Genome-wide In Silico Identification of Transcriptional Regulators Controlling the Response to Salt Stress in Soybean (Glycine max (L.) Merr.)

Article · February 2021

CITATIONS
0

READS
58

11 authors, including:

Vijayata Singh
Central Soil Salinity Research Institute
39 PUBLICATIONS **154** CITATIONS
SEE PROFILE

Satish Sanwal
Central Soil Salinity Research Institute
76 PUBLICATIONS **441** CITATIONS
SEE PROFILE

Zeetendra Singh
Central Soil Salinity Research Institute
13 PUBLICATIONS **2** CITATIONS
SEE PROFILE

Giriraj Kumawat
ICAR-Indian Institute of Soybean Research, Indore
35 PUBLICATIONS **571** CITATIONS
SEE PROFILE

Journal of Soil Salinity and Water Quality 12(2), 277-289, 2020

Genome-wide *In Silico* Identification of Transcriptional Regulators Controlling the Response to Salt Stress in Soybean (*Glycine max* (L.) Merr.)

Vijayata Singh[1], Satish Kumar Sanwal[1], Zeetendra Singh[1], Giriraj Kumawat[2], Shivakumar M[2], Gyanesh Kumar Satpute[2], Sanjay Gupta[2], Balwinder Singh Gill[3], Nisha Walia[3], Mohan Lal Jakhar[4], Parbodh Chander Sharma[1]* and Jogendra Singh[1]*

[1]ICAR-Central Soil Salinity Research Institute, Zarifa farm, Karnal-132001, Haryana, India
[2]ICAR- Indian Institute of Soybean Research, Khandwa Road, Indore-452001, Madhya Pradesh, India
[3]Punjab Agricultural University, Ludhiana-141004, Punjab, India
[4]SKN Agriculture University, Jobner-303328, Rajasthan, India
**Corresponding author E-mail: jogendra.singh@icar.gov.in/pcsharma.knl@gmail.com*

Abstract

Comprehensive characterization of networks of gene expression regulators is key challenge of functional genomics. Soybean genomic sequences can be used to frame model for binding sites of fundamental transcription factors, and RNA seq data. It has been manifested that the reverse genetics approach can reveal gene regulator networks in soybean, which explains regulatory mechanisms for salt response from expression of gene patterns. In this study, the sequences of protein is functionally validated salt tractable genes is used to key out soybean orthologous with the help of '*blastP*' search against *Glycine max* genome database. Thirty- eight salt responsive genes allocate among 39 *cis*-acting elements of soybean is elucidated by the '*PlantCARE*' database. Gene promoter analysis of salt tractable genes indicated the presence of biotic and abiotic stress responsive, light, hormone, plant tissues specific and gene specific *cis*-regulatory elements in soybean. A total, 38 salt tractable genes in soybean is entrusted for *GO* terms based on the '*Ensembl Plants*' database. These *GO* terms are sum up into the three main categories (biological process, cellular component, and molecular function) and 62 sub-categories. These *GO* annotations comprise a general profile signature of gene expression for salt tractable genes in soybean. Our results explained about the identified soybean salt tractable genes differentially respond in plant growth, development and gene regulation during biotic and abiotic stress conditions. The data obtained from this study contribute to a better enlightening of the role of the regulatory and functional pathway of salt tolerant genes in soybean.

Key words: Soybean, *In silico* identification, Salt stress, *Cis*-regulatory elements

Abbreviations: ABA: Abscisic acid, GA: Gibberellic acid, ABAREs: Abscisic acid responsive elements, GBREs: Gibberellins-responsive element, AUREs: Auxin responsive elements, SalREs: Salicylic acid responsive elements, MeJAREs: Methyl jasmonate responsive elements, TID: Transcript ID, pI: Isoelectric point, Mw: Molecular weight, Da: Dalton, CREs: *Cis*-regulatory elements, TFB: Transcription factor binding sites, HRE: Hormone responsive element, LRE*:* Light responsive element

Introduction

Salt stress is one of the most prominent abiotic stresses around the globe. It has pernicious effects on survival, yield and life cycle crop (Zhu, 2016). Breeding of salt stress tolerant crop genotypes is the best practical way of curtailing such problems. There are two types of approaches followed for salinity tolerance breeding; (i) elevating yield of salt endurable cultivars, and (ii) transfer of salt bearable genes to prominent genotypes (Singh J *et al.*, 2018, 2019a, b; Singh V *et al.,* 2018. In the first case, traditional cultivars of salinity affected areas are improved for increasing productivity without affecting their salt endurance ability while the other approach accompany to transfer salt tolerance genes from locally adapted (salt tolerant) cultivars to high yielding one (Kumawat *et al.,* 2020; Singh *et al.*, 2014; Singh *et al.*, 2018). Therefore, the physio-molecular mechanisms viz.,

salt responsive genes, network and their regulatory pathways, through which plants adapt better for salinity stress, need further investigation to pace up the improvements in crop productivity worldwide, especially on saline land (Singh *et al.*, 2020; Singh *et al.*, 2019, 2020)

A large number of transcription factors including promoters are notable to control the expression of targeted genes in various signal transduction cascades in plants (Venter and Botha, 2004). The TFBs or *cis*-regulatory elements are determined the conspicuous timing and location of transcriptional activity. The TFBs are found notably in the long non-coding sequence upstream of a gene (Chaboute *et al.,* 2002). These regulatory motifs unified into distinct *cis*-regulatory modules are obligatory for a specific expression pattern (Babu *et al.,* 2004). *CREs* is part of noncoding DNA that regulate the transcription of genes nearby. These are crucial components of genetic regulatory networks by controlling various aspects of developmental biology. Thus, the discerning of regulatory motifs and their confederation is an important step to better understanding of expression and regulation of gene.

The sortilege of *cis*-regulatory sites in a non-coding DNA sequence is developed by many databases such as *PlantCARE* (Lescot *et al.,* 2002) and *PLACE* (Higo *et al.,* 1999). The promoter regions of candidate *CREs* can be distinguished by searching against the known elements in the databases. Therewith, unique *CREs* could also be detected with unknown transcription factor binding sites using the representative sequence of the promoters of co-expressed genes (Helden, 2003). In consequence, the analysis of *in silico* promoter with the help of bioinformatics has become more attractive and feasible (Chareerat *et al.,* 2009).

