daughter gene keeps only a fraction of regulatory elements as compared to the ancestral origin (Lynch and Force, 2000). Protein sequences analysis of the three classes of wheat GST in this study revealed higher conservation in the sequences of TaGSTU than TaGSTF. Structural alignment of amino acid sequence of *T. aestivum, O. sativa, H. vulgare, Z. mays* and *G. max*, revealed high variable C'-terminal, the large domain (86-220 residues) with specific substrate-binding site (H-site) was completely helical. This implied that differences in the C'-terminal domain might be accountable for differences in substrate specificity between the GST classes (Wilce and Parker, 1994). The plant GST family presents a conundrum for functional genomics. The genome and EST databases have allowed us to classify GSTs and study their evolution and sequence diversity, while crystallographic studies have provided powerful insights into their structural biology. Phylogenetic tree shows relationship between selective sequences of GSTs. The phylogenetic tree shows a distinct class-specific clustering of sequences, indicating clear class specific-sequences difference. The *T. aestivum* sequences were showing similarity with *O. sativa* and *Z. mays* in all classes of GST. However, high sequence similarity was noticed in the Zeta-GST of plant species. *Brachypodium distachyon*, a newly sequenced plant genome showed close evolutionary relationship with *T. aestivum*. A positive correlation was observed between the phylogenetic relationship of members of the GST super family. The 3D-structure of Tau, Phi and Zeta was predicted by homology modeling. Three-dimensional structure of the GST is essential for defining the active site and also for designing, improving, and docking of small ligands to the complex target protein. In the absence of the experimental structure, this model will provide foundation for elucidating structure and functional relationship amongst wheat GSTs. The expression levels of *GST* genes in various tissues were estimated by assessing the relative number of ESTs per tissue type (Audic and Claverie, 1997). No EST information was available for TaGSTZ. The TaGSTF5 and TaGSTF4 showed expression in nine tissues while TaGSTF4 showed equal expression in sheath and root. Moreover, the expression was the highest for TaGSTF4 expression detected in 10 tissues with highest expression in sheath, root, cell culture and flower. They are useful for gene discovery and in the analysis of genes expression patterns (Clifton et al., 2009). The result revealed that GST genes of different classes from wheat were controlled strictly in different organ and growth stages. Therefore, we speculated that these GST may play an important role in growth and development in wheat. The molecular characterization of GSTs in wheat using experimental approaches is in progress in our laboratory.

Materials and methods

Source of data

Nucleotide (ntd) sequences for different GST subclasses were retrieved from the NCBI (National Center for Biotechnology Information) nucleotide databank (http://www.ncbi.nih.nim.gov/). More sequences of wheat GST gene families were collected by, using the Basic Local Alignment Tool (BLASTN; http://blast.ncbi.nlm.nih.gov/Blast.cgi). Complete ntd of Tau (TaGSTU), Phi (TaGSTF), and Zeta (TaGSTZ) GST protein sequences of *T. aestivum* were considered for BLAST homology searches against the plant species.

