Analysis of ACGT *Cis* Elements Across Plant Genomes and Its characterization Using Protein Phosphatase 2C (*PP2C*) like Promoter from *Arabidopsis thaliana*

THESIS

Submitted in partial fulfilment
of the requirements for the degree of
DOCTOR OF PHILOSOPHY

by

Purva

Under the Supervision of
Prof. Rajesh Mehrotra

BIRLA INSTITUTE OF TECHNOLOGY AND SCIENCE, PILANI

2015

BIRLA INSTITUTE OF TECHNOLOGY AND SCIENCE, PILANI

CERTIFICATE

This is to certify that the thesis entitled "**Analysis of ACGT *Cis* Elements Across Plant Genomes and Its characterization UsingProtein Phosphatase 2C (*PP2C*) like Promoter from *Arabidopsis thaliana***" submitted by **Purva (ID.No. 2008PHXF414P)** for the award of Ph.D. Degree of the Institute embodies original work done by her under my supervision.

Signature of the Supervisor

Name : **RAJESH MEHROTRA**

Date:

Designation: **Associate Professor (Head, Dept. Of Biological Sciences, BITS Pilani)**

Dedicated to my parents
and Sadhana mausi

Contents

Preface	**1-2**
Acknowledgement	3-5
Abbreviations	6-8
Chapter1: Introduction & Review of Literature	9-37
Chapter 2: Materials & Methods	38-82
Chapter 3: Results & Discussion	83-125
Chapter 4: Summary	126-130
Chapter 5: References	131-148

Preface

Gene expression in multicellular organisms is a fine tuned process which is regulated at transcriptional, post-transcriptional, translational and post-translational levels. Transcriptional regulation is the initial step of gene regulation and plays a very critical role in the activation and repression of gene expression (Zou et al., 2011). Gene expression is mediated through interaction between *cis* regulatory elements and transcription factors or DNA binding proteins. *Cis* regulatory elements are defined as non-coding DNA sequences that provide the binding sites for transcription factors and are clustered in the upstream region of genes (Ong and Corces 2011). Mutation in *cis* regulatory elements affects the activity of the promoter region which leads to divergence. Divergence in the *cis* regulatory elements' function is also responsible for evolution (Wittkopp and Kalay 2012). The full length promoter is delineated into proximal and distal region which possess regulatory sequences such as enhancers, silencers, insulators and *cis* regulatory elements (Lee and Young 2000; Hernandez-Garcia and Finer 2014). Enhancers which are present in upstream, downstream or within introns are also classified as regulatory elements (Kleinjan and Van-Heyningen 2005; Wittkopp and Kalay 2012). Core promoter region which is also a *cis* regulatory element is enough to kick start the basal transcription process (Molina and Grotewold 2005).

Transcription initiates with the assembly of RNA polymerase and associated/auxiliary transcription factors like TFIIA, TFIIB, TFIID, TFIIE, TFIIF, and TFIIH on the core promoter (Lee and Young 2000; Reese 2003; Thomas and Chiang 2006; Muller et al., 2007; Bhattacharjee et al., 2013). This complex is known as pre-initiation complex (PIC) which leads to the activation of the basal transcription. The activity of full length promoter is based on the relative position of the core promoter and combinatorial interaction between the *cis* regulatory elements present in the proximal and distal regions of promoter (Hernandez-Garcia and Finer 2014). The combinatorial interaction may be synergistic or antagonistic between the various *cis* regulatory elements present in the full length promoter region (Chaturvedi et al., 2006). The distant *cis* regulatory elements come in the close vicinity of core promoter by DNA folding mediated by conformational change in the DNA and chromatin structure (Bulger and Groudine 2011). In higher organisms, DNA is wound around the histone octamers to form nucleosomes which are a basic unit of

chromatin. The access of transcription factors to *cis* regulatory elements is restricted by nucleosomes (Li et al., 2007). Soufi et al., (2015) showed however pioneer transcription factors access silent chromatin and initiate cell fate changes, using diverse type of DNA binding domain. The entry of transcription factors is made possible by sliding of nucleosomes done by chromatin remodelers or post translational modification of nucleosome (Becker and Horz 2002; Narlikar et al., 2002; Saha et al., 2006). Constitutive genes have an open promoter and this condition is referred to nucleosome-depleted region (NDR). The NDR has been found to be present in the upstream of the transcription start site (TSS). The transcription factor binding sites are not occupied by the nucleosomes (Cairns 2009). NDR as defined by Cairns (2009) is a gradient of depletion of nucleosome region and this region is rich in poly dA: dT sequences. Regulated or inducible genes have closed or covered promoters where nucleosome occupies the proximal promoter region or large areas of the promoter. In closed promoters, nucleosomes block the interaction between transcription factors and *cis*-regulatory elements (Cairns 2009). Pan et al., (2010) showed that transcription factors recognise specific *cis* regulatory elements among the pool of *cis* regulatory elements in the genome with the help of flanking and core sequence.

The work is presented in five major sections:

1. *Introduction and literature review: This Section summarizes information on stress responsive cis elements and transcription factors.*
2. *Material and Methods: This Section describes details of in-silico and various experimental procedures and techniques, which were employed in the present study.*
3. *Results and Discussions: This Section gives details of various observations made and the results obtained from the in silico studies and experiments performed. The results are analyzed in a broader perspective and the inferences drawn thereof, are discussed in this section, taking into consideration the observations made in earlier publications.*
4. *Conclusions: This Section summarizes the major findings of the present study and the conclusion drawn therefrom.*
5. *References: This Section consists of a list of references cited in the present study.*

Acknowledgement

Keeping in mind, what my Nani had taught me in my childhood, firstly and fore mostly I send a silent prayer to the supreme power in this cosmic world, for blessing me with this success and always strengthening me with hope and will, to carry on with all my might.

"The three great essentials, to achieve anything worthwhile are, first, hard work; second, stick-to-itiveness; and third, common sense." — *Thomas A. Edison*

This quote always reminds me of my mentor and guide, Dr. Rajesh Mehrotra, to whom I would like to express my heartfelt gratitude; for playing an exceptional role in making all my hard work fruitful. I would also like to thank him for all his timely encouragements and for being supportive and patient towards me and my research. I also want to use this medium, to convey my utmost regard for him, for giving me the opportunity to work under his able guidance and an honour of being his student.

Sir, your numerous advices on both my research as well as my career are invaluable. You have taught me a great deal about science and life, both consciously and subconsciously. I truly appreciate all your criticisms, concerns and contributions in terms of precious hours and ideas, which have resulted in an experience full of productive and a thought-provoking Ph.D. The enthusiasm, you have had in this entire course of journey, was contagious and motivational for me, even during the tough times, in the pursuit of a good research. I could not be more proud of my academic lineage and I hope that when time allows, I am also able to pass on the same values and virtues in the way you have.