Materials and Methods

Genome-wide identiûcation of salt perceptive genes in soybean

Protein sequences of salt responsive genes are identified in different crops such as *Arabidopsis thaliana*, durum wheat (*Triticum turgidum* ssp. *durum*), *Nicotiana tabacum*, *Oryza sativa*, *Triticum aestivum*, *Solanum lycopersicum*, and *Glycine max* were downloaded from NCBI (*https://www.ncbi.nlm.nih.gov*) and *UniProt* (*http://www.uniprot.org*) database. These protein sequences are used as query and blast search is performed against soybean (*G. max*) genome database using *Ensembl Plants* (*http://plants.ensembl.org/Glycine_max/Tools/Blast/Results?tl=JQdmOrzWNyaer45o-19627944*). The *ExPASy* tool (*https://web.expasy.org/compute_pi/*) used to compute soybean homologs sequence, TID, pI, Mw and amino acid length of salt responsive genes. Then, the biological process, cellular component and molecular function of the salt responsive genes are predicted through '*Ensembl Plants*' (*http://plants.ensembl.org/Glycine_max/Gene/Ontologies/cellular_component?db=core;g=GLYMA_09G218600;r=9:44167828-44170729;t=KRH39775*). The sub-cellular localization of the salt responsive genes was predicted through *TargetP 1.1* server (*http://www.cbs.dtu.dk/services/TargetP/*) and *Cell-PLoc* 2.0 (*http://www.csbio.sjtu.edu.cn/bioinf/plant-multi/*).

Cis-acting regulatory element prediction in promoter regions of thirty eight salt tractable genes

The promoter sequence consists of 1000 bp upstream sequences of a total of 38 salt responsive genes are downloaded from the soybean (*G. max*) genome database. The sequences put in to *PlantCARE* database program (*http://bioinformatics.psb.ugent.be/webtools/plantcare/html/*) disclosed the phenomena of a large number of different *CREs* in upstream sequences of salt responsive genes, which have different functions in sessile plants which are usually exposed to various biotic and abiotic stress conditions. *PlantCARE* database can provide locations for known *CREs*, enhancers and repressors and also provide tools for *in silico* analysis of promoter sequences (Lescot *et al.,* 2002).

Results and Discussion

Genome-wide identiûcation of salt perceptive genes in soybean

The elucidated soybean salt perceptive genes with their TID, pI, Mw, protein size, sub-cellular

localization, biological process, cellular component and molecular function are given in Table 1. There are large variations obtained in molecular weight, 5763.35 Da *(GmASR1)* to 119025.18 Da (*GmSOS1*); while pI ranged 3.9 (*GmOsHRD*) to 9.56 (*GmCaM4)* and protein size 51 amino acids (*GmASR1*) to 1073 amino acids (*GmSOS1*). These genes are predicted to localize in different sub-cellular compartments *viz.,* cell membrane, cell membrane & nucleus, chloroplast, chloroplast & nucleus, cytoplasm, endoplasmic reticulum, nucleus and vacuole (Fig. 1).

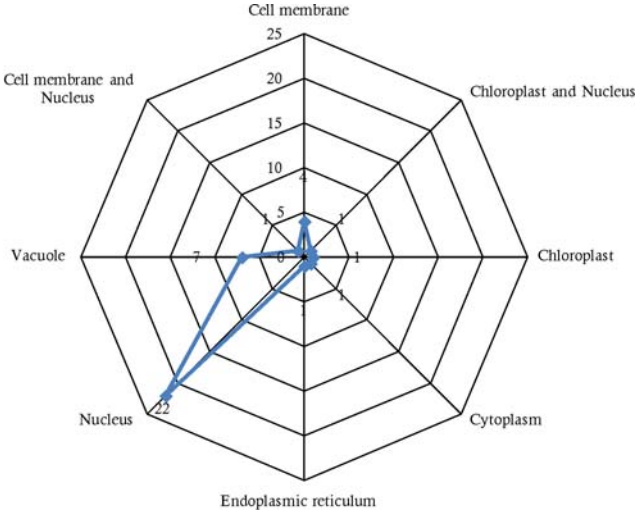


Fig. 1. Radar diagram showing the number of predicted soybean salt responsive genes and their localization in different sub-cellular compartments

The promoter regions of predicted soybean salt perceptive genes comprised of 1000 bp upstream of the translation start site. The sequences submitted to *PlantCARE* database revealed the occurrence of a large number of different *cis*-motifs in upstream sequences of salt responsive genes, which have different functions in sessile plants, and are usually exposed to various biotic and abiotic stresses. The *cis*-motifs occurred differentially in all the predicted 38 soybean genes (Fig. 2). The *cis*-motifs CAAT-box and TATA- Box were found at high frequency (38 times) in the most of the salt responsive genes whereas, other *cis*-motifs; Box 4 (35 times); ERE (29 times); TCT-motif (23 times); ARE (22 times); ABRE and G-Box (20 times); GT1 motif (18 times); TGACG-motif (12 times); CGTCA-motif (11 times); I-box, GCN4-motif and TCA-element (10 times) were also predominantly found in the promoter region of these predicted genes.

Based on the *cis*-motif in the upstream region of the salt responsive genes, the sequences were categorized into elements of HRE, condition specific, LRE, plant tissues specific, regulation specific, promoter and enhancer regions element, and transcription element. The functions of these *cis*-motifs were predicted (Table 2). TATA box and CAAT box are common elements found in all the promoters which have importance in the initiation of the transcription process. The *HREs* composed of five subgroups viz., ABAREs, GBREs, AUREs, SalREs and MeJAREs. The condition specific element was composed of four subgroups: drought-inducibility responsive element, defense and stress responsiveness element, low-temperature responsiveness element and anaerobic induction element. The plant tissues specific element was composed of two subgroups;

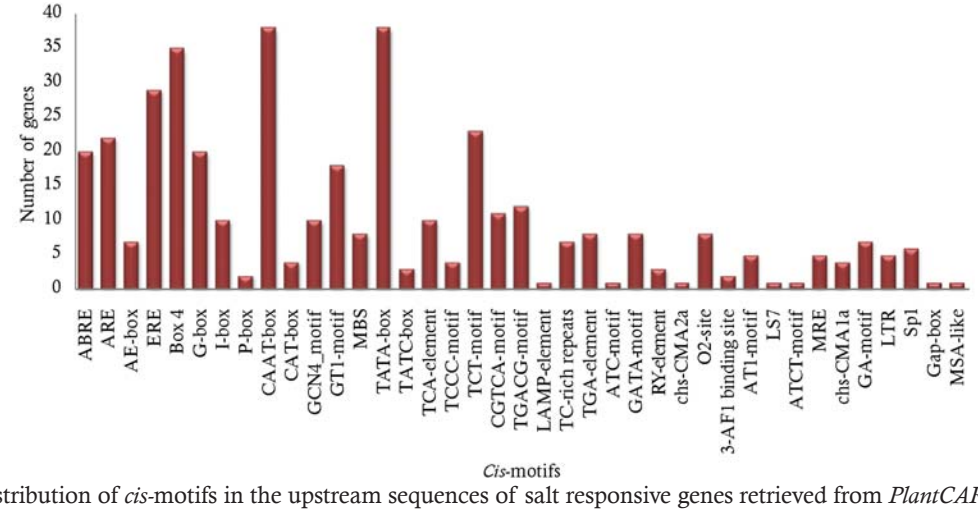


Fig. 2. Distribution of *cis*-motifs in the upstream sequences of salt responsive genes retrieved from *PlantCARE* database

