Analele Științifice ale Universității „Alexandru Ioan Cuza”, Secțiunea Genetică și Biologie Moleculară
TOM XVIII, Fascicula 3, 2017

COMPUTATIONAL APPROACHES TO IDENTIFY REGULATORS OF STRESS INDUCIBLE GENES OF RAB FAMILY IN *S. LYCOPERSICUM*

CHAITALI ROY[1*]

Received: 19 July 2017 / Revised: 1 August 2017 / Accepted: 14 August 2017 / Published: 17 October 2017

Keywords: gene expression; promoter; tomato; salinity stress; *cis*-regulatory elements; transcription factor
Abstract: Adverse environmental conditions with increasing soil salinization have become a major problem in realizing the potential yields of crops. Plants deploy various mechanisms to deal with salt stress facilitated by the expression of several stress related genes. One such group is the late embryogenesis abundant (LEA) protein gene family which has various functions including protection of cellular structures from dehydration. To investigate the molecular basis of high tolerance of Punjab Keshari cv of tomato, we performed a comprehensive analysis of the proximal promoters of candidate members *Rab1A, Rab1B, Rab1C, Rab11a* of *Rab* gene family which are essential components of stress response. Genes and protein structures were also studied to identify the active sites. The completion and enhanced annotations of tomato genome sequence has provided the opportunity for genome-wide in-depth analysis of gene expression. Analysis suggests presence of various common stress related *cis*-acting elements (CREs) in the putative promoters of members of this gene family. Our results strengthen the idea that plants are comprised of numerous genes with an exceptionally wide range of functionalities along with huge number of promoters and regulatory elements and more remains to be unravelled. With advances in bioinformatics, functional genomics will aid in understanding the molecular and physiological basis to improve the salinity tolerance for sustainable crop production.

INTRODUCTION

High salinity is a prevalent abiotic stress limiting the productivity and the geographical distribution of plants. During the past decade, our understanding of plant response to salt stress has been greatly improved by studying model plants. Salt tolerance is believed to be achieved by many different genes involved in different pathways, such as ion compartmentation, ion extrusion, ion selectivity, compatible solute synthesis and reactive oxygen species (ROS) scavenging (Blumwald et al, 2000; Zhu, 2003, 2016; Munns and Tester, 2008). Cultivated tomato, one of the most important vegetable crops in the world, is moderately sensitive to salinity. Abiotic stress is the negative impact of non-living factors on plants in a specific environment. Cold, salinity and drought are among the major abiotic stresses that adversely affect plant growth and productivity. In fact, these abiotic stresses represent the main cause of crop failure worldwide, leading to average yields of major crops. Stress signals are transmitted through the plasma membrane (many times due to hormones such as abscisic acid, cytokinins, or ethylene). The signal is transmitted within the cell by secondary messengers (e.g. ROS, Ca^{2+}, or IP molecules). The complex interplay between many different kinds of transcription factors (TFs) within the nucleus is responsible for changes in gene expression levels in response to a given form of stress. The expression of numerous plant genes is regulated by abiotic environmental stresses such as drought, high salinity and cold (Shinozaki, 2007).

Plants have developed several physiological and biochemical strategies to adapt or tolerate osmotic stress conditions and deal with salt injury. Reduction of sodium ions in the cytoplasm and the accumulation of compatible, low molecular weight protective compounds called osmolytes have been suggested as the two major mechanisms that underlie the adaptation or tolerance process (Hasegawa et al, 2015). In addition to such metabolic changes, a large set of plant genes commonly called *LEA* (Late Embryogenesis Abundant) are transcriptionally activated, which leads to the accumulation of novel proteins in the vegetative tissues of plants under osmotic stress (Skriver and Mundy, 1990). It is generally assumed that these stress induced LEA proteins might play a protective role in tolerance which is essential for plant survival during episodes of various stress conditions. However, direct and clear experimental evidence supporting the exact functions of LEA proteins is still lacking and the physiological roles of many stress-responsive genes remain largely unknown or poorly understood. Studies conducted with several indica varieties of rice show that the levels of Group 2 LEA proteins (also known as dehydrins) and Group 3 LEA proteins in roots were significantly higher or induced by abscisic acid (ABA) and salt stress only in salt tolerant varieties as compared to salt sensitive varieties (Moons et al, 1995). Our earlier communication (Roychowdhury et al, 2007) showed that *Rab16A* gene is expressed at a much higher and constitutive level in salt tolerant rice cultivar while the transcript is almost undetectable in salt sensitive variety even upon salt stress (Kim et al, 2012, 2015). This probably gives an indication of the involvement of Group 2 and Group 3 *LEA* genes in salt tolerance.

The elucidation of transcriptional regulation in plant genes is important area of research for plant scientists, following the mapping of various plant genomes, such as *A. thaliana*, *O. sativa*, *S. lycopersicum* and *Z. mays*. A variety of bioinformatic servers or databases of plant promoters have been established. The combinatorial interaction of TFs is

important in regulating the gene group that is associated with the same expression pattern. Therefore, a tool for detecting the co-regulation of TFs in a group of gene promoters is required. Investigations on TFs and their corresponding CREs in promoters have attracted much attention from researchers of gene regulation. However, defining all functional binding sites within an identified promoter is difficult, and the existence of some additional binding sites should be assumed (Wray, 2003).

In the present study, we have analyzed GTPases on the basis of sequence, motif identification, protein sequences, and active site prediction. Such computational analyses provide insight into the structure and function of target biomolecules. These results may aid to build up a view of these Rab proteins and their role in plant stress response. Here we report the putative promoters of 4*Rab* genes that contain stress-responsive CREs which may be responsible for regulating the expression of these genes under abiotic stress conditions.

