In silico analysis of whole-genome of *Solanum lycopersicum* for alpha-crystallin domains associated with heat stress tolerance

M.K. Chandra Prakash, Reena Rosy Thomas and Papiya Mondal
Section of Economics & Statistics
ICAR-Indian Institute of Horticultural Research
Hessaraghatta Lake Post, Bengaluru - 560089, India
E-mail: mk_chandraprakash@yahoo.com

ABSTRACT

Living organisms alter their gene-expression patterns to withstand stressful conditions. Drought, salinity, heat and chilling are potent abiotic stresses causing an alteration in gene expression. Among these, high temperature stress stimulates Heat Shock Transcription Factors (HSF) which activate heat shock promoters, thus turning on the heat shock genes. Heat shock proteins are, therefore, products of heat shock genes and are classified as per their molecular weight, including small heat shock proteins (sHsps). Hsps are chaperones playing an important role in stress tolerance. These consist of a conserved domain, flanked by N- and C-terminal regions termed the alpha-crystallin domain (ACD), and are widely distributed in living beings. Their role as chaperones is to help the other proteins in protein-folding and prevent irreversible protein aggregation. The conserved domains in sHsps are essential for heat-stress tolerance and for their molecular chaperone activity, enabling plant survival under increasing temperatures, leading to adaptations needed for coping with extremes climatic conditions. The present study focusses on identification of ACDs in the whole-genome of *Solanum lycopersicum*. A multinational consortium, International Tomato Annotation Group (ITAG), funded in part by the EU-SOL Project, provides annotation of the whole genome of *S. lycopersicum* available in the public domain. We used several *in silico* methods for exploring alpha-crystallin domains in all the chromosomes of *S. lycopersicum*. Surprisingly, these ACDs were found to be present in all the chromosomes excepting Chromosome 4; these are highly conserved in sHsps and are related to heat tolerance.

Key words: *Solanum lycopersicum*, alpha-crystallin domain (ACD), small heat shock proteins (sHsps), *in silico*, heat stress

INTRODUCTION

Alpha-crystallin domain (ACD) belongs to the class of small heat-shock proteins (sHsps) functioning as a molecular chaperone, preventing undesired protein-protein interactions and assisting in refolding of denatured proteins (Liberek *et al,* 2008). The term 'molecular chaperone' is normally used for describing the function of alpha-HSPs, as, these bind to and stabilize misfolded conformers of proteins, and, facilitate refolding of proteins *in vivo* (Parsell and Lindquist, 1993). To protect irreversible disaggregation of proteins, the chaperone activity of alpha-sHsps is limited to binding unstable intermediates (Jorg and Elizabeth 1994). Most ACDs have some common structural and functional features, including the molecular chaperone activity. Alpha-crystallins merge into a highly synergistic and adaptable multi-chaperone network to secure protein-quality control in the cell (Franz, 2002). Productive delivery and refolding of misfolded proteins into their native state demands close cooperation with other cellular chaperones. Further, alpha-Hsps have a significant role in stabilizing the cell membrane.

ACDs are conserved regions found in sHsps of all the three domain of life (bacteria, archaea and eukarya), suggesting their wide importance. There are different classes of ACD proteins comprising classical sHsps and, likely, chaperones. The α-crystallin domains are fundamental building blocks for most sHsps, consisting of several beta strands accountable for dimer formation arranged into two beta-sheets (Tariq *et al*, 2010). Mostly, varying the sHsps acquires less conserved N- and C-terminal extensions, whereas, greater sequence similarities can be seen in conserved ACDs (Eisenhardt, 2013). Unfortunately, very little information is available on ACDs. However, some examples indicate that this family is of great significance. Recently, in *Arabidopsis,* one ACD protein was reported

to be an essential element of a specific resistance-mechanism against systemic spreading of the tobacco etch virus (Whitham *et al*, 2000). It was reported that both sHsps and α-crystallins show ATP-independent molecular chaperone activity and, under heat stress, interact with partially unfolded polypeptides to prevent unspecific aggregation of protein substrates (Jakob *et al*, 1993 and Tyedmers *et al*, 2010).

The tomato genome: Tomato Genome Consortium started its work in the year 2003. Genome of the tomato (*Solanum lycopersicum* L.) was sequenced completely and published in 2012 (ftp://ftp.sgn.cornell.edu/genomes/ Solanum_lycopersicum/annotation/). Size of the tomato genome is 950 megabases (Mb), divided into 12 linear molecules, each containing different chromosomes. The shortest is Chromosome 6, with 46,041,647 nucleotides; the longest is Chromosome 1, with 90,304,255 nucleotides. Average length of the chromosome is 65 million nucleotides of DNA sequence. In the present study, ACDs available in whole-genome of *S. lycopersicum* have been explored.