Table 1. List of identified salt responsive genes in soybean

S. N.	Gene name	TID	pI	MW (Da)	Size (aa)	Sub-cellular localization	Biological process	Cellular component	Molecular function
1	*GmbZIP132*	KRH21958	5.56	17498.97	155	Nucleus	Transcription, DNA-template, response to karrikin	Unknown	Sequence-specific DNA binding
2	*GmCYP707A*	KRH39775	9.3	29165.21	255	Endoplasmic reticulum	Abscisic acid metabolic process, sterol metabolic process, Oxidation-reduction process	Membrane, Integral component of membrane	Mono-oxygenase activity, iron ion binding, Oxido-reductase activity, acting on paired donors, with incorporation or reduction of molecular oxygen, Heme binding
3	*GmAKT1*	KRH03745	6.66	90079.18	805	Nucleus	Ion transport, potassium ion transport, regulation of ion transmembrane transport	Membrane, integral component of membrane	Ion channel activity, voltage-gated potassium channel activity, protein binding
4	*GmCaM4*	KRH38989	9.56	12352.83	112	Nucleus	Ion transport, chloride transport, transmembrane transport	Unknown	DNA binding
5	*GmCLC1*	KRH57669	9.02	78829.94	713	Cell membrane	Unknown	Membrane, integral component of membrane	Voltage-gated chloride channel activity
6	*GmOLPa*	KRH27960	4.17	16575.56	154	Vacuole	Defense response	Extracellular region	
7	*GmSOS1*	KRH42480	6.43	119025.18	1073	Cell membrane	Unknown	Unknown	Unknown
8	*GmSBP65*	KRH22300	6.06	60091.35	573	Nucleus	Unknown	Unknown	Unknown
9	*GmZLDE-2*	KRH48466	6.54	17478.35	172	Cytoplasm	Unknown	Unknown	Unknown
10	*GmFDL19*	KRG94991	8.87	27803.06	250	Nucleus	Regulation of transcription, DNA-template	Nucleus	DNA-binding transcription factor activity
11	*GmNHX1*	KRH34011	7.84	53615.77	486	Vacuole	Ion transport, cation transport, sodium ion transport, regulation of pH, response to salt stress, regulation of stomatal closure, proton transmembrane transport	Vacuolar membrane, plasma membrane, Membrane, integral component of membrane	Antiporter activity, sodium: proton antiporter activity, Heme binding antiporter activity
12	*GmNHX2*	KRH50930	9.05	63189.59	570	Vacuole	Cation transport, sodium ion transport, response to salt stress, proton transmembrane transport, regulation of pH, potassium ion homeostasis	Vacuolar membrane, plasma membrane, membrane, integral component of membrane	Antiporter activity, solute: proton antiporter activity, potassium: proton antiporter activity
13	*GmNHX3*	KRH02191	9.35	49874.01	448	Vacuole	Cation transport, sodium ion transport, response to salt stress, proton transmembrane transport, regulation of pH, potassium ion homeostasis	Vacuolar membrane, plasma membrane, membrane, integral component of membrane	Antiporter activity, sodium: proton antiporter activity, potassium: proton antiporter activity, solute: proton antiporter activity

S. N.	Gene name	TID	pI	MW (Da)	Size (aa)	Sub-cellular localization	Biological process	Cellular component	Molecular function
14	*GmNHX4*	KRH68649	9.37	54255.40	4.87	Vacuole	Cation transport, sodium ion transport, response to salt stress, proton transmembrane transport, regulation of pH, potassium ion homeostasis	Plasma membrane, membrane, integral component of membrane	Antiporter activity, sodium: proton antiporter activity, potassium: proton antiporter activity, solute: proton antiporter activity
15	*GmNHX5*	KRH11690	5.24	51948.78	474	Vacuole	Cation transport, sodium ion transport, response to salt stress, proton transmembrane transport, regulation of pH, potassium ion homeostasis	Endosome, plasma membrane, membrane, integral component of membrane	Antiporter activity, sodium: proton antiporter activity, potassium: proton antiporter activity, solute: proton antiporter activity
16	*GmNHX6*	KRG96912	9.43	54223.43	4.87	Vacuole	Cation transport, sodium ion transport, response to salt stress, proton transmembrane transport, regulation of pH, potassium ion homeostasis, regulation of stomatal closure	Vacuolar membrane, plasma membrane, Membrane, integral component of membrane	Antiporter activity, sodium: proton antiporter activity, potassium: proton antiporter activity, solute: proton antiporter activity
17	*GmHRD*	KRH67858	3.9	13494.85	129	Nucleus	Transcription, DNA-template, regulation of transcription, DNA-template	Nucleus	DNA binding, DNA-binding transcription factor activity
18	*GmSIP*	KRH47115	8.54	21341.65	194	Cell membrane, Nucleus	Unknown	Plasma membrane, Membrane, integral component of membrane	Unknown
19	*GmANAC019*	KRH27200	9.25	27997.27	247	Nucleus	Regulation of transcription, DNA-template	Nucleus	DNA binding
20	*GmANAC055*	KRH22137	8.85	31756.36	283	Nucleus	Regulation of transcription, DNA-template	Nucleus	DNA binding, DNA-binding transcription factor activity
21	*GmANAC072*	KRH22137	8.85	31756.36	283	Nucleus	Regulation of transcription, DNA-template	Nucleus	DNA binding, DNA-binding transcription factor activity
22	*GmGSK1*	KRH25798	9.13	36320.35	314	Nucleus	Protein phosphorylation	Nucleus	Nucleotide binding, protein kinase activity, protein serine/ threonine kinase activity, ATP binding
23	*GmMYB3R2*	KRH14930	5.63	19322.69	168	Nucleus	Unknown	Nucleus	DNA binding
24	*GmCIPK*	KRH00424	7.81	43173.58	383	Nucleus	Protein phosphorylation, signal transduction,	Nucleus cytoplasm	Nucleotide binding, protein kinase activity, protein serine/ threonine kinase activity, ATP binding, transferase activity

