INTERNATIONAL JOURNAL OF AGRICULTURE & BIOLOGY
ISSN Print: 1560–8530; ISSN Online: 1814–9596
18–0688/201x/00–0–000–000
DOI: 10.17957/IJAB/15.0809
http://www.fspublishers.org

Full Length Article

Characterization of Wheat Cell Wall Invertase Genes Associated with Drought Tolerance in Synthetic-derived Wheat

Maria Khalid[1], Zoya Khalid[2,3], Alvina Gul[1,4], Rabia Amir[1] and Zubair Ahmad[5]
[1]*Atta-ur-Rehman School of Applied Biosciences (ASAB), National University of Science and Technology (NUST), Islamabad, Pakistan*
[2]*Department of Biological Sciences and Bioengineering, Faculty of Engineering and Natural Sciences (FENS), Sabanci University, Istanbul, Turkey*
[3]*National University of Medical Sciences, The Mall Road, Rawalpindi, Pakistan*
[4]*School of Integrative Plant Science, Cornell University, Ithaca, NY 14850, USA*
[5]*Crop Science Institute, Maize Program, National Agriculture Research Centre, Islamabad, Pakistan*
[*]For correspondence: alvina_gul@yahoo.com; zoyakhalid@sabanciuniv.edu

Abstract

Cell wall invertase (CWI) gene hydrolyzes sucrose into glucose and fructose that supports pollen development. Down regulation of CWI is responsible for drought induced pollen sterility which ultimately reduces grain yield by limiting grain numbers. In bread wheat, CWI gene has been localized on chromosomes 4A, 5B and 5D which have conserved WECPDF domain. In synthetic derived wheat (SYN-DER) diversity panel, 123 accessions had Hap-4A-C haplotype at *TaCwi-A1* which was significantly associated with 1000 grain weight (TGW) and other agronomic traits under both well-water and water-limited conditions. On the other hand, Hap-5D-C haplotype was fixed at *TaCwi-D1* in synthetic derivatives. Previous studies identified high sequence conservation at *TaCwi-B1*, however sequencing of this gene in diverse SYN-DER identified several mutations putatively transferred from durum parents of synthetic hexaploid wheats. The non-synonymous substitutions observed in *TaCwi-B1* in the conserved domain (WECPDF) were Glu372Lys, Glu372Gly, Pro374Gln, Asp375Thr, while Phe376Leu, Tyr377Thr, Val379Cys variants were observed in the neighboring region. *In silico* analysis revealed that these point mutations sequentially and structurally influenced the biological function of *TaCwi-B1* protein. All the identified mutations caused poor hydrolysis of sucrose followed by improper pollen development which had implications in wheat drought adaptability. In addition, G320C allelic variant was found in high percentage (54%) in SYN-DERs. The association analysis confirmed that SNP *TaCwi-B1-G* enhanced TGW and grain yield in SYN-DERs. Our results significantly enhance the understanding of gene function affecting drought adaptability in wheat. © 2018 Friends Science Publishers

Keywords: Cell wall invertase; Drought; Synthetic wheat; Point mutations; Non-synonymous variants
Abbreviations: CWI, Cell wall invertase; WW, Well water; WL, Water limited; DH, Days to heading; PH, Plant Height; TGW, Thousand grain weight; SOD, Superoxide dismutase; GpS, Grain per spike; SpPS, Spikelets per spike; TN, Tiller number

Introduction

Wheat (*Triticum aestivum* L.) is an important constituent of food security given that ~20% of the calories to world population. It is extensively grown crop globally that adapted to a broad range of environments (Reynolds *et al.*, 2012). Drought is currently the leading threat to wheat based food security systems, and afflicts 42% of global wheat production. This problem will get more severe with the climate change associated with drier environment (Acuna *et al.*, 2015). The prediction demonstrated further increase in population inhabitant up to 8 to 10 billion and drastic increase in rainfed agriculture by 2050 owing to extreme consequences of climate change (De Marsily, 2007). Genetic improvement in drought stress needs sources of traits identification related to drought tolerance and genes introgression underlying target traits to cultivars. The challenge for implementing this strategy in breeding programs is the identification of the most suitable target traits in a time-efficient and cost-effective way for different drought scenarios (Passioura, 2012).

Synthetic hexaploid wheat (SHW) evolved by hybridization of tetraploid wheat *T. turgidum* L. and wild ancestor *Aegilops tauschii* Coss. are widely exploited genetic resources for the improvement of bread wheat by identifying useful genes in primary SHW and introgression of such desirable genes into bread wheat via synthetic derivatives (SYN-DER) or advanced derivatives synthetic backcross

To cite this paper: Khalid, M., Z. Khalid, A. Gul, R. Amir and Z. Ahmad, 201x. Characterization of wheat cell wall invertase genes associated with drought tolerance in synthetic-derived wheat. *Int. J. Agric. Biol*. 00: 000–000

derived lines (SBLs). It is reviewed that 30% of yield improvement is associated with SHW as they contribute in wheat genetic resources for drought tolerance (Trethowan, 2014). Similar findings on performance of SBLs cultivars under rain-fed environments have been reported earlier (Tang *et al.*, 2014). There is possibility that yield advantages in SHW are due to favorable alleles related to yield related traits, demonstrated to be preferentially retained in SBLs (McIntyre *et al.*, 2014: Bhattacharyya *et al.*, 2018).