Binding of TFs to CREs or transcription factor binding sites (TFBSs) are involved in transcriptional regulation of gene expression. Discovering *cis*-regulatory elements has been an important research challenge for some years (FitzGerald, 2004; Hernandez-garcia, 2014). With the availability of genome sequences, the computational approach provides an alternate low-cost method to find CREs that can effectively deal with a large number of genes (Bussemaker, 2000; Cora`, 2005). In recent years, many computational approaches have been proposed to find putative CREs. To evaluate the accuracy, computationally discovered CREs are compared with the known TFBSs from public databases or published literatures. This type of evaluation, however, does not validate or associate the putative CREs with biological functions. Another disadvantage of the other studies is that they did not investigate the relationships between the putative CREs. The expression of a gene is usually not regulated by a single TF, but by clusters of TFs that might bind to different CREs. Therefore by exploring the combinatorial regulation of gene expression, one can obtain a better understanding of the complex gene regulation machinery. This study will endeavour to determine the coexistence of putative CREs and will also take advantage of phylogenetic footprinting to improve the prediction accuracy in the search of CREs.

MATERIALS AND METHODS

Biocomputation: Methodology

The selection of the DNA sequences for the identification of Transcription Factor Binding Sites (TFBS) is driven by the typical location of the binding sites.The detailed information of stress-related *cis*-element sequences and annotation is found in Table 1/ 2, whereas the position and abundance of all *cis*-elements predicted to localize in promoter regions of targeted *Rab* genes are shown in Figure 3. Based on our previous transcriptomic study where *Rab* genes were found upregulated in tomato plant under salinity stress, all complete mRNA coding sequences of tomato *Rab* genes (Table1 and Fig.1) were collected from the RefSeq database of NCBI (http://www.ncbi.nlm.nih.gov). Constructed gene structures with noncoding and coding regions have been shown in Figure 1. In order to recognize the upstream promoter region, nucleotide sequences of 1.0 Kbp extending 5′ from the genes' translation start site were identified (https://solgenomics.net). Genome sequences and gene coordinates stored at The Institute for Genomic Research (TIGR), Sol Genomic Network (SGN), NCBI were used for the analyses. 1.0 Kbp of the 5′ upstream promoter region of each gene was scanned for the presence of *cis*-acting regulatory elements involved in abiotic stress signalling pathways using PLACE, PlantPan and Plant CARE program (www. Plant CARE. Com/encyclopedia). A number of 4 *Rab* genes (*Rab1A/1B/1C and Rab11a*) were nominated for the network study based on containing the most varied motifs involved in abiotic stress tolerance. In order to identify the varied CREs in the putative promoter region and analyze co-expressed genes, the MEME and PlantPan web tools were used. The domain architectures of the deduced proteins (Fig. 3) were identified based upon their sequence alignment using various tools like SMART, MAFFT, MUSCLE and ClustalX/ClustalW.

Table 1. List of genes under study with annotations.

Gene name *Solanum lycopersicum*	**Accssn. No. of NCBI**	**Sizes of the cDNAs base pair**	**Source tomato sp. & Annotation**
Rab1A	U38464	611	LA1221, Small GTP binding protein
Rab1B	U38465	611	LA1221, Small GTP binding protein
Rab1C	U38466	611	LA1221, Small GTP binding protein
Rab11a	AJ245570	656	Ailsa Craig, Rab11GTPase putative role in secretion in cell wall

Table 2. List of regulatory motifs with consensus sequences. Motif logo generated using MEME (http://meme-suite.org/) Nucleotides: A-Adenine; C-Cytosine; G-Guanine; T-Thymine; M-A or C; R-A or G;W-A or T; S-C or G; Y-C or T; K -G or T; B -C., G or T; D -A, G or T; H -A, C or T; V -A, C or G; N –Any

***cis*-Regulatory Elements**	**Consensus Sequence**	**Motif logo**
ABRE	MACGYGB	
CE1	SSBCACCSV	
LTRE	CCGAC	
WRKY	WTGACH	
MYB	CNGTTR	
MYC	CANNTG	
ELRE	TTGACC	
DPB	ACACNNG	
GATA	WGATAR	

Analysis by PLACE Database (PLACE) (Higo et al, 1999) and PLANTPAN 2.0

Characterisation of 1 kb upstream sequences: Normally, putative regulatory regions have been located in within 500 bp and 1 kb upstream of the transcription or translation start point (Goda, 2004; Thijs et al, 2001). Therefore, considering that the analysis of intergenic regions revealed a median length of 1 kbp, the computational prediction of TFBSs was made using 1 kbp long sequences upstream of the start of a coding region. The default length of 1 kb was applied in all cases, including intergenic regions smaller than 1 kb, because it was not discarded that some coding sequences exert regulatory actions on a neighboring gene, moreover if the intergenic sequence of the neighboring gene is too short.

Identifying cis-regulatory elements in the plant genome: Many databases are there having collections of numerous TFs and found useful for the prediction of TFBSs in the promoter regions of plants. For instance, PLACE (Higo, 1999) is a database that collects various *cis*- and *trans*- acting regulatory DNA elements. TRANSFAC (Matys, 2006) is a database of TFs, including genomic binding sites and DNA-binding profiles. JASPAR (Bryne, 2008) is an open-access database of annotated, high-quality, matrix-based TFBS profiles for multicellular eukaryotes. Athena (O'Connor, 2005) is a database, which contains 30,067 predicted *Arabidopsis* promoter sequences and consensus sequences for 105 previously characterized TFBSs and provides analysis on over-represented TFBSs occurring in multiple promoters. PlnTFDB (Riano-Pachon, 2007) is an integrative plant transcription factor database that provides a web interface to access large sets of TFs of several plant species. AGRIS (Davuluri, 2003) contains an *Arabidopsis thaliana* transcription factor database (*At*TFDB) consisting of approximately 1,770 *Arabidopsis* TFs and their sequences (protein and DNA) grouped into around 50 families

with information on available mutants in the corresponding genes. AGRIS (Davuluri, 2003) integrates a variety of tools to determine TFs and their putative binding sites on all genes to reconstruct transcriptional regulatory networks in *Arabidopsis*. DATF (Guo, 2005) stores information on 3D structural templates, EST expression, TFBSs and nuclear location signals (NLSs) of known and predicted *Arabidopsis* transcription factors. PlantCARE (Lescot, 2002) is a database of plant CREs and a portal to tools for the *in silico* analysis of promoter sequences.