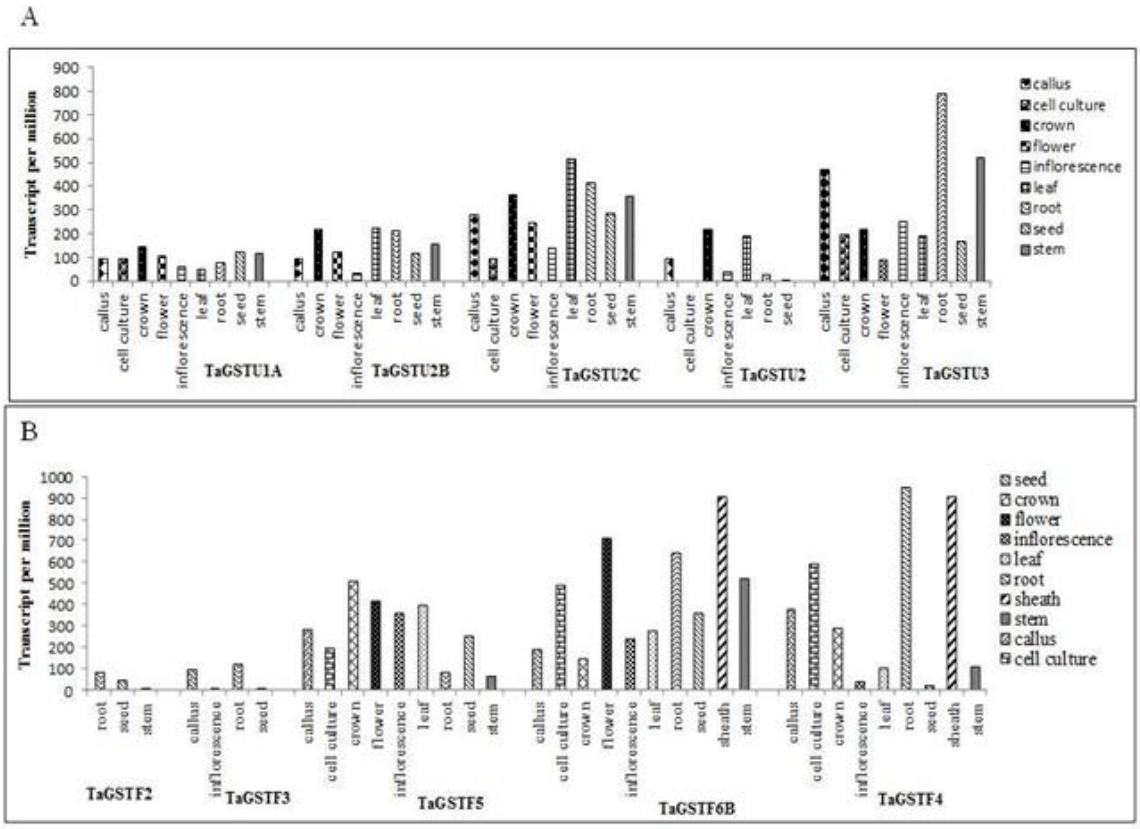

Fig 5. Tissue-specific expression of the (A) TaGSTU and (B) TaGSTF genes of *T. aestivum*. The expression profile was determined by analyzing the EST counts based on UniGene of T. aestivum each GST gene. Gene expression is enumerated as transcripts per million (TPM).

Conserved Motif Scanning, Position weight matrices, Information Content

MEME (http://meme.sdsc.edu/meme/intro.html) program was used to screen a single conserved motif in the GST classes with default parameters except that the length of the motif was set more stringently to vary between 8 and 15 bp to limit the lengths of detected motifs. For each motif, we constructed corresponding position weight matrix (PWM). The consensus sequence for a PWM was constructed by choosing the nucleotide with the highest frequency of occurrence at each position. Log-odd score for each motif sequence was calculated by usual method of taking logarithm to base 2 of ratio of frequencies at each position of the matrix against a background frequency of 0.25 and to study the variability at each position. Each position in the motif has an associated information content that can be derived from the frequencies of the 4 ntd, using information theory. For a random variable that can take on four values with four probabilities, pi, the information content is - $\sum pi \lg pi$. Information content can be used to study the variability (or level of conservation) at each position, varying from 0 (one possibility is certain) to 2 (all possibilities are likely equal). All the above calculation was done with the help of Perl script generated in the lab. Given a PWM of a given motifs, a log-odds scoring matrix was constructed to score the motifs.

Identification of the 5' regulatory region of three classes of wheat GST genes

Protein sequences of TaGSTU, TaGSTF, and TaGSTZ were retrieved through National Center for Biotechnology Information (NCBI, http://www.ncbi.nlm.nih.gov/) was used for computational analysis. Each GST gene of wheat were scanned for the presence of putative *cis*-acting regulatory elements identical with or similar to the registered in Plant CARE (http://bioinformatics.psb.ugent.be/webtools/plantcare/cgi-bin/CallMat IE55.htpl), and PLACE (http://www.dna.affrc.go.jp/PLACE/signalscan.html).