S. N.	Gene name	TID	pI	MW (Da)	Size (aa)	Sub-cellular localization	Biological process	Cellular component	Molecular function
25	*GmDREB1*	KRH35383	5.04	16019.88	145	Nucleus	Transcription, DNA-template	Nucleus	DNA binding, DNA-binding transcription factor activity
26	*GmOsiSAP8*	KRH23151	8.58	11906.28	112	Nucleus	Unknown	Unknown	DNA binding, zinc ion binding, metal ion binding
27	*GmSnac1*	KRH27200	9.25	27997.27	247	Nucleus	Regulation of transcription, DNA-template	Nucleus	DNA binding
28	*GmHKT1*	KRH55676	8.7	57781.42	514	Cell membrane	Cation transport, sodium ion transmembrane transport, cation transmembrane transport	Plasma membrane	Cation transmembrane transporter activity, sodium ion transmembrane transporter activity
29	*GmASR1*	KRH08687	6.36	5763.35	51	Cell membrane	Unknown	Unknown	Unknown
30	*GmHPS*	KRH57371	7.46	10656.19	112	Nucleus	Unknown	Unknown	Unknown
31	*GmNAC2a*	KRH01889	9.2	39277.44	343	Nucleus	Regulation of transcription, DNA-template	Nucleus	DNA binding
32	*GmSnRK2.4*	KRH00424	7.81	43173.58	383	Nucleus	Protein phosphorylation, signal transduction	Nucleus, cytoplasm	Nucleotide binding, protein kinase activity, protein serine/ threonine kinase activity, ATP binding, transferase
33	*GmSRHP*	KRH16681	9.14	10095.59	91	Chloroplast, Nucleus	Unknown	Unknown	Unknown
34	*GmWRKY19*	KRH74951	8.95	33759.00	300	Nucleus	Regulation of transcription, DNA-template	Nucleus	DNA binding, DNA-binding transcription factor activity, sequence-specific DNA binding
35	*GmWRKY2*	KRH72765	7.74	57368.80	520	Nucleus	Regulation of transcription, DNA-template	Nucleus	DNA binding, DNA-binding transcription factor activity, sequence-specific DNA binding
36	*GmSOD*	KRH70388	4.76	28996.12	251	Chloroplast	Superoxide metabolic process, removal of superoxide radicals, oxidation-reduction process	Chloroplast nucleoid	Superoxide dismutase activity, oxidoreductase activity, metal ion binding
37	*GmDhn1*	KRH60791	6.09	17270.18	154	Nucleus	response to water, cold acclimation, response to abscisic acid	Cytosol, membrane	Unknown
38	*GmASR1*	KRH08687	6.36	5763.35	51	Nucleus	Unknown	Unknown	Unknown

TID= Transcript ID; pI= Isoelectric point; Mw= Molecular weight; Da= Dalton; aa= amino acid

Table 2. The function of *cis*-motifs in the promoter of salt responsive genes retrieved from *PlantCARE*

Types of motif	Sequence	Function
Hormone responsive element		
ABRE	ACGTG	*cis*-acting element involved in the abscisic acid responsiveness
TATC-box	TATCCCA	*cis*-acting element involved in gibberellins responsiveness
TGA-element	AACGAC	Auxin responsive element
TCA-element	CCATCTTTTT	*cis*-acting element involved in salicylic acid responsiveness
CGTCA-motif	CGTCA	*cis*-acting regulatory element involved in the MeJA responsiveness
TGACG-motif	TGACG	*cis*-acting regulatory element involved in the MeJA responsiveness
Condition specific element		
MBS	CAACTG	MYB binding site involved in drought-inducibility
TC-rich repeats	ATTCTCTAAC	*cis*-acting element involved in defense and stress responsiveness
LTR	CCGAAA	*cis*-acting element involved in low-temperature responsiveness
ARE	AAACCA	*cis*-acting regulatory element essential for the anaerobic induction
Light responsive element		
GT1-motif	GGTTAA	light responsive element
I-box	CCATATCCAAT	light responsive element
3-AF1 binding site	TAAGAGAGGAA	light responsive element
Sp1	GGGCGG	light responsive element
MRE	AACCTAA	MYB binding site involved in light responsiveness
TCCC-motif	TCTCCCT	part of a light responsive element
TCT-motif	TCTTAC	part of a light responsive element
LAMP-element	CTTTATCA	part of a light responsive element
GATA-motif	AAGATAAGATT	part of a light responsive element
chs-CMA2a	TCACTTGA	part of a light responsive element
LS7	CAGATTTATTTTTA	part of a light responsive element
chs-CMA1a	TTACTTAA	part of a light responsive element
GA-motif	ATAGATAA	part of a light responsive element
Gap-box	CAAATGAA(A/G)A	part of a light responsive element
AT1-motif	AATTATTTTTTATT	part of a light responsive module
AE-box	AGAAACAA	part of a module for light response
Box 4	ATTAAT	part of a conserved DNA module involved in light responsiveness
ATC-motif	AGTAATCT	part of a conserved DNA module involved in light responsiveness
ATCT-motif	AATCTAATCC	part of a conserved DNA module involved in light responsiveness
Plant tissues specific element		
CAT-box	GCCACT	*cis*-acting regulatory element related to meristem expression
GCN4_motif	TGAGTCA	*cis*-regulatory element involved in endosperm expression
Regulation specific element		
O2-site	GATGACATGG	*cis*-acting regulatory element involved in zein metabolism regulation
RY-element	CATGCATG	*cis*-acting regulatory element involved in seed-specific regulation
MSA-like	T/C)C(T/C)AACGG (T/C)(T/C)A	*cis*-acting element involved in cell cycle regulation
Promoter and enhancer regions element		
CAAT-box	CAAAT	common *cis*-acting element in promoter and enhancer regions
Transcription element		
TATA-box	TATAA	core promoter element around 30 of transcription start

meristem responsive element and endosperm expression responsive element. However, the regulation specific element was composed of three subgroups; *zein* metabolism responsive element, seed-specific regulation element and cell cycle regulation element.

Comprehensive genome-wide analysis of soybean salt perceptive genes manifested the abiotic stress particular expression pattern, analyzed the conserved *cis*-acting elements in the promoter region. These genes were localized in different sub-cellular compartments (Table 1),

thereby suggested a wider cellular localization and function. In this paper, 38 salt responsive genes have been identified and the promoter analysis showed the presence of conserved *CREs* regulating the expression of the salt perceptive gene in response to light, abiotic and biotic stresses.

These *cis*-acting elements responsible for molecular switches in response to environmental stress signals due to biotic and abiotic stress on plants. Light is a most prominent factor which controls varied life processes such as growth, development and stress responses in plants. In the present study, it has been interpreted that all the upstream sequences of four salt perceptive genes have light-responsive *cis*-elements which could have some preface in defense mechanism and control overexpression. Four genes (*GmbZIP132*, *GmANAC055, GmANAC072* and *GmCaM4*) were involved in the light response. The amount of absorbed light energy by plants is used for photosynthetic metabolism and the remaining energy called excess excitation energy have several major functions such as optimization of energy status, minimization of reactive oxygen species and as a source of information about seasonal changes (Karpinski *et al.,* 2003).