In computational biology, genes can be detected by comparing genomes of related species. These detect evolutionary pressure for conservation, especially identification of conserved regions (including markers) as, these are conserved evolutionarily across species, and include solanaceous crops (Reena *et al*, 2013). Conservedness relies heavily on sequence-similarity, whose run-time grows with square of the number and length of the aligned sequence, demanding noteworthy computational resources (Nagar and Hahsler 2013).

MATERIAL AND METHODS

Identification of conserved domains in genes is one of the important steps in understanding the genome of any species. Sequence-similarities provide evidence for functional and structural conservation, along with evolutionary relationships, between sequences. Specifically, sHsp sequences are highly conserved across species, despite evolutionary pressure (Chandraprakash *et al,* 2013). Comparative analysis is a key method by which functional elements are identified. Ligand-binding sites of proteins and active sites of enzymes are the most highly conserved protein sequences. To identify conserved elements in a desired gene, the same sequence from several species should be aligned and common areas should be identified.

In our work, several *in silico* methods were used for showing the presence of conserved domains, especially

ACDs belonging to sHsp of *S. lycopersicum*. Published sHsp sequences were used for comparative analysis to identify similar regions in the whole-genome of *S. lycopersicum*. The identified, similar regions in tomato genomic sequences were obtained. These sequences were uploaded onto NCBI batch CD search program in an appropriate format to be processed as batch files. NCBI batch CD search program is one of the programs used for searching the input sequences against Conserved Domain Database (CDD). The output generated (CDD) is a tab-delimited list of conserved-domain hits, found on each protein, against input query sequences of *S. lycopersicum*. From the output file, matching sequences of α-crystallin domain were clustered chromosome-wise. These ACDs were found to be highly conserved and available in almost all the chromosomes in multiple copies.

RESULTS AND DISCUSSION

Generally, highly-conserved proteins are needed in fundamental cellular functions, stability or reproduction. Conservation of the protein structure is indicated by presence of functionally equivalent amino acid residues (not necessarily identical) and structures between analogous parts of the protein structure. Chandraprakash *et al* (2013) reported HSPs to be evolutionarily conserved in solanaceous crops. Defined by conserved alpha- crystalline domains, a sequence of 90 amino acid residues (approximately) constitutes small heat shock proteins (also known as α-Hsps) (MacRae, 2000). Most multiple α-Hsps translated in plants are housed in different cellular compartments, to prevent them from interacting with each other (Franz, 2002). However, engineering a single transcribed gene is not of much use, as, more than one stress-responsive genes may be necessary for survival of a plant under extreme conditions. At the genomic level, understanding the expression of a specific protein under a particular abiotic stress can provide a base for recognizing genes (Reena *et al,* 2013). Expression of sHsps is seen in response to various kinds of abiotic stresses, including extreme temperatures, oxidative stress, osmotic stress, etc.

In the present study, published sequences matched against the available *S. lycopersicum* database revealed 61 Alpha-crystallin domains to be widely distributed throughout the whole-genome of *S. lycopersicum,* except in Chromosome 4. A genome-wide analysis for presence of ACDs revealed these to be conserved in sHsps. A chromosome-wise distribution of ACDs in the whole-genome of *S. lycopersicum* is shown in Fig. 1.

144
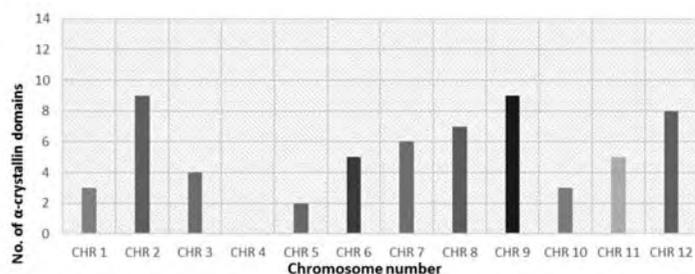
Fig. 1. Distribution of Alpha-crystallin domains (ACD) in *Solanum lycopersicum* chromosomes

Fig. 2. Quaternary structure of a typical ACD

Table 1. Distribution of α-crystalline domains (ACD) in the whole-genome of *Solanum lycopersicum*

Sl. No.	Chromosome No.	Total number of ACDs	Nature of protein	Hit-value range
1	CHR 1	3	sHSP	599-734
2	CHR 2	9	sHSP	523-959
3	CHR 3	4	sHSP	649-765
4	CHR 4	-	-	-
5	CHR 5	2	sHSP	507-1033
6	CHR 6	5	sHSP	667-8824
7	CHR 7	6	sHSP	597-1101
8	CHR 8	7	sHSP	791-948
9	CHR 9	9	sHSP	616-1007
10	CHR 10	3	sHSP	520-780
11	CHR 11	5	sHSP	595-933
12	CHR 12	8	sHSP	625-941

In silico analysis in wheat had shown the presence of alpha crystalline domain (Kumar *et al,* 2012). In the human genome, α-crystallin–related small heat shock proteins are dispersed over nine chromosomes (Kappé *et al,* 2003). Chromosome-wise distribution of alpha crystallin domains (ACDs) in *S. lycopersicum,* along with the nature of protein and matching similarity-values, is presented in Table 1.