Wheat genome is very large and has complex vital crop having hexaploid genome with 80% highly repetitive DNA and size containing ~16 Gb (IWGSC, 2014). However, recent advancement in genomics has made it possible to clone and characterize important genes and study their function. Several genes related to drought tolerance have been cloned recently and their function is described (Edae *et al.*, 2013). Cell wall Invertase (CWI) is an enzyme for the development of sink tissue and is associated with carbon partitioning. Water stress makes the pollen sterile and down regulates CWI gene. This gene has function to hydrolyze the sucrose into fructose and glucose, which are important for the pollen development. It plays an important role in crop yield. It is localized at 4A, 5B and 5D chromosome; however, lack of diversity at D genome is the main hindrance to study their mechanism over there (Webster *et al.*, 2012). Strong allele selection occurred in breeding, polyploidization and domestication (Jiang *et al.*, 2014). It was reported that high gene expression of cell wall invertase in Arabidopsis, maize and rice increased the carbon partitioning and CWI activity and improved grain yield in transgenic maize. Hence, it is associated with grain size and number which are the important determinants of grain yield and quality as well (Li *et al.*, 2013). Cell wall invertase was also used in making high yield hybrid maize cultivar by selecting those cultivars having high CWI expression and make inbred lines to increase grain yield in them (Bi *et al.*, 2018).

Protein mutations occur due to various genetic and environmental factors. Non synonymous mutations can cause structural changes that ultimately lead to loss of protein function. In this study, mutations were analyzed if they are affecting the structure and function of CWI protein. *In silico* methods were utilized for analyzing the functionally important point mutations in CWI protein. Both *in silico* sequential and structural features were used to assess the impact of non-synonymous variants. The features like protein stability changes which is a central issue in protein folding and misfolding were analyzed in particular. Additionally, the phylogenetic analysis was also carried out to explore the evolutionary aspects of the protein sequences used in this work.

Materials and Methods

Germplasm

Two hundred and nine advanced lines derived from synthetic hexaploid wheats and elite bread wheat cultivars were used in the present study (Afzal *et al.*, 2017). Synthetic derivatives (SYN-DERs) were developed by primary synthetic hexaploid wheat crossed with advanced lines and improved cultivars from CIMMYT and Pakistan. Initial collection was 800 that were further reduced to 171 by continuous selection in National Agriculture Research Centre (NARC), Pakistan. Additionally, the SHW and bread wheat parents of these advanced lines and durum parents of the SHW were also used to identify the source of alleles.

Phenotyping

All the accessions were evaluated under polyethylene glycol (PEG) induced osmotic stress condition in controlled laboratory environment. Wheat seeds were sterilized by using 1% solution of mercuric chloride solution and placed for 5 min and then washed thoroughly with deionized water. Seeds were germinated first in wet filter papers placed in petri plates for two days. Seedlings were transferred to pot containing peat moss for further growth at room temperature. After 6 days, uniform height seedlings were than transplanted to hydroponic cultured boxes (8 × 8 × 12 cm) containing Hoagland's nutrient solution for 10 days at irradiance of 130 μmol m^{-2} s^{-1}, 22–25°C air temperature and 12-h photoperiod. Hydroponic solution was supplemented with a fresh medium each day and was replaced completely with fresh nutrient solution after 6 days. After sixteen days, seedlings were subjected to osmotic stress by gradually adding PEG 6000. Hydroponic solution was aerated by air pumps (Tetra Blacksburg, 115 V). In case of control condition, roots were maintained in nutrient solution only for comparison purpose. After stress imposition, relative water content (RWC) was measured according to (Turner, 1981) using the following equation:

$$RWC = \frac{(FM - DM)}{(SM - DM)}$$

Where, FM is the fresh mass of leaf, DM is the dry mass after drying leaves at 75°C for 2 h and TM is the turgid mass. Shoot length from the base of the plant to the tip of the main tiller and root length data were recorded.

A field experiment was also setup with an alpha lattice design in two environmental conditions, *i.e.*, well water (WW) and water limited (WL) conditions, details of the experiment are referenced in (Afzal *et al.*, 2017). Accessions were screened during 2014-2015 and 2015-2016 in Barani Agriculture Research Institute (BARI), Chakwal, Pakistan. Chakwal is rainfed area located at 575 m above sea level. Soil is well drained contain moderately fine particles, with pH 7.6, non-saline and slightly calcareous with 0.32 dS/m electrical conductivity (Islam *et al.*, 2013). In WW condition, all genotypes were planted with two replications in field. Three irrigations were provided to WW plants and maintained soil moisture with 100% field capacity till harvesting. In WL conditions, genotypes were panted in polyethylene tunnel

with two replications. Tunnel was supported by iron frame to protect from rain. Irrigation was done up to tillering stage. All genotypes were sown in 2 m length row with 30 cm row spacing in both conditions. Physiological, agronomic and biochemical traits were measured accordingly.