I would also like to acknowledge, members of my Doctoral Advisory Committee members, Dr. Sandhya Mehrotra and Dr B. Vani, for their valuable guidance and contribution to the furtherance of this research, insightful discussion, scientific advice, knowledge and also for their time and effort in checking the manuscript in short time period.

I offer my gratitude, to Director Prof A.K. Sarkar, Prof. S.K. Verma (Dean, Academic & Research Division) BITS-Pilani, Pilani Campus, for allowing me the usage of institute's

infrastructure. I also thank my alma mater, BITS Pilani for providing me with this and various other opportune possibilities which resulted in the culmination of this research.
Special thanks to the Head of the Department of Biological Sciences, Dr. Rajesh Mehrotra, DRC members Dr. Jitendra Panwar, Dr. Ashis Kumar Das, Dr. Shibashsis Chowdhary, Dr. Lalita Gupta, Dr. Sandhya Mehrotra and Dr. Prabhat Nath Jha for their time to time co-operation and words of encouragement and also for giving a period of time to conclude this manuscript. All the faculty members of Department of Biological Sciences were kind enough to extend their help and encouragement at various phases of research, whenever I approached them, and I do hereby acknowledge all of them. I would also like to acknowledge all my teachers during the course work for their consistent help and support.

"But if the while I think on thee, dear friend. All losses are restored and sorrows end. A friend is one that knows you as you are, understands where you have been, accepts what you have become, and still, gently allow you to grow." - *William Shakespeare.*
Heartfelt thanks to my colleagues and friends, Arpit Bhargava, Gagandeep Singh Saggu and Shobha Mehra thank you for listening, offering me advice, for their untimely help, motivation, which drives me to give my best and moral support. My special gratitude to Gurpreet Kaur Sidhu and Panchsheela Nogia for being with me in thick and thin of life, I find myself lucky to have friends like them in my life.

I also thank my fellow colleagues, Dr. Garima Gupta, Dr. Navin Jain, Dr. Satish Kumar, Zarna Rajesh Pala, Vandana, Nisha Jangir, Asha Jhakar,Vidushi Asati, Chetna Sangwan, Zaiba Hasan Khan, and R.V. Dilip for their immense support.
I also thank Brijesh Kumar, Amit Kumar Yadav, Prashali Bansal, Suvendu Kumar and Ujjwal Banerjee who continuously helped me with the experiments, and incented me to strive towards my goal. I am glad to have all of them around me and lifting my spirits in the difficult situations. To each one of you, I could not have asked for anyone better.

I would like to acknowledge Dr. Rakesh Tuli, (former executive director) and Dr. Akhilesh Tyagi (Executive Director at present) NABI-Mohali, for their generosity in permitting me to use their experimental and lab facilities.

I am thankful to Dr. Siddharth Tiwari, Anshu Alok, Jitesh Kumar and Simran Kaur for their co-operation and assistance while I was working at NABI-Mohali.

The special mention of my seniors Dr. Deepak Pakalpati, Dr. Narayan Kumar, Dr. Amit Subuhi and P.A. Boopathi who have taught me the lab culture. My thanks to the staff of Department of Biological Sciences for their innumerable help. A special word of gratefulness to Mr. Naresh Kumar Saini, Mr. Raghuveer, Mr. Mahipal and Mr. Manoj Kumar who have helped me with several official matters.

Words cannot express how grateful I am to my parents, Dr. Mani Ram Choudhary and Sumitra for all of the sacrifices that they have made in-order for me to be able to accomplish this feat. I am deeply obliged to my maternal grandparents- Mr Jawahar Singh Meel and Ramki Meel and my grandmother-Shanti Devi, Susheela, Dr. Vinod Kumar Budania, Lt Col Surender Baliwal, Mr Madhukar Meel and Sunita Meel and my siblings, for their love and tireless support since the inception of my research. I don’t imagine a life without their love, concern and blessings. Your prayers and strong belief in me had helped me go on and sustain thus far. Without them, this thesis would never have been accomplished.

A special thanks to my in-laws Mr Vidhyadhar Bhalothia and Manohari Devi for their words of encouragement and support during my thesis time.

I dedicate this thesis to my beloved mausi- Sadhna, whose role in my life is, and will always remain immense. For, it was because of her faith and confidence in my abilities, that I kept working through the hurdles and failures that I had faced during this entire time.

This last word of acknowledgment I would like to dedicate to my dear husband Deepak, who compromised, adjusted and has made these past couple of years, the best. Thank you, for being my pillar of strength.

I would also like to communicate my gratitude to UGC/BSR for financial aid.

Date: Purva

Abbreviations

A	Adenine
ABA	Abscisic acid
ABI1/2	ABA insensitive ½
ABF	ABRE-binding factor
ABI5	ABA-insensitive 5
ABRE	ABA responsive element
AHG1/3	ABA-hypersensitive germination 1/3
AREB	ABA responsive element binding protein
BME	β mercaptoethanol
Bp/ kb	Base pair/ kilo base pair
BSA	Bovine serum albumin
bZIP	Basic leucine zipper protein
C	Cytosine
$CaCl_2$	Calcium chloride
CaMV	Cauliflower mosaic virus
CTAB	Cetyl trimethyl ammonium bromide
DEPC	Diethyl pyrocarbonate
DNA	Deoxyribonucleic acid
DNase	Deoxyribonuclease
dNTP	Deoxy nucleoside triphosphate
DTT	Dithiothreitol
EDTA	Ethylene diamine tetra acetic acid
E.coli	*Escherichia coli*
ET	Ethylene
g	Gram
GTE	Glucose-Tris-EDTA
h	Hour
HAB1/2	Homology to ABA 1/2
IPTG	Isopropyl β-D thiogalactopyranoside

JA	Jasmonic acid
kDa	kilo Dalton
LBA	Luria Bertani agar medium
LB	Luria Bertani broth meium
LEA proteins	Late embryogenesis abundant protein
LMP agarose	Low melting point agarose
M	Molar
Mg	Milligram
$MgCl_2$	Magnesium chloride
Min	Minute
MS	Murashige and Skoog's medium
MU	4-Methyl-umbelliferone
MUG	4-Metyl-umbelliferyl β-DD-glucuronide
NaCl	Sodium chloride
NaOH	Soddium hydroxide
Ng	Nanogram
OD	Optical density
OSRK1	*Oryza* SnRK2-related kinase 1
PP2C	Protein phosphatase 2C
PYL	PYR1 like PYR1 pyrabactin resistance 1
PCR	Polymerase chain reaction
PEG	Polyethylene glycol
RNA	Ribonucleic acid
RNase-A	Ribonuclease-A
Rab	Responsive to ABA
RCAR1	Regulatory component of ABA receptor 1
ROS	Reactive oxygen species
RT-PCR	Real time PCR
SA	Salicylic acid
SDS	Soddium dodecyl sulphate
Sec	Second
Snrk2	Sucrose non fermenting 1-related protein kinase 2