Genome-wide, this ACD is sited maximally in Chromosomes 2 and 9, followed by Chromosomes 12, 8, 7, 6, 11, 3, 1, 10 and 5, with the exception of Chromosome 4. ACDs in Chromosome 6 had maximum matching similarity. Genome-wide analysis of ACDs revealed these genes to be enriched on several chromosomes, majorly, 15% in Chromosomes 2 and 9.

ACDs are unusually abundant and cover a segment in heat stress induced proteins termed small heat shock proteins, ranging in size from ~ 17 to 30 kDa. These are highly conserved sequences of around 90 amino acids, found extensively in all the domains of life. Representation of a typical ACD structure with two conserved regions that form a sandwich of two β pleated-sheets is shown in Fig. 2. The ubiquitous nature of ACDs implies that these domains are of great significance and help plants combat heat-stress and render longevity.

ACKNOWLEDGEMENT

REFERENCES

Chandraprakash, M.K., Reena Rosy Thomas, Krishna Reddy, M. and Sukhada Mohandas. 2013.Molecular evolutionary conservedness of small heat shock protein sequences in Solanaceae crops using *in silico* methods. *J. Hortl. Sci.*, **8**:82-87

Eisenhardt, B.D. 2013. Small heat shock proteins: recent developments. *Biomol. Concepts*,**4**:583-595

Franz Narberhaus. 2002. Alpha-crystallin-type heat shock proteins: socializing minichaperones in the context of a multichaperone network. *Microbiol. Mol. Biol. Rev.*, **66**:64-93

Jakob, U., Gaestel, M., Engel, K. and Buchner, J. 1993. Small heat shock proteins are molecular chaperones. *J. Biol. Chem.,* **268**:1517-20

Jorg Becker and Elizabeth A. Craig. 1994. Heat-shock proteins as molecular chaperones. *European J. Biochem.*, **219**:11-23

Kappé, G., Franck, E., Verschuure, P., Boelens, W.C., Leunissen, J.A.M. and de Jong, W.W. 2003. The human genome encodes 10 α-crystallin-related small heat shock proteins: HspB1-10. *Cell Stress Chaperones,* **8**:53-61

Kumar, R.R., Singh, G.P., Sharma, S.K., Singh, K., Goswami, S. and Rai, R.D. 2012. Molecular cloning of HSP17 gene (sHSP) and their differential expression under exogenous putrescine and heat shock in wheat

145

J. Hortl. Sci.
Vol. 10(2):143-146, 2015

(*Triticum aestivum*). *African J. Biotech.*, **11**:16800 -16808

Liberek, K., Lewandowska, A. and Zietkiewicz, S. 2008. Chaperones in control of protein disaggregation. *EMBO J.*, **27**:328-335

MacRae, T.H. 2000. Structure and function of small heat shock/alpha-crystallin proteins: established concepts and emerging ideas. *Cell Mol. Life Sci.,* **57**:899-913

Nagar, A. and Hahsler, M. 2013. Fast discovery and visualization of conserved regions in DNA sequences using quasi-alignment. *BMC Bioinformatics*, 14 Suppl., **11**:S2

Parsell, D.A. and Lindquist, S. 1993. The function of heat shock proteins in stress tolerance: degradation and reactivation of damaged proteins. *Annu. Rev. Genet.*, **27**:437-497

Reena Rosy Thomas, Chandraprakash, M.K., Krishna

Reddy, M., Sukhada Mohandas and Riaz Mahmood. 2013. Microsatellite identification in Solanaceae crops associated with Nucleoside Diphosphate Kinase (NDK) specific to abiotic stress tolerance through *in silico* analysis. *J. Hortl. Sci.*, **8**:195-198

Tariq Mahmood, Waseem Safdar, Bilal Haider Abbasi and Saqlan Naqvi, S.M. 2010. An overview on the small heat shock proteins. *African J. Biotech.*, **9**:927-939

Tyedmers, J., Mogk, A. and Bukau, B. 2010. Cellular strategies for controlling protein aggregation. *Nat'l. Rev. Mol. Cell Biol.*, **11**:777-788

Whitham, S.A., Anderberg, R.J., Chrisholm, S.T. and Carrington, J.C. 2000. *Arabidopsis* RTM2 gene is necessary for specific restriction of Tobacco Etch Virus and encodes an unusual small heat shock like protein. *Pl. Cell*, **12**:569-582

146

J. Hortl. Sci.
Vol. 10(2):143-146, 2015