Access to PlantPAN is via a web interface, freely available to all interested users, at http://PlantPAN.mbc.nctu.edu.tw. CREs published in PLACE and AGRIS were downloaded. The visual representation of the motif logos was constructed (Table 2 and 3). There are so many recent techniques being used for analysis of stress response in plants. Transcriptome is the set of all mRNAs/transcripts produced under a given set of conditions. Transcriptomic data (Roy, 2015) was taken from our previous work for this study. Transcriptomics is always considered as a step next to genomics in the study of biological systems.

RESULTS AND DISCUSSIONS

Upstream and protein sequences of 4 Rabs:

To better understand plant stress responses, transcript profiling experiments have been successfully employed for many different abiotic and biotic stresses. Based on our previous differential expression of 10 stress responsive genes, we carried out comprehensive searches for the stress related CREs in the upstream regions of 4 genes (*Rab1A, Rab1B, Rab1C and Rab11a*). To fully understand the role of the 4 genes under study in tomato plants, it is therefore essential to characterize these genes. Studying the structure of genes of interest is crucial in biology and can provide important clues concerning gene evolution. Accordingly, we analyzed the promoters of the up-regulated genes to identify potential TF-binding motifs important for activating stress resistance. We also selected elements as more abundant CREs in promoter regions of these genes without mention of common CREs such as CAAT-box and TATA-box. The both of them have fundamental role in initiation of transcription process and are found in all the promoter regions of genes.

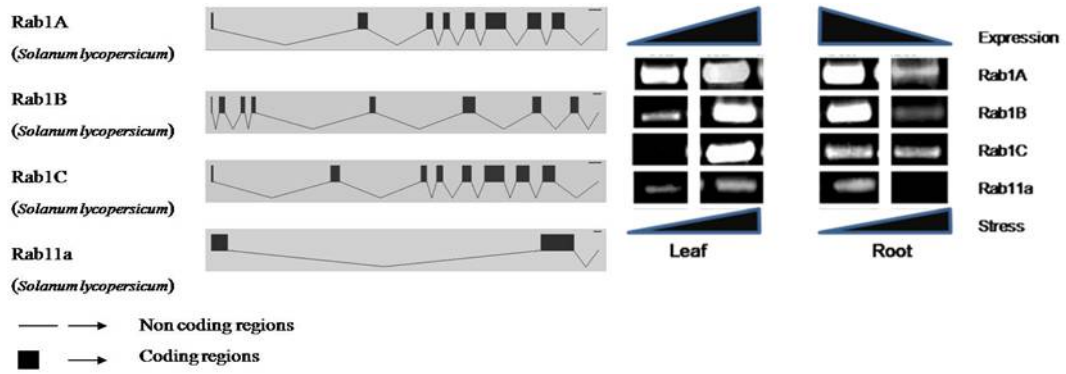

Figure 1. A schematic representation of four candidate genes with coding exon and noncoding intron. The expression profile at the right side showing expression pattern of 4 *Rab* genes(*Rab1A, Rab1B, Rab1C* and *Rab11a*) in tomato(var. Punjab Keshari) under salt stress.

Genes of LEA proteins have been identified in many plant species, and at least six different groups of LEA proteins have been defined on the basis of expression pattern and sequence. Although a series of Group 2 *LEA* genes have been isolated and their accumulation

correlated with exogenous stress (Roychowdhury et al, 2007), The Gr2 proteins, also called dehydrin or RAB (responsive to ABA) proteins (Close et al, 1989). Group 2 proteins are characterized by up to three sequence motifs, known as the K-domain (lysine-rich), the Y-domain (DEYGNP) and the S-segment (poly-serine stutter). In our previous findings (Roy, 2015) all the 4 *Rabs* were up regulated (higher expression detected in cv Punjab Keshari than cv Pusa Ruby of tomato) in salt treated tomato leaves only (down regulated in treated roots). It is known that protein functions are closely related to their structures. Predicted amino acid sequences (Fig. 2) of the Rab protein were aligned using multiple sequence alignment software (MUSCLE). The degree of similarity was lower for the Rab1A than of Rab1B and Rab1C plant GTP-binding proteins; Similarly, lower sequence identities were observed between 11a and 1A/B/C. The four regions (Fig. 2 boxed area) required for nucleotide binding (Bednarek, 1994), characteristic of Rab proteins in sequence and spacing are present in these Rab proteins. These sequence segments, GDQSVGKTS, WDTAG , NKTD and ETSA were found identical. The effector region (Fig.2 SF1-SF4) for the genes above, sequences are completely alike. All information about gene structure, motif distribution, and protein size of *Rab* genes under study support the conserved *Rab* genes in the same groups shown in Figure 2.

Increasing evidence indicates that genes with the similar expression patterns may contain the same regulatory elements in their promoters (Wittkopp et al, 2012; Uygun et al, 2016). In general, signalling networks involve *cis*-elements working with their cognate TFs. The tomato genome was sequenced as the cornerstone of an International Solanaceae Genome Initiative, a project that aims to develop the family Solanaceae as a model for systems biology for understanding plant adaptation and diversification. The tomato genome comprises approximately 950 Mb of DNA. The sequencing of the tomato genome and sequencing of the wild relative were achieved and published in the SGN database (http://solgenomics.net/tomato/). The putative promoter regions of the 4 genes under study retrieved from SGN showed several putative environment stimulus responsive CREs when scanned through PlantPan/PLACE. The detailed information of stress-related cis-element sequences and annotation is found in Table1 and 2, and the position and abundance of all CREs predicted to localize in promoter regions of tomato *Rab* genes are shown in Fig 3. We found surprising differences in the numbers of *cis*-elements in their upstream regions. The evaluated CREs represented a subset of 49 *cis*-elements retrieved from the databases PLACE, SOL Genomics and AGRIS corresponded varying oligonucleotides. This study reports TFBSs correspond to nine different consensus oligonucleotides including ABRE, WBOX, CRTDRE, MYB, MYC, DPB, GATA, LTRE, ELRE.