SOS	Salt overly sensitive
START	Starr Related lipid transfer
T	Thymine
TAE	Tris- acetate- EDTA
TE	Tris-EDTA
TFs	Transcription factors
Tris	Tris (hyddroxymethyl) aminomethane
U	Unit
UV	Ultra-violet
X-gal	5-bromo-4-chloro-3-indolyl-β-D-galactopyranoside
X-gluc	5-bromo-4-chloro-3-indolyl glucuronidde
μ	Micro

INTRODUCTION

CHAPTER-1

1.0 *Introduction*

With population explosion, there is an increased need to focus on the improvement of plant productivity. A report by Polizel et al., (2011) has shown that Brazil, the second leading producer of soybean worldwide, has reported 20% decrease in their gross output due to drought in the years 2003-2004 and 2004-2005. World Agricultural Supply and Demand Estimates (WASDE) (October 11, 2012) report states that the production of major crops like corn, soybean, sorghum, wheat, etc., has reduced to 27.6 % due to severe drought in 2012 in United States (*http://www.ers.usda.gov/topics/in-the-news/us-drought-2012-farm-and-food-impacts.aspx*). According to the WASDE (2012) report, this trend is increasing year by year. In correspondence to a report a 6% decrease in wheat production is predicted with every degree rise in temperature (Asseng et al., 2014). The rapid increase in population has created a global food demand which is a huge challenge due to limited cultivable land. Furthermore, frequent atypical weather conditions are greatly affecting plant productivity all across the globe. To deal with this comprehending the molecular mechanisms as to how plants combat stress conditions is of prime importance in improving the stress tolerance in plants. Plants are constantly challenged by several abiotic and biotic stresses, affecting both their growth and productivity. Abiotic stresses occur due to unfavourable changes in the environment such as drought, extreme temperature, salinity stress, osmotic stress etc., whereas biotic stresses are concerned with negative effects of viruses, bacteria, fungi, nematodes, herbivores and insects. These stresses affect the plant growth and development. Immobility of plants forces them to adopt unique mechanisms to cope with their surroundings and ever changing conditions in the given environment (Mehrotra et al., 2014).

Plants adapt themselves to changing conditions through morphological, physiological and biochemical processes which are ultimately governed by the molecular and cellular processes (Kim et al., 2010; Mehrotra et al., 2014). Plants have a fine tuned protection system involving phytohormones - abscisic acid (ABA), salicylic acid (SA), jasmonic acid (JA) and ethylene (ET) which help the plants in adapting themselves to abiotic and biotic stress (Fujita et al., 2006). To restore the homeostatic equilibrium of the

plant, ABA in case of abiotic stress and SA, JA and ET for biotic stress, act by imparting changes in gene expression and epigenetic adaptations such as histone modification, chromatin remodelling and phosphorylation /de-phosphorylation signal cascade pathways (Kim et al., 2010).

ABA accumulates under stress conditions in plants and plays a significant role in seed dormancy, seed desiccation, guard cell opening-closing, stress responses and tolerance in plants (Busk and Pages 1998; Shinozaki et al., 2003; Yamaguchi-Shinozaki and Shinozaki 2006; Nakashima et al., 2009b). In recent years, it has been shown that ABA is involved in plant–pathogen interactions, resistance to pathogen infection and interaction with SA, JA/ ET mediated pathways (Mauch-Mani and Mauch 2005; Melotto et al., 2006; de Torres-Zabala et al., 2009). Plants employ a group of signalling components, comprising of receptors, protein kinases and phosphatases, secondary messengers like nitric oxide (NO), reactive oxygen species (ROS) and cytosolic free calcium ions (Ca^{2+}) assigned to sense and transduce a broad range of stress inducing signals. This results in the activation of signalling cascade leading to changes in molecular mechanisms in the plant system (Neill et al., 2002; Xiong et al., 2002). Various components participating in the regulatory networks are involved in response to different stresses. There is an interconnected network of components which are responsible for crosstalk between several pathways (Fujita et al., 2006). This is also revealed from the *~omics* data which states that certain stress response associated components appear to participate in both abiotic and biotic stress signalling (Walley et al., 2007). A significant number of genes in *Arabidopsis thaliana,* found to be up-regulated in response to salinity stress, are also found to be expressed in response to biotic stress (Ma et al., 2006). This shows convergence of the pathways. Mitogen-activated protein kinase (MAPK), Ca^{2+} ions, ROS, phospholipids, mitochondrial functions, vesicle trafficking and apoptosis are induced under biotic as well as abiotic stresses (Ma and Bohnert 2007). The whole genome expression analysis revealed that a significant number of genes are commonly regulated under abiotic and biotic stress conditions (Ma and Bohnert 2007; Shaik and Ramakrishna 2013, 2014). Ma and Bohnert (2007) have identified 197 commonly up regulated genes in response to SA, JA and ET.

To withstand unwanted cues, many pathways are activated by ABA, SA and JA hormones leading to activation of stress- responsive genes whose products provides protection against stress by production of important metabolic proteins, osmoprotectant and transcription factors involved in the regulation of other genes involved in stress

induced signalling pathways and thus, help the plants in adapting themselves under unfavourable conditions (Yamaguchi-Shinozaki and Shinozaki 2006). Numerous genes are induced in response to stress and many of these genes are commonly up- and/or down-regulated in response to both abiotic and biotic stresses. Seki et al., (2002) have studied the expression of 7000 *Arabidopsis* genes under drought, cold and high-salinity stresses. Using a full-length cDNA microarray it was observed that 277, 53 and 194 genes were induced in response to drought, cold and salinity stress respectively. Further Seki et al., (2002) have also identified 79, 89 and 71 down-regulated genes in response to drought, high-salinity and cold stress respectively through microarray analysis. Out of these stress- induced genes, commonly 40 genes for transcription factors were up-regulated in response to drought, cold or salt stress. From the cDNA microarray analysis, it was identified that the following transcription factor family genes viz dehydration responsive element binding protein (DREB), ethylene responsive element binding protein, zinc finger family, WRKY, myeloblastosis (MYB-R_2R_3), basic helix loop helix (bHLH), basic leucine zipper (bZIP), NAC (NAM, ATAF1/2 and CUC2) and homeo-domain were induced under above mentioned conditions.

Using microarray profile, hundreds of abiotic-stress inducible genes in rice were identified by Rabbani et al., (2003). From comparative genomic studies, hundreds of inducible promoters were identified to induce stress- responsive gene expression for generating transgenic plants (Gorantla et al., 2007). 73 abiotic stress inducible genes have been identified using cDNA microarrays and RNA blot analysis (Rabbani et al., 2003). Rabbani et al., (2003) have studied stress responsive *cis*-regulatory elements in the 2 kb 5' promoter region for generating stress-inducible promoters.

