Indian Journal of Experimental Biology
Vol. 59, May 2021, pp. 316-327

Molecular cloning and *in silico* analysis of heat stress responsive gene *ClpB*1 from *Ziziphus nummularia* genotypes

Kishor Prabhakar Panzade[1,2], Harinder Vishwakarma[1], Om Prakash Awasthi[3] & Jasdeep Chatrath Padaria[1]*

[1]ICAR-National Institute for Plant Biotechnology, New Delhi, India
[2]Division of Molecular Biology and Biotechnology, PG School, ICAR-Indian Agricultural Research Institute, New Delhi, India
[3]Division of Fruits and Horticulture Technology, ICAR-Indian Agricultural Research Institute, New Delhi, India

Received 24 December 2019; *revised 21 April 2021*

Heat stress is one of the most destructive abiotic stresses which adversely affect crop plants, resulting in reduced potential yield. Plants that are able to tolerate heat stress possess an intrinsic mechanism which needs to be unravelled at molecular level so as to decipher the role of gene and metabolic pathways involved in heat stress tolerance. To understand the molecular mechanism of heat stress tolerance, studies on isolation and characterization of gene for abiotic stress tolerance, *ClpB*1 were performed in *Ziziphus nummularia* (Burm. f.) Wight & Arn, an inherently abiotic stress tolerant plant. Differential expression studies of gene *ClpB*1 by qRT-PCR in contrasting genotypes of *Z. nummularia* (genotype Jaisalmer: heat tolerant and genotype Godhra: heat sensitive) was carried out. CDS (Coding DNA sequence) of gene *ClpB*1 from the genotypes *Z. nummularia* J and *Z. nummularia* G were cloned and characterized. These genes *ZnJClpB*1 (ACNO: MN398267) and *ZnGClpB1* (ACNO: MN398268) showed 1.09 and 2.3% dissimilarity at nucleotide and amino acid level, respectively. Computational based analysis revealed the presence of larger functional AAA lid 9 domains in ZnJClpB1 as compared to ZnGClpB1. Phylogenetic relationship and structure modeling was performed to understand isoform type and basic molecular functioning and of gene *ZnClpB*1 from *Z. nummularia* genotypes. Possibly, it is the first report on cloning, characterization and comparative *in silico* based analysis of gene *ZnClpB1* in *Z. nummularia*. Gene *ZnClpB*1 would be a prospective resource for developing abiotic stress tolerant crops by transgenic or breeding approach.

Keywords: Abiotic stress, *Jhar Beri*, Lotebush, qRT-PCR

Ziziphus nummularia (Burm. f.) Wight & Arn., belonging to *Rhamnaceae* family, has the congenital capability to grow and tolerate different abiotic stresses including heat, drought, salinity, chilling and metal toxicity[1,2]. *Z. nummularia* grows in a hot and dry area of north-western plains, central and peninsular regions of India[2]. Commonly known as Lotebush, *Z. nummularia* is locally called *Jhar Beri* (Hindi). It is valued for its ability to grow under adverse climatic conditions especially high temperature[2]. Plants are unable to relocate to avoid stress but they develop various defense mechanisms such as synthesis of various signaling molecules, TFs (Transcription factors), carbohydrates, osmolytes, heat shock proteins (HSPs) to struggle against the stress[3].

A number of genes have been identified that are associated with response to heat stress, which includes regulatory genes such as heat shock factors (*HSFs*)[4], conserved WRKYGQK motif (*WRKY*)[5], NAM/ATAF/CUC (*NAC*) {no apical meristem (NAM), cup shaped cotylodeons (CUC)}[6], and structural genes such as heat shock proteins (*HSPs*), Ca-dependent protein kinases (*CDPKs*), late embryogenesis abundant proteins (*LEA*)[7] and ascorbate peroxidase (*APX*)[8], etc. Heat shock proteins are a group of conserved proteins, present widely in prokaryotes and eukaryotes and are known to play an essential role in cell homeostasis under both non-stress and stress conditions as molecular chaperones[9]. Depending on their molecular weight, functions, sequence homology, plant HSPs were grouped into five classes, namely small HSP, HSP60, HSP70, HSP90, and HSP100 family[10]. It has been observed that caseinolytic protease B/heat shock protein (ClpB/HSP100) is fundamental in regulating thermotolerance in plants[11]. The family Clp belongs to a superfamily of AAA^+ (ATPase associated with diverse cellular activities) proteins[12]. Class I ATPases contain two ATP binding domains (ClpB, ClpC, ClpD, ClpA,) and class II Clp ATPases contain one ATP binding domain (ClpM, ClpN, ClpY and ClpX)[13]. These proteins

*Correspondence:
Telefax: +91 11 25841787 Ext. 270
E-Mail: jcpadaria@gmail.com, lab206nrcpb@gmail.com

play a crucial role in protecting plant cells under heat stress conditions. Protein *ClpB*1 is involved in various plant defense functions as regulating the activity of protein complexes, unfolding proteins for presentation to proteases and facilitating the refolding of denatured protein[14]. In maize, it was found that *ClpB*1 was required for basal and induced heat stress tolerance[15]. Similarly, a study in maize has established that *ClpB* plays a vital role in heat stress tolerance[16]. Gene *ClpB*1 has been observed to be necessary for the enhancement of heat stress tolerance in prokaryotes such as cyanobacteria and *E. coli*. Thus, it indicates that the function of HSP100/ ClpB1 remains highly conserved during the evolution[17].

Functional genomics approach has made it easier to identify genes that are differentially expressed in response to any environmental stress as well as decipher the gene structure and function. Studies related to conserved sequences, features, localization or differential expression pattern of any gene are some considerations that strengthen the characterization of a isolated/cloned gene. In the present study, we tried cloning, expression analysis and *in silico* characterization of gene *HSP100*/*ClpB*1 from two contrasting genotypes of *Z. nummularia* to understand structure and function.

Materials and Methods

Plant materials, growth conditions and heat stress treatment

Seeds of *Z. nummularia* genotype Jaisalmer (CIAH-J) (national identity, IC0598427; and registration number INGR15011) and Godhra (CIAH-G) varying in their tolerance to abiotic stress were obtained from CIAH, Bikaner, India[18]. After seed surface sterilization, seeds were sown in plastic pots (15″) filled with soilrite and grown under glasshouse condition, at National Phytotron Facility, IARI, New Delhi (temperature 22°±2°C, relative humidity 70-75%, under day length of 12 h). At the seedling stage (30 days after sowing), the plants were exposed to heat stress at 42°C for 2 and 6 h, in a growth chamber. Heat stress was given by raising the temperature gradually 1°C per 10 min until the temperature reached 42°C[19]. Relative humidity of 70-75% and normal light conditions were maintained. Leaf samples were collected with three biological replicates after 2 and 6 h of heat stress and quickly frozen in liquid nitrogen to be stored at −80°C for further use. Plants grown under no heat stress conditions were used as control and samples collected from them.

RNA isolation and cDNA synthesis

Total RNA was extracted from the leaf of pooled samples of 2 h heat-stressed using Trizol method[20] and purified using MN (Macherey-Nagel, Germany) kit. To remove any genomic DNA contamination, TURBO DNase (Ambion, USA) was used according to the manufacturer's instructions. cDNA was prepared from 5 µg of total RNA using SuperScript™ III first-strand cDNA synthesis kit (Invitrogen, USA).

