Plant Omics Journal POJ

POJ 6(2):150-156 (2013) ISSN:1836-3644

In-silico prediction of an uncharacterized protein generated from heat responsive SSH library in wheat (*Triticum aestivum* L.)

Jasdeep Chatrath Padaria*, Deepesh Bhatt, Koushik Biswas, Gagandeep Singh, Rajkumar Raipuria

Biotechnology and Climate Change Laboratory, National Research Centre on Plant Biotechnology, PUSA Campus, New Delhi- 110012, India

***Corresponding author: jasdeep_kaur64@yahoo.co.in**

Abstract

Wheat is exposed to various abiotic stresses at different stages of its life cycle leading to severe decline in productivity. With rapid climate changes, high temperature stress is a major limitation to wheat production. Certain cultivars of wheat display a tolerant response to heat stress. Studies on differential expression in response to heat stress leads to identification of genes involved in molecular mechanism of thermo tolerance. Large-scale differential display analysis generates a large number of transcripts, of which a few are stress responsive whereas, many are of unknown or uncharacterized functional identity. The present study was done to identify a transcript of uncharacterized function obtained from heat responsive subtractive library generated from anthesis stage of thermo-tolerant wheat cv. Raj3765. Real time PCR analysis showed a four-fold increase in expression of the identified transcript at a stress of 37°C at the anthesis stage, indicating its role in facilitating the plant to cope the deleterious effects of high temperature at anthesis stage. Protparam tool analysis revealed that the leucine (Leu) is dominant amino acid present in the sequence, involving 15.5% of total amino acids. *In-silico* analysis revealed the existence of conserved domain region similar to leucine rich repeat (LRR) motif, an important DNA-binding domain. The presence of LRR motif in the protein predicted from the transcript under study indicates that this protein has a role as a signaling molecule involved in stress responses. Functional validation of the identified transcript in a model plant system shall confirm its role in heat stress tolerance.

Keywords: Wheat cv. Raj 3765; Heat stress; Suppressive Subtractive Hybridization; qPCR; Homology modeling; *In-silico* analysis.
Abbreviations: BLAST-Basic Local Alignment Search Tool; CDD-Conserved Domain Database; cDNA-complimentary DNA; cv-cultivar; EST-Expressed Sequence Tag ; GFP- Green Fluorescent Protein; GO-Gene Ontology; LRR-Leucine rich repeat; NCBI-National Center for Biotechnology Information; PCR- Polymerase Chain Reaction; PDB-Protein Data Bank; qPCR- quantitative PCR; SAVES-Structural Analysis and Verification Server; SSH-suppressive subtraction hybridization; TIGR- The Institute for Genomic Research.

Introduction

Bread wheat (*Triticum aestivum* L.) is one of the most widely grown wheat species in India, occupying 37% of the total cultivated land. Gradual increase in temperature causes detrimental effect on physiological and biological processes resulting in drastic reductions in grain yield and quality (Asseng et al., 2011). Continued heat stress is a problem in about 7 mha while terminal heat is a problem in about 40% of the irrigated wheat growing areas of the world (Joshi et al., 2007). Heat stress occurs when the ambient temperature is higher than the threshold temperature. When the rate of evaporation exceeds the pace at which moisture is being replaced through water taken up by the plant, the plant becomes dehydrated. Genotypes tolerant to heat stress are known and their enhanced tolerance is due to the altered expression of stress-inducible genes that protect and maintain the function and structure of cellular components (Sanghera et al., 2011). As compared with the other methods for analyzing altered gene expression, such as mRNA differential display (Liang and Pardee, 1992), serial analysis of gene expression (Velculescu et al., 1997), and cDNA microarray (Chu et al., 1998); suppressive subtractive hybridization (SSH) (Diatchenko et al., 1996) is a productive and efficient approach for identifying and characterizing both known and unknown genes involved in complex developmental processes and stress responses (Fernandez et al., 2007; Namasivayam and Hanke, 2006; Tsuwamoto et al., 2007). In recent years, SSH method is widely employed as a method of differential expression of genes in corn (Zhang et al., 2004), rape (Wu et al., 2007), wheat (Li et al., 2008; Zhang et al., 2005), rice (Jiang et al., 2009), and cucumber (Terefe and Tatlioglu, 2005). The present study aimed at identifying heat responsive genes at anthesis stage in thermo tolerant wheat cv Raj3765 (Rane et al., 2007) by construction and sequence analysis of a subtractive cDNA library. Genes that were up-regulated under heat stress at anthesis stage were enriched in the SSH library. Sequencing and annotation analysis of the SSH library provided a basis for obtaining full-length genes related to anthesis stage under heat stress from cv. Raj3756 a thermo tolerant wheat genotype. A number of sequences identified as uncharacterized in the SSH library, prove to be very rare heat stress responsive transcripts (Gorantla el al., 2007). A 252bp EST sequence of uncharacterized protein from the generated library was selected for functional and structural analysis. qPCR studies confirmed the heat responsive stage specific expression of the transcript for uncharacterized protein. *In-silico* characterization of the raw sequence of the identified stress-inducible transcript was performed. The present investigation implies homology modeling to deduce a three dimensional structure of the uncharacterized protein under study, followed by validation

and comparative modeling of the obtained structure for its conformational stability and further biological analysis.