Genotyping

DNA was extracted from all accessions by using CTAB method (Doyle and Doyle, 1987). The KASP primers CWI_4A_1523 and CWI-5D-312 were used to identify haplotypes at *TaCwi-A1* and *TaCwi-D1*, respectively. The primer sequences and related information are provided in Table 2. Primer mixture for 5µL reaction included 2.2 μL of 10-20 ng/μL DNA template, 2.5 μL of 2 x KASP master mixture and 0.056 μL primer mixture. PCR was performed in 384 well (S1000, Thermal cycler, USA). PCR cycling was performed using following protocol: hot start at 95°C for 15 min, followed by ten touchdown cycles (95°C for 20 s; touchdown 65°C-1°C per cycle 25 s) further followed by 30 cycles of amplification (95°C for 10 s; 57°C for 60 s). Extension step was unnecessary as amplicon is less than 120 bp. After PCR amplification, plate was read in BioTek Synergy H1 microplate reader.

Cloning and Sequencing of *TaCwi-B1* Gene

The coding sequence of *TaCwi-B1* was cloned to identify new polymorphism in SYN-DER. In total, 96 SYN-DER accessions with varying RWC were selected for cloning and sequencing. Briefly, the reference cell wall invertase gene sequence (AF030240.1) was used as a query to BLAST wheat genome survey sequence database specific for chromosome 5B (http://i.versailles.inra.fr/blast/blast.php). The resulting contig (4492773) with highest similarity was downloaded and aligned with reference sequence using Geneious Pro software version 4.8.3. The aligned contig sequence was verified to have WECPDF domain which is characteristic of cell wall invertase gene. A gene specific primer pair (CWI-B1-F2/CWI-5B-R) was designed to clone 1200 bp coding region flanking the WECPDF domain. Purified DNA fragments were ligated into pEASY-T1 cloning vector (Beijing TransGen Biotech, Beijing, China) following manufacturer's protocol. *Trnas*1-T1 phage resistant chemically competent cells (CD501), TrnasGene Biotech, Beijing, China) were prepared as outlined in manual. Ligation reactions were used for cell transformation by keeping tubes in ice for 30 min followed by incubation at 42°C for 1 min. L.B medium 800 μL was then added and tubes were incubated at 37°C with 200 rpm. After that, all mixture was spread on plates with antibiotic (Ampicillin 1 mL/L) added LB media and incubated in dark at 37°C for 16 h. Sterile pipette tips were used to pick single bacterial colonies and placed into five different Eppendorf tubes with 500 μL LB media (with 1 mL/L Ampicillin). Samples were incubated at 37°C at 250 rpm for 7 h to ensure cell dispersal and growth. Colony PCR was done by taking 2 μL of the sample as PCR template to confirm the relevant gene. After confirmation, three replicates of each sample were sent for sequencing to Shanghai Sangon Biological Engineering Technology & Services Co., Ltd.

Nucleotide Sequences Retrieval and Alignment of *TaCwi-B1*

Firstly, the nucleotide sequences were analyzed for the possible indels observed. All sequences were aligned with the reference sequence (AF030240.1) and potential substitutions were noted. The CLUSTAL X was used for aligning the sequences (Larkin *et al.*, 2007).

Analysis of DNA Indels with novo SNP of *TaCwi-B1*

The Novo SNP was used for analyzing the variations observed in nucleotide sequences. Novo SNP uses reference sequence and trace files as input and generates the probable variations with a score that filters the non-synonymous substitutions with the synonymous ones. The tool novo SNP is available online (Weckx *et al.*, 2005).

Protein Sequence Retrieval and Alignment of *TaCwi-B1*

All the nucleotide sequences were translated into protein sequences by using CLUSTAL X. Full-length protein sequences were aligned to the reference sequence *TaCwi-B1* using CLUSTAL X with default BLOSUM scoring matrices. The aligned sequences were properly trimmed in order to avoid unnecessary gaps in the alignment and to preserve the conservation of functionally important motifs (Librado and Rozax, 2009).

Protein Variation Effect Analyzer (PROVEAN) of *TaCwi-B1*

The variants extracted from multiple sequence alignment were first checked through PROVEAN tool (Choi and Chan, 2015). This tool predicts the non-synonymous substitutions that are affecting the biological function of the protein. These mutations were selected to carry out the rest of the analysis.

Protein Structure Prediction of *TaCwi-B1*

To predict the 3D structure of the protein first step is to look for the homologous proteins that provide a significant hit over a fair length of the target protein. For CWI no such templates were available hence we chose to perform *ab-initio* threading methods for the structure prediction by using Phyre2 (Kelley *et al.*, 2015) and I-TASSER (Zhang, 2008) webservers.

Prediction of Disordered Regions of *TaCwi-B1*

The intrinsic unstructured regions that are not predicted as flexible regions are defined as disordered regions.

Disordered regions possess important biological function that includes post-translational modifications and binding to other proteins. Many studies have reported about how mutations are affecting the folding, stability of proteins. In our analysis we included two tools IUPRED (Dosztányi *et al.*, 2005) and MetaPrDOS (Ishida and Kinoshita, 2008). The threshold is set as 0.5 so if the residue is above the threshold it is considered as disordered.

Prediction of Protein Stability Changes of *TaCwi-B1*

The non-synonymous variations can cause stability changes in the protein structure. In order to analyze this factor, I-mutant was utilized which was reported by Capriotti *et al.* (2005). This tools takes in mutant residue and its position and outputs the stability change value that shows either decrease or increased stability change upon mutation.