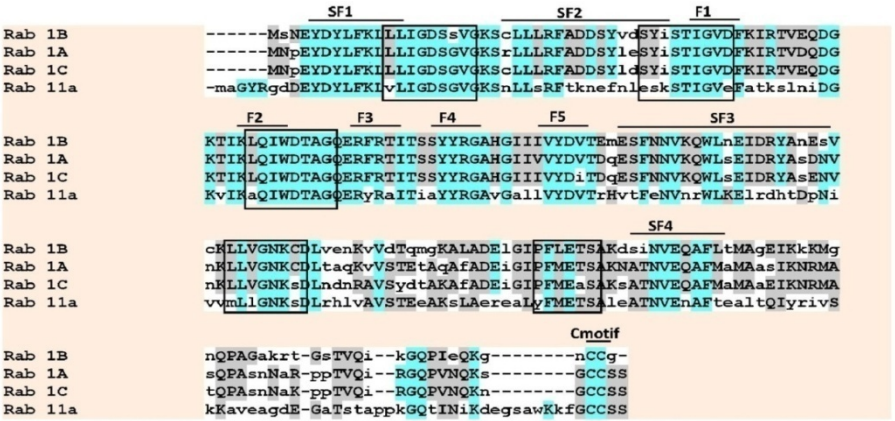

Figure 2. Multiple Sequence Alignment of Rab1A, Rab1B, Rab1C and Rab11a with MUSCLE software. Similar residues are coloured as the most conserved one (according to BLOSUM62). Average BLOSUM62 score: Max: 3.0 Mid: 1.5 Low: 0.5

GDP/GTP binding domains (G1–G5) are boxed which are important for guanine nucleotide interactions and GTP hydrolysis. Other five Rab family-specific regions (F1-F5) conserved for the Rab family and the four Rab subfamily specific regions (SF1-SF4), which are suggested to be recognized by Rab subfamily effectors. A Cys–Cys motif (C motif) at the carboxyl-terminal of 4 Rabs is essential for prenylation, a crucial modification necessary for its action as a molecular switch regulating intracellular vesicular transport and correct association with membranes.

The CREs elements present in the putative upstream region are described below (Fig.3 and Table 3):

ABRE: ABRE (GCCACGTGCA): ABA response complex consisting of ABRE element and a novel coupling element CE1 (TGCCACCGG) was sufficient for high-level ABA induction. The interaction between these 2 elements determined the specificity in ABA regulated gene expression as detected from barley (Shen and Ho, 1995). In our upstream analysis, ABRE elements were found in 5' flanking region between position –900 and –200 from transcription start site in *Rab* genes except in *Rab11a*.

LTRE: (CCGAC): is a low temperature responsive element (LTRE) found in *cor15a* gene in *Arabidopsis thaliana* . This element involved in cold induction of *BN115* gene from the winter *Brassica napus* and light signalling mediated by phytochrome for cold or drought induced gene expression in *Arabidopsis thaliana* (Jiang et al, 1996). In our analytical study it was identified as a single occurence in the flanking region of *Rab1B* gene only at position in between -900 to –1000 from transcription start site. ABRE (Abscisic Acid Responsive Element) and DRE/LTRE (for Dehydration Responsive Element/Low Temperature Responsive Element) are the two major *cis*-acting elements involved in the regulation of gene expression in response to osmotic stress in ABA-independent and ABA-dependent signalling pathways respectively (Yamaguchi-Shinozaki, 2005).

WBOX/WRKY: (T/A)TGAC(T/A): The WRKY domain forms a unique wedge-shaped structure that enters perpendicularly in the major groove of the DNA strand. WRKY protein domains interact with the (T/A)TGAC(T/A) *cis*-element, also called the W-box. In our study this group of *cis*-elements were noticed scattered in all the 4 genes with rich number. (TTTGACT) was found in the Parsley WRKY3 gene promoter that required for elicitor responsiveness. WB box and WC box constitute a palindrome WRKY1 protein binding site that play an important role in the

regulation of early defence response gene (Eulgem et al, 2000). They are recognized specifically by salicylic acid (SA)-induced WRKY DNA binding proteins.

MYC: (CANNTG): MYC recognition site found in the promoters of the dehydration–responsive gene (Abe et al, 2003). 2 to 6 of this element was observed in the upstream sequences of the candidate genes in the 5’ flanking region between position –200 and –1100 from transcription start site.

MYB: (CNGTTR): Plant MYB proteins ATMYB1 and ATMYB2, both isolated from *Arabidopsis*; ATMYB2 is involved in regulation of genes that are responsive to water stress in *Arabidopsis*. The maize cl gene was the first plant transcription factor described that encoded a MYB protein, and is involved in the regulation of anthocyanin biosynthesis in seed development (Paz-Ares et al, 1987). Our analysis shows presence of MYB mostly in *Rab1C* and few in *Rab1B* too.

Elicitor-responsive element (ElRE): It is W-box-like (TTGACC) (Rushton et al, 1996). V K Srivastava reported (2014), ELRECOREPCRP1 a well defined component of promoter region responsible for defence signalling with defined regulatory elements in pathogen- and wound induced signalling. Here this element was noticed in *Rab11a* only.