The specificity and diversity of gene expression is due to the recruitment of transcription factors to *cis* regulatory elements. *Cis* regulatory elements are small nucleotide sequences and are functional DNA elements, clustered in a specific manner in the upstream region of the gene. They associate with transcription factors and co-regulators and form large assemblies responsible for gene regulation. Chromatin remodelling determines expression kinetics of functionally related promoters (Cairns 2009). *Cis* regulatory elements are buried under the nucleosome and are not accessible to transcription factors (Pan et al., 2010). The change in the DNA conformation depends upon the kind of protein complex that binds to DNA. The transcription process is initiated by the recognition of the core promoter by the RNA polymerase. The basal transcriptional machinery initiates basal gene expression. Along with transcription factors, *cis* regulatory

elements are equally important for establishing conspicuous temporal and spatial transcriptional activity. Functional elucidation of underlying mechanism of *cis* regulatory elements is very important in understanding how cells recognize signals, respond to environment and bring about molecular changes. Sawant et al., (2001) have designed an artificial promoter for high level gene expression in plants. Various *cis* regulatory elements were assembled in this artificial promoter to synthesize a constitutive promoter (Sawant et al., 2001). Zou et al., (2011) have identified distinct *cis* regulatory elements which respond to abiotic and biotic stresses using co-expression data. Since the number of *cis* regulatory elements is more than the transcription factors, the latter get to choose from a large number of *cis* regulatory elements. Upto 3000 transcription factors and 7,00,000 possible *cis* regulatory elements are reported in human genome (Levine and Tjian 2003). 10% of human genome codes for transcription factors (Sikorski and Buratowski 2009) whereas in *Arabidopsis thaliana* only 5% of genome encodes transcriptional factors (Riechmann et al., 2000).

The selectivity in recognition of *cis* regulatory elements by transcription factors depends on the core sequence of *cis* regulatory elements as well as the flanking regions of the core sequence. A single nucleotide substitution or change in the flanking sequence or the core sequence of *cis* regulatory elements can alter the selective binding of transcription factors. A *cis* regulatory element may be recognized by two members of the same family. As in case of ACGT, this core is recognized by members of bZIP family. Although the transcription factor binding sites are similar, transcription factors vary in their sizes. In this case, selectivity is based on the size and spacer sequence between *cis* regulatory elements. Short spacer between the two *cis* regulatory elements is selectively bound by the small sized transcription factor, i.e., number of domains will be less in it, whereas longer spacer sequences are bound by the larger sized transcription factor, i.e., number of domains will be more. The different size spacer sequence between two *cis* regulatory elements enforces the selective binding. Since the larger size transcription factors cannot bind to the smaller size spacer sequence, it leads to steric hindrance. When smaller size transcription factors bind to the bigger spacer sequence it leads to lower stability and lack of co-operativity. Transcription factors are comprised of multiple domains. These domains vary with respect to their number, size, structure, sequence and organization (Pan et al., 2010).

Transcription factors show high tolerance for *cis* regulatory element's sequence diversity. A different set of *cis* regulatory elements and transcription factors are involved

in different stress responsive gene expressions (Shinozaki and Yamaguchi-Shinozaki 2007). Many inducible genes involved in stress response and tolerance have been identified using microarray and gene chip technology (Seki et al., 2002; Shinozaki et. al., 2003; Vogel et al., 2005). Transcription factors and *cis* regulatory elements are the key point of transcription regulation. The products of these genes give tolerance to plants to cope up with the stress condition. The contribution of individual *cis* regulatory element needs to be studied in order to understand the kinetics of the gene expression.

1.1 *Review of literature*

Plants encounter a wide range of environmental cues during their life cycle. To endure abiotic and biotic stress, plants have evolved mechanisms by which they increase their tolerance through physical adaptations, molecular and cellular changes that begin at the onset of stress. The foremost step to activate the molecular responses is to perceive the stress signal and transmit the stress signal through a signal transduction pathway. The various pathways eventually get activated which lead to physiological changes, such as guard cell closure, or to the expression of stress inducible genes and result into molecular and cellular modifications. Various physiological and molecular changes are brought upon due to the transmission of signals from external to internal environment of the cell, aided by phosphorylation and dephosphorylation processes. ABA, SA and JA and ET are the phytohormones which play a significant role during abiotic and biotic stresses (Fujita et al., 2006; Umezawa et al., 2010).

At the onset of abiotic stress, ABA-independent or /and ABA-dependent signalling pathways get activated (Yamaguchi-Shinozaki and Shinozaki, 2005). ABA dependent signalling has SnRK2s, kinases as positive regulators (Fujii and Zhu, 2009; Fujita et al., 2009; Nakashima et al., 2009a) and the PP2Cs, as negative regulators (Hirayama and Shinozaki 2007). Molecular changes are brought upon by interaction between the SnRK2s activated ABA responsive element binding protein (AREB) (belongs to bZIP transcription factor family), and their *cis* regulatory element, present upstream of the stress- inducible genes or ABA dependent genes (Figure 1-1). In ABA-dependent pathway, ABA-inducible gene promoter region has a conserved *cis* regulatory element ABA responsive element (ABRE) (PyACGTGG/TC) (Bray, 1994; Giraudat et al., 1994;