Analyzing Point Mutations by Hope Server in *TaCwi-B1*

Hope is a webserver that analyses the structural effects of the mutations from a protein sequence. Selected mutations were provided to the Hope server to records the results (Venselaar *et al.*, 2010).

Phylogenetic Analysis of *TaCwi-B1*

To examine the relationship of CWI proteins and investigate evolutionary history of this protein among different species of Grasses family. Phylogenetic analysis was constructed by using Geneious 11.0.4. by the neighbor joining (N-J) method and phylogenetic tree was drawn. Protein sequences of *A. tauschii, Hordeum vulgare, Brachypodium distachyon, T. uratu and Zea mays* were retrieved and nine sequences of wheat including reference sequence were subjected to phylogenetic analysis.

Statistical Analysis

Allelic effects of markers were compared by restricting maximum likelihood (REMLMM). Genotypes served as random effect and gene served as fixed effect. Least significant tests (LSD) are used to compare means with threshold probability $P<0.05$ for multiple haplotypes in a given trait. For statistical significance, t-test was used to check allelic effect of single KASP marker.

Results

Allelic Variations at *TaCwi-A1, TaCwi-B1, TaCwi-D1* and their Effects on Phenotypes

In diversity panel, 58% accessions had Hap-4A-C haplotype which is favorable to yield components (Fig. 1 and Table 1). *TaCwi-A1* was associated with days to heading (DH), grains per spike (GpS), plant height (PH), proline, thousand grain weight (TGW) in both well-watered and water-limited condition. The KASP marker CWI-5D-312 was used to identify haplotypes at *TaCwi-D1* and results indicated that Hap-5D-C haplotype was fixed in diversity panel and none of the accession had contrasting Hap-5D-G haplotype. Previously, no sequence variation was observed at *TaCwi-B1* in bread wheat (Jiang *et al.*, 2014): therefore, this gene was sequenced in diversity panel to identify any new variations probably contributed by synthetic hexaploid wheats. The results are given below.

Analysis of DNA Indels in *TaCwi-B1* with novoSNP

Two variations were observed at the position 153 (C153del) and 320 (G320C). Both were analyzed by novo SNP. In 54% SYN-DERs the variant G320C was observed significantly. The scores generated by this tool identifies the variation at position 320 (G320C) as deleterious one. However, variation at position 153 C153del is neutral. Wild type (TaCwi-B1-G) showed high thousand grain weight (TGW) as compared to mutated (TaCwi-B1-C) as shown in Fig. 1.

Mutations Analysis by PROVEAN of *TaCwi-B1*

A highly conserved motif in the B genome was previously defined as WECPDF. The variants identified in this conserved region and in the surroundings were Glu372Gly, Glu372Lys, Pro374Gln, Asp375Thr, Tyr377Thr, Val379Cys, Ala380Val, Phe376Leu and Trpp371del. Few other substitutions observed in the sequence are Ser107Thr, Arg108Thr, Val111Ile, Ile115Leu, Gly121Ser, Gln132Pro. All these substitutions were analyzed through PROVEAN tool. The results generated predicted all mutations of the conserved domain as deleterious, while other substitutions were predicted as neutral. The results are tabulated in Table 3.

Protein Structure Prediction of *TaCwi-B1*

We predicted the 3D structure of *TaCwi-B1* by using Phyre2. Phyre2 modelled 324 residues with 100% confidence covering 38% of the whole sequence based on the template available. Furthermore, as the structure predicted by Phyre2 covered less percentage of the whole target sequence. I-TASSER uses top ten threading templates that covers the larger percentage of the entire target sequence. Second model was picked to predict by I-TASSER as it provides the highest C-score and covers 60% of the entire sequence with 2.99 Z-score. In addition to that the predicted model was also analyzed by the values generated as normalized B factor from I-TASSER. The negative value means the residues are more stable. We analyzed the WECPDF domain and found that all these residues were more stable in the structure as tabulated in Table 5. Hence mutations in this region can cause serious damage to the structure and function of the *TaCwi-B1* protein. Structures are shown in Fig. 2.

Table 1: Phenotypic validation of KASP assays for *TaCwi-A1* under well-watered and water-limited conditions

		WW			WL		
Gene	Trait[a]	P [b]	R2 [c]	Estimate effect	P	R2	Estimate effect
TaCwi-A1	DH	0.03287	0.02212	-0.91968	0.000132	0.06932	11.1087
	GpS	0.03281	0.02214	3.90401			
	PH	0.000259	0.06257	4.79532			
	Proline	0.01348	0.02913	-26.499			
	TGW	0.01502	0.02823	-2.3943	0.02869	0.02325	-2.1635

[a] DH: days to heading; GpS: grains per spike; PH: plant height; TGW: thousand grain weight
[b] Significance at $P>0.05$
[c] R^2: Phenotypic variance explained by marker

Table 2: Primer sequences and allelic information of individual KASP assays used in current