Categorization of *cis*-motifs in the upstream regions of salt responsive genes

C*is*-motifs in the upstream regions of salt responsive genes, retrieved from *PlantCARE* database were categorized based on their regulatory function (Fig. 3). The *cis*-motifs AE-box, Box 4, I-box and TCT-motif were found in the regulatory regions of *GmbZIP132*, *GmANAC055*, *GmANAC072* and *GmCaM4* genes that are associated with the light response. *Cis*-motif *ABRE* was found in the regulatory regions of *GmbZIP132*, *GmCYP707A*, *GmCLC1*, *GmOLPa*, *GmSBP65*, *GmZLDE-2*, *GmNHX3*, *GmANAC019*, *GmANAC055*, *GmANAC072*, *GmGSK1*, *GmOsiSAP8*, *GmSnac1*, *GmASR1*, *GmNAC2a*, *GmSRHP*, *GmWRKY2*, *GmSOD*, *GmDhn1* and *GmASR1* genes that are associated to the abscisic acid-responsive element. *Cis*-motif *MBS* was found in the regulatory regions of *GmCYP707A*, *GmNHX4*, *GmHRD*, *GmSIP*, *GmANAC055*, *GmANAC072*, *GmASR1* and *GmASR1* genes that are associated to the drought-inducibility responsive element. *Cis*-motif *TC*-rich repeats was found in the regulatory regions of *GmAKT1*, *GmCLC1*, *GmFDL19*, *GmNHX3*, *GmOsiSAP8*, *GmSRHP* and *GmWRKY19* genes that are

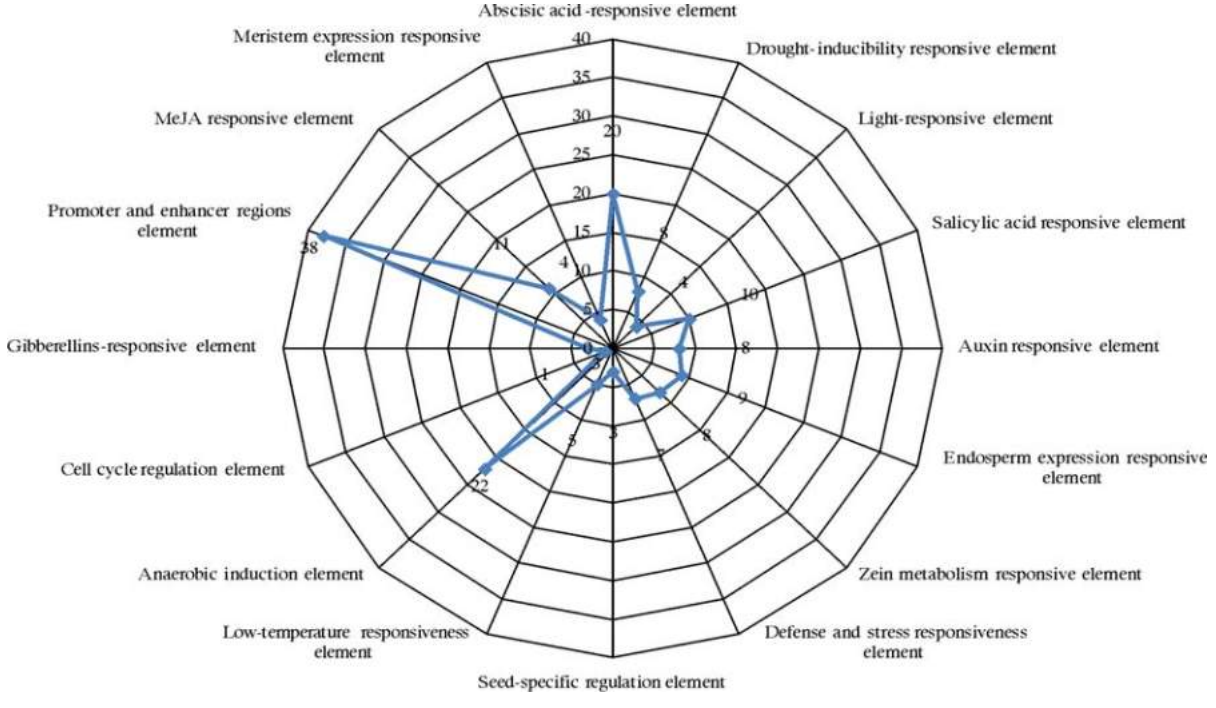


Fig. 3. Categorization of *cis*-motifs in the predicted salt responsive genes with specific functions

associated with the defense and stress responsiveness element.

Cis-motif *LTR* was found in the regulatory regions of *GmNHX2*, *GmANAC019*, *GmMYB3R2*, *GmSnac1* and *GmNAC2* genes that are associated to the low-temperature responsiveness element. *Cis*-motif *TATC*-box was found in the regulatory regions of *GmCaM4*, *TaHPS* and *TaWRKY19* genes that are associated to the gibberellins-responsive element. The CAT-box was found in the regulatory regions of *GmCLC1*, *GmFDL19*, *GmASR1* and *GmASR1* genes that are associated to the meristem expression responsive element while, TCA-element was found in the regulatory regions of*GmbZIP132*, *GmCaM4*, *GmOLPa*, *GmSBP65*, *GmNHX5*, *GmHRD*, *GmANAC055*, *GmANAC072*, *GmOsiSAP8* and *GmSOD* genes that are associated to the salicylic acid responsive element.

The CAAT-box were found in the regulatory regions of *GmbZIP132*, *GmCYP707A*, *GmAKT1*, *GmCaM4*, *GmCLC1*, *GmOLPa*, *GmSOS1*, *GmSBP65*, *GmZLDE-2*, *GmFDL19*, *GmNHX1*, *GmNHX2*, *GmNHX3*, *GmNHX4*, *GmNHX5*, *GmNHX6*, *GmHRD*, *GmSIP*, *GmANAC019*, *GmANAC055*, *GmANAC072*, *GmGSK1*, *GmMYB3R2*,*GmCIPK*, *GmDREB1*, *GmOsiSAP8*, *GmSnac1*, *GmHKT1*, *GmASR1*, *GmHPS*, *GmNAC2a*, *GmSnRK2.4*, *GmSRHP*, *GmWRKY19*, *GmWRKY2*, *GmSOD*, *GmDhn1* and *GmASR1* genes that are associated to the promoter and enhancer regions element. *Cis*-motif TGA was found in the regulatory regions of *GmAKT1*, *GmCaM4*, *GmHRD*, *GmANAC019*, *GmOsiSAP8*, *GmSnac1*, *GmWRKY19* and *GmWRKY2* genes that are associated to the auxin response. The *GCN4*_motif was found in the regulatory regions of *GmbZIP132*, *GmSOS1*, *GmSBP65*, *GmNHX1*, *GmNHX2*, *GmNHX3*, *GmANAC019* and *GmANAC055*, *GmANAC072* and *GmSnac1* genes that are associated to the endosperm expression responsive element.