Protein characterization and Sub-cellular localization

Based on program of ExPASy, (http://web.expasy.org/compute_pi/) theoretical pI was calculated for all the classes of wheat GST. The folding states were predicted by FoldIndex program (http://bioportal. weizmann.ac.il/fldbin/findex). Three different protein targeting prediction programs were used to estimate the putative subcellular locations of the candidate proteins: IPSORT (http://hc.ims.u-tokyo.ac.jp/iPSORT/), TARGETP (http:// www.cbs.dtu.dk/services/TargetP/), and PREDOTAR (http://urgi.versailles.inra.fr/predotar/). Proteins solubility was predicted at web server of University of Oklahoma using the statistical model assuming the protein is being over expressed in *coli* (http://biotech.ou.edu/).

Gene ontology

Gene ontology (GO) annotation was performed by using AmiGO (http://www.geneontology.org/) on different classes of GST genes to explore their biological process, cellular component and molecular functions.

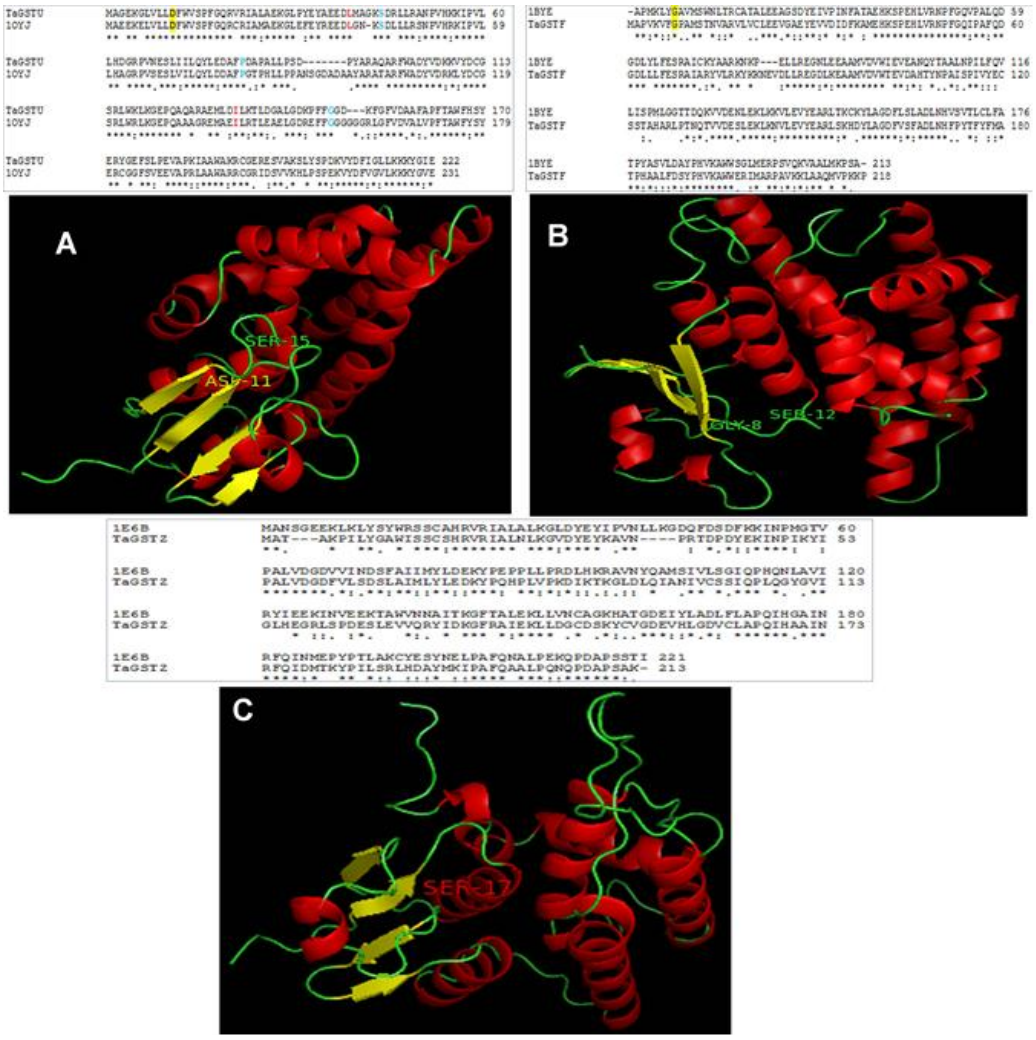

Fig 6. (A) Pairwise alignment of Tau from *T. aestivum* and the PDB ID: 1OYJ. (B) Phi GST and the PDB ID: 1BYE, (C) Zeta GST and the PDB ID: 1E6B. Dash represents insertion and deletion; conserved residues involved in metal binding are in blue, ligand binding residue are in red and catalytic residue are in yellow. Ser residues are been labeled in the protein structure proven catalytically essential in GST catalysis. Alpha helix and beta strands are elements of secondary structure represented as rods and arrow.

Multiple sequence alignment and phylogenetic analysis

To study the phylogenic analysis, tree was generated using homologous GST sequences**.** The percentage of amino acid sequence identity was determined using the PRANK (Loytynoja and Goldman, 2010) and T-Coffee program (Notredame et al., 2000)**.