DPB: DPB (ACACNNG): is a novel class of bZIP TFs. DPBF-1 and 2 binding core sequences were found in carrot (*Daucus carota*) *Dc3* gene promoter. Dc3 expression is normally embryo specific and also can be induced by ABA (Kim et al, 1997). Our findings exhibit lesser of such CREs in *Rab1A, Rab1B, Rab1C* in between the region –700 to –800 whereas none was detected in *Rab11a*.

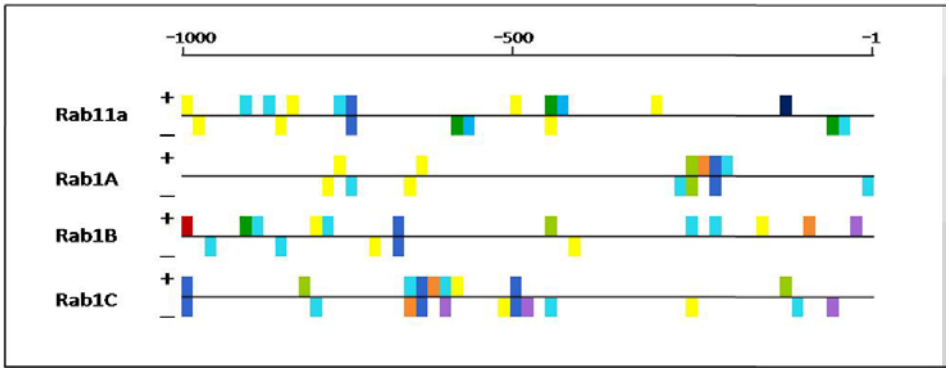

Figure 3. Cis-regulatory elements: **ABRE-, CE1-, LTRE-, WRKY-, MYB-, MYC-, ELRE-, DPB-, GATA-**
Positions of stress-responsive predictive *cis*-elements in the putative promoters of Rab family genes (tomato) with respect to start codon (ATG). The 1000 bp putative promoter regions of corresponding *Rab* genes were used for analysis, using colour code. Nine selected stress responsive *cis*-elements are: ABA responsive element (ABRE), WRKY binding site (W-box), CE1, LTRE, DPB, ELRE, MYC, MYB and GATA

GATA Box: (ATGATAAGG, WGATAR): Based on sequence analysis, Grob and Stuber (1987) identified a sequence motif 5’ -ATGATAAGG-3’ that is present in many *LHC* and *rbcS* genes from a number of plant species. Presence of core GATA motif in the set of relative modules in the promoters of many light regulated genes has been reported. Also, GATA motifs are known

to be often associated with other LREs including G-boxes. Plant GATA factors typically contain a single zinc finger. All the upstream sequences here show predominant presence of this motif.

CE1: (TGCCACCGG): ABRE alone is not sufficient, and another *cis*-element known as "coupling element (CE)" is required for full range ABA-regulation of gene expression. CE1 is necessary for the ABA-regulation of *HVA22* gene (Shen, 1996). In *RD29A* gene, DRE (Dehydration-responsive element, TACCGACAT) functions as a coupling element to ABRE (Narusaka, 2003). This CRE was only identified in the upstream region of *Rab11a* only at –150 site.

Variance and co-occurences of CREs: Considerable variation in the number of CREs exists across the upstream regions of *Rab1A, Rab1B* and *Rab1C.* The ABA-responsive element (ABRE; PyACGTG/TC) is a well-studied CRE involved in ABA-induced gene expressions (Fujita, 2011; Hattori, 2002)**.** ABRE-binding Protein/factors (AREB/ABF) have positive effects on the osmotic stress tolerance of plants (Kang, 2002; Kim, 2004a,b; Yoshida, 2015). Presence of ABREs in the promoters of the genes under study indicated a role for ABA-signalling in the regulation of their expressions. Absence of ABRE in Rab11a may supported by the reports which say that, one exception to this is the MYB- and MYC-regulated RD22 gene, which lacks any ABREs (Abe, 1997). However, all of these arrangements allow for regulation by varying combinations of TFs whose identities are determined by the availability of specific regulators and binding sites. In the upstream analysis of our selected *Rab* genes the most frequently found elements are namely MYB, MYC, GATA and WBOX. A rather large MYB family, reported from different plants, plays a variety of key roles in the regulation of gene expression, and is also related to transcriptional responses to hormones during seed development and germination. For example, the maize *C1* gene regulates the expression of genes that are involved in the biosynthesis of anthocyanin pigments in the aleurone (Cone et al, 1986; Paz-Ares et al, 1986, 1987; Hattori et al, 1992). Putative MYB binding sites involved in the up-regulation of genes in plants overexpressing AtMYB2 (Abe, 1997), and regulatory sequences involved in the regulation by cold stress (Hannah, 2005). Interestingly, the MYC and MYB consensus sequences were reported in the *rd22* promoter (*Arabidopsis*), which do not contain any typical ABRE recognition site (Yamaguchi-Shinozaki and Shinozaki, 1993). However, both AtMYC2 and AtMYB2 genes are induced by drought and ABA treatment, suggesting that AtMYB2 may regulate cooperatively with AtMYC2 in another regulatory system other than the ABRE-bZIP regulatory system in the ABA signalling pathway in vegetative tissues and seeds under drought and salt stress (Iwasaki et al, 1995; Abe et al, 1997, 2003). It is worth mentioning that an important element like CRT/DRE was found totally missing in the upstream region of tomato *Rab* genes (*Rab11a, Rab1A, Rab1B, Rab1C*).

Table 3. Co-occurrence of transcription factor binding sites in gene group of *Rab1A,Rab1B, Rab1C* and *Rab11a* of *Solanum lycopersicum*. 1kbp Upstream sequences of the genes were scanned and *cis*-regulatory elements along with their frequency were searched from different databases like TRNSFAC/PLACE/PLANTPAN/PLANTTFDB. Elements from both the strands of the upstream sequences were considered for comparison.