Results and Discussion

Construction of subtractive suppressive hybridization (SSH) libraries and EST analysis

Heat stress responsive forward and reverse EST library of Indian wheat genotype cv. Raj 3765 was generated by subjecting the plant to 37°C heat stress at anthesis stage. Anthesis is a crucial stage before grain filling and has vital role in wheat productivity. Based on random selection 784 clones were sequenced. After single pass sequencing of selected positive clones of the forward library, a total number of 109 good ESTs were obtained and assembled into 19 contigs and 59 singlets/singletons by DNASTAR software. During Gene Ontology by TIGR Blast of the generated library, a total of 2 ESTs were obtained showing homology with uncharacterized protein (Data attached as 'Supplementary Table'). Of these, the EST (252bp), that was of the longest conserved domain was selected for functional annotation. In NCBI-BLAST search (BLASTX), 78% sequence homology was identified to LRR like protein of *Arabidopsis thaliana* (Accession No. NP175397). Sequence alignment analysis of the identified EST and *Arabidopsis thaliana* LRR like protein (Accession No. NP175397) revealed the presence of conserved domains showing homology with LRR specific signatures (Fig 1). Identified EST was translated in BioEdit software and a 84 amino acid sequence length was obtained that did not showed any significant match with known annotated protein in related online databases (Fig 2). For predicting function of this uncharacterized protein, *in-silico* analysis of structural and functional characteristics of the protein was carried out. During sequence analysis the presence of conserved repeats suggested that the protein belongs to Leucine rich repeat (LRR) family (Kobe and Kajava, 2001) by using Pfam server (Punta et al., 2012). For finding signature sequence that uncharacterized protein was screened for the presence of conserved domains using sequence similarity search with close orthologs available in various protein databases using the web tools as CDD (Marchler-Bauer et al., 2011) followed by INTERPROSCAN (Zdobnov and Apweiler, 2001) and FFPred (Lobley et al., 2008) which search the defined conserved domains in the sequence. Presence of LRR specific signature sequence using CDD analysis further confirmed that our uncharacterized protein under study belongs to leucine rich repeat receptor like protein. Thus underlying the temporal expression of this protein, which is stage specific, verifies its vital role in survival under heat stress.

Stress treatment and real time PCR analysis

Real-time quantitative polymerase chain reaction (qPCR) technique was employed for analyzing gene expressions at different developmental stages attributable to its accurate fold detection. Depending on the appropriate internal control genes, the relative qPCR method estimates fold change of expression difference between target and reference genes relative to a control condition through $2^{-\Delta\Delta Ct}$ calculation. Endogenous reference seems to be a priority for normalization; however the choice of reference gene needs to be experimentally validated under a particular experimental condition (Schmittgen and Zakrajsek, 2000). Thus, a certain gene needs to be carefully assessed for stable expression in the experimental system under study before using it as a reference gene. Different housekeeping genes (β actin, Glyceraldehyde-3 phosphate dehydrogenase, 18S rRNA etc) have successfully been used as reference genes for normalization of qPCR analysis (Bas et al., 2004). For qPCR analysis of the transcript in the present study, actin gene was used as a reference gene after several rounds of semi quantitative PCR using cDNA from four developmental stages of thermotolerant wheat cv. Raj3765 and a heat susceptible cv. HD2967 for assorting varietal differences. The availability of a series of stage-specific promoters with diverse specificities would allow precise control of transgene expression for the biotechnological improvemnt of crop plants. To precisely monitor the expression level of a gene various transcript profiling approaches are used to examine gene expression patterns. Real time PCR seems to be one of them in which we can monitor distinct patterns of expression at various stages of development. After stress imposition real time PCR was employed to confirm the temporal expression of this uncharacterized protein in various developmental stages. At anthesis stage, as compared to the level of control, there was an increase of 2 fold and 4 fold in the expression levels of unknown transcript at 37°C and 42°C stress respectively. However, there was no change in the expression level of transcript under study at anthesis stage of heat stress susceptible genotype, cv. HD2967. A comparative analysis of expression of the same at different developmental stages in cv. Raj3765 indicated that there was no significant change in the expression level at both 37°C and 42°C at different developmental stages including seedling, tillering and stem elongation stages (Fig 3). The higher magnitude of transcript expression at anthesis stage of cv. Raj 3765 confirms that the expression was growth-stage and genotype specific. This may help the plant to cope with the deleterious effects of high temperature at anthesis stage that is critical for grain growth and development.