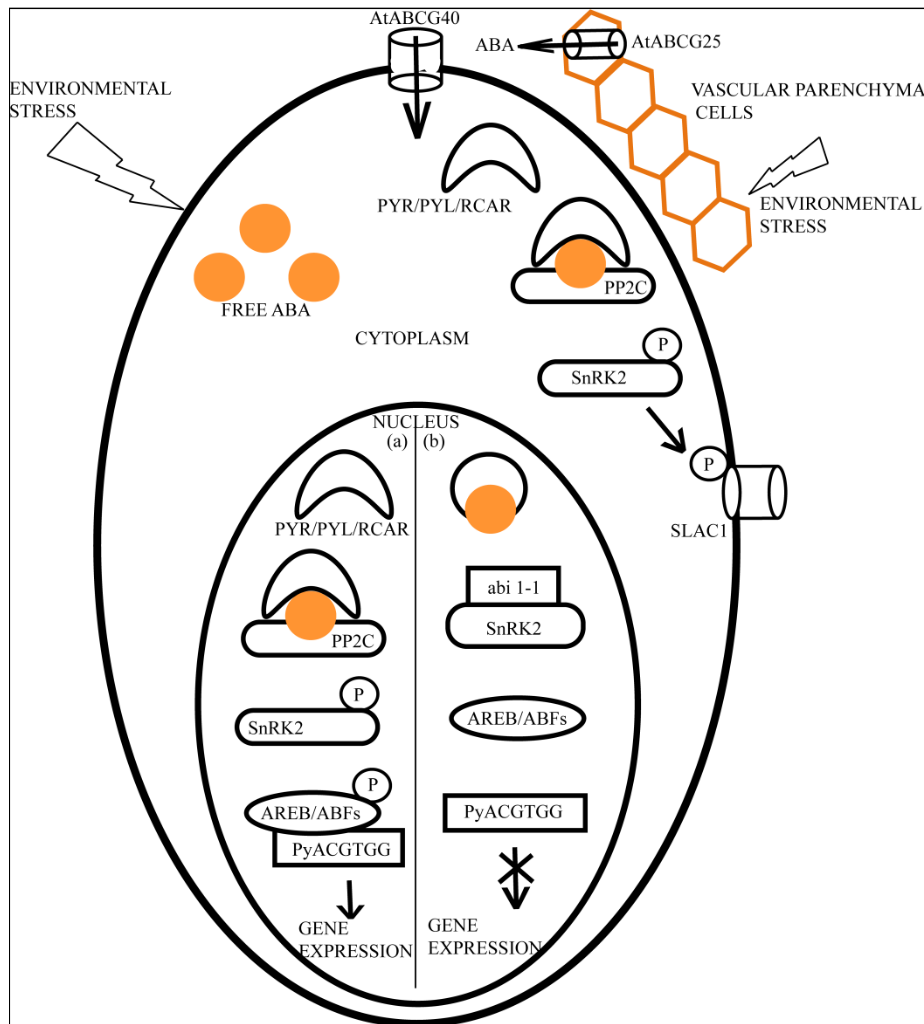


Figure 1-1 The pictorial representation of ABA dependent signalling as shown in both cytoplasm and nucleus. During abiotic stress conditions and developmental conditions the synthesis of ABA increases in vascular parenchyma cells and transported to targeted cells through AtBCG25 (from parenchyma cells) and AtBCG40 transporters (to target cells). The components involved in ABA signalling pathway are – receptor PYR/PYL/RCAR, a negative regulator PP2C, a positive regulator SnRK2s and AREB/ABFs factors. In the presence of ABA, a complex is formed ABA-PYR/PYL/RCAR-PP2C which prevents PP2C inhibitory activity. SnRK2s gets phosphorylated and further activates AREB/ABFs which induces the stress responsive gene expression. In ABA insensitive condition, ABA insensitive 1-1 (Abi1-1) interacts with SnRK2, hence inhibits its interaction with ABA-PYR/PYL/RCAR complex thereby preventing gene expression. (Adopted from Mehrotra et al., 2014.)

Maruyama et al., 2012). The expressions of ABA-responsive genes require more than one ABRE or a combination of an ABRE and a coupling element (CE) for optimum gene expression (Marcotte et al., 1989; Shen and Ho, 1995; Shen et al., 1996; Hobo et al., 1999; Narusaka et al., 2003; Nakashima and Yamaguchi-Shinozaki 2013). The genome wide analysis of *Arabidopsis thaliana* and rice (*Oryza sativa*) reveals that ABRE core motifs are uniformly distributed throughout the genome (Zhang et al., 2005; Go´mez-Porras et al., 2007).

Major transcription factor families involved in ABA-dependent pathway is AREB/ABFs which binds to the ABRE sequence. The AREB/ABF members of group A of bZIP transcription factor have three N-terminal and one C-terminal conserved domains in their structure. The basic domain is positioned at C-terminal. The members of group A have 3 to 4 leucine repeats as mentioned in Table 1-2 (Jakoby et al., 2002). Under abiotic stress conditions AREB/ABFs regulate the ABA responsive or inducible gene expression. The post transcriptional activation by phosphorylation of AREB/ABF is regulated in the presence of ABA (Fujita et al., 2005; Yoshida et al., 2010) which happens at more than one site within the conserved domains of AREB/ABFs (Furihata et al., 2006). The phosphorylation of AREB/ABFs is done by SnRK2s in the ABA-dependent signalling network (Fujita et al., 2009; Nakashima et al., 2009a; Umezawa et al., 2013). In response to abiotic stresses, such as dehydration, salt stress and ABA treatment in vegetative tissues, AREB1/ABF2, AREB2/ABF4 and ABF3 are induced (Fujita et al., 2005) and their gain-of-function transgenic mutant *Arabidopsis* plants have shown enhanced drought stress tolerance (Kang et al., 2002; Kim et al., 2004; Fujita et al., 2005).

Fujita et al., (2005) have observed that transgenic *Arabidopsis* plants over expressing deleted and active forms of AREB1 showed enhanced drought tolerance and hypersensitivity towards ABA. In another study it has been observed that over expression of AREB1 gives improved drought tolerance in rice and soybean (Oh et al., 2005; Barbosa et al., 2013). Yoshida et al., (2010) have reported that ABA is required for full activation of AREB/ABF transcription factors in *Arabidopsis* and rice. The microscopic studies have shown that AREB1, AREB2, and ABF3 are found to be localized in nuclei of rice and *Arabidopsis.* AREB2, ABF3 and AREB1 are activated by SRK2D, SRK2E, and SRK2I in ABA signalling pathway.

Bimolecular fluorescence complementation (BiFC) analysis revealed that AREB1, AREB2, and ABF3 form the homo- or hetero- dimers with each other. The *areb1 areb2 abf3* triple mutant *Arabidopsis* plants have shown reduced drought tolerance in

comparison with the single and double mutants and wild type plants (Yoshida et al., 2010). AHG1, AHG3, HAI1, HAI2 and HAI3 members of group-A PP2C were found to be down-regulated in the *areb1areb2 abf3* triple mutant under water stress conditions (Schweighofer et al., 2004; Xue et al., 2008; Yoshida et al., 2010). Findings of PYL/PYR/RCAR receptors further help in understanding ABA perception and signal transduction (Ma et al., 2009; Park et al., 2009; Umezawa et al., 2010; Weiner et al., 2010; Nakashima amd Yamaguchi-Shinozaki 2013). It was revealed that PP2Cs (negative regulator), SnRK2s (positive regulator), RCAR/PYR/PYL ABA receptors, ABA receptors, AREB/ABFs and ABRE are key components of the ABA (dependent) signalling pathway in land plants (Umezawa et al., 2010; Nakashima and Yamaguchi-Shinozaki 2013; Mehrotra et al., 2014).