Primer name	Primer sequence 5' to 3'	Purpose
CWI-B1-F2	F: GAGTGTGGTAAATTGCAAGTTG R: GTTCGACATCAACGGTTGCT	Purpose to resequence *TaCwi-B1*
CWI-4A-1523	F1: GAAGGTGACCAAGTTCATGCTTTTATTTAAAATTTGATGAACTTTTCATAAAC F2: GAAGGTCGGAGTCAACGGATTTTTATTTAAAATTTGATGAACTTTTCACAAAT R: CATCGAATTGAAGAAAAGTTCACGC	KASP assay for SNP at 1523 bp at TaCwi-A1
CWI-5D-312	F1: GAAGGTGACCAAGTTCATGCTTAGAGGAGATCAAGTCATTGCGTGC F2: GAAGGTCGGAGTCAACGGATTTAGAGGAGATCAAGTCATTGCGTGG R: GGCCTTGTTGCTGACATTGACGTGCT	KASP assay for SNP at 312 bp in TaCwi-D1

Table 3: Variants predicted from PROVEAN in sequence of wheat cell wall invertase gene (*TaCwi-B1*) and effect on stability upon mutation of cell wall invertase (*TaCwi-B1*) predicted from I-Mutant in synthetic derived wheat in synthetic-derived wheat

Variant	PROVEAN Score	Prediction	Stability Change (I-Mutant)	Effect
Glu372Lys	-3.908	Deleterious	ΔΔG = -0.87367226	Destabilizing
Glu372Gly	-6.830	Deleterious	ΔΔG = -1.0654503	Destabilizing
Pro374Gln	-7.225	Deleterious	ΔΔG = -0.96078728	Destabilizing
Asp375Thr	-5.984	Deleterious	ΔΔG= -0.44249054	Destabilizing
Tyr377Thr	-6.147	Deleterious	ΔΔG= -1.5475124	Destabilizing
Val379Cys	-4.144	Deleterious	ΔΔG= -0.82590609	Destabilizing
Ala380Val	-1.972	Neutral	ΔΔG= -0.39137295	Destabilizing
Ser107Thr	0.355	Neutral	ΔΔG= -0.325	Destabilizing
Val111Ile	-0.270	Neutral	ΔΔG= -0.265	Neutral
Ile115Leu	0.234	Neutral	ΔΔG= -0.365	Neutral
Gly121Ser	-0.575	Neutral	ΔΔG= -0.245	Neutral
Gln132Pro	1.270	Neutral	ΔΔG= -0.367	Neutral
Trp371del	-19.920	Deleterious	ΔΔG= -0.435	Likely Destabilizing
Phe376Leu	-5.439	Deleterious	ΔΔG = -0.98579368	Destabilizing
Arg108Thr	0.378	Neutral	ΔΔG =-1.0654539	Destabilizing

Disorder Prediction of *TaCwi-B1*

The variations in consideration occurring in the conserved domain did not lie in the disordered regions as predicted by both softwares. Fig. 2 shows the disordered regions graphically.

Protein Stability Changes upon Mutation in *TaCwi-B1*

The stability of the protein structure was analyzed by I-Mutant tool for all the mutations identified in *TaCwi-B1* protein sequence. I-Mutant showed that all the deleterious substitutions in WECPDF domain are also destabilizing the structure of the protein while some neutral variants was also found in sequence. The results are tabulated in Table 4. Also the mutated structure is generated through PyMOL shown in Fig. 3.

Point Mutations Analysis by Hope Server in *TaCwi-B1*

All the identified variants were predicted as affecting the protein structure by disrupting the protein-protein interactions. If a small sized amino acid substituted by a large sized amino acid or vice versa, it can cause drastic conformational changes in the protein. Hence these features should be considered for analyzing the structural effect upon mutations induced.

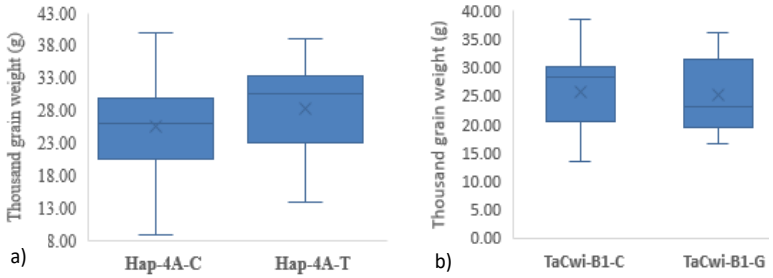

Fig. 1: Allelic effects and haplotypes of *TaCwi-A1* and *TaCwi-B1* genes on thousand grain weight. The x-axis represents the accession group with specific alleles and y-axis represents the phenotypic value in units mentioned for relevant traits