3D structure prediction using homology modeling

Generation of stress responsive ESTs is a well accepted approach to identify genes involved in stress tolerance response. Attributes similar to leucine-rich repeat sequence were exhibited by transcript under study on alignment analysis with known leucine rich repeat receptor like proteins. Prediction of secondary structure of this amino acid sequence was performed by using various online tools. By using the amino acid sequence of heat inducible transcript in Gor IV method we predict that alpha helix region is 14.29%, extended strand region is 16.67% and Random coil region is 69.05%. From CATH online server, it was confirmed that fold consists of (3)-alph-beta subunits, (3.80)-alpha-beta Horseshoe, (3.80.10) Leucine rich repeat and (3.80.10.10) ribonuclease inhibitor. Conserved domain search again revealed that our sequence resembles leucine rich repeat receptor like protein which was performed using CDD (Marchler-Bauer et al., 2011). ProtParam results exhibit the physicochemical parameters of uncharacterized protein (Gasteiger et al., 2005). There are 84 amino acids in the sequence, its molecular weight was 9420.8 and theoretical pI was 5.63. The maximum number of amino acids present in the sequence were Leucine (15.5%) followed by Proline (10.7%) and Aspargene (10.7%) (Table 1). The total numbers of negatively charged residues (Aspartic acid + Glutamic acid) were eight while the total numbers of positively charged residues (Arginine + Lysine) were seven. Since there is one ambiguous residue (X) in the sequence, atomic composition cannot be computed. The instability

Table 1. DNA sequence of primers used in the present study.

Gene/Transcript	Primer Sequence
Transcript for uncharacterized protein	F 5' ACCTCAGCAACAACAAGCTAACCG3'
	R 5' ATCGGGAAGTGGCTGTTCTAGGTT3'
Actin gene	F 5' GAAGCTGCAGGTATCCATGAGACC3'
	R5' AGGCAGTGATCTCCTTGCTCATC3'

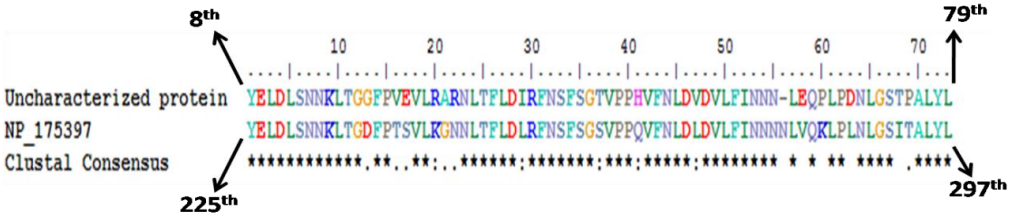

Fig 1. Sequence alignment of translated amino acid sequence deduced from identified EST sequence for uncharacterized protein indicating LRR specific signature sequence of *Arabidopsis thaliana*. Arrows indicate the position of amino acids of the predicted protein sequences where the conserved domain shared maximum homology.

index is computed to be 40.68 which classified the protein as unstable. The grand average of hydropathicity (GRAVY) was calculated to be -0.193. The structure for the uncharacterized protein was deduced by homology modeling. The structural information was obtained by template from PDB. Psi-Blast analysis of the predicted protein of the identified EST revealed homology to leucine rich repeat receptor like proteins present in PDB database with maximum resemblance to 3RGX chain A. The identity of this template was 46% which was good score to begin modeling. The homology modeling was done using ESyPred-3D. Finally structure was visualized using PyMOL. The model is finalized, with validated results showing Ramachandran plot (Fig 4). Comparative analysis of the model was performed using Profunc (Laskowski et al., 2005) and Dali Server (Holm and Park, 2000) which predicted function of protein on the basis of its 3D structure. On superimposing the predicted protein structure of the EST under study against template 3RGX chain A, a root mean square value of 1.352 was obtained indicating the deduced structure to be stable (Fig-5). Sequence and structure analysis proved that the uncharacterized protein obtained as an important rare transcript from anthesis stage wheat SSH library under heat stress is a member of leucine rich repeat receptor like protein.

Materials and Methods

Plant material, growth conditions and stress treatment

Wheat (*Triticum aestivum*) cv. Raj3765, a thermotolerant genotype (Rane et al., 2007) was grown in National Phytotron facility (Kumar et al., 2012) maintained at ±28°C day:night temperature in a 16:8 h photoperiod (Lu et al., 1991). The plants were arranged in a randomized complete block design within each phytotron chamber. At anthesis stage, heat stress was imposed by transferring 70 days-old plants to different glass-houses maintained at ambient 28 °C (control), 37°C (stress) and 42°C (stress) for 30 min, 2h, 4h and 6 h separately. After each heat stress treatment, leaf samples from control and stressed plants were collected, frozen in liquid nitrogen and stored at -80°C until RNA isolation.