The O_2-site was found in the regulatory regions of *GmCLC1*, *GmNHX3*, *GmANAC055*, *GmANAC072*, *GmCIPK*, *GmDREB1*, *GmSnRK2.4* and *GmSRHP* genes that are associated with the *Zein* metabolism responsive element. *Cis*-motifs CGTCA and TGACG were found in the regulatory regions of *GmCYP707A*, *GmCLC*, *GmOLPa*, *GmSOS1*, *GmSBP65*, *GmGSK1*, *GmOsiSAP8*, *GmASR1*, *GmSRHP*, *GmSOD* and *GmASR1* genes that are associated to the *MeJA* responsive element. The *cis*-motif ARE was found in the regulatory regions of *GmCLC1*, *GmOLPa*, *GmZLDE-2*, *GmNHX1*, *GmNHX3*, *GmNHX4*, *GmNHX5*, *GmHRD*, *GmANAC019*, *GmANAC055*, *GmANAC072*, *GmCIPK*, *GmOsiSAP8*, *GmSnac1*, *GmHKT1*, *GmASR1*, *GmHPS*, *GmSnRK2.4*, *GmWRKY19*, *GmWRKY2*, *GmSOD* and *GmASR1* genes that are associated with the anaerobic induction element. *Cis*-motif MSA-like was found in the regulatory regions of *GmWRKY19* genes that are associated to the cell cycle regulation element. The RY-element was found in the regulatory regions of *GmCaM4*, *GmHKT1*, and *GmDhn1* genes that are associated with the seed-specific regulation element.

In the salt stress response, MBS core sequence assists in modulation of MYB motif and plays a coupled role in controlling drought and salt stress induction. MYB protein performs a prime role in transcriptional inducement of ABA-inducible gene under regulation in higher salt concentrations. The gene *GmMYB76* from *G. max*; *AtMYB2* and *AtMYB7* genes from *A. thaliana* are popular to manage salt stress (Abe *et al.,* 2003; Yanhui *et al.,* 2006; Liao *et al.,* 2008). The *cis*-motif O_2 site encoding *bZIP* transcription factor imparts significant role in salt stress regulation in *A. thaliana* via *ABF3* gene (Choi *et al.,* 2000). Plant hormone gibberellins are involved in controlling various aspects of plant growth, including the germination of seeds, stem growth, leaf growth, flowering and fruit ripening. *GmCaM4*, *TaHPS* and *TaWRKY19* genes are associated to the GAREs however, abscisic acid, conditioning plant developmental processes, including seed and bud dormancy, organ size and stomatal closure and are very important for plants in response to environmental stresses like drought, salinity, cold, freezing, heat and heavy metal ion tolerance. Our predicted genes include *GmbZIP132*, *GmCYP707A*, *GmCLC1*, *GmOLPa*, *GmSBP65*, *GmZLDE-2*, *GmNHX3*, *GmANAC019*, *GmANAC055*, *GmANAC072*, *GmGSK1*, *GmOsiSAP8*, *GmSnac1*, *GmASR1*, *GmNAC2a*, *GmSRHP*, *GmWRKY2*, *GmSOD*, *GmDhn1* and

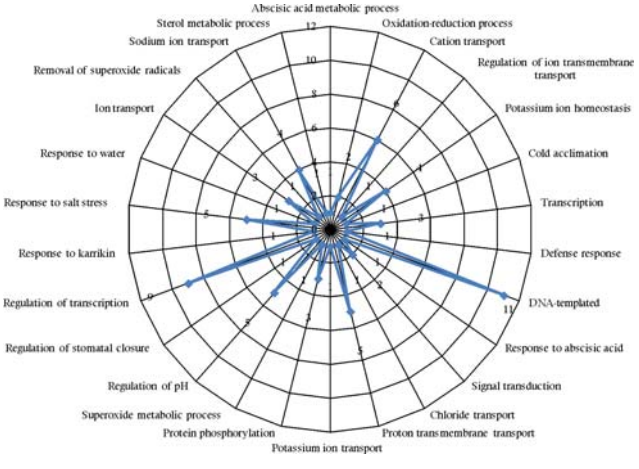
GmASR1 are associated to the ABAREs. These also regulate the expression of many genes that might function in dehydration tolerance in both vegetative tissues and seeds with *ABA*-dependent /independent gene expression, respectively, in osmotic and cold stress responses. Drought and high salinity cause plants to produce high levels of ABA (Pandey *et al.,* 2015).

GO (Gene Ontology) assignments

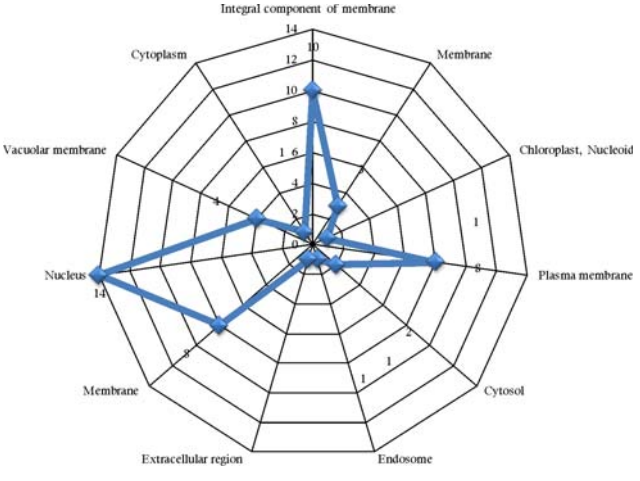
The gene ontology is widely used to standardize representation of genes across species and provide a controlled vocabulary of terms describing gene products. In present investigation, 38 salt responsive genes predicted in soybean for GO terms based on the *Ensembl Plants* database have been assigned. These GO terms were summarized into the three main categories; i) biological process; ii) cellular component, and iii) molecular function; and 62 subcategories (Fig. 4). The number of genes in each category and assigned function could be described as:-

(i) The biological process comprised of DNA templates (11 genes) and transcription regulation (9 genes), and were the most dominant subcategories. Further, the biological process subcategorized into cation transport (6 genes), response to salt stress (5 genes), proton trans-membrane transport (5 genes), and regulation of pH (5 genes).