Differential expression analysis of *ZnClpB*1 under heat stress

Gene *ClpB*1 from genotype Godhra (*ZnGClpB*1) and Jaisalmer (*ZnJClpB*1) were selected and their pattern of expression in response to heat stress at 42°C was analyzed at intervals of 2 and 6 h. On the basis of gene sequence obtained from RNASeq Data, the gene-specific primers were designed using Integrated DNA Technology (IDT) software (www.idtdna.com) and got synthesized from IDT (forward primer 5′-CCCATACA GTTAGTCCCTCTGA-3′, reverse primer 5′-TGTACA GCCAACAGCTGATATAA-3′. qRT-PCR analysis was carried out using cDNA synthesized from total RNA isolated (genotypes Jaisalmer and Godhra). The expression level of gene under heat stress was normalized with the internal reference gene *actin*[21]. The total volume of the reaction mixture was set as 20 µL containing 200 ng template cDNA, 400 nM of each primer and µL of 10 X KAPA SYBR qPCR master mix buffer. The qRT-PCR reaction was carried out in LightCycler® 480 II (Roche, Germany). Conditions of PCR amplification includes 94°C for 5 min then 40 cycles of 94°C for 20 s, and 60°C for 30 s and 72°C for 30 s. The relative expression level was calculated using the $2^{-\Delta\Delta Ct}$ method[22]. Samples from plants not subjected to heat stress were used as control. All qRT-PCR reactions were carried out in three technical replicates and three biological replications.

Construct preparation and transformation in *E. coli* DH5α cells

The CDS (coding DNA sequence) of gene *ZnJClpB*1 from genotype Jaisalmer and *ZnGClpB*1 from genotype Godhra were amplified by PCR using gene-specific primers that included *Xba*I and *Sal*I restriction sites (forward 5′-GCTCTAGAATGAATC CAGACAAATTCACTCATAAG-3′, reverse 5′-ACG CGTCGACCTATTCTTCCATCTCCTCATCTTC-3′) for further restriction digestion and directional cloning experimentation. Gene *ClpB*1 was amplified by PCR in a three-step program (initial denaturation at 94°C for 5 min, then 30 cycles of 94°C for 40 s, 60°C for 30 s and 72°C for 120 s). The obtained amplicon was analyzed on

1.2% agarose gel and then eluted from the gel using QIAquick gel extraction kit (Qiagen, USA). The specific amplicon of *ClpB*1 gene was ligated to pGEM®-T Easy vector (Promega, USA) using Rapid DNA ligation kit (Thermo Scientific™, USA). The ligated product was transformed to *E. coli* strain DH5α by heat shock method and the transformed cells were selected on Luria–Bertani (LB) agar plate containing ampicillin (100 mg/L), 5-Bromo-4-chloro-3-indolyl-β-D-galacto-pyranoside (X-Gal) (20 mg/L) and isopropyl-β-D thiogalactoside (IPTG) (1.0 mM) were used for blue/white screening. Isolation of plasmid from the positive clones was carried out using GenElute™ plasmid miniprep kit (Sigma-Aldrich, USA) and restriction digestion was performed to verify the existence of desirable insert. The recombinant plasmids carrying the CDS *ClpB*1 was sequenced by primer walking. The genes were accordingly named as *ZnGClpB*1 and *ZnJClpB*1 obtained from Godhra and Jaisalmer genotype, respectively.

In silico studies of *ZnJClpB*1 and *ZnGClpB*1

Physicochemical properties and subcellular location analysis

The vector sequences were removed by using the Chromas version 2.6.6 software (Technelysium, AU). The gene sequences obtained were searched for homology in the NCBI database using BLASTn program. CDS sequences were used for ORF prediction using the ExPASy software tool (http://us.expasy.org/tools/dna.html)[23]. Different protein parameters like amino acid composition, the total number of negatively and positively charged residues, hydropathicity index, atomic composition and instability index was estimated for ZnGClpB1 and ZnJClpB1 using the Protparam tool (http://web.expasy.org/cgi-bin/protparam). A prediction of the subcellular location of ClpB1 protein was performed using CELLO v.2.5 (http://cello.life.nctu.edu) software tool[24].

Prediction of post translational modifications and transmembrane helices

Netphos 3.1 server (http://cbs.dtu.dk/services/Netphos/) was employed for the prediction of phosphorylation site and kinase-specific prediction. YinOYang 1.2 (http://www.cbs.dtu.dk/services/YinOYang/) and NetNGly 1.0 server (http://www.cbs.dtu.dk/services/NetNGlyc/) was used for the identification of O-linked and N-linked glycosylation sites respectively, of the ClpB1 protein25. TMHMM server (http://www.cbs.dtu.dk/services/TMHMM/) was used to estimate the trans-membrane helix region.

Secondary structure and 3D structure prediction, modeling and Ramachandran plot.

Prediction of amino acid composition, α-helices, β-strand and coils was estimated for ClpB1 protein by using PSIPRED tool[26] (http://bioinf.cs.ucl.ac.uk/psipred/) and GOR4 server (http://npsa-pbil.ibcp.fr/cgi bin/npsa_automat pl?page=npsa_gor4.html)[27]. The 3D protein structure prediction was carried out using, Phyre2 tool (http://www.sbg.bio.ic.ac.uk/phyre2)[28]. The prediction of the active site and identification of ligand binding site and type of ligand binding to the ClpB1 protein generated model was joined up with a ligand by 3D Ligand site prediction server (http://www.sbg.bio.ic.ac.uk/3dligandsite)29. The quality of conformations of the obtained 3D structure was examined using the Ramachandran plot in terms of amino acid residues percentage in favourable regions. Ramachandran plot was determined using Vader 1.8 (http://vadar.wishartlab.com/)[30] and Rampage (http://mordred. bioc.cam.ac.uk/~rapper/rampage.php) tool[31].

Motif and domains prediction

The different conserved motifs present in ClpB1 protein sequence was predicted using MEME (Multiple Em for motif elicitation) servers with default parameters[32]. Orthologous genes to *ClpB*1 from various plant species were downloaded from the NCBI database and analysis was performed to find the relationship based on nucleotide and amino acid sequences. The functional domain prediction of ZnGClpB1 and ZnJClpB1 protein along with different plant species was carried out using the InterProScan tool (https://www.ebi.ac.uk/interpro/search/sequence-search)[33].

Phylogenetic analysis

Multiple sequence alignment of ZnJClpB1 and ZnGClpB1 with ClpB1 of other plant species was performed using MUSCLE with default parameters. Phylogenetic relationship analysis performed using the neighbour-joining method with the Poisson model using MEGA v7.0. software[34]. The internal node stability was assessed by the bootstrap value of 1000 replicates.

Results and Discussion

Heat stress-responsive differential expression analysis of gene *ZnClpB*1

qRT-PCR analysis of *ZnJClpB*1 showed increased expression of 5.24 fold levels, whereas expression of only 2.45 folds was observed in case of *ZnGClpB*1 after 2 h of heat stress condition. However, after 6 h

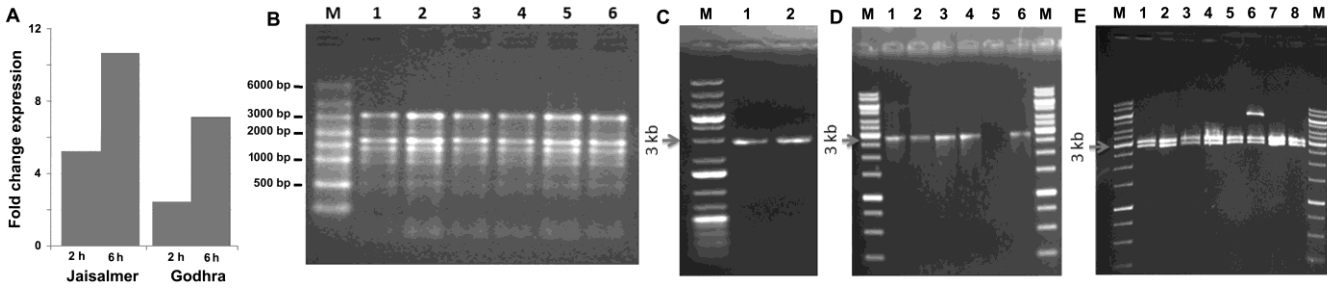

Fig. 1 — (A) Real time expression analysis of *ZnJClpB*1 in Jaisalmer and *ZnGClpB*1 in Godhra genotype; (B) Total RNA extracted. Lane 1-3 from Jaislmer and Lane 4-6 from Godhra; (C) PCR amplification of full length *ClpB*1 gene in *ZnJClpB*1 from Jaislmer (Lane 1) and Godhra (Lane 2) genotype; (D) Colony PCR confirmation of selected colonies.: Colony PCR amplification of full length *ZnJClpB*1 from Jaislmer (Lanes 1 &2), and Godhra (Lanes 3 & 4) genotype; (E) Restriction digestion confirmation of isolated plasmid. Lanes 1-4, *ZnJClpB1* and Lanes 5-8, *ZnGClpB*1.

of heat stress, *ZnJClpB*1 and *ZnGClpB*1 showed 10.65 and 6.88 fold change expression, respectively (Fig. 1A).