1.1.1 PYR/PYL/RCAR receptors

PYR/PYL/RCAR receptors belong to START domain/Bet V allergen super family proteins. Genome wide analysis of *Arabidopsis* has revealed the presence of 14 highly conserved PYR/PYL/RCAR protein encoding genes, out of which 9 receptor proteins majorly interact with ABA insensitive 1 (ABI1) (Ma et al., 2009; Nishimura et al., 2009a; Park et al., 2009). Park et al., (2009) and Ma et al., (2009) identified PYR/PYL/RCAR receptor as a direct ABA receptor. In the presence of ABA, PYR/PYL/RCAR receptor–ABA complex forms which interact with PP2C. This trimeric complex, results in inhibition of phosphatase activity of PP2Cs. Resolution study reveals the ABA binding with PYR1/PYL1/PYL2 receptors. This event leads to closure of proline cap which in turn limits the enzyme present in the cavity resulting in the prevention of contact with the solvent while the lock seals the lid over ABA. The ABA and PYR/PYL/RCAR receptor binding brings conformational changes, thereby creating a new site and PP2C‘s interaction with PYR/PYL/RCAR receptor further stabilizes the structure. This interaction shields the catalytic site of PP2C, leading to inactivation of the activity (Park et al., 2009; Ma et al., 2009; Nishimura et al., 2009a; Santiago et al., 2009a). The disruption of ABA-PYR1 interaction by mutating the residues which make direct contact with ABA reduces ABA sensitivity in the mutant, thus highlighting the importance of proline cap and leucine lock (Nishimura et al., 2009a).

The role of PYR/PYL/RCAR has been deciphered in case of quadruple mutants of *pyr1, pyl1 pyl2* and *pyl4* which are found to be insensitive towards ABA (Park et al., 2009), whereas over expression of RCAR1/PYL1, PYL5/ RCAR8 or PYL8/RCAR3 leads to more tolerance towards drought stress by an increase in response towards ABA in *Arabidopsis* (Ma et al., 2009; Santiago et al., 2009b; Saavedra et al., 2010). PYR/RCAR interaction with ABI1 is found to be ABA dependent as in case of PYR1 and PYL1-4 or constitutive as observed in case of PYL5-12 (Nishimura et al., 2009a). BiFC and co-immunoprecipitation assays have established PYL5 as an HAB1 antagonist. PYL5 inhibits HAB1, ABI1 and ABI2 functions in an ABA dependent manner by activating ABA signalling cascade (Park et al., 2009, Ma et al., 2009). Even in case of ABI1 and ABI2, HAB1 activity is inhibited by stereo-specific RCAR1, RCAR3, RCAR8, RCAR11 and RCAR12 receptors in the presence of ABA, but not observed in the absence of receptors (Ma et al., 2009; Nishimura et al., 2009a; Park et al., 2009; Mehrotra et al., 2014).

1.1.2 The PP2Cs as negative regulators of ABA signalling cascade

Phosphatases are broadly divided in two major families based on their substrate specificity i.e. serine/threonine phospho protein phosphatase (PPP) family and phospho tyrosine phosphatase family (PTP). The PPP family is further classified into two groups PP1 and PP2 based on their pharmacological properties. Depending on the requirements of divalent ions for their activity, the PP2 group is again sub-divided into PP2A, PP2B and PP2C. PP2A does not require divalent ions whereas PP2B and PP2C require Ca^{2+} and Mn^{2+}or Mg^{2+}, respectively, for their activity. PP1, PP2A and PP2B have shown sequence similarities in their catalytic subunit, thus falling into PPP family. PP2Cs, pyruvate dehydrogenase phosphatase and other Mg^{2+} dependent phosphatases have been grouped under a separate PPM family (Luan 2003).

Based on phylogenetic analysis, PP2C isoforms are classified into 13 subfamilies i.e. A-L in *Arabidopsis thaliana* (Xue et al., 2008). The members of subfamily A are global negative regulators of ABA signalling pathway (Schweighofer et al., 2004; Umezawa et al., 2010; Weiner et al., 2010; Nakashima amd Yamaguchi-Shinozaki 2013). PP2Cs are monomeric enzymes and the largest protein phosphatase family known in plants. The catalytic and regulatory domains of PP2C enzyme are present on the same polypeptide. 80 and 90 members of PP2Cs have been found in *Arabidopsis* and rice

respectively (Xue et al., 2008; Singh et al., 2010). Members of PP2Cs viz ABA-insensitive (ABI) mutants (ABI) 1, ABI2, HAB1 (homology to ABI1) HAB2, AHG1/AHG3 (ABA-hypersensitive germination I) are negative regulators of ABA dependant pathway (Schweighofer et al., 2004; Xue et al., 2008; Mehrotra et al., 2014). The role of PP2Cs was evident from the double or triple PP2C knockout mutants which have shown more hypersensitivity towards ABA (Nishimura et al., 2007, Rubio et al., 2009).

The expression of PP2Cs varies in different tissues. ABI1 is expressed in vegetative tissues, seeds and guard cells and are found both in cytosol and nucleus, whereas AHG1 and AHG3/AtPP2CA are expressed mostly in seeds and localized in the nucleus (Yoshida et al., 2006, Nishimura et al., 2007, Umezawa et al., 2009). In the absence of ABA or unstressed conditions, ABI1, ABI2 and HAB1 inactivates the members of SnRK2 family – namely SnRK2.2, 2.3, and 2.6 by dephosphorylation. Tougane et al., (2010) have isolated and characterized *MpABI1*, an ortholog of *ABI1* from liverwort *Marchantia polymorpha.* The transient overexpression of *MpABI1* in *Marchantia polymorpha* and *Physcomitrella patens* cells have shown suppression of ABA-induced expression of the wheat promoter fused to the *β-glucuronidase* gene. Transgenic *Physcomitrella patens* expressing *MpABI1* and its mutant construct, *MpABI1-d2* lacking the amino-terminal domain, have shown reduced freezing, osmotic stress tolerance and ABA induced late embryogenesis abundant-like boiling-soluble proteins. *MpABI1* was found to be a negative regulator of ABA signalling in liverworts. The group A PP2Cs are identified as regulators of intrinsic desiccation tolerance in the moss, *Physcomitrella patens*. The enhancement of desiccation tolerance has been observed in the presence of ABA and in absence of PP2Cs (Komatsu et al., 2013). Studies by Komatsu et al., (2013) have revealed that group A PP2Cs evolved early in land plants.