** When necessary, the alignments were manually adjusted using the GeneDOC program (Nicholas et al., 1997). Phylogenetic trees were constructed using the neighbor-joining (NJ) distance method (Saitou and Nei, 1987) and diagrams of phylogenetic trees were drawn with MEGA v4.0 software (Tamura et al., 2007). Evolutionary distances were calculated following the Poisson correction evolutionary model, and the excessive gaps/ missing data were handled using the pair-wise delete option.

Comparison between 3D structures of the Tau, Phi, Zeta GSTs of wheat

The protein sequence of wheat GST classes was subjected to Modeller9v8 software (http://www.salilab.org/modeller/) for 3D model prediction. The minimization of the modeled structure was carried out by GROMOS96 software, incorporated in Swiss Pdb Viewer (Guex and Peitsch, 1997). The stereo-chemical quality and accuracy of the predicted model was evaluated using Ramachandran plot calculated via Procheck (Laskowski et al., 1993). Finally the model was visualized using pymol (DeLano, 2004).

Expression model analysis

The wheat UniGene sets, derived from over one million ESTs in the NCBI GenBank offer a platform for identifying differentially expressed genes in wheat tissues. Using a data mining tool known as Digital Differential Display (DDD), the expression profile was determined by analyzing the EST counts based on UniGene of *T. aestivum* GST classes for the various tissues (http://www.ncbi.nlm.nih.gov/UniGene/UGOrg.cgi?TAXID=4565) on the basis of transcript per million (Eujayl and Morris, 2009). EST profiles show approximate gene expression patterns as inferred from EST counts and the cDNA library sources. The EST collection is derived from a variety of different cDNA libraries, capturing genes expressed in different tissue types and developmental stages or expressed during pathogen-elicited responses**.** Libraries known to be normalized, subtracted, or otherwise biased were removed, but for a variety of reasons,

EST counts may not be a true indication of gene activity. The EST expression profile was calculated by