Amplification and cloning of gene *ClpB*1 from contrasting genotypes

Total RNA was extracted from leaves of seedling which were subjected to heat stress (42°C for 2 and 6 h), and its quality checked on 1% agarose gel (Fig. 1B). Synthesized cDNA from the extracted total RNA was used as a template for PCR amplification of *ClpB*1 CDS. The amplified amplicon of 2.7 kb was detected on 1.2 % agarose gel just below 3 kb band of DNA marker (Fig.1C). The PCR amplified amplicon was ligated with pGEM-T easy sequencing vector and transformed in *E. coli* strain DH5α with a transformation efficiency of 4.4×10^4 colony forming unit (CFU)/mg. Colony-PCR of recombinant cells showed specific single band (2.7 kb) equal to the size of *ClpB*1 CDS on 1.2% agarose gel (Fig. 1D). Restriction digestion analysis of positive clones revealed a desired insert of 2.7 kb of gene *ClpB*1 (Fig. 1E).

Full-length *ZnJClpB*1 and *ZnGClpB*1 cDNA and ORF prediction

Sequencing of the cloned gene *ClpB*1 revealed that CDS of *ZnJClpB*1 (from genotype Jaisalmer) is 2739 bp long and CDS of *ZnGClpB*1 (from genotype Godhra) is 2733 bp. Homology search revealed that CDS of gene *ClpB*1 shared 80-95% similarities with other identified gene *ClpB1/HSP100* from other plant species such as *Ziziphus nummularia*, *Morus notabilis*, *Quercus suber*, *Prunus avium*, *Gossypium arboreum* and *Vitis vinifera*. Based on the BLAST results, it was observed that *ZnJClpB*1 and *ZnGClpB*1 showed highest similarity (95.38%) with *Ziziphus jujuba* ClpB1 (LOC101778759). The annotated CDS sequences were designated as *ZnJClpB*1 (from Jaisalmer genotype) and *ZnGClpB*1 (from Godhra genotype, Deposited in gene bank (GenBank accession number: MN398267 and MN398268, respectively). Alignment between these two genes revealed that they were 2.3% dissimilar overall.

Physicochemical properties and subcellular location analysis

The prediction of subcellular location of protein ClpB1 carried out by the CELLO server illustrated that the proteins ZnJClpB1 and ZnGClpB1 were localized in the cytoplasmic region. Both ZnJClpB1 and ZnGClpB1 protein have leucine (>11%) as the highest and cysteine and tryptophan were the lowest amino acid residue. Differences between total numbers of positively and negatively charged amino acid residues were more in ZnGClpB1, which accounts for pI of ZnGClpB1 was more basic than the ZnJClpB1. The total number of negatively charged residues (Asp+Glu) were 146 while total number of positively charged residues (Arg+Lys) were 134 in ZnGClpB1, which may be the reason for its basic pI (8.71). The GRAVY (Grand average of hydropathicity) values and instability index were –0.390 and 38.65 for ZnJClpB1 and -0.385 and 37.96 for ZnGClpB1, respectively. This suggests the stable and hydrophobic nature of ZnJClpB1and ZnGClpB1 protein. All predicted parameters are presented in Table 1.

Post-translational modification

Post-translational modification such as phosphorylation, *O*-linked and *N*-linked glycosylation of amino acid residues play a significant role in modulating the functioning of the protein. Twenty and sixteen potential *O*-linked glycosylation sites were found in ZnJClpB1 and ZnGClpB1, respectively. We found that alteration in a few positions of *O*-linked glycosylation site. *O*-linked glycosylation of intracellular proteins has a crucial role in response to different abiotic stresses especially oxidative stress[13]. Two potential *N*-linked glycosylated sites were predicted in

Table 1 — Theoretically predicted all parameter of *ZnJClpB1* and *ZnGClpB1* protein

Parameters	*ZnJClpB1*	*ZnGClpB1*
Number of amino acids	912	910
Molecular weight	101364.86	100960.10
Theoretical pI	6.31	5.98
number of negatively charged residues	144	146
number of positively charged residues	137	134
Formula:	$C_{4431}H_{7291}N_{1303}O_{1365}S_{22}$	$C_{4407}H_{7247}N_{1299}O_{1368}S_{20}$
Total number of atoms	14412	14341
Instability index:	38.65(stable protein)	37.96 (stable protein)
Aliphatic index	96.90	97.21
Grand average of hydropathicity (GRAVY)	–0.390	–0.385

ZnJClpB1 at position 11 and 844 amino acid residues, while these sites were present in ZnGClpB1 at position 11 and 843 amino acid residues (Table 2). Position of second *N*-linked glycosylation site was altered by a single amino acid. We also found that the absence of any transmembrane helix in ZnClpB1 proteins.

Phosphorylation of tyrosine, threonine, and serine amino acid residues play a significant role in modulating the functioning of the protein. 21 different kinases {CKI (Casein kinase I), ATM (ataxia-telangiectasia mutated), CaM-II (Ca^{2+}/Calmodulin-Dependent kinase II), CKII, (Casein Kinase II) DNAPK (DNA dependent protein kinase), INSR (Insulin receptor tyrosine kinase), GSK3 (Glycogen synthase kinase 3), EGFR (Epidermal Growth Factor Receptor), PKB (Protein kinase B), PKA (Protein kinase A), PKC (Protein kinase C), RSK (Receptor tyr kinase), PKG (Protein kinase G), cdc2 (Cell division cycle protein kinase), cdk5 (Cyclin-dependent Kinase 5), SRC (Sarcoma family kinases), p38MAPK Mitogen-activated protein kinase and CKII, unspecified. Sixty five phosphorylating sites were present in ZnJClpB1 and ZnGClpB1 but their positions varied.

Secondary, 3D structure, ligand biding sites prediction and Ramachandran plot

Secondary structure of ZnJClpB1 protein produced by GOR4 (Garnier-Osguthorpe-Robson) method showed an alpha helix region of 60.20% (549), extended strand region of 8.44% (77) and random coil region of 31.36% (286), while in case of ZnGClpB1 an alpha helix region of 63.81% (580), extended strand region of 7.59% (69) and a random coil region of 28.60% (260). Secondary structure of ZnClpB1 (ZnGClpB1 and ZnJClpB1) generated through PSIPRED tool is given in Fig. 2 A and C. Analyzed and assessed peptide and protein structures through the Vader tool. All statistics generated given in Table 3.