List of *cis*-elements	Rab1A	Rab1B	Rab1C	Rab11a
ABRE	2	1	2	-
CE1	-	-	-	1
DPB	1	1	2	-
ELRE	-	1	-	3
GATA	4	4	3	7
LTRE	-	-	1	-
MYB	-	1	3	-
MYC	2	2	6	2
WRKY	4	6	5	6

Rab genes under study were found simultaneously up regulated in salt treated leaves and down regulated in treated roots and as reported in our earlier report (Roy, 2015). Because co-expressed genes tend to behave similarly, they are expected to be co regulated. Under the simplifying assumption that this co regulation occurs at the transcriptional level, co-expressed genes should contain similar CREs in their promoter regions. As a consequence, these yet unknown CREs will be statistically overrepresented in the intergenic regions of the co expressed genes in comparison with their frequent occurrence in a set of unrelated sequences. Co-occurrence of TFBSs means that these CREs do not act alone but in networks, along with many other elements within the promoter (Nguyen, 2006). In Table 3, presentation shows co-occurrence of *cis*-elements like ABRE, ACGT, WBOX, WRKY, MYB/MYC and GATA. The underlying assumption here is that since proteins which all take part in the same biochemical process or pathway are basically to a large degree localized in the same cell compartment, and that since many of these genes are active at just about the same time, then this means that all or most of the genes should be regulated by common regulatory factors, and therefore most or all contain similar CREs (Walhout, 2006). Here the statistically significant co-occurrence of motifs with each other is what is used to define the regulation of a set of genes, which usually make up a genetic regulatory network. In this way, even those motifs can be detected which have a low occurrence, if they co-occur relatively many times along with another motif. For example, the program RiCES (Rice *Cis*-Element Searcher) uses likelihood statistics in order to discover significant relationships between pairs of motifs (Doi, 2008). A number of scientists have studied synergistic effects of pairs of motifs acting in concert with each other in order to increase the level of expression of genes that they co-occur in, compared to gene expression where only one of the motifs are present. These include studies of the ABRE and DRE elements (Zhang, 2005; Nakashima et al, 2014).

CONCLUSIONS

Present computational work is based on simple assumption that genes with similar expression profiles are considered “co-expressed” as well as co-regulated. Four *Rab* genes in the

current study were found up-regulated in tomato plant (cv. Punjab Keshari) under salinity stress in our previous comparative transcriptional experiment. These four candidate genes found to contain similar CREs in their putative promoter regions. Conserved domains in the amino acid sequences show their functional and structural conservation during evolution. Various stress related putative CREs including ABA, salt, cold and dehydration were commonly identified with variable frequencies. Two known TF binding motifs, the W-box and the GATA-box, were found to be significantly enriched in putative promoters of these up-regulated genes. The CREs under study show their highest abundance in the upstream region of *Rab1C*. This observation is in agreement with our previous data demonstrating highest accumulation of this transcript in leaves and roots of tomato plant under severe salt stress. Number of CREs can be critically important in attraction of cognate TFs indicating that more copies of motifs led to greater promoter activity.

This information may aid to manipulate the cell function through preparation of synthetic promoters using different motifs. Such knowledge is also important in understanding the cross-talks between distinctive signalling pathways and the underlying regulatory machinery of cellular stress response.

REFERENCES

Abe, M., Katsumata H., Komeda, Y., Takahashi, T., (2003): Regulation of shoot epidermal cell differentiation by a pair of homeodomain proteins in *Arabidopsis*. Development, 130, 635–643.

Abe, H., Yamaguchi-Shinozaki, K., Urao, T., Iwasaki, T. and Shinozaki, K., (1997): Role of MYC and MYB homologs in drought- and abscisic acid-regulated gene expression. Plant Cell, 9, 1859–1868.

Bednarek, S.Y., Reynolds, T.L., Shroeder, M., Grabowski, R., Hengst, L., Gallwitz, D., Raikhel, N. V., (1994**):** A small GTP-binding protein from *Arabidopsis thaliana* functionally complements the yeast ypt6 nu11 mutant. Plant Physiol, 104, 591–596.

Blumwald , E,, Aharon, G. S., Apse, M. P., (1994): Sodium transport in plant cells. Biochim Biophys Acta, 1465, 140–:51.

Bussemaker, H. J., et al., (2000): Building a dictionary for genomes: identification of presumptive regulatory sites by statistical analysis. Proc Natl Acad Sc USA, 97: 10096–10100.

Close, T. J., Kort, A. A., Chandler, P. M., (1994): A cDNA-based comparison of dehydration-induced proteins (dehydrins) in barley and corn. Plant Molecular Biology, 13, 95–108.

Cone, K. C., Burr, F. A., Burr, B., (1986): Molecular analysis of the maize anthocyanin regulatory locus C1. Proc Natl Acad Sci USA, 83: 9631–9635.

Corà, D., Herrmann, C., Dieterich, C., Cunto, F. Di., Provero, P., Caselle, M., (2005): Ab initio identification of putative human transcription factor binding sites by comparative genomics. BMC Bioinformatics, 6:110

Davuluri, R. V., et al., (2003): AGRIS: Arabidopsis gene regulatory information server, an information resource of Arabidopsis cis-regulatory elements and transcription factors. BMC Bioinformatics, 4, 25.

Doi, K., Yasui, H., Yoshimura, A., (2008): Genetic variation in rice. Curr Opin Plant Biol, 11(2), 144–8.

Eulgem, T., Somssich I. E., (2007): Networks of WRKY transcription factors in defense signalling. Curr Opin Plant Biol, 10(4), 366–371.

FitzGerald , R. J., Murray, B. A., Walsh, D. J., (2004): Hypotensive peptides from milk proteins. J Nutr, 134(4), 980S–8S.