Transcript per million=Gene EST/ ∑Total EST in pool

Conclusion

Discovery of cis-regulatory elements in gene promoters is a highly challenging research issue in computational molecular biology. This computational analysis focuses on non-coding regions of GST genes and numbers of putative motifs at the 5' regulatory regions were identified. This method provides a competent approach indicating if a gene might be regulated by a given TF, and a structure for ongoing analysis of the transcriptional gene regulation network. Though, transcription is affected by joint involvement of many regulators that affects the RNA Polymerase II selectivity and its binding, in response to various environmental stimuli (Zhu, 1996; Lloyd et al., 2001). More significantly, apparent and full knowledge of the interaction between *cis*-acting regulatory elements and their related TFs would allow a better understanding of transcriptional gene regulation. It also make a foundation for restructuring of gene regulatory networks which in turn will help to improve biological research and provide ways to study and model cell response to various stimuli. The present study gave a detailed insight into the structural and molecular evolution of glutathione in wheat. With the assistance of a well-defined structure and annotations, we can predict protein functional and binding sites, which can help in understanding what biological role it fulfils. We used homology modeling to solve the 3D structure of GST, an important detoxification enzyme from wheat. Studies on the phylogeny and expression patterns of Tau, Phi and Zeta GSTs in wheat are becoming one of the fertile research areas, which could help us to make clear the molecular genetic basis for the wheat genetic improvement and also provide the functional gene resources for transgenic research.

Acknowledgements

The authors thankfully acknowledge the financial support for Agri Bioinformatics Promotion Program by Bioinformatics Initiative Division, Department of Information Technology, Ministry of Communications & Information Technology and Indian Council of agricultural research (ICAR grant No. DWR/RP/10-5.3) Government of India, New Delhi.

References

Andrews CJ, Skipsey M, Townson JK, Morris C, Jepson I, Edwards R (1997) Glutathione transferase activities toward herbicides used selectivelyin soybean. Pestic Sci 51: 213-222.

Armstrong R N (1997) Structure, catalytic mechanism, and evolution of the glutathione transferases, Chem Res Toxicol. 10: 2-18

Audic S, Claverie JM (1997) The significance of digital gene expression profiles. Genome Res. 7: 986–995

Bjellqvist B, Righetti PG, Gianazza E, Gorg A, Westermeier R, Postel W (1982) Isoelectric-focusing in immobilized pH gradients - principle, methodology and some applications. J Biochem Biophys Methods. 6:317-339

Blackburn AC, Coggan M, Tzeng HF, Lantum H, Polekhina G, Parker MW, Anders MW, Board PG (2001) GSTZ1d: a new allele of glutathione transferase zeta and maleylacetoacetate isomerase. Pharmacogenetics 11: 671-678

Board PG, Coggan M, Chelvanayagam G, Easteal S, Jermiinm LS, Schulte GK, Danley DE, Hoth LR, Griffor MC, Kamath AV, Rosner MH, Chrunyk BA, Perregaux DE, Gabel CA, Geoghegan KF, Pandit J (2000) Identification, characterization, and crystal structure of the omega class glutathione transferases. J Biol Chem. 275: 24798-24806

Bonjean AP, Angus WJ (2001) The world wheat book: a history of wheat breeding. Intercept, Andover Wilce MCJ, Parker MW (1994)

Bossolini E, Wicker T, Knobel P A, Keller B (2007) Comparison of orthologous loci from small grass genomes *Brachypodium* and rice: implications for wheat genomics and grass genome annotation. The Plant J. 49:704-717

Buetler TM, Eaton DL (1992) Glutathione S-transferases: amino acid sequence comparison, classification and phylogenetic relationship. J Environ Sci Health C. 10: 181–203.

Busk PK, Pages M (1998) Regulation of abscisic acid induced transcription. Plant Mol Biol 37: 425-435

Clifton SW, Mitreva M (2009) Strategies for under taking expressed sequence tag (EST) projects. Methods Mol Biol. 533: 13-32

Coelho LP, Peng T, Robert FM (2010) Quantifying the distribution of probes between subcellular locations using unsupervised pattern unmixing. Bioinformatics 26: i7-i12

Dean JV, Gronwald JW, Eberlein CV(1991) Induction of glutathione-S- transferase isozymes in sorghum by herbicide antidotes. Plant Physiol. 92:467-473

Degenhardt J, Tobin EM (1996) A DNA binding activity for one of two closely defined phytochrome regulatory elements in an Lhcb promoter is more abundant in etiolated than in green plants. Plant Cell 8:31–41

DeLano WL (2002**)** The PyMOL Molecular Graphics System**.** San Carlos, CA: DeLano Scientific