Table 2 — Deduced O-linked glycosylation and N-linked glycosylation sites on *ZnJClpB1* and *ZnGClpB1*

O-linked glycosylation				N-linked glycosylation			
ZnJClpB1		*ZnGClpB1*		*ZnJClpB1*		*ZnGClpB1*	
P	R	P	R	P	R	P	R
88	T	88	T	11	NGTL	11	NETL
136	S	136	S	844	NSTV	843	NSTV
543	S	552	S	-	-	-	-
799	T	798	T	-	-	-	-
877	S	905	S	-	-	-	-
913	T	906	T	-	-	-	-

[P, Position; R, Residue]

Table 3 — Statistics generated through Vader tool

Statistics	*ZnJClpB1*	*ZnGClpB1*
phipsi core	789 (86%)	790 (86%)
Phipsi allowed	96 (10%	84 (9%)
phipsi generous	19 (2%)	9 (1%)
phipsi outside	8 (0%)	0 (0%)
omega core	871 (95%)	873 (96%
omega allowed	32 (3%)	27 (3%)
omega generous	3 (0%)	0 (0%)
omega outside	6 (0%)	9 (1%)
Free energy of folding	-754.96	-883.89
buried charges	22	0
Res 95% buried	187	454

The 3D structure of ZnJClpB1 and ZnGClpB1 protein was predicted based on template c1qvrB, which belongs to *Thermus thermophilus* (strain HB8) chaperone. 857 residues of ZnJClpB1 (94% of our sequence) have been modelled at >90% accuracy and 855 residues (94%) residues of *ZnGClpB*1 (94%) modelled at >90% accuracy by the single highest scoring template (Fig. 2 B and D) The model of the dimensions (Å) of ZnJClpB1 were X: 108.166 Y:80.193 Z:129.270 and Model dimensions (Å) of ZnGClpB1 were X:110.311 Y:81.501 Z:116.648. Further, ClpB1 was predicted for ligand binding sites. Active sites as shown in Fig. 3 A and C, consists of unchanged Arg, Val, Gly, Val, Gly, Lys, Thr, Glu, Leu, Lys, Val, Leu, and Gln residues in both the genes ZnJClpB1 and ZnGClpB1 but positions of

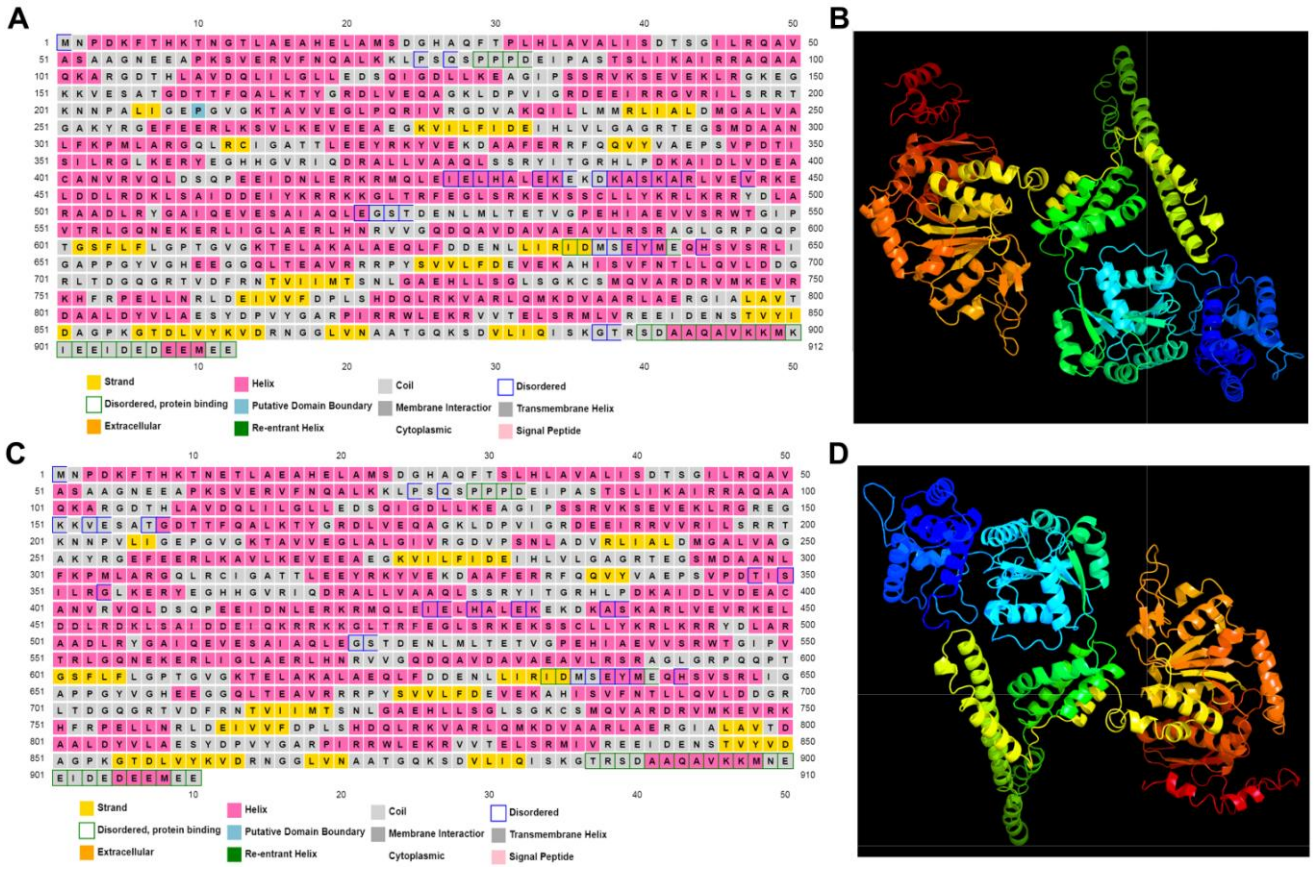

Fig. 2 — Sequence plot of secondary structure of (A) *ZnJClpB*1; & (C) *ZnGClpB*1; and Tertiary structure of (B) *ZnJClpB*1 & (D) *ZnGClpB*1

amino acid residues are varying (Table 4). These amino acid residues were interacting with ligands CDP, AMG, AMP, ADP, and ATP. Furthermore, Ramachandran plot generated through Vader and Rampage tool given in Fig. 3 B and D. In ZnJClpB1 number of residues in favored region 833 (91.5%), number of residues in the allowed region of 55 (5.9%), number of residues in the outlier region of 23 (2.5%). While in ZnGClpB1 number of residues in the favoured region is 817 (90.0%, number of residues in the allowed region of 70 (7.7%), the number of residues in the outlier region of 21 (2.3%).

Table 4 — *ZnJClpB1* and *ZnGClpB1* docking positions and amino residues to a predicted ligand

ZnJClpB1		*ZnGClpB1*	
Position	Amino acid	Position	Amino acid
572	ARG	571	ARG
573	VAL	572	VAL
574	VAL	573	VAL
611	GLY	610	GLY
612	VAL	611	VAL
613	GLY	612	GLY
614	LYS	613	LYS
615	THR	614	THR
616	GLU	615	GLU
617	LEU	616	LEU
619	LYS	618	LYS
778	VAL	777	VAL
782	GLN	780	LEU
-	-	781	GLN

Motifs, Domains prediction, and multiple sequence alignment

To study the conserved regions and structural characteristics of the ClpB1 proteins, their conserved motifs and domains were observed, and their deduced protein sequences were analyzed. The conserved motifs present in ZnJClpB1 and ZnGClpB1 sequences predicted using MEME (Multiple Em for motif elicitation) server with default parameters. We found that LDDLRDKLSAIDDE sequence in motif 10, were conserved in ZnClpB1 and ZjClpB1 protein (Fig. 4 A and B). Furthermore, we analyze the protein sequences for domain prediction and found that two ClpN amino-terminal domain, ATPase (AAA) domain, AAA lid domain (ClpA/ClpB), AAA domain (Cdc48 subfamily) and C-terminal a D2-small domain