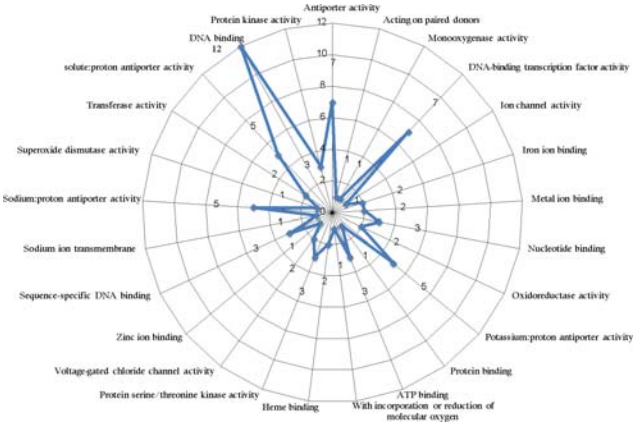
(ii) The cellular component comprised of nucleus (14 genes), integral component of membrane (10 genes), plasma membrane (8 genes) and membrane (8 genes), were the most highly represented subcategories.



(a) Biological process



(b) Cellular component



(c) Molecular function

Fig. 4 Distribution of gene ontology of salt responsive genes in soybean according to (a) biological process; (b) cellular component and (c) molecular function retrieved from '*Ensembl Plants'* database

(iii) The molecular function was comprised mainly DNA binding function (12 genes), DNA binding transcription factor activity (7 genes), antiporter activity (7 genes), potassium: proton antiporter activity (5 genes), sodium: proton antiporter activity (5 genes) and solute: proton antiporter activity (5 genes). These GO annotations represent a general gene expression profile signature for salt responsive genes in soybean.

The TGA-element was obtained in the regulatory regions of *GmAKT1*, *GmCaM4*, *GmHRD*, *GmANAC019*, *GmOsiSAP8*, *GmSnac1*, *GmWRKY19* and *GmWRKY2* genes connected with the auxin response. Auxin is very important for root formation, apical dominance, tropism, and senescence (Yazaki *et al.,* 2003). The estimated genes *GmbZIP132*, *GmCaM4*, *GmOLPa*, *GmSBP65*, *GmNHX5*, *GmHRD*, *GmANAC055*, *GmANAC072*, *GmOsiSAP8* and *GmSOD* are associated to the salicylic acid response. Role of SA in plant growth and development is evidenced from the fact that this hormone regulates processes such as seed germination, vegetative growth, photosynthesis, respiration, thermogenesis, flower formation, seed production, and senescence (Vicente and Plasencia, 2011). Salicylic acid regulates signal mediating plant response to abiotic stresses such as drought (Munné-Bosch and Peñuelas, 2003; Chini *et al.,* 2004), chilling (Janda *et al.,* 1999; Kang and Saltveit, 2002) heavy metal tolerance (Metwally *et al.,* 2003; Yang *et al.,* 2003; Freeman *et al.,* 2005) heat (Larkindale and Knight, 2002; Larkindale *et al.,* 2005) and osmotic stress (Borsani *et al.,* 2001). *GmCaM4*, *GmHKT1* and *GmDhn1* genes are found to be associated with the seed-specific regulation of salt stress response in soybean. Seed-specific promoters isolated from genes with restricted or enhanced expression during seed development are the most often reported spatiotemporal promoters (Rao *et al.,* 2014; Tsai, 2003). These promoters have a wide range of applications including tissue-specific targeting of industrial and pharmaceutical compounds, and development of transgenic seeds with improved nutritional quality and better functional quality of soybean and production of recombinant proteins (Kawakatsu and Takaiwa, 2010).

Conclusions

The present results revealed that the identified soybean salt perceptive genes perform distinct functions in plant growth development and regulation of gene expression during biotic and abiotic stress conditions. The data obtained from this study contribute to a better understanding of the salt responsive genes in soybean, and provide the basis for further studies to dissect the regulatory network and function of these genes during plant growth and development as well as in response to environmental stimuli. It is evident that the *cis*-element based gene finding approach is effective and has high prediction accuracy and is applicable to different organisms and different type of genes. With more information on genes available in the different database, we expect this cost-effective and accurate approach to be widely applied to various targeted gene finding problems in the future.

References

Abe H, Urao T, Ito T, Seki M, Shinozaki K and Yamaguchi-Shinozaki K (2003) Arabidopsis AtMYC2 (bHLH) and AtMYB2 (MYB) function as transcriptional activators in abscisic acid signaling. *The Plant Cell* **15**: 63–78.

Babu MM, Luscombe NM, Aravind L, Gerstein M and Teichmann SA (2004) Structure and evolution of transcriptional regulatory network. *Current Opinion in Structural Biology* **14**: 283–291.

Borsani O, Valpuesta V and Botella A (2001) Evidence for a role of salicylic acid in the oxidative damage generated by NaCl and osmotic stress in *Arabidopsis* seedlings. *Plant Physiology* **126**: 1024–1030.

Chaboute ME, Clement B and Phillips G (2002) S phase and meristem-specific expression of tobacco RNR1 b gene is mediated by an E2F element located in the 5¢ leader sequence. *The Journal of Biological Chemistry* **277**: 17845–17851.

Chareerat M, Pradit P and Surin P (2009) *Insilico* promoter analysis of photoperiod-responsive genes identified by DNA microarray in rice (*Oryza sativa* L.). *Kasetsart Journal (Natuaral Science)* **43**: 164–177.

Chini A, Grant JJ, Seki M, Shinozaki K and Loake GJ (2004) Drought tolerance established by enhanced expression of the *CCI-NBS-LRR* gene, *ADR1*, requires salicylic acid, EDS1 and ABI1. *The Plant Journal* **38(5)**: 810-22.

Choi HI, Hong JH, Ha JO, Kang JY and Kim SY (2000) ABFs, a family of ABA-responsive element binding factors. *The Journal of Biological Chemistry* **275(3)**: 1723–1730.

Freeman JL, Garcia D, Kim D, Hopf AM and Salt DE (2005) Constitutively elevated salicylic acid signals glutathione-mediated nickel tolerance in Thlaspi nickel hyper accumulators. *Plant Physiology* **137**:1082–1091.

Helden JV (2003) Regulatory sequence analysis tools. *Nucleic Acids Research* **31**: 3593–3596.

Higo K, Ugawa Y, Iwamoto M and Korenaga T (1999) Plant cis-acting regulatory DNA elements (PLACE) database. *Nucleic Acids Research* **27**: 297–300.

Janda T, Szalai G, Tari I and Páldi E (1999) Hydroponic treatment with salicylic acid decreases the effects of chilling injury in maize (*Zea mays* L.) plants. *Planta* **208**: 175–180.

Kang HM and Saltveit ME (2002) Chilling tolerance of maize, cucumber and rice seedling leaves and roots are differentially affected by salicylic acid. *Physiologia Plantarum* **115**: 571–576.

Karpinski S, Gabrys H, Mateo A, Karpinska B and Mullineaux PM (2003) Light perception in plant disease defence signalling. *Current Opinion in Plant Biology* **6:** 390–396.