Dixon DP, Davis BG, Edwards R (2002) Functional divergence in the glutathione transferase superfamily in plants. Identification of two classes with putative functions in redox homeostasis in *Arabidopsis thaliana*. J Biol Chem. 277: 30859-30869

Dixon D, Cole DJ, Edwards R (1997) Characterization of multiple glutathione transferases containing the GST I subunit with activities toward herbicide substrates in maize (*Zea mays*). Pestic Sci. 50: 72–82

Dixon DP, McEwen AG, Lapthorn AJ, Edwards R (2003) Forced evolution of a herbicide detoxifying glutathione transferase. J Biol Chem. 278: 23930-23935

Droog FNJ, Hooykaas PJJ, van der Zaal EJ (1995) 2,4-Dichloro-phenoxylaceticacid and related chlorinated compounds inhibit two auxin regulated type-III tobacco glutathione-S-transferases. Plant Physiol. 107:1139-1146

Durrant WE, Rowland O, Piedras P, Hammond-Kosack KE, Jones JDG (2000) cDNA-AFLP reveals a striking overlap in race-specific resistance and wound response gene expression profiles. Plant Cell 12:963–977

Edwards R, Cole DJ (1996) Glutathione transferases in wheat (*Triticum*) species with activity toward fenoxaprop-ethyl and other herbicides. Pestic Biochem Physiol. 54: 96-104

Edwards R, Dixon DP, Walbot V (2000) Plant glutathione -S-transferases: enzymes with multiple functions in sickness and in health. Trends Plant Sci. **5**: 193-8

Esmaeili M, Shahrtash M, Moosavi F, Mohsenzadeh S, Mohabatkar H (2009) Plant glutathione S-transferase function. The 6th National Biotechnol. Congress of Iran.

Eujayl I, Morris C (2009) Identification of differentially expressed Uni-Genes in developing wheat seed using Digital Differential Display. J Cereal Sci. 49: 316-318

Franklin AE, Cande WZ (1999) Nuclear organization and chromosome segregation. Plant Cell 11:523–534

Frova C (2006) Glutathione transferases in the genomics era: New insights and perspectives. Biomolecular Eng. 23:149-169

Garciarrubio A, Legaria JP, Covarrubias AA (1997) Abscisic acid inhibits germination of mature Arabidopsis seeds by limiting the availability of energy and nutrients. Planta. 203:182-187.

Gilmartin PM, Sarokin L, Memelink J, Chua NH (1990) Molecular light switches for plant genes. Plant Cell 2:369–378

Goldsbrough AP, Albrecht H, Stratford R (1993) Salicylic acid-inducible binding of a tobacco nuclear protein to a 10 bp sequence which is highly conserved amongst stress-inducible genes. Plant J. 3:563-571

Gomez-Cadenas A, Zentella R, Walker-Simmons MK, Ho TH (2001) Gibberellin/abscisic acid antagonism in barley aleurone cells: site of action of the protein kinase PKABA1 in relation to gibberellin signaling molecules. Plant Cell. 13:667-79.

Gubler F, Jacobsen J V (1992) Gibberellin-responsive elements in the promoter of a barley high-pl α-amylase gene. Plant Cell 4:1435-1441

Guex N, Peitsch MC (1997**)** SWISS-MODEL and the Swiss-Pdb Viewer: an environment for comparative protein modeling. Electrophoresis 18:2714-2723

Halliwell B, Gutteridge JMC (2006) Free Radicals in Biology and Medicine, Ed 4. Clarendon Press, Oxford.

Hansch R, Mendel RR (2009) Physiological functions of mineral micronutrients (Cu, Zn, Mn, Fe, Ni, Mo, B, Cl). Curr Opin Plant Biol. 12 :259–266

Harrison RG (1999) Recombinant protein solubility prediction. University of Oklahoma.