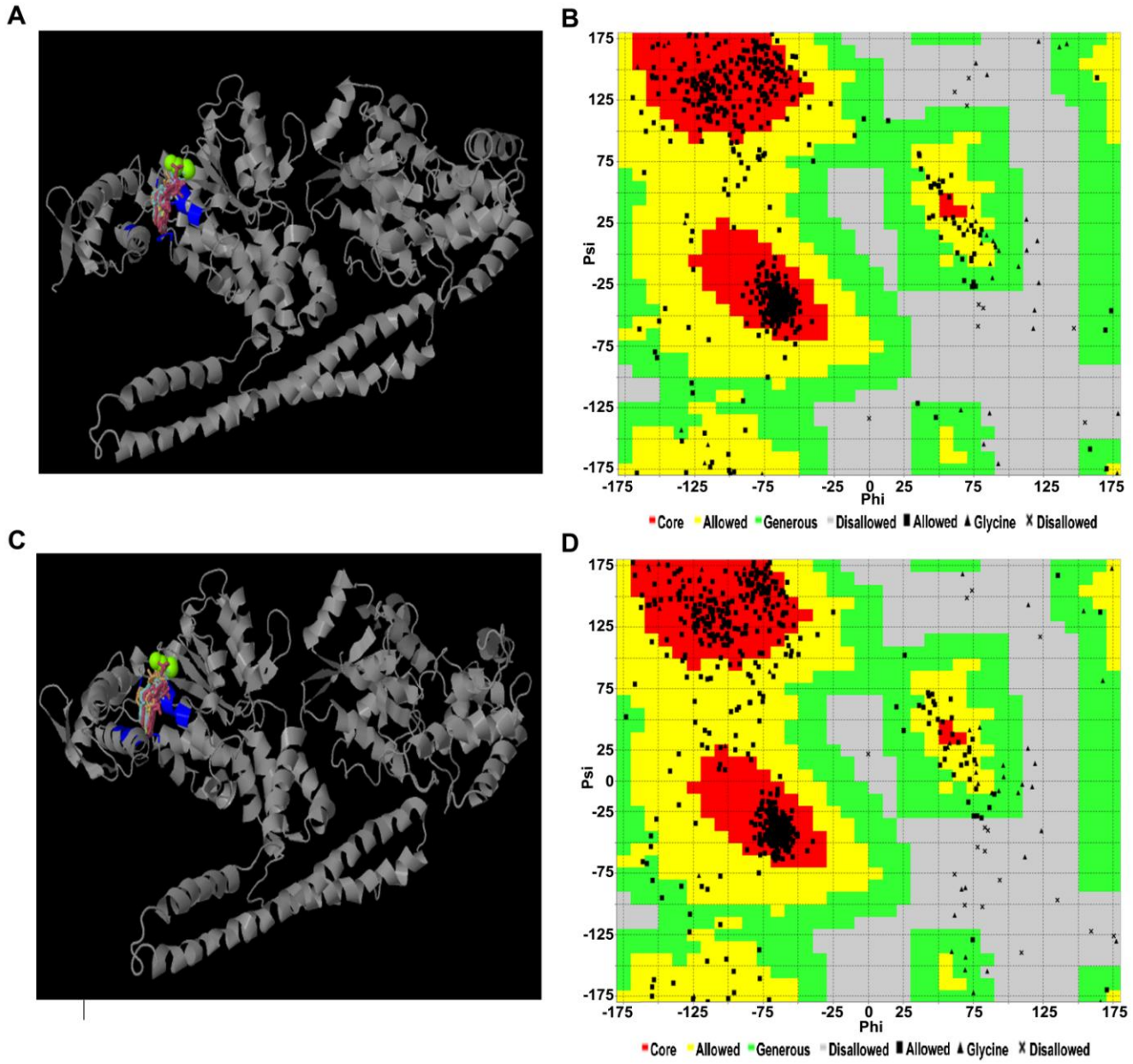

Fig. 3 — Presentation of docking to a predicted ligand (A & C); and Ramchandran plots (B & D) of ClpB1. [Docking to a predicted ligand in (A) *ZnJClpB1* and (B) *ZnGClpB1*. (coloured portion shows active site interaction with ligand). Ramchandran plot of (C) *ZnJClpB1* and (D) *ZnGClpB1* protein]

which was specific to Clp class members. We observed that all domains were highly conserved in ZnJClpB1 and ZnGClpB1 sequences along with orthologous protein sequences except AAA lid domain (ClpA/ClpB) were found larger in ZnJClpB1. Multiple sequence alignment of ZnJClpB1 and ZnGClpB with its orthologue from *M. notabilis. Z. jujuba* and *A. thaliana*, revealed that domain AAA, AAA lid 9, AAA2 and ClpB D2-small were highly conserved than two ClpN domains (Fig. 5). Each domain of ZnJClpB1 and ZnGClpB1 is highly conserved than its orthologue sequences. The extra larger size of the ZnJClpB1 AAA lid 9 domains is due to the presence of the LDDLRDKLSAIDDE sequence (Fig. 6 A and B) LDDLRDKLSAIDDE sequence present in all ClpB1 protein but in ZnClpB1 highly conserved. In ZnJClpB1 it is included within the AAA lid domain (ClpA/ClpB) while in the case of

ZnGClpB1 it is present outside the domain. While in its orthologues protein it is seen with variations. MSA suggests these sequences were specific to ZnClpB1 protein. That large domain might be the reason for the higher expression level under heat stress.

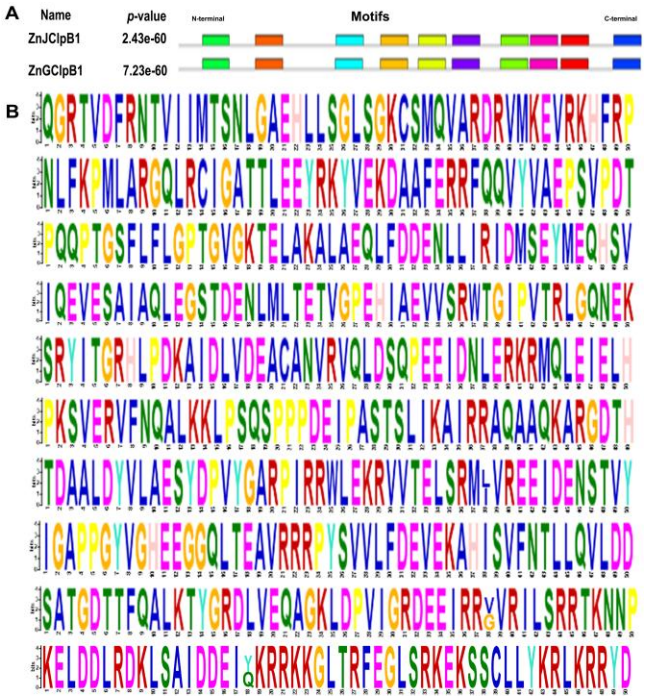

Fig. 4 — (A) Conserved motif sites prediction of ClpB1 protein. Protein terminals are shown as N and C-terminal. p value is given against each plant species. B) Conserved motif sequences in *ZnJClpB*1 and *ZnGClpB*1

Phylogenetic relationship analysis

Although ClpB1 is a type of protein encoded by a gene family, it is present in all plant species. We investigate the evolutionary distance and relationship

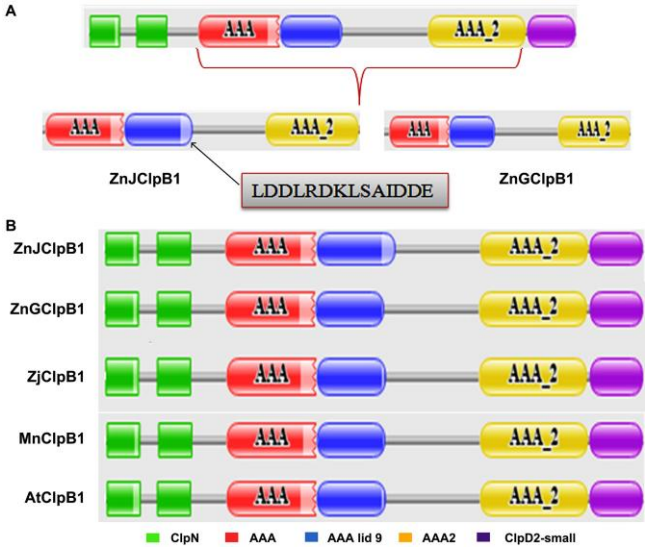

Fig. 6 — (A) AAA lid 9 domain in ZnJClpB1 larger than ZnGClpB1 due to presence of conserved sequence; and (B) Different domains present in *Z. nummularia* (ZnJClpB1and ZnGClpB1), *Z. jujube*, *M. notabilis*, and *A. thaliana* class I Clp ATPases