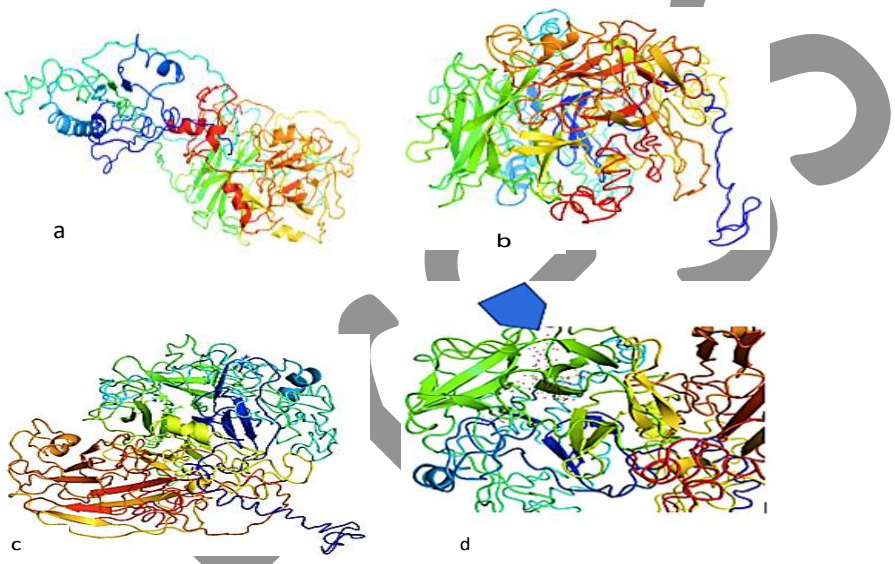

Fig. 2: The *TaCwi-B1* protein structures predicted by (a) Phyre2 (b) and I-TASSER (c) mutated structure generated by PyMOL (d) zoomed view of the mutations in the *TaCwi-B1* protein structure

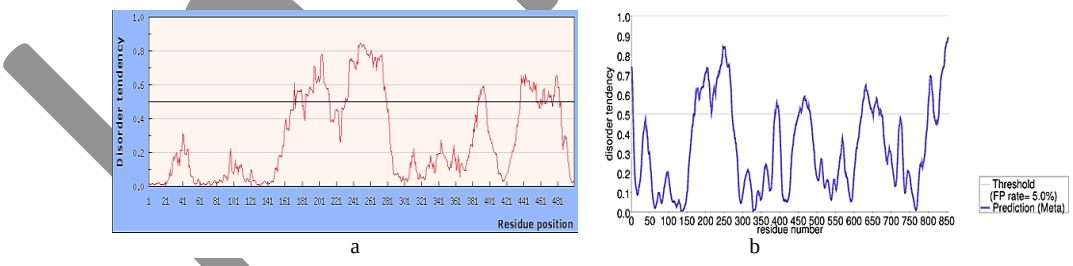

Fig. 3: Intrinsic Disordered Regions. a: The results generated from IUPred b: The graph generated from MetaPrDOS. The grey line is the threshold. Our domain lies below the line

Phylogenetic Analysis of *TaCwi-B1*

The tree is generated into two major clades showing reference protein sequence (*TaCwi-B1*) and other wheat sequences in one major clade. However, *B. distachyon*, *A. tauschii*, Z. *mays*, *T. urata* and *H. vulgare* in another clade. Branch labels showed bootstrap values as shown in Fig. 4.

Discussion

In this study, although haplotypes were identified only in *TaCwi-A1*, there may be polymorphism exist within *TaCwi-D1* in synthetic derivative as it has been shown in recombinant inbred lines RILs (Rasheed *et al.*, 2016). In *TaCwi-A1*, two SNPs formed haplotypes (TAAAC/CAAAT) named as *Hap-4A-C* and *Hap-4A-T*.

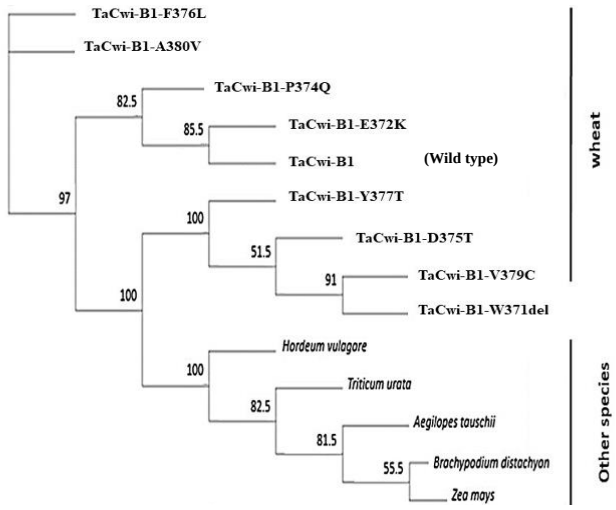

Fig. 4: Phylogenetic analysis of *TaCwi-B1* gene of *T. aestivum* (wild type and mutated), *B. distachyon*, *A. tauschii*, *Z. mays*, *T. uratu* and *H. vulgare*

These haplotypes were linked with jasmonate regulated plant defense at *Hap-4A-C* but lacked in *Hap-4A-T*. Previous studies showed that Hap-4A-C was associated with high grain number under rainfed condition (Jiang *et al.*, 2014). In this study Hap-4A-C is associated in both conditions. It was reported that by constitutive expression of CWI driven by cauliflower mosaic virus (CaMV) 35S promoter improved grain yield up to 145.3% when compared to wild life in maize (Li *et al.*, 2013). Hence, CWI has a role in increase in grain yield.