Fujita, Y., Fujita, M., Shinozaki, K., Yamaguchi-Shinozaki, K., (2011): ABA-mediated transcriptional regulation in response to osmotic stress in plants. J Plant Res, 124, 509–525.

Goda, H., Sawa, S., Asami, T., Fujioka, S, Shimada, Y., et al., (2004): Comprehensive comparison of auxin-regulated and brassinosteroid-regulated genes in *Arabidopsis*. Plant Physiol, 134, 1–19.

Grob, U., Stüber, K., (1987): Discrimination of phytochrome dependent light inducible from non-light inducible plant genes. Prediction of a common light-responsive element (LRE) in phytochrome dependent light inducible plant genes. Nucleic Acids Res, 15(23), 9957–9973.

Guo, H. S., Xie, Q., Fei, J. F., Chua, N. H., (2005): MicroRNA directs mRNA cleavage of the transcription factor NAC1 to downregulate auxin signals for *Arabidopsis* lateral root development. Plant Cell, 17, 1376–1386.

Hasegawa, P. M., (2013): Sodium (Na^+) homeostasis and salt tolerance of plants. Environ Exp Bot, 92, 19–31.

Hattori, T., Vasil, V., Rosenkrans, L., Hannah, L. C., McCarty, D. R., Vasil, I. K., (1992): The viviparous-1 gene and abscisic acid activate the C1 regulatory gene for anthocyanin biosynthesis during seed maturation in maize. Genes Dev, 6, 609–618.

Hattori, T., Totsuka, M., Hobo, T., Kagaya, Y., Yamamoto-Toyoda, A., (2002): Experimentally determined sequence requirement of ACGT-containing abscisic acid response element. Plant Cell Physiol, 43, 136–140.

Hannah, M. A., Heyer, A. G., Hincha, D. K., (2005): A global survey of gene regulation during cold acclimation in Arabidopsis thaliana. PLoS Genet.; 1: e26

Hernandez-garcia, C. M., Finer, J. J., (2014): Identification and validation of promoters and -acting regulatory elements. Plant Science, 217, 109-119.

Higo, K., et al., (1999): Plant cis-acting regulatory DNA elements (PLACE) data.base. Nucleic Acids Res, 27, 297–300.

Iwasaki, T., Yamaguchi-Shinozaki, K. and Shinozaki, K., (1995): Identification of a cis-regulatory region of a gene in *Arabidopsis thaliana* whose induction by dehydration is mediated by abscisic acid and requires protein synthesis. Mol Gen Genet, 247, 391–398.

Jiang, C., Iu, B., Singh, J., (1996): Requirement of a CCGAC *cis*-acting element for cold induction of the *BN115* gene from winter *Brassica napus*. Plant Mol Biol, 30, 679–684.

Kang, J. Y., Choi, HI., Im, M. Y., Kim, S.Y., (2002): Arabidopsis basic leucine zipper proteins that mediate stress-responsive abscisic acid signalling. Plant Cell, 14, 343–357.

Kim, T. K., Lagrange, T., Wang, Y. H., Griffith, J. D., Reinberg, D., Ebright, R. H., (1997): Trajectory of DNA in the RNA polymerase II transcription pre-initiation complex. Proc Natl Acad Sci ,USA., 94, 12268–12273.

Kim, S., Choi, H. I., Ryu, H. J., Park, J. H., Kim, M. D., Kim, S. Y., (2004a): ARIA, an *Arabidopsis* arm repeat protein interacting with a transcriptional regulator of abscisic acid-responsive gene expression, is a novel abscisic acid signalling component. Plant Physiol, 136, 3639–3648.

Kim, S., Kang, J. Y., Cho, D. I., Park, J. H., Kim, S. Y., (2004b): ABF2, an ABRE-binding bZIP factor, is an essential component of glucose signalling and its over expression affects multiple stress tolerance. Plant J, 40,75–87.

Kim, H., Hwang, H., Hong, J. W., Lee, Y. N., Ahn, I. P., Yoon, I. S., Yoo, S. D., Lee, S., Lee, S. C., Kim, B. G., (2012): A rice orthologue of the ABA receptor, sPYL/RCAR5, is a positive regulator of the ABA signal transduction pathway in seed germination and early seedling growth. Journal of Experimental Botony, 63, 1013–024.

Lescot, M., Dehais, P., Thijs, G., Marchal, K., Moreau, Y., Van de Peer, Y., Rouze, P. and Rombauts, S., (2002): PlantCARE, a database of plant *cis*-acting regulatory elements and a portal to tools for in silico analysis of promoter sequences. Nucleic Acids Res, 30, 325–327.

Matys, V., et al., (2006): TRANSFAC and its module TRANSCompel: transcriptional gene regulation in eukaryotes. Nucleic Acids Res, 34, D108–D110.

Moons, A., Bauw, G., Prinsen, E., V an Montagu, M., Van der Straeten, D., (1995): Molecular and physiological responses to abscisic acid and salts in roots of salt-sensitive and salt-tolerant indica rice varieties. Plant Physiol, 107, 177–186.

Munns, R. & Tester, M., (2008): Mechanisms of salinity tolerance. Annual Review of Plant Biology, 59, 651–681.

Nakashima, K., Yamaguchi-Shinozaki, K and Shinozaki, K., (2014): The transcriptional regulatory network in the drought response and its crosstalk in abiotic stress responses including drought, cold, and heat. Front Plant Sci, 5:170.

Narusaka, Y., Nakashima, K., Shinwari, Z. K., Sakuma, Y., Furihata, T., Abe, H., Narusaka, M., Shinozaki, K., Yamaguchi-Shinozaki, K., (2003): Interaction between two *cis*-acting elements, ABRE and DRE, in ABA-dependent expression of *Arabidopsis rd29A* gene in response to dehydration and high-salinity stresses. Plant J, 34 (2), 137–148.