Isolation of total RNA and construction of subtracted cDNA libraries

Total RNA from control and stressed leaves collected from anthesis stage at different time periods, as mentioned above, were isolated using Trizole Reagent (Invitrogen, Carlsbad, CA, USA) as per the manufacturer protocol. Subtracted cDNA library was constructed by using CLONTECH PCR-Select cDNA subtraction kit (CLONTECH Laboratories, USA) following manufacturer's instructions. In brief, tester (stressed) and driver (control) double stranded cDNAs were prepared from 2 µg mRNA. Tester and Driver cDNAs were separately digested with RsaI to obtain shorter blunt ended molecules. Two tester populations were created by ligating two aliquots of diluted tester cDNA with two different adaptors (adaptors 1 and 2R), separately. First hybridization was performed by the following procedure. Each tester population was mixed with an excess of digested driver cDNA. The samples were heat denatured and allowed to anneal at 68°C for 8 h. The two samples from the first hybridization reaction were mixed together, and more denatured driver cDNA was added for further overnight hybridization to enrich differentially expressed genes. Differentially expressed cDNAs, with different adaptor sequences at two ends, were selectively amplified by PCR and a second PCR was done with nested primers to further reduce the background. The subtracted and enriched DNA fragments were directly cloned into pGEM-T Easy T/A cloning vector (Promega, USA). Competent cells of *Escherichia coli* DH5α were transformed with the ligation mix and plated on LB-agar plates containing ampicillin, isopropylthio-β-D-1 thiogalactopyranoside (IPTG), and 5-Bromo-4-chloro –indolyl- β-D galactopyranoside (X-gal) for blue-white selection (Sambrook and Russel, 2001).

Sequencing and EST analysis

Sequencing and EST analysis was performed for all 784 clones obtained using universal primers. Each sequence was screened for overall base quality and vector, mitochondrial, ribosomal, and *E. coli* sequences were removed using VecScreen tool available online. All sequences of the generated and assembled EST library were examined for homology to sequences in the NCBI database (Pruitt et al., 2005) using BLASTX (Altschul et al., 1990). For Gene Ontology (GO) analysis, BLASTX was done against TIGR rice genome annotation project database to identify Locus IDs from corresponding rice homologs (Ouyang et al., 2007).

Stress treatment and real time PCR analysis

Heat stress was imposed on plants of thermo tolerant wheat genotype cv. Raj3765 at important developmental stages viz.

Table 2. ProtParam analysis for fetching the physicochemical parameters of predicted protein of the transcript with uncharacterized function obtained in heat responsive SSH library of wheat cv. Raj 3765.

Amino acid present	No. of residues of amino acids	Percentage of amino acid residues in the predicted protein
Alanine (A)	4	4.8%
Arginine (R)	6	7.1%
Asparagine (N)	9	10.7%
Aspartate (D)	5	6.0%
Cysteine (C)	0	0.0%
Glutamine (Q)	1	1.2%
Glutamate (E)	3	3.6%
Glycine (G)	5	6.0%
Histidine (H)	1	1.2%
Isoleucine (I)	2	2.4%
Leucine (L)	13	15.5%
Lysine (K)	1	1.2%
Methionine (M)	0	0.0%
Phenylalanine (F)	6	7.1%
Proline (P)	9	10.7%
Serine (S)	5	6.0%
Threonine (T)	4	4.8%
Trptophan (W)	1	1.2%
Tyrosine (Y)	2	2.4%
Valine (V)	6	7.1%
Pyrrolysine (O)	0	0.0%
Selenocysteine (U)	0	0.0%

XAWSRPRYELDLSNNKLTGGFPVEVLRARNLTFLDIRFNSFSGTVPPHVFNLDVD VLFINNNLEQPLPDNLGSTPALYLPGRPA

Fig 2. Translated amino acid sequence of annotated EST for uncharacterized protein generated in heat responsive SSH library of wheat cv Raj 3765