WECPDF was the most conserved domain in *TaCwi-B1* genome (Webster *et al.*, 2012). The multiple nucleotide and protein sequence alignment of B genome sequences of SYN-DERs showed few non-synonymous substitutions shown in Fig. S1 and S2. All the substitutions were analyzed by exploiting sequence and structure derived features. Most of the deleterious mutations were observed in the conserved domain WECPDF starting from position 370 to 376 while few mutations were confined in the neighborhood of this domain. For the mutation Glu372Gly wild type and mutant residue differs in size as glycine is smaller in size than glutamic acid. Further the Glycine is a flexible amino acid hence can disturb the rigidity of the protein at position 372 (Yan and Sun, 1997). The effect of mutation is also evaluated on the contacts made by the mutant. The wild type residue makes a hydrogen bond and salt bridge with Lysine at position 318. As the residue gets mutated into glycine, it affects hydrogen bond formation and ionic interactions made by wild type. For the second mutation Pro374Gln the mutant residue is bigger in size than wild type. The hydrophobicity measure is also different on the two residues. The wild type residue was buried in the core of the protein as identified in the structure. As the size of mutant residue is big it will not fit and will disrupt the structure and function of the protein.

In the mutation Asp375Thr, the mutated amino acid is smaller in size and is neutral charge than the wild type which is negative charged (Ajun *et al.*, 2009; Akram *et al.*, 2011). As the wild-type residue forms a salt bridge with Lysine at position 411, the difference in charge introduced by mutation will disturb the ionic interaction. Loss of charge after mutation also disturbs the interactions with other residues in other domains. The mutation Phe376Leu causes an empty space in the core of the protein as in this case a large sized amino acid is mutated into smaller sized with difference in hydrophobicity measures. In case of Ala380Val, Valine is bigger in size than Alanine. The wild type residue is located on the surface of the protein hence it can clearly disrupt protein-protein interactions on local and global level. The mutations Arg108Thr, Tyr377Thr and Val379Cys are expected to disrupt the interactions of the protein as in both cases the mutated amino acids are smaller in size and will cause conformational changes in the protein. In Arg108Thr the mutation introduces the hydrophobic residue which can lead to the loss of hydrogen bond and disrupting the protein folding. The mutation was predicted as neutral by PROVEAN but it is decreasing the stability of the protein as predicted. Hence this substitution needs further verification in order to define it as high risk mutation.

Taken together all the results, it has been inferred that all high risk SNPs identified in this study were either effecting the folding of protein or disrupting the stability of the folded protein. For protein to perform its function properly, it has to attain proper conformation at first. The destabilization of native structure will result in loss of function in most cases. Hence, in this case mutated *TaCwi-B1* would no longer be able to perform its function (Bross *et al.*, 1999). Up to the best of our knowledge, no study has ever been conducted before this that identified the missense mutations in TaCwi gene which are crucial for the structure and function of the protein.

Furthermore, phylogenetic analysis result revealed that these proteins have high level of conservation during evolution and performed similar functions. Hence, further phenotypic experiments and expression analysis experiments may reveal the association of this mutation with drought and explore susceptibility and resistance of water stress.

Conclusion

The CWI gene is associated with grain yield related traits in well-watered and water-limited conditions. Synthetic derivative has a potential to tolerate drought tolerance as it has predominant favorable allele in *TaCwi-A1* gene. However, mutations in *TaCwi-B1* gene particularly in the conserved region are predicted as high risk variants that affects the structure and function of *TaCwi-B1*.

References

Acuna, A., R. Mason, N. Subramanian and D. Hays, 2015. Meta-analysis of wheat QTL regions associated with adaptation to drought and heat stress. *Crop Sci*., 55: 447–492

Afzal, F., B. Reddy, A. Gul, M. Khalid, A. Subhani, K. Shahzadi, U. Quraishi, A. Ibrahim and A. Rasheed, 2017. Physiological, biochemical and agronomic traits associated with drought tolerance in a synthetic derived wheat diversity panel. *Crop Past. Sci*., 68: 213–224

Ajun, W., S. Yan, G. Li and L. Huili, 2009. Preparation of aspirin and probucol in combination loaded chitosan nanoparticles and in vitro release study. *Carbohydr. Polym*., 75: 566–574

Akram, M., M. Asif, M. Uzair, A. Naveed, M.A. Madni, D.S. Ali Shah, Z. Ul Hassan and A. Ullah, 2011. Amino acids: A review article. *J. Med. Plants Res*., 5: 3997–4000

Bhattacharyya, A., P. Datta, P. Chaudhuri and B. Barik, 2018. Nanotechnology- A new frontier for food security in socio economic development. *Appl. Sci. Bus. Econ*., 1: 23–28

Bi, Y.J., Z.C. Sun, J. Zhang, E.Q. Liu, H.M. Shen, K.L. Lai, S. Zhang, X.T. Guo, Y.T. Sheng, C.Y. Yu, X.Q. Qiao, B. Li and H. Zhang. 2018. Manipulating the expression of a cell wall invertase gene increased grain yield in mazie. *J. Plant Growth Regul*., 84: 37–43

Bross, P., T.J. Corydon, B.S. Andresen, M.M. Jorgensen, L. Bolund and N. Gregersen, 1999. Protein misfolding and degradation in genetic diseases. *Hum. Mutat*., 14: 186–198

Choi, Y. and A.P. Chan, 2015. PROVEAN web server: a tool to predict the functional effect of amino acid substitutions and indels. *Bioinformatics*, 31: 2745–2747

De Marsily, G., 2007. An overview of the world's water resources problems in 2050. *Ecohydrol. Hydrobiol*., 7: 147–155

Dosztányi, Z., V. Csizmok, P. Tompa and I. Simon, 2005. IUPred: web server for the prediction of intrinsically unstructured regions of protein based on estimated energy content. *Bioinformatics*, 21: 3433–3434