Kawakatsu T and Takaiwa F (2010) Cereal seed storage protein synthesis: fundamental processes for recombinant protein production in cereal grains. *Plant Biotechnology Journal* **9**: 939–953.

Kumawat G, Maranna S, Gupta S, Rachana T, Agarwal N, Singh V, Vangala R, Subhash C, Kamble V, Nataraj V, Bharati A, Sharma MP, Jadhav PV, Milind BR, Satpute GK and Bhatia VS (2020) Identification of novel genetic sources for agronomic and quality traits in soybean using mutli-trait allele specific genic marker assays. *Journal of Plant Biochemistry and Biotechnology*. https://doi.org/10.1007/s13562-020-00580-x

Larkindale J and Knight MR (2002) Protection against heat stress-induced oxidative damage in Arabidopsis involves calcium, abscisic acid, ethylene, and salicylic acid. *Plant Physiology* **128(2)**: 682–695.

Larkindale J, Hall JD, Knight MR and Vierling E (2005) Heat stress phenotypes of Arabidopsis mutants implicate multiple signaling pathways in the acquisition of thermo tolerance. *Plant Physiology* **138**: 882–897.

Lescot MP, Déhais G, Thijs K, Marchal Y, Moreau YV, Peer P, Rouzé 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 Research* **30**: 325–327.

Liao Y, Zou HF, Wang HW, Zhang WK, Ma B and Zhang JS (2008) Soybean GmMYB76, GmMYB92, and GmMYB177 genes confer stress tolerance in transgenic Arabidopsis plants. *Cell Research* **18**: 1047–1060.

Metwally A, Finkemeier I, Georgi M and Dietz KJ (2003) Salicylic acid alleviates the cadmium toxicity in barley seedlings. *Plant Physiology* **132**: 272–281.

Munné-Bosch S and Peñuelas J (2003) Photo- and antioxidative protection, and a role for salicylic acid during drought and recovery in field grown *Phillyrea angustifolia* plants. *Planta* **217(5)**: 758–766.

Pandey P, Ramegowda V and Senthil KM (2015) Shared and unique responses of plants to multiple individual stresses and stress combinations: physiological and molecular mechanisms. *Frontiers in Plant Science* **6**: 723.

Rao SS, El-Habbak MH, Havens WM, Singh A, Zheng D, Vaughn L, Haudenshield JS, Hartman GL, Korban SS and Ghabrial SA (2014) Overexpression of *GmCaM4* in soybean enhances resistance to pathogens and tolerance to salt stress. *Molecular Plant Pathology* **15(2)**: 145–160.

Singh D, Singh CK, Singh YP, Singh V, Singh R and Tomar RSS (2018) Evaluation of cultivated and wild genotypes of *Lens* species under alkalinity stress and their molecular collocation using microsatellite markers. *PLoS ONE* **13(8):** e0199933.

Singh D, Singh CK, Tomar R SS, Sharma S, Karwa S, Pal M, Singh V, Sanwal SK and Sharma PC (2020) Genetics and molecular mapping for salinity stress tolerance at seedling stage in lentil (*Lens culinaris* Medik). *Crop Science* **(1–13)**. https://doi.org/10.1002/csc2.20030

Singh J, Sastry EVD and Singh V (2014) Effect of salinity on genetic architecture of fruit yield and its contributing traits in tomato (*Lycopersicon esculentum* Mill.). *Journal of Soil Salinity and Water Quality* **6(1):** 42–51.

Singh J, Sharma PC, Singh V and Kumar P (2019b) Predicted model to reveal the mechanism of salt tolerance in *Brassica juncea***.** *Journal of Soil Salinity and Water Quality.* **11(1)**: 18–30.

Singh J, Singh V and Sharma PC (2018) Elucidating the role of osmotic, ionic and major salt responsive transcript components towards salinity tolerance in contrasting chickpea (*Cicer arietinum* L.) genotypes. *Physiology and Molecular Biology of Plants* **24(3)**: 441–453.

Singh J, Singh V, Vineeth TV, Kumar P, Neeraj and Sharma PC (2019 a) Differential response of Indian Mustard *(Brassica juncea* L., Czern & Coss) under salinity: photosynthetic traits and gene expression. *Physiology and Molecular Biology of Plants.* **25(1):** 71–83.

Singh V, Sanwal SK, Kumawat G, K M Shiva, Satpute GK, Gill BS, Panwar S, Singh Jogendra and Sharma PC (2020) Assessing the effect of salt stress on soybean [*Glycine max* (L.) Merrillis] genotypes using AMMI and GGE biplot analysis. *Journal of Soil Salinity and Water Quality* 12**(1):** 95–100.

Singh V, Singh AP, Bhadoria J, Giri J, Singh J, Vineeth TV and Sharma PC (2018) Differential expression of salt-responsive genes to salinity stress in salt-tolerant and salt-sensitive rice (*Oryza sativa* L.) at seedling stage. *Protoplasma* **255 (6):** 1667–1681.

Singh V, Singh J, Kumar P, Banyal RK and Sharma PC (2019) *Prosopis juliflora* and *Vachellia nilotica*: boon for salt-

affected land and livelihood security-a review. *Journal of Soil Salinity and Water Quality.* **11(1)**: 108–116.

Tsai SN (2003) *Cloning and Characterization of Ion Transporters Genes from a Salt tolerant Soybean Variety.* Dissertation, the Chinese University of Hong Kong.

Venter M and Botha FC (2004) Promoter, analysis and transcription profiling: Integration of genetic data enhances understanding of gene expression. *Physiologia Plantarum* **120**: 74–83.

Vicente M and Plasencia J (2011) Salicylic acid beyond defence: Its role in plant growth and development. *Journal of Experimental Botany* **62**: 3321-38.

Yang ZM, Wang J, Wang SH and Xu LL (2003) Salicylic acid-induced aluminium tolerance by modulation of citrate efflux from roots of *Cassia tora* L. *Planta* **217(1)**: 168–174.

Yanhui C, Xiaoyuan Y, Kun H, Meihua L, Jigang L, Zhaofeng G, Zhiqiang L, Yunfei Z, Xiaoxiao W, Xiaoming Q, Yunping S, Li Z, Xiaohui D, Jingchu L, Xing WD, Zhangliang C, Hongya G and Li JQ (2006) The MYB transcription factor superfamily of Arabidopsis: expression analysis and phylogenetic comparison with the rice MYB family. *Plant Molecular Biology* **60**: 107–124.

Yazaki J, Kishimoto N, Nagata Y, Ishikawa M, Fujii F and Hashimoto A (2003) Genomics approach to abscisic acid and gibberellin-responsive genes in rice. *DNA Research* **10**: 249–261.

Zhu JK (2016) Abiotic stress signaling and responses in plants. *Cell* **167**: 313–324.

Received: June 05, 2020; Accepted: October 02, 2020