seedling, tillering, stem elongation and anthesis stages by transferring the plants to different glass-houses maintained at ambient 28°C (control), 37°C (stress) and 42°C (stress) for 30 min, 2h, 4h and 6h separately. Similar heat stress was also imposed on plants of thermo-susceptible wheat genotype cv. HD2967 at anthesis stage to evaluate the varietal differences. After each heat stress treatment, leaf samples from control and stressed plants were collected, frozen in liquid nitrogen and stored at -80°C until RNA isolation. Primers were designed for differential expression studies by real time PCR of the transcript of uncharacterized protein obtained from annotation of heat responsive SSH library generated from anthesis stage of wheat cv. Raj3765. Primers were also designed from the available database for amplification of actin gene used as endogenous control in the real time PCR studies (Table1). The expected amplicon for the EST under study and actin gene was 176bp and 151bp respectively. For carrying out real time PCR studies, RNA was first reverse transcribed into cDNA using reverse transcriptase enzyme. Total cDNA was synthesized using Superscript cDNA kit (Invitrogen, Carlsbad, CA, USA) using 150 nanogram of DNase treated total RNA and oligo dT as primer by following the manufacturer's instruction. For optimizing the annealing temperature for real time PCR, semi quantitative PCR amplification was carried to amplify the transcript for uncharacterized protein and actin gene using cDNA as a template for optimizing the specific annealing temperature before real time PCR. PCR was set at 30 cycles in order to reflect a semi-quantitative way of estimating the abundance of transcripts of uncharacterized protein under study and actin. Real-time PCR were performed and analyzed subsequently (Bhatt et al., 2013) using LightCycler real time PCR (Roche Applied Science, Indianapolis, IN, USA.

3D structure prediction using homology modeling

The secondary structure of the uncharacterized protein was predicted using Gor IV Server (Garnier et al., 1996) and BLAST analysis was used to compare the query sequence with the database sequence to find its homologues (Altschul et al., 1990). CATH server was employed to carry out the hierarchical classification of protein domains (Orengo et al., 1997). Conserved domain search was performed using CDD according to (Marchler-Bauer et al., 2011). The physicochemical parameters of hypothetical protein were evaluated using ProtParam (Gasteiger et al., 2005). For homology modeling, initially a suitable template was searched using PDB (Protein data bank) PSI-BLAST (Altschul et al., 1997; Berman et al., 2000). Since, the uncharacterized protein sequence showed resemblance with PDB-ID 3RGX Chain A, therefore it was selected as a suitable template. Amino acid sequence alignment of target and template proteins and rough 3D models (06 models) were constructed by using different homology modeling servers like CPH models (Nielsen et al., 2010), Swiss models (Arnold et al., 2006), ESyPred (Lambert et al., 2002), FUGUE (Shi et al., 2001). Finally on the basis of model stability best model was selected from ESyPred. All the generated structures of the protein models were validated using SAVES (Structural Analysis and Verification Server), which uses PROCHECK (Laskoswki et al., 1993), WHAT_CHECK (Hooft et al., 1996), ERRAT (Colovos and Yeates, 1993), VERIFY_3D (Luthy et al., 1992) and PROVE (Pontius et al., 1996) with specially inspecting Psi/Phi values of Ramachandran plot obtained from PROCHECK analysis followed by checking an acceptable average 3D-ID score. The structure for the uncharacterized protein was deduced by homology modeling. The structural information was obtained by template from PDB. The homology modeling was done using ESyPred-3D. Finally structure was visualized using PyMOL (DeLano, 2002). Comparative structural and functional analysis of the modeled 3D structure of the uncharacterized protein was carried out with all known structures present in PDB database by using Profunc (Laskowski et al., 2005) and Dali Server (Holm and Park, 2000).

Conclusion

The heat stress responsive SSH library at anthesis in wheat lead to a transcript that did not show homology with any known sequence in available databases. Functional validation using qPCR approach confirmed that the sequence obtained was an important transcript having a vital role under heat stress. *In-silico* analysis revealed knowledge of the structure and function of the deduced uncharacterized protein of the identified transcript. Comparative analysis of the obtained EST sequence provided valuable information about conserved domains and protein family of the uncharacterized protein. Modeling studies confirmed that its function is similar to the one predicted during sequence analysis and the protein belongs to LRR family. Further studies are required for sequencing of the upstream region to identify promoters and enhancers involved in gene expression. Cloning of the