Hayes JD, Flanagan JU, Jowsey IR (2005) Glutathione transferases. Ann Rev Pharm Toxicol. 45:51-88

Hertz GZ, Stormo GD (1999) Identifying DNA and protein patterns with statistically significant alignments of multiple sequences. Bioinformatics.15:563–577

Hunatti AA, Ali BR (1990) Glutathione S-transferase from oxadiazon treated chickpea. Phytochemistry 29: 2431-2435

Lam E, Chua NH (1989) ASF-2: a factor that binds to the cauliflower mosaic virus 35S promoter and a conserved GATA motif in Cab promoters. Plant Cell 1:1147–1156

Lamoureux G L, Gouot J M, Davis D G, Rusness D G(1981) Pentachloro-nitrobenzene metabolism in peanut. 3. Metabolism in peanut cell suspension cultures. J Agric Food Chem .29: 996-1002

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

Lim LP, Burge CB (2001) A computational analysis of sequence features involved in recognition of short introns. Proc Natl Acad Sci. 98:11193–11198

Lloyd G, Landini P, Bushby S (2001) Activation and repression of transcription initiation in bacteria. Essays Biochem. 37:17–31

Luscher B, Eisenman RN (1990) New light on Myc and Myb. Part II. Myb. Genes Dev.4:2235–2241

Lynch M, Force A (2000) The probability of duplicate gene preservation by subfunctionalization. Genetics.154:459-473

Mohsenzadeh S, Esmaeili M, Moosavi F, Shahrtash M, Saffari B and Mohabatkar H (2011) Plant glutathione S-transferase classification, structure and evolution. Afr J Biotechnol 10:8160-8165

Murray JD, Karas BJ, Sato S, Tabata S, Amyot L, Szczyglowski K (2007) A cytokinin perception mutant colonized by *Rhizobium* in the absence of nodule organogenesis. Science 315: 101–104

Neuefeind T, Huber R, Dasenbrock H, Prade L, Bieseler B (1997) Crystal structure of herbicide-detoxifying maize glutathione-*S*-transferase-I in complex with lactoylglutathione: evidence for aninduced-fit mechanism**.** J Mol Biol. 274: 446-453

Nicholas KB, Nicholas HB, Jr, Deerfield DW II (1997) GeneDoc: analysis and visualization of genetic variation, EMBnet News 4 1–14

Notredame C, Higgins DG, Heringa J (2000) T-Coffee: A novel algorithm for multiple sequence alignment. J Mol Biol. 302: 205-217

Offen W, Martinez-Fleites C, Yang M, Kiat-Lim E, Davis BG, Tarling CA, Ford CM, Bowles DJ, Davies GJ (2006) Structure of a flavonoid glucosyltransferase reveals the basis for plant natural product modification. EMBO J. 25: 1396–1405

Ogawa M, Hanada A, Yamauchi Y, Kuwahara A, Kamiya Y, Yamaguchi S (2003) Gibberellin biosynthesis and response during Arabidopsis seed germination. Plant Cell 15:1591–1604

Prilusky J, Felder C E, Ben-Mordehai T Z, Rydberg EH, Man O, Beckmann JS, Silman I, Sussman JL (2005) FoldIndex: a simple tool to predict whether a given protein sequence is intrinsically unfolded. Bioinformatic 21: 3435–3438

Qui P (2003) Computational approaches for deciphering the transcriptional regulatory network by promoter analysis. Bio silico 1:125–133

Quinn JM , Barraco P, Eriksson M, Merchant S (2000) Coordinate copper- and oxygen-responsive Cyc6 and Cpx1. J Biol Chem. 275:6080–6089

Reinemer P, Prade L, Hof P, Neuefeind T, Huber R, Zettl R, Palme K, Schell J, Jkoelln I, Bartunik H, Bieseler B (1996) Three-dimensional structure of glutathione S-transferase from *Arabidopsis theliana* at 2.2 A resolution: structural characterization of herbicide conjugating plant glutathione S-transferases and a novel active site architecture**.** J Mol Biol. 255: 289-309