Nguyen, D. H. and D'haeseleer, P., (2006): Deciphering principles of transcription regulation in eukaryotic genomes. Molecular Systems Biology, 2, 2006–0012.

O'Connor, T. R., et al., (2005): Athena: a resource for rapid visualization and systematic analysis of Arabidopsis promoter sequences. Bioinformatics, 21, 4411–4413.

Paz-Ares, J., Wienand, U., Peterson, P. A. and Saedler, H., (1986): Molecular cloning of the *c* locus of *Zea mays*: a locus regulating the anthocyanin pathway. Embo J, 5, 829–833.

Paz-Ares, J., Ghosal, D., Wienand, U., Peterson, P. A., Saedler, H., (1987): The regulatory c1 locus of *Zea mays* encodes a protein with homology to myb proto-ncogene products and with structural similarities to transcriptional activators. EMBO J, 6, 3553–3558.

Riano-Pachon, D. M., Ruzicic, S., Dreyer, I., Mueller-Roeber, B., (2007): PlnTFDB: an integrative plant transcription factor database. BMC Bioinformatics, 8, 42.

Roy, C., (2015): A comparative molecular analysis of two genotypes of *Solanum lycopersicum* in respect of the expression of selected salt responsive genes in leaves and roots. Biotechnologia, 96, 307–316.

Roychoudhury, A., Roy, C., Sengupta, D. N., (2007): Transgenic tobacco plants overexpressing the heterologous *lea* gene *Rab16A* from rice during high salt and water deficit display enhanced tolerance to salinity stress. Plant Cell Report, 10, 299–371.

Rushton, P. J., Torres, J. T., Parniske, M., Wernert, P., Hahlbrock, K., Somssich, I. E., (1996): Interaction of elicitor-induced DNA-binding proteins with elicitor response elements in the promoters of parsley PR1 genes. EMBO Journal, 155690–5700.

Shen, Q. and Ho, T-HD., (1995): Functional dissection of an abscisic acid (ABA)-inducible gene reveals two independent ABA-responsive complexes each containing a G-box and a novel *cis*-acting element. Plant Cell, 7, 295–307.

S**hen Q, Zhang P, Ho T-HD., (**1996): Modular nature of abscisic acid (ABA) response complexes: composite promoter units that are necessary and sufficient for ABA induction of gene expression in barley. Plant Cell. ; 8**:** 1107–1119.

Shinozaki, K.,Yamaguchi-Shinozaki, K., (2007): Gene networks involved in drought stress response and tolerance.J. Exp Bot, 58,221-227.

Srivastava, V. K., Raikwar, S., Tuteja, N., (2014): Cloning and functional characterization of the promoter of *PsSEOF1* gene from Pisum sativum under different stress conditions using Agrobacterium-mediated transient assay. Plant Signal Behav, 9, e29626.

Shriver, K., Mundy, J., (1990): Gene expression in response to abscisic acid ancl osmotic stress. Plant Cell, 2, 503–512.

Thijs, G., Lescot, M., Marchal, K., Rombauts, S., De Moor, B., Rouzé, P., Moreau, Y., (2001): A higher-order background model improves the detection of promoter regulatory elements by Gibbs sampling. Bioinformatics, 17(12), 1113–22.

Uygun, S., Peng, C., Lehti-Shiu, M. D., Last, R. L., Shiu, S-H., (2016): Utility and Limitations of Using Gene Expression Data to Identify Functional Associations. PLoS Comput Biol, 12(12): e1005244.

Walhout, A. J. M., (2006): Unravelling transcription regulatory networks by protein-dNA and protein-protein interaction mapping. Genome Res, 16, 1445–1454.

Wittkopp, P. J. & Kalay, G., (2012): *Cis*-regulatory elements: molecular mechanisms and evolutionary processes underlying divergence. Nature Reviews Genetics, 13, 59–69.

Wray, G. A., Hahn, M. W., Abouheif, E., Balhoff, J. P., Pizer, M., Rockman M.V., Romano, L.A., (2003): The evolution of transcriptional regulation in eukaryotes. Mol Biol Evol, 20**,** 1377–1419.

Yamaguchi-Shinozaki, K. and Shinozaki, K., (1993):The plant hormone abscisic acid mediates the drought-induced expression but not the seed-specific expression of rd22, a gene responsive to dehydration stress in *Arabidopsis thaliana*. Mol Gen Genet, 238, 17–25.

Yamaguchi-Shinozaki, K. and Shinozaki, K., (2005): Organization of *cis*-acting regulatory elements in osmotic- and cold-stress-responsive promoters. Trends in plant science, 10 (2), 88–94.

Yoshida, .T, Fujita, Y., Maruyama, K., Mogami J., Todaka, D., Shinozaki, K., Yamaguchi-Shinozaki, K., (2015): Four Arabidopsis AREB/ABF transcription factors function predominantly in gene expression downstream of SnRK2 kinases in abscisic acid signalling in response to osmotic stress. Plant Cell Environ, 38: 35–49.

Zhang, W., et al., (2005): *Cis*-regulatory element based targeted gene finding: genome wide identification of abscisic acid and abiotic stress-responsive genes in *Arabidopsis thaliana*. Bioinformatics, 21, 3074–3081.

Zhu, J. K., (2016): Abiotic Stress Signaling and Responses in Plants. Cell, 167(2), 313–324.

Zhu, J. K., (2003): Regulation of ion homeostasis under salt stress. Curr Opin Plant Biol, 6, 441–445.

Acknowledgments: Author is thankful to the Director of Bose Institute. Thanks also go to the Chairperson of the Division of Plant Biology.

[1]Division of Plant Biology, Bose Institute, 93/1 Acharya Prafulla Chandra Road, Kolkata-700009, India
[*] croy.bi@gmail.com