1.1.3 SnRK2s – as positive player of ABA signalling pathway

SnRKs (sucrose non fermenting1 (SNF1) related kinases) are monomeric protein kinases that are involved in stress signal transduction in plants. These can be divided into three sub divisions: SnRK1, SnRK2 and SnRK3 (Kulik et al., 2011; Mehrotra et al., 2014). SnRK1 regulates the energy metabolism of plants in response to sucrose levels, hypoxia, salinity, pathogen attack, etc. (Hey et al., 2010). SnRK3s, also known as CBL Interacting Kinases (CIPKs), respond to salinity regulating pathways, namely, Salt Overly Sensitive

(SOS) which controls expression of several salt stress specific genes (Kulik et al., 2011). Further, SnRK2 family has been divided into three subgroups based on their affinity towards ABA viz. group I, group II and group III. The affinity towards ABA depends upon the C-terminal domain II. Group I kinases do not respond to ABA, Group II does not or weakly responds to ABA while Group III strongly responds to ABA. There are 10 SnRK2 members present in *Arabidopsis* and are named as SRK2A-SRKJ or SnRK2.1-SnRK2.10 (Kulik et al., 2011; Mehrotra et al., 2014). A SnRK2 from rice, OSRK1 auto-phosphorylates the myelin basic protein (MBP) and histone protein which are activated or induced under dehydration stress (Johnson et al., 2002). Direct action of PP2C is observed on sub class III of SnRK2 kinases. It has been observed that the SnRK2 gets phosphorylated in the presence of ABA whereas this is absent in *abi1-1* mutants. The ABA-responsive activation of SnRK2 is regulated by PP2Cs when ABA is not present or plants are not exposed to stress (Umezawa et al., 2009).

In response to ABA dependent pathway, SnRK2D/SnRK2.2, SRK2E/OST1/SnRK2.6 and SRK2I/SnRK2.3 are major positive regulators. SRK2E/OST1/SnRK2.6 is expressed in guard cells and leads to stomatal closure in the presence of ABA in water deficient conditions whereas SRK2D/SnRK2.2 and SRK2I/SnRK2.3 are expressed in seeds and vegetative tissues. A triple mutant in *Arabidopsis thaliana* not only disrupted stomatal closure but also down regulated ABA and water stress induced genes resulting in ABA insensitivity and reduced drought tolerance (Yoshida et al., 2002, 2006; Fujita et al., 2009; Fujii and Zhu 2009; Nakashima et al., 2009a). This mutant also showed up-regulation of JA dependent genes. Researchers have also reported down-regulation of almost all genes encoding dehydration-responsive late embryogenesis abundant (LEA) proteins and group A PP2Cs under water stress conditions in this mutant indicating the essentiality of these three genes in controlling seed development and dormancy (Nakashima et al., 2009a).

ABA and osmotic stress activate SRK2E/OST1 mediated stomatal closure by two independent pathways namely ABA and Osmotic Stress dependent pathways as shown by experiments on ABA-insensitive or ABA-deficient mutants. OST1 has also been shown to positively regulate stress response genes like *rd29B* and *rd22* (Yoshida et al., 2002). SRK2E/OST1 in *Arabidopsis* also interacts with ABI1/ABI2, where both have distinct roles in controlling its activity. Experimental results have indicated dual roles of ABI1 in ABA dependent pathway. It acts as a negative regulator of SRK2E in ABA signalling and as a positive regulator in activation of SRK2E in

low humidity. In the absence of ABA or unstressed condition ABI1, ABI2, and HAB1 dephosphorylate members of SnRK2 family - namely the SnRK2.2, -2.3, and -2.6. Various members of SnRK2s family from different plant species playing role in different abiotic stresses have been reported, as mentioned in Table 1-1.

Table 1-1 Illustration of SnRK2 members from different plant species and their function in different stress conditions

SnRK2	Function	Plant	Reference
PKAB1	Stomatal closure	Wheat	Anderberg and Walker Simmons 1992
SPK1&SPK4	Dehydration & high salinity	Soybean	Yoon et al., 1997
NtOSAK	Osmotic stress	Tobacco	Mikolajczyk et al., 2000
SPK1& SPK2	Osmotic stress	Soybean	Monks et al., 2001
SRK2E/OST1/SnRK2.6	Stomatal response	*Arabidopsis*	Yoshida et al., 2002
SAPK 1-10 (Stress/ABA activated Protein Kinase)	Osmotic stress	Rice	Kobayashi et al., 2004
ZmSPK1	Roots, leaves, vessels & reproductive organ	Maize	Zou et al., 2006
SnRK2.2/3/6	Dehydration, seed maturation & germination	*Arabidopsis*	Fujii et al., 2007
SnRK2D4 & SRK21	Seed germination & root growth	*Arabidopsis*	Nishimura et al., 2009a

ABA signalosome complex includes interaction of ABA with its receptor PYR/PYL/RCAR to form the ABA- PYR/PYL/RCAR complex which in turn binds with PP2C and inhibits its phosphatase activity. SnRK2 phosphorylates serine /threonine residues at R-X-X-S/T sites in the conserved regions of AREB/ABFs. The auto

phosphorylation activity by SnRK2 occurs at Ser-175 residue in its kinase activation loop. The Ser-175 residue is dephosphorylated in the absence of ABA (Umezawa et al., 2010, Mehrotra et al., 2014).

1.1.4 Gene Expression of stress inducible genes

The molecular changes brought in the plant system during stress conditions are reflected by gene expressions. The activation and inactivation of selective genes in response to stress cues is a tightly regulated process, which leads to synthesis of specific proteins in precise amount. The changes induced at molecular level in the plant system occur due to varied stress conditions and stress responsive gene expression. The interaction between the functional *cis* regulatory elements and transcription factors are responsible for temporal and spatial gene expression. The transcription process is set in motion by recognition of specific *cis* regulatory elements known as core promoter by its basal transcription factors and associated transcription factors. This forms the pre initiation complex (PIC) which turns on the basal transcription process. Core promoter architecture plays a distinct role in regulation of light-mediated responses in plant genes (Srivastava et al., 2014). Yoshiharu et al., (2011) studied three types of core promoters- TATA promoters, TATA less promoters and GA rich promoters and analyzed that TATA sequence is rich in stress responsive promoters whereas GA and TATA less promoters are abundant in constitutive gene promoters.

Different regulatory elements such as enhancers, silencers, insulators and other *cis* regulatory elements which contribute to the fine tuning of gene expression, are present upstream to the core promoter and are located in the two virtual regions of the promoter, namely proximal and distal regions (Lee and Young 2000; Hernandez and Finer 2014). The fine tuned transcription is the result of the fully functional promoter which is obtained by the communication of the *cis* regulatory elements present in both the proximal and distal regions. Different sets of *cis* regulatory elements are present in the promoter region and they are associated with different transcription factors or co-transcription factors to activate the promoter region. A single transcription factor can elicit various biological responses from different promoter architectures depending on the kind of additionally recruited transcription factors or co-factors. The binding between the diverse sets of *cis*-regulatory elements and their cognate transcription factors and their interaction forms a complex which further facilitates various biological responses from a

limited pool of transcription factors and promoter architectures. Transcription factors tolerate a certain degree of variability in their *cis* regulatory binding sites, as some transcription factors binds to diverse sequences (same core sequence) with varying binding affinities (Ben-Tabou and Davidson 2007; Pan et al., 2010).