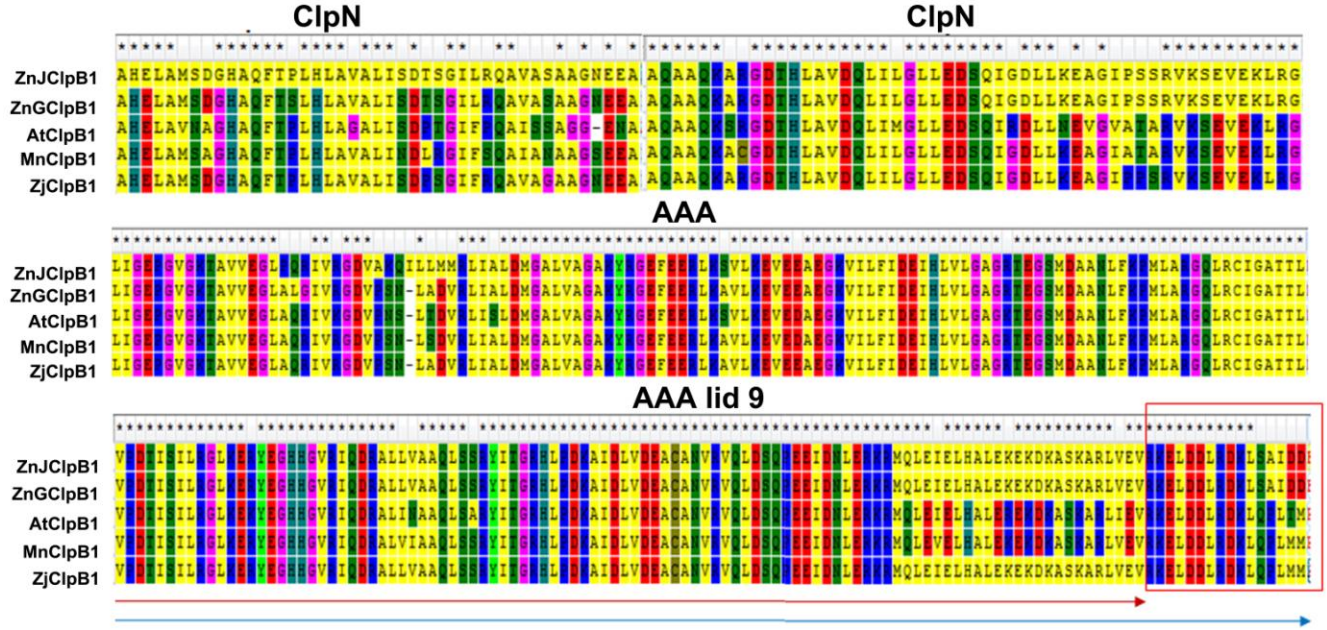

Fig. 5 — Amino acid sequence alignment of the ClpB1 domains with its orthologous. Conserved motifs are marked. Conserved region of AAA lid 9 domain in Jaisalmer and Godhra are indicated by red box. Region included in AAA lid 9 domain of *ZnJClpB*1 indicated by blue arrow. Region included in AAA lid 9 domain of *ZnGClpB*1, At*ClpB*1, Mn*ClpB*1, *ZjClpB*1 indicated by red arrow

of ZnJClpB1 and ZnGClpB1 with its orthologs genes. Phylogenetic relationship analysis of *Z. nummularia* class I Clp ATPases (ZnJClpB1 and ZnGClpB1) with its orthologous genes from *Z. jujuba*, *A. thaliana* and *M. notabilis* proteins allowed us to classify and annotate the ZnJClpB and ZnGClpB proteins, designated as ZnJClpB1 and ZnGClpB1. The phylogenetic tree of all the class I Clp ATPases protein sequences, the class I Clp ATPases was divided into three groups (ClpB, ClpD and ClpC), which was consistent with a phylogenetic relationship in rice[35]. Among this group, ZnGClpB1 is closely related to ZjClpB1 than ZnJClpB1. It reveals that ZnGClpB1 and ZnJClpB1 have the most recent common ancestors with the ClpB1 gene from *Ziziphus jujube* (Fig. 7).

It has been established that the phylogenetic relationships of heat shock protein (HSPs) are related to their subcellular localizations and members of subfamilies are named according to their protein localization of this group[35]. Our phylogenetic analysis of the ZnClpB1 showed results were consistent with the predictions of subcellular localization which again proves the above conclusions.

Z. nummularia, an underutilized fruit crop, is xerophytic in nature and therefore inherently tolerant to a variety of abiotic stresses such as heat drought and salinity. It is an excellent genetic resource to identify and isolate candidate genes involved in stress tolerance[1]. For isolation and characterization of differentially expressed genes, transcriptome profiling (RNAseq) has been established to be a dominant approach for identifying biotic and abiotic stress (drought, heat, salinity, etc.) responsive genes in plants[36]. In our analysis, thermo tolerant *Z. nummularia* genotype Jaisalmer and thermo sensitive genotype Godhra subjected to heat stress at 42°C for different time periods (2 and 6 h) and normally grown genotype Godhra and Jaisalmer used as a control was to obtain a broad variety of genes showing differential expression in response to heat stress were observed. This group of the plant was chosen to obtain a broad variety of heat-responsive genes expressing at two different temperatures. Expression of gene *ClpB*1 was 2.79 and 3.77 folds higher in Jaisalmer as compared to the Godhra after 2 and 6 h of heat stress respectively. Furthermore, a rapid increase in gene *ClpB*1 expression is an indication of a protective mechanism against protein denaturation due to heat stress[13].

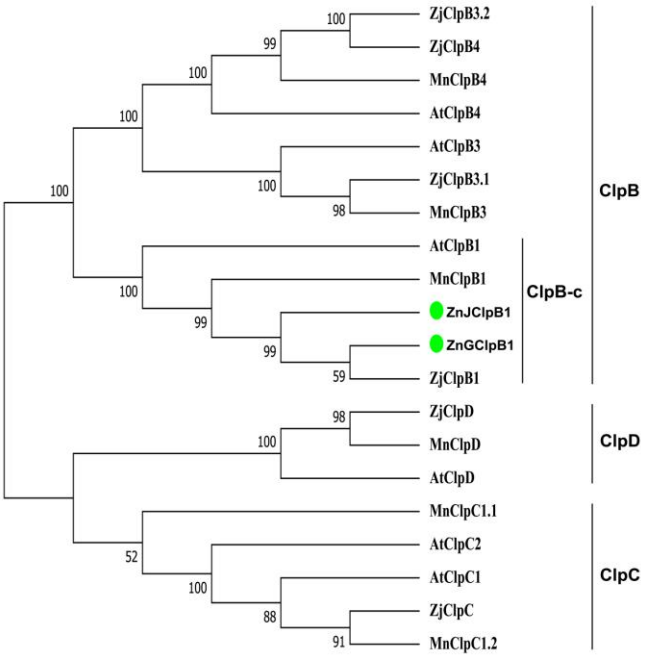

Fig. 7 — Phylogenetic analysis of ZnJClpB1 and ZnGClpB1 gene with its orthologous. [ZjClpB1 (XP_015883323.1) ZjClpB3 (XP_015879764.1) ZjClpB4.1 XP_015884727.1 ZjClpB4.2 (XP_024935320.1) ZjClpC (XP_015865748.1) ZjClpD (XP_015888444.1). AtClpB1 (AT1G74310), AtClpB3 (AT5G15450), AtClpB4 (AT2G25140) AtClpC1 (AT5G50920), AtClpC2 (AT3G48870) AtClpD (AT5G51070), MnClpB1 (XP_010105828.1), MnClpB3 (XP_024016890.1) MnClpB4 (XP_010090988.1) MnClpC1.1 (XP_024023142.1) MnClpD (XP_010099524.1)]