Riechmann JL , Heard J, Martin G, Reuber L, Jiang C, Keddie J, Adam L, Pineda O, Ratcliffe OJ, Samaha RR, Creelman R, Pilgrim M, Broun P, Zhang JZ, Ghandehari D, Sherman B K, Yu G-L (2000) Arabidopsis transcription factors: genome-wide comparative analysis among eukaryotes. Science 290: 2105-2110

Romano ML, Stepheson GR, Tal A, Hall JC (1993) The effect of monooxygenase and glutathione S-transferase inhibitors on themetabolism of diclofop-methyl and fenoxaprop-ethyl in barley and wheat. Pesticide Biochem Physiol. 46: 181-189

Ross CA, Liu Y, Shen QJ (2007) The WRKY gene family in rice (Oryza sativa). J Integr Plant Biol. 49:827–842

Rushton PJ , Somssich IE (1998) Transcriptional control of plant genes responsive to pathogens. Curr Opin Plant Biol. 1:311–315

Saitou N, Nei M (1987) The neighbor-joining method: a new method for reconstruction of phylogenetic trees. Mol Biol Evol. 4: 406-425

Schneider TD, Stephens RM (1990) Sequence logos: a new way to display consensus sequences. Nucleic Acids Res. 18:6097–6100

Shao H, He X, Achnine L, Blount JW, Dixon RA, Wang X (2005) Crystal structures of a multifunctional triterpene/flavonoid glycosyl transferase from *Medicago truncatula*. Plant Cell 17: 3141–3154

Singhal SS, Tiwari NK, Ahmad H, Srivastava SK, Awasthi YC (1991) Purification and characterization of glutathione S-transferase from sugarcane leaves. Phytochemistry 30: 1409-1414

Stalberg K, Ellerstom M, Ezcurra I, Ablov S, Rask L (1996) Disruption of an overlapping E-box/ABRE motif abolished high transcription of the napA storage-protein promoter in transgenic *Brassica napus* seeds. Planta.199: 515–51

Tamura K, Dudley J, Nei M, Kumar S (2007) MEGA4: Molecular Evolutionary Genetics Analysis (MEGA) software version 4.0. Mol Biol Evol. 24: 1596–1599

Thatcher LF, Carrie C, Andersson CR, Sivasithamparam K, Whelan J, Singh KB (2007) Differential gene expression and subcellular targeting of Arabidopsis glutathione S-transferase F8 is achieved through alternative transcription start sites. J Biol Chem. 282: 28915–28928

Thom R, Cummins I, Dixon DP, Edwards R, Cole DJ, Lapthorn AJ (2002) Structure of a Tau class glutathione S-transferase from wheat active in herbicide detoxification. Biochemistry. 41: 7008-7020

Thom R, Dixon DP, Edwards R, Cole DJ, Lapthorn A (2001) The structure of a Zeta class glutathione *S*-transferase from *Arabidopsis thaliana*: characterization of a GST with novel active site architecture and a putative role in tyrosine catabolism. J Mol Biol. 308: 949-962

Vieweg MF, Fruhling M, Quant HJ, Baumlein HJ, Pühler A, Küster H, Perlick AM (2004) The promotor of the *Vicia faba* L. hemoglobin gene *VfLb29* is specifically activated in the infected cells of root nodules and of mycorrhizal roots of different legumes and the non-legume *Nicotiana tabacum*. Mol Plant Microbe Inter. 17**:** 62–69

Wilce MCJ, Parker MW (1994) Structure and function of glutathione -S-transferases. Biochem Biophy Acta. 1205:1-18

Wyeth WW, Albin S (2004) Applied bioinformatics for the identification of regulatory elements. Nat Rev Genet. 5:276-287

Xie Z, Zhang Z L, Zou X, Huang J, Ruas P, Thompson D, Shen QJ (2005) Annotations and functional analyses of the rice WRKY gene super family reveal positive and negative regulators of abscisic acid signaling in aleurone cells. Plant Physiol. 137:176–189

Zhu Q (1996) RNA polymerase II-dependent plant in vitro transcription systems. Plant J. 10:185-188