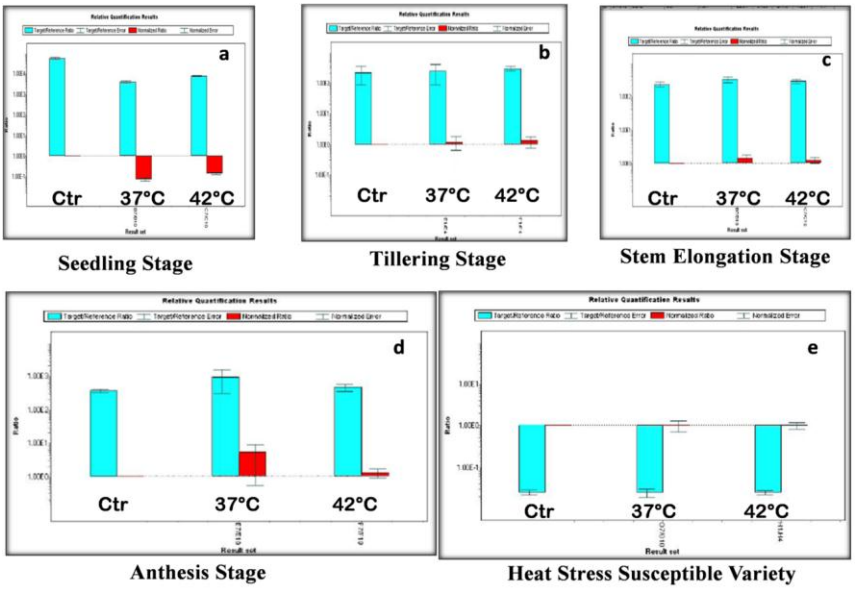

Fig 3. Differential expression analysis by qPCR studies : qPCR analysis of the identified EST with uncharacterized function at different development stages of thermo tolerant wheat cv. Raj 3765 viz Seedling **(a);** Tillering**(b);** Stem Elongation **(c) ;** Anthesis stage**(d)** and at anthesis stage of thermo susceptible wheat cv. HD2967**(e).** The relative expression value of each gene was normalized to an endogenous control actin and calculated using the CT (Cycle threshold) method. Line above bars represents Mean ±standard error.

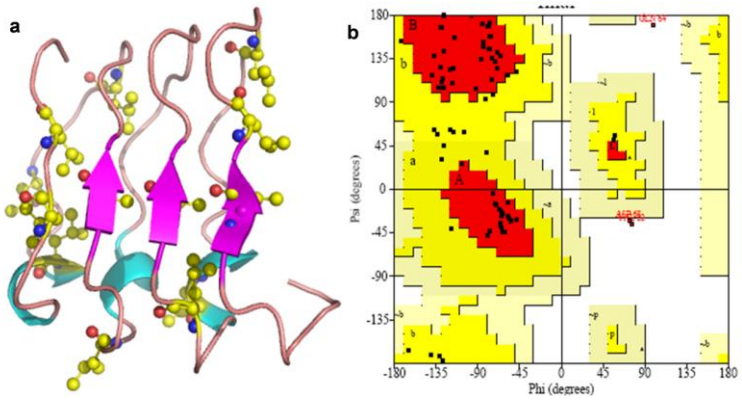

Fig 4. Predicted protein model and validation**. (a)** Final model with LRR (Ball and Sticks). Cyan colour represents helix, pink colour is sheets, dirty violet is loops, yellow colour is carbon, blue is nitrogen and red is oxygen in leucine chemical structure. (**b)** Ramachandran Plot using PROCHECK: 65.2% residues in most favored region (A,B,L); 30.3% residues in additional allowed region (a,b,l,p), there are only 4.5% residues in disallowed region.

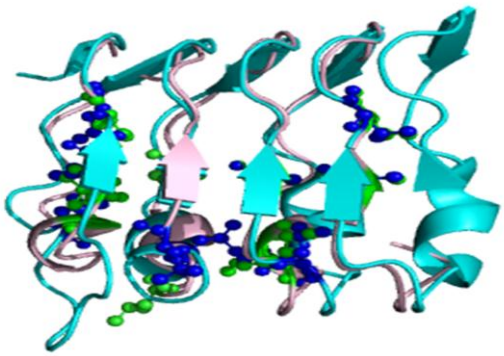

Fig 5. Superimposed model: Superimposing the predicted protein structure of the EST under study (Light pink) against its template 3RGX chain A (cyan) having LRR repeats (ball and sticks model) showing EST under study (green) and 3RGX chain A (blue).

identified full length gene in a binary vector and its functional validation in a model plant system is essential for development of thermo-tolerant transgenic wheat.

Acknowledgements

The financial support provided by Indian Council of Agricultural Research, India under National Initiative for Climate Resilient Agriculture is duly acknowledged. Authors are thankful to Project Director, NRCPB for providing facilities to carry out the above work. Authors are also thankful to Director, IARI for providing facilities for plant stress experiments at National Phytotron Facility, IARI, New Delhi.

References

Altschul SF, Gish W, Miller W, Myers E, Lipman DJ (1990) Basic local alignment search tool. J Mol Biol. 215(3):403-410

Altschul SF, Madden TL, Schaffer AA, Zhang J, Zhang Z, Miller W, Lipman DJ (1997) Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. Nucleic Acids Res. 25:389-3402

Arnold K, Bordoli L, Kopp J, Schwede T (2006) The SWISS-MODEL Workspace: A web-based environment for protein structure homology modeling. Bioinformatics. 22:195-201

Asseng S, Foster IAN, Turner NC (2011) The impact of temperature variability on wheat yields. Global Change Biol. 17(2):997-1012

Bas A, Forsberg G, Hammarstrom S, Hammarstrom ML (2004) Utility of the Housekeeping Genes 18S rRNA, β-Actin and Glyceraldehyde-3-Phosphate-Dehydrogenase for Normalization in Real-Time Quantitative Reverse Transcriptase-Polymerase Chain Reaction Analysis of Gene Expression in Human T Lymphocytes. Scand J Immunol. 59:566-573