Doyle, J.J. and J.L. Doyle, 1987. A rapid DNA isolation procedurefor smal quanitities of fresh leaf tissue. *Phytochem. Bull*., 19: 11–15

Edae, E., P. Byrne, H. Manmathan, S. Haley, M. Moragues, M. Lopes and M. Reynolds, 2013. Association mapping and nucleotide sequence variation in five drought tolerance candidate genes in spring wheat. *Plant Genet*., 6: 1–13

Ishida, T. and K. Kinoshita, 2008. Prediction of disordered regions in proteins based on the meta approach. *Bioinformatics*, 24: 1344–1348

Islam, M., M. Uddin and M. Sattar, 2013. A comparative economic study on BARI GOM-24 and BARI GOM-23 production in a selected area of Dinajpur district. *Prog. Agric*., 23: 123–132

Jiang, Y., Q. Jiang, C. Hao, J. Hou, L. Wang, H. Zhang, S. Zhang, X. Chen and X. Zhang. 2014. A yield associated gene TaCWI in wheat: its function, selection and evolution in global breeding revealed by haplotype analysis. *Theor. Appl. Genet*., 128: 131–143

Kelley, L.A., S. Mezulis, C.M. Yates, M.N. Wass and M.J.E. Sternberg, 2015. The phyre 2 web portal for protein modelling, prediction and analysis. *Nat. Protocols*, 10: 845–858

Larkin, M.A., G. Blackshields, N.P. Brown, R. Chenna, P.A. McGettigan, H. McWilliam, F. Valentin, I.M. Wilm, R. Lopez, J.D. Thompson, T.J. Gibson and D.G. Higgins, 2007. Clustal W and clustal X version 2.0. *Bioinformatics*, 23: 2947–2948

Li, B., H. Liu, Y. Zhang, T. Kang, L. Zhang, J. Tong, L. Xiao and H. Zhang, 2013. Constitutuve expression of cell wall invertase genes increases grain yield and starch content in maize. *Plant Biotechnol. J*., 11: 1080–1091

Librado, P. and J. Rozas, 2009. DnaSP v%: a software for comprehensive analysis of DNA polymorphism data. *Bioinformatics*, 25: 1451–1452

IWGSC, 2014. A chromosome based draft sequence of the hexaploid bread wheat (*Triticum aestivum*) genome. *Science*, 345: 1251788

McIntyre, C.L., A. Rattery, A. Kilian, M.F. Dreccer and R. Shorter, 2014. Preferential retention of chromosome regions in derived snthetic wheat lines: a source of novel alleles for wheat improvement. *Crop Pasteur Sci*., 65: 125–138

Passioura, J.B., 2012. Phenotyping for drought tolerance in grain crops: When is it useful to breeders? *Funct. Plant Biol*., 39: 851–859

Rasheed, A., W. Wen, F. Gao, S. Zhai, H. Jin, J. Liu, Q. Guo, Y. Zhang, S. Dreisigacker, X. Xia and Z. He, 2016. Development and validation of KASP assays for genes underpinning key economic traits in wheat. *Theor. Appl. Genet*., 10: 1843–1860

Reynolds, M., J. Foulkes, R. Furbank, S. Griffiths, J. King, E. Murchie, M. Parry and G. Slafer, 2012. Achieving yield gains in wheat. *Plant Cell Environ*., 35: 1799–1823

Tang, L., H. Zhai, X. Luo, Z. Wang, L. Cui and X. Bai, 2014. Overexpression of *Glycine soja* WRKY20 enhances both drought and salt tolerance in transgenic alfalfa (*Medicago sativa* L.). *Plant Cell Tiss. Org. Cult*., 118: 77–86

Trethowan, R.M., 2014. Delievering drought tolerance to those who need it: from genetic resource to cultivar. *Crop Pasteur Sci*., 65: 645–654

Turner, N.C., 1981. Techniques and experimental approaches for the measurement of plant water status. *Plant Soil*, 58: 339–366

Venselaar, H., T.A. Beek, R.K. Kuipers, M.L. Hekkelman and G. Vriend, 2010. Protein structure analysis of mutations causin inheritable diseases. An e-science appraoch with life scientist friendly interfaces. *BMC Bioinformatics*, 11: 548

Webster, H., G. Keeble, B. Dell, J. Fosu-Nyarko, Y. Mukai, P. Moolhuijzen, M. Bellgard, J. Jia, X. Kong, C. Feuillet and F. Choulet, 2012. Genome level identification of cell wall invertase genes in wheat for the study of drought tolerance. *Funct. Plant Biol*., 39: 569–579

Weckx, S., J. Del-Favero, R. Rademakers, L. Claes, M. Cruts, P. De Jonghe, C. Van Broeckhoven and P. De Rijk, 2005. NovoSNP, a novel computational tool for sequence variation discovery. *Genom. Res*., 15: 436–442

Yan, B.X. and Y.Q. Sun, 1997. Glycine residue provide flexibility for enzyme active sites. *J. Biol. Chem*., 272: 3190–3194

Zhang, Y., 2008. I-TASSER server for protein 3D structure prediction. *BMC Bioinformatics*, 9: 32–40

(Received 12 May 2018; Accepted 02 July 2018)