The differentially expressed *ClpB*1 transcripts from available heat stress responsive transcriptome data, which were obtained and, full-length CDS was cloned using PCR. *ClpB*1 has been isolated from several other crops like *A. thaliana*, cotton, and barley[37]. The most basic function of *ClpB* can be well defined by the fact that *ClpB* genes are required for high-temperature stress tolerance in *E. coli*, cyanobacteria and other prokaryotes[38]. Research in the past years has offered detailed knowledge about plant heat-stress responsive genes and their mechanism. It is appeared that heat shock protein, *ClpB* has a crucial role in adjusting the thermotolerance trait of plants[38]. Moreover, overexpression of *ClpB* in *Arabidopsis* and rice resulted in the improvement of thermotolerance[39]. In view of the above logic, the CDS of *ZnJClpB*1 stands as a potentially valuable candidate gene for the development of heat stress-tolerant transgenic crop plants under the unpredictability of climate in future.

In silico characterization of *ZnClpB*1 with regard to, motif analysis, secondary structure, 3D structure,

and active sites deduced by homology modeling and thus protein-substrate interaction analysis by docking study, multiple sequence alignment, construction of the phylogenetic tree. Cellular localization shows that *Ziziphus nummularia* has ClpB-c types of *ClpB* encoding isoform of a gene. Plants are known to have three isoforms of *ClpB,* localized to mitochondria (ClpB-m), chloroplast (ClpB-p) and cytoplasm (ClpB-c). Arabidopsis also has three isforms of *ClpB* i.e AtClpB1/AtClpB-c, ClpB-p (plastid) and ClpB-m (mitochondria). Mishra & Grover[40] have observed that AtClpB1/AtClpB-c was more involved in imparting heat stress tolerance than the other two ClpB proteins. *ClpB-c* was found essential for sustaining under heat stress, and *ClpBs* was not much significant[40]. Taken as a whole, the understanding in plant *ClpB* biology is additionally focused on *ClpB-c*. From the above arguments, it is clear that plant *ClpB*-c is one of the most crucial genes that direct the thermotolerance in plants.

The refined 3D structure of protein ZnClpB1 from both the genotypes was successfully developed and their active site residues were recognized. The 3D structures offer valuable information associated to identification of active sites and molecular function[41]. Searched PDB PSI-BLAST for identifying its template indicating the maximum identity of >90% which can be appeared as a high quality score of modelling. Variability at protein level observed with a position of phosphorylation, glycosylations on protein ZnClpB1 and amino acid residues inactive sites were variable among ZnJClpB1 and ZnGClpB1. Overall similarity at nucleotide level between *ZnJClpB1* and *ZnGClpB1* were 98%. This variability at protein and DNA results indicates the differences between *ZnJClpB1* and *ZnGClpB1*. These phosphorylated and glycosylations sites of ClpB1 might have a role in regulating response to abiotic stresses[42]. Moreover, phosphorylation also influenced other cellular activities like splicing, protein-protein interaction[43]. Highly conserved motifs present within an α-helical AAA+ lid domain of ZnJClpB1 which alter its structure may responsible for better functionality than ZnJClpB1. A-helical AAA+ lid domain occurs at the C-terminus of AAA domains. The helical bundle has a functional role in mediation of subunit interactions in oligomeric protein complexes and the formation of a lid to the nucleotide-binding site[44].

A phylogenetic study using the close relative as *Morus notabilis* and out-group taken as *A. thaliana* sequences indicated that class I Clp ATPase sequences of *ZnClpB1* are significantly conserved[45]. The tree represents three major clades, each one corresponding to ClpD, ClpC and ClpB proteins (Fig. 1). Proteins ClpB were further classified into mitochondrial, chloroplastic and cytoplasmic isoforms based on localization. Phylogenetic study of the CssHSP, CsHSP70 and CsHSP90 families of *Camellia sinensis* demonstrate results were consistent with the predictions of subcellular localization of protein[46].

Conclusion

Gene Zn*ClpB*1 from genotypes Jaisalmer and Godhra of *Z. nummularia* was analyzed for quantitative real-time expression. CDS was cloned, sequenced and structural differences between two CDS/protein were predicted through *in-silico* analysis. These identified structural differences may be modulating the expression level of gene Zn*ClpB*1 in both genotypes. Gene *ZnJClpB*1 expressed more than 5 fold, which signifies the role of *ZnJClpB*1 in heat stress tolerance. This potential genomic resource may be used to impart thermotolerance in heat stress susceptible plant species either by transgenic or breeding approach. In the present investigation, genes *ZnJClpB*1 and *ZnGClpB*1 not only illustrate diversity in gene evolution but also structural evolution. We have widely investigated the gene C*lpB*1 family in different plant species. Our results, represent a foundation for understanding the structural roles and evolutionary relationship of genes C*lpB*1 in thermotolerance. Furthermore, the detailed investigation on regulation of gene *ZnClpB*1 could create a way for a better understanding of thermotolerance mechanism in this plant species. It may offer a significant impact on blueprint of advanced crop improvement programs.

Acknowledgment

Authors are thankful to Dr. Palaiyur Nanjappan Sivalingam, ICAR-Central Institute for Arid Horticulture, Bikaner for providing seed material. Award of fellowship to the first author by PG School-IARI and Department of Biotechnology (DBT), Government of India is duly acknowledged.

Conflict of interest

Authors declare no conflict of interests.

References

1 Sabir MA, Rasheed F, Zafar Z, Khan I, Nawaz MF, ul Haq I & Bilal M, A consistent CO_2 assimilation rate and an enhanced root development drives the tolerance mechanism in *Ziziphus jujuba* under soil water deficit. *Arid Land Res Manag*, 34 (2020) 392.

2 Pandey A, Rakesh S, Radhaman J & Bhandari DC, Exploring the potential of *Ziziphus nummularia* (Burm. f.) Wight et Arn. from drier regions of India. *Genet. Resour. Crop Evol*, 57 (2010) 929.

3 Hassan MU, Chattha MU, Khan I, Chattha MB, Barbanti L, Aamer M, Iqbal MM, Nawaz M, Mahmood A, Ali A & Aslam MT, Heat stress in cultivated plants: nature, impact, mechanisms, and mitigation strategies—a review. *Plant Biosyst*, 155 (2021) 211.

4 Liu HC, Liao HT & Charng YY, The role of class A1 heat shock factors (HSFA1s) in response to heat and other stresses in Arabidopsis. *Plant Cell Environ*, 34 (2011) 738.

5 Dang FF, Wang YN & Yu L, CaWRKY40, a WRKY protein of pepper plays an important role in the regulation of tolerance to heat stress and resistance to Ralstonia solanacearum infection. *Plant, Cell Environ*, 36 (2013) 757.

6 Rahman H, Ramanathan V, Nallathambi J, Duraialagaraja S & Muthurajan R, Over-expression of a NAC 67 transcription factor from finger millet (*Eleusine coracana* L.) confers tolerance against salinity and drought stress in rice. *BMC Biotechnol*, 16 (2016) 35.

7 Liu Y, Zhang H, Han J, Jiang S, Geng X, Xue D & Wang J, Functional assessment of hydrophilic domains of late embryogenesis abundant proteins from distant organisms. *Microb Biotechnol*, 12 (2019) 752.

8 Padaria JC, Vishwakarma H, Biswas K, Jasrotia RS & Singh GP, Molecular cloning and *in-silico* characterization of high temperature stress responsive pAPX gene isolated from heat tolerant Indian wheat cv. Raj 3765. *BMC Res Notes*, 1 (2014) 7.