Berman HM, Westbrook J, Feng Z, Gilliland G, Bhat TN, Weissig H, Shindyalov IN, Bourne PE (2000) The Protein Data Bank. Nucleic Acids Res. 28:235-242

Bhatt D, Saxena SC, Jain S, Dobriyal AK, Majee M, Arora S (2013) Cloning, expression and functional validation of drought inducible ascorbate peroxidase (*Ec-apx1*) from *Eleusine coracana*. Mol Biol Rep. 40(2):1155-1165

Chu S, Derisi J, Eisen M, Mulholland J, Botstein D, Brown PO, Herskowitz I (1998) The transcriptional program of sporulation in budding yeast. Science. 282:699-705

Colovos C, Yeates TO (1993) Verification of protein structures: patterns of nonbonded atomic interactions. Protein Sci. 2:1511-1519

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

Diatchenko L, Lau YF, Campbell AP, Chenchik A, Moqadam F, Huang B, Lukyanov S, Lukyanov K, Gurskaya N, Sverdlov ED (1996) Suppression subtractive hybridization: a method for generating differentially regulated or tissue-specific cDNA probes and libraries. Proc Natl Acad Sci. 93:6025-6030

Fernandez L, Torregrosa L, Terrier N, Sreekantan L, Grimplet J, Davies C, Thomas MR, Romieu C, Ageorges A (2007) Identification of genes associated with flesh morphogenesis during grapevine fruit development. Plant Mol Bio. 63:307-323

Garnier J, Gibrat JF, Robson B (1996) GOR method for predicting protein secondary structure from amino acid sequence. Method Enzymol. 266:540-553

Gasteiger E, Hoogland C, Gattiker A, Duvaud S, Wilkins MR, Appel RD, Bairoch A (2005) Protein identification and analysis tools on the expasy server. In: Walker JM (ed) The proteomics protocols handbook, Humana Press, New Jersey

Gorantla M, Babu PR, Reddy-Lachagari VB , Reddy AMM, Wusirika R, Bennetzen JL, Reddy AR (2007) Identification of stress-responsive genes in an indica rice (*Oryza sativa* L.) using ESTs generated from drought-stressed seedlings. J Exp Bot. 58(2):253-265

Holm L, Park J (2000) DaliLite workbench for protein structure comparison. Bioinformatics. 16(6):566-567

Hooft RW, Vriend G, Sander C, Abola EE. (1996) Errors in protein structures. Nature. 381:272

Jiang DG, Li J, Wu P, Liu ZL, Zhang CX (2009) Isolation and characterization of a microsporocyte specific gene, OsMSP, in rice. Plant Mol Biol Rep. 27:469-475

Joshi AK, Ortiz-Ferrara G, Crossa J, Singh G, Sharma RC, Chand R, Parsad R (2007) Combining superior agronomic performance and terminal heat tolerance with resistance to spot blotch (*Bipolaris sorokiniana*) of wheat in the warm humid Gangetic Plains of South Asia. Field Crop Res. 103:53-61

Kobe B, Kajava AV (2001) The leucine-rich repeat as a protein recognition motif. Curr Opin Struc Biol. 11(6):725-732

Kumar RR, Sharma SK, Gadpayle KA, Singh K, Sivaranjani R, Goswami S, Rai RD (2012) Mechanism of action of hydrogen peroxide in wheat thermotolerance-interaction between antioxidant isoenzymes, proline and cell membrane. Afr J Biotechnol. 11(78):14368-14379

Lambert C, Leonard N, DeBolle X, Depiereux E (2002) ESyPred3D: Prediction of proteins 3Dstructures. Bioinformatics. 18(9):1250-1256

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

Laskowski RA, Watson JD, Thornton JM (2005) ProFunc: a server for predicting protein function from 3D structure. Nucleic Acids Res. 33:89-93

Li HX, Zhang LY, Zhang GS, Niu N, Zhu ZW (2008) Construction of cDNA library from fertility-related genes of male sterile wheat with *Aegilops kotschyi* cytoplasm by SSH. Acta Agron Sin. 34:965-971

Liang P, Pardee A (1992) Differential display of eukaryotic messenger RNA by means of the polymerase chain reaction. Science. 257:967-971

Lobley AE, Nugent T, Orengo CA, Jones DT (2008) FFPred: an integrated feature-based function prediction server for vertebrate proteomes. Nucleic Acids Res. 36:297-302

Lu CS, Sharma HC, Ohm HW (1991) Wheat anther culture: effect of genotype and environmental conditions. Plant Cell Tiss Org. 24(3):233-236

Luthy R, Bowie JU, Eisenberg D (1992) Assessment of protein models with three-dimensional profiles. Nature. 356:83-85

Marchler-Bauer A, Lu S, Anderson JB, Chitsaz F, Derbyshire MK, Deweese-Scott C, Fong JH, Geer LY, Geer RC, Gonzales NR, Gwadz M, Hurwitz DI, Jackson JD, Ke Z, Lanczycki CJ, Lu F, Marchler GH, Mullokandov M, Omelchenko MV, Robertson CL, Song JS, Thanki N, Yamashita RA, Zhang D, Zhang N, Zheng C, Bryant SH (2011) CDD: a Conserved Domain Database for the functional annotation of proteins. Nucleic Acids Res. 39:D225-D 229

Namasivayam P, Hanke D (2006) Identification of differentially expressed sequences in pre-embryogenic tissue of oil seed rape by suppression subtractive hybridization (SSH). Plant Cell Tiss Org. 86:417-421

Nielsen M, Lundegaard C, Lund O, Petersen TN (2010) CPHmodels-3.0 - Remote homology modeling using structure guided sequence profiles Nielsen. Nucleic Acids Res. 38:576-581

Orengo CA, Michie AD, Jones S, Jones DT, Swindells MB, Thornton JM (1997) CATH-a hierarchic classification of protein domain structures. Structure. 5(8):1093-1108

Ouyang S, Zhu W, Hamilton J, Lin H, Campbell M, Childs K, Thibaud-Nissen F, Malek RL, Lee Y, Zheng L, Orvis J, Haas B, Wortman J, Buell CR (2007) The TIGR Rice Genome Annotation Resource: improvements and new features. Nucleic Acids Res. 35:D883-D887

Pontius J, Richelle J, Wodak SJ (1996) Deviations from standard atomic volumes as a quality measure for protein crystal structures. J Mol Biol. 264:121-136

Pruitt KD, Tatusova T, Maglot DR (2005) NCBI reference sequences (RefSeq): a curated non-redundant sequence database of genomes, transcripts and proteins. Nucleic Acids Res. 35:D61-D65

Punta M, Coggill PC, Eberhardt RY, Mistry J, Tate J, Boursnell C, Pang N, Forslund K, Ceric G, Clements J, Heger A, Holm L, Sonnhammer EL, Eddy SR, Bateman A, Finn RD (2012) The Pfam protein families database. Nucleic Acids Res. 40:D290-D301

Rane J, Pannu RK, Sohu VS, Saini RS, Mishra B, Shoran-Crossa J, Vargas M, Joshi AK (2007) Performance of yield and stability of advanced wheat genotypes under heat stress environments of the Indo-Gangetic plains. Crop Sci. 47(4):1561-1573

Sambrook J, Russell DW (2001) Molecular cloning In: A laboratory manual (3rd edn), Cold Spring Harbor Laboratory Press, Cold Spring Harbor, New York

Sanghera GS, Wani SH, Hussain W, Singh NB (2011) Engineering Cold Stress Tolerance in Crop Plants. Curr Genomics. 12(1):30-43

Schmittgen TD, Zakrajsek BA (2000) Effect of experimental treatment on housekeeping gene expression: validation by real-time, quantitative RTPCR. J Biochem Bioph Meth. 46:69-81

Shi J, Blundell TL, Mizuguchi K (2001) FUGUE: sequence-structure homology recognition using environment-specific substitution tables and structure-dependent gap penalties. J Mol Biol. 310:243-257

Terefe D, Tatlioglu T (2005) Isolation of a partial sequence of a putative nucleotide sugar epimerase, which may involve in stamen development in cucumber (*Cucumis sativus* L.). Theor Appl Genet. 111:1300-1307

Tsuwamoto R, Fukuoka H, Takahata Y (2007) Identification and characterization of genes expressed in early embryogenesis from microspores of *Brassica napus*. Planta. 225:641-652

Velculescu VE, Zhang L, Zhou W, Vogelstein J, Basrai MA, Bassett DE, Hieter P, Vogelstein B, Kinzler KW (1997) Characterization of the yeast transcriptome. Cell. 88:243-251

Wu JY, Shen JR, Mao XZ, Liu KD, Wei LP, Liu PW, Yang GS (2007) Isolation and analysis of differentially expressed genes in dominant genetic male sterility (DGMS) *Brassica napus* L. using subtractive PCR and cDNA microarray. Plant Sci. 172:204-211

Zdobnov EM, Apweiler R (2001) InterProScan-an integration platform for the signature-recognition methods in InterPro. Bioinformatics. 17(9):847-848

Zhang Y, Liu Y, Feng R, Zhang Y. (2004) Analysis of differential expression of thermo-sensitive genic-male-sterile gene in maize. Chinese J Trop Crop. 25(2):66-70

Zhang ZZ, Ma ZQ, Liu DJ (2005) Suppression subtractive hybridization analysis of Ms2 near-isogenic lines of wheat reveals genes differentially expressed in spikelets and anthers. J Plant Physiol Mol Biol. 31:175-182