9 Hartl FU & Hayer-Hartl M, Converging concepts of protein folding *in vitro* and *in vivo*. *Nat Struct Mol Biol*, 16 (2009) 574.

10 Gupta SC, Sharma A, Mishra M, Mishra RK & Chowdhuri DK, Heat shock proteins in toxicology: how close and how far? *Life Sci*, 86 (2010) 377.

11 Frickey T & Lupas AN, Phylogenetic analysis of AAA proteins. *J Struct Biol*, 146 (2004) 2.

12 Neuwald AF, Aravind L, Spouge JL & Koonin EV, AAA+: A class of chaperone-like ATPases associated with the assembly, operation, and disassembly of protein complexes. *Genome Res*, 9 (1999) 27.

13 Schirmer EC, Glover JR, Singer MA & Lindquist S, HSP100/Clp proteins: a common mechanism explains diverse functions. *Trends Biochem Sci*, 21 (1996) 289.

14 Zolkiewski M, Zhang T & Nagy M, Aggregate reactivation mediated by the Hsp100 chaperones. *Arch Biochem Biophy*, 520 (2012) 1.

15 Nieto-Sotelo J, Martinez LM, Ponce G, Cassab, GI, Alagon A, Meeley RB, Ribaut JM & Yang R, Maize HSP101 plays important roles in both induced and basal thermotolerance and primary root growth. *Plant Cell*, 14 (2002) 1621.

16 Nieto-Sotelo J, Kannan KB, Martinez LM & Segal C, Characterization of a maize heat shock protein 101 gene HSP101 encoding a ClpB/Hsp100 protein homologue. *Gene*, 230 (1999) 187.

17 Agarwal M, Katiyar-Agarwal S & Grover A, Plant Hsp100 proteins: structure, function and regulation. *Plant Sci*, 163 (2002) 397.

18 ICAR News, A science and technology news letter, 20 (2014) 2.

19 Vishwakarma H, Junaid A, Manjhi J, Singh GP, Gaikwad K & Padaria JC, Heat stress transcripts, differential expression, and profiling of heat stress tolerant gene TaHsp90 in Indian wheat (*Triticum aestivum* L.) cv C306. *PloS One*, 13 (2018) e0198293.

20 Chomczynski P & Sacchi N, Single-step method of RNA isolation by acid guanidinium thiocyanate-phenol-chloroform extraction. *Anal Biochem*, 162 (1987) 156.

21 Sun HF, Meng YP, Cui GM, Cao QF, Li J & Liang AH, Selection of housekeeping genes for gene expression studies on the development of fruit bearing shoots in Chinese jujube (*Ziziphus jujube* Mill.). *Mol Biol Rep*, 36 (2009) 2183.

22 Livak KJ, & Schmittgen TD, Analysis of relative gene expression data using real-time quantitative PCR and the 2^{-} ΔΔCT method. *Methods*, *25* (2001) 402.

23 Gasteiger E, Hoogland C, Gattiker A, Duvaud S, Wilkins MR, Appel RD & Bairoch A, Protein Identification and Analysis Tools on the ExPASy Server. In: *The Proteomics Protocols Handbook*. (Ed: John M. Walker; Humana Press, New Jersey, USA), 2005, 571.

24 Yu CS, Chen YC, Lu CH & Hwang JK, Prediction of protein subcellular localization. *Proteins: Struct Funct Bioinf*, 64 (2006) 643.

25 Blom N, Gammeltoft S, & Brunak S, Sequence and structure-based prediction of eukaryotic protein phosphorylation sites. *J Mol Biol*, 294 (1999) 1351.

26 Jones DT, Protein secondary structure prediction based on position-specific scoring matrices. *J Mol Biol*, 292 (1999) 195.

27 Garnier J, Gibrat JF & Robson B, GOR method for predicting protein secondary structure from amino acid sequence. *Methods Enzymol*, 266 (1996) 540.

28 Kelley LA, & Sternberg MJE, Protein structure prediction on the web: a case study using the Phyre server. *Nat Protoc*, 4 (2009) 363.

29 Wass MN, Kelley LA & Sternberg MJ, 3DLigandSite: predicting ligand-binding sites using similar structures. *Nucleic Acids Res*, 38 (2010) 469.

30 Leigh W, Anuj R, Haiyan Z, Hassan M, Robert F, Boyko, Brian DS & David SW, VADAR: a web server for quantitative evaluation of protein structure quality. *Nucleic Acids Res*, 31 (2003) 3316.

31 Newman KS, *Rampage:* the social roots of school shootings. (Basic Books, New York, USA), 2004.

32 Timothy LB, Mikael B, Fabian AB, M artin F, Charles EG, Luca C, Jingyuan R, Wilfred WL & William SN, MEME SUITE: tools for motif discovery and searching, *Nucleic Acids Res*, 37 (2009) 202.

33 Philip J, David B, Hsin-YC, Matthew F, Weizhong L, Craig M, Hamish M, John M, Alex M, Gift N, Sebastien P, Antony FQ, Amaia S, Maxim S, Siew YY, Rodrigo L & Sarah H, InterProScan 5: genome-scale protein function classification. *Bioinformatics*, 30 (2014) 1236.

34 Kumar S, Nei M, Dudley J & Tamura K, MEGA: a biologist-centric software for evolutionary analysis of DNA and protein sequences. *Brief Bioinform*, 9 (2008) 299.
35 Amanjot S, Upasana S, Dheeraj M & Anil G, Genome-wide analysis of rice *ClpB*/*HSP100*, *ClpC* and *ClpD* genes. *BMC Genomics*, 11 (2010) 95.
36 Wang Z, Gerstein M & Snyder M, RNA-Seq: A revolutionary tool for transcriptomics. *Nat Re Genet*, 10 (2009) 57.
37 Lee U, Rioflorido I, Hong S, Larkindale J, Water E & Vierling E, The Arabidopsis ClpB/Hsp100 family of proteins: chaperones for stress and chloroplast development. *Plant J*, 49 (2006) 115.
38 Mishra RC & Grover A, ClpB/Hsp100 proteins and heat stress tolerance in plants. *Crit Rev Biotechnol*, 36 (2016) 862.
39 Queitsch C, Hong SW, Vierling E & Lindquist S, Heat shock protein 101 plays a crucial role in thermotolernace in *Arabidopsis*. *Plant Cell*, 12 (2000) 479.
40 Mishra RC, & Grover A, Voyaging around ClpB/Hsp100 proteins and plant heat tolerance. *Proc Indian Natl Sci Acad*, 85 (2019) 791.
41 Katiyar A, Lenka SK, Lakshmi K, Chinnusamy V & Bansal KC, *In silico* characterization and homology modeling of thylakoid bound ascorbate peroxidase from a drought tolerant wheat cultivar. *Genom Proteom Bioinf*, 7 (2009) 185.
42 Ogiso H, Kagi N, Matsumoto E, Nishimoto M, Arai R, Shirouzu M & Yokoyama S, Phosphorylation analysis of 90 kDa heat shock protein within the cytosolic arylhydrocarbon receptor complex. *Biochem*, 43 (2004) 15510.
43 Zhang XN, & Mount SM, Two alternatively spliced isoforms of the Arabidopsis SR45 protein have distinct roles during normal plant development. *Plant Physiol*, 150 (2009) 1450.
44 Miller JM, & Enemark EJ, Fundamental characteristics of AAA+ protein family structure and function. *Archaea*, 2016.
45 Liu MJ, Zhao J, Cai QL, Liu GC, Wang JR, Zhao ZH, & Luo LH, The complex jujube genome provides insights into fruit tree biology. *Nat Comm*, 5 (2014) 1.
46 Chen J, Gao T, Wan S, Zhang Y, Yang J, Yu Y, & Wang W, Genome-Wide Identification, Classification and Expression Analysis of the HSP Gene Superfamily in Tea Plant (*Camellia sinensis*). *Int J Mol Sci*, 19 (2018**)** 2633.