A number of MADS box family members binds to CArG elements and variable sequence of CArG elements identified as $CC(A/T)_6GG$ (Folter and Angenent 2006). The WRKY protein binds to W box TTTGAC(C/T) (Ishiguro S et al., 1994; Rushton et al., 1995; de Pater et al., 1996; Rushton et al., 1996; Eulgem et al., 2000). MYB transcription factors binds to variable sequences such as (GTT(A/T)GTT(G/A), G(G/T)T(A/T)GGT(G/A) (Romero et al., 1998) and (A/C)ACC(A/T)A(A/C)C (Sablowski et al., 1994). For example the bZIP proteins preferentially binds to palindromic or pseudo-palindromic hexamers with an ACGT core *cis* regulatory elements such as the G-box (CACGTG), A box (TACGTA) and C-box (GACGTC) (Foster et al., 1994). The bZIP proteins binding the G-box (CACGTG) are known as G-box binding factors.

Cellular regulatory systems can take advantage of diverse flanking sequences, same core sequence in the sites to get transcription regulation and accomplish the variability of gene expression. Conversely, the majority of the genome comprises non-specific but weak affinity binding sites for particular transcription factors. The *cis*-regulatory elements that are bound by transcription factors with high affinity are the members of the same family (Izawa et al., 1993; Foster et al., 1994). The interaction of *cis* regulatory elements and transcription factors lead to the activation, enhancement or suppression of gene expression. The transcription level is a regulated process which depends on following factors:

- Availability and activation of transcription factors (Ben-Tabou and Davidson 2007)
- The type, number and position of *cis* regulatory elements present in the promoter region (Mehrotra et al., 2005; Mehrotra and Mehrotra 2010)
- The interaction between the *cis* regulatory elements, which regulates the transcription in a constricted manner (Sawant et al., 2005).

A huge portion of plant genome is involved in transcription process. For example, in *Arabidopsis thaliana* an estimated number of 1,346–2,290 putative transcription factor genes are likely to be involved in regulating stress-responsive transcription process

(Davuluri et al., 2003; Guo et al., 2008). The *Oryza sativa* (rice) genome codes for nearly 2300 transcription factors (Riano-Pachon et al., 2007). Reichmann et al., (2000) have made the comparative analysis of transcription factors among the *C. elegans*, *S. cerevisiae*, *Arabidopsis thaliana* and *Drosophila melanogaster* and have studied many transcription factors family which are present only in the plant kingdom or their prevalence is higher in plants viz WRKY, bHLH, MYB, MYC, NAC, bZIP, DREB, ERF. According to plant transcription factor data base PlnTFDB, 2451 distinct protein sequences are arranged in 81 gene families of *Arabidopsis thailana* (Riaño-Pachón et al., 2007). The stress responsive transcription factor database reports 3150, 1118, 1716 stress responsive genes in the *Arabidopsis thaliana, Oryza sativa* sub japonica and *Oryza sativa* sub indica respectively (Naika et al., 2013).

469 *cis* regulatory elements identified in the entire plant kingdom are documented in the PLACE data base (Higo et al., 1999). The size of *cis* regulatory elements varies from 3 to 33 nucleotides in length. Many *cis* regulatory elements and their cognate transcription factors have been studied with the help of EMSA, DNA foot printing and microarray techniques. The interaction among different stress inducible *cis* regulatory elements function in cross-talk between different gene networks (Yamaguchi-Shinozaki and Shinozaki 2005). The synergistic and antagonistic interactions of *cis* regulatory elements in the promoter region are responsible for the gene expression (Sawant et al., 2005). The major stress responsive regulatory elements found in plant system are – abscisic acid-responsive element (ABRE), CGTA motif, coupling element (CE), C repeat element (CRT), dehydration response element (DRE), GCC box, heat shock element (HSE), inducer (ICEr), low temperature element (LTRE), MBS, MYCR, NACR, TC-rich repeats, TCA elements and W-box, which help the plant to withstand the abiotic and biotic stresses. Multiple *cis* regulatory elements are present in the promoter which interact with multiple transcription factors/ co-factors, stablishing a co-ordination and this fine tuned process results into a specific transcriptional activity.

Numerous plant genes get induced in response to drought, high salinity and cold stress. The stress responsive or inducible gene products are functionally classified into two major groups, first group includes those which provide tolerance against stress, viz chaperones, LEA proteins, enzymes for osmolyte biosynthesis and detoxification enzymes, and the other group includes those that activate gene expression and signal transduction in stress response, such as protein kinases, phosphatases, transcription factors and enzymes of phospholipid metabolism. There are some genes which are

induced in response to ABA and are known as ABA- dependent, whereas some stress-responsive genes are not induced in response to ABA and are known as ABA-independent (Figure 1-2). Promoter analysis of the drought and cold inducible genes reveals the existence of ABA- dependent and ABA- independent pathway (Figure 1-2). Further analysis discloses that they have ABRE, DRE, C-repeat, ICER, LTE *cis* regulatory elements, hence indicating the existence of ABA -independent or -dependent pathway. Some of the promoters have only ABA -dependent or ABA -independent *cis* regulatory elements, while some have combination of both ABA- independent and ABA -dependent.

The necessary *cis* regulatory elements of ABA -independent pathway include DRE, C-repeat and LTE CREs. Some of the gene products are common in both the pathways. As several genes are induced by drought, high salinity and cold stress, this suggests that there is cross-talk between these stress-signalling pathways, emphasising the involvement and interaction of different *cis* regulatory elements. The computational analysis of upstream region of 20 *Arabidopsis DREB*-genes revealed the presence of diverse *cis* regulatory elements. Nine major types of significant *cis* regulatory elements related to stress conditions were found in these *AtDREB* genes. These *cis* regulatory elements are - ABRE, (CGTCA, TGACG) MeJA response, MBS, HSE, TC rich, TCA, LTR and ERE motifs. Out of these elements, ABRE element is found to be present predominantly in the upstream region of *DREB* genes. This shows that the presence of ABRE elements in the promoters of *DREB* genes indicates the role of ABA-signaling in the regulation of *DREB* gene expressions (Sazegari et al., 2015).

In addition to these major pathways, other regulatory factors including the NAC and MYB/MYC (R2-R3) (myeloblastosis/ myelocytomatosis) factors are also involved in abiotic stress-responsive gene expression. The NAC transcription factor family has been shown to interact with *cis* regulatory elements present in the *erd1* (early response to dehydration stress) promoter under osmotic stress and induce transcription in response to both ABA-dependent and ABA–independent pathway (Figure 1-2) (Tran et al., 2004).
Zou et al., (2011) identified 1,215 putative *cis* regulatory elements (pCREs) in the putative promoter regions of up- and/or down-regulated stress responsive genes. Similar analysis was carried out in control promoter sequences of those genes which do not show considerable changes under stress cues. From this study, pCREs involved in abiotic and biotic stress responsive gene regulation are classified into two motif super families- α and β. As much as 19.6% of pCREs implicated in transcriptional regulation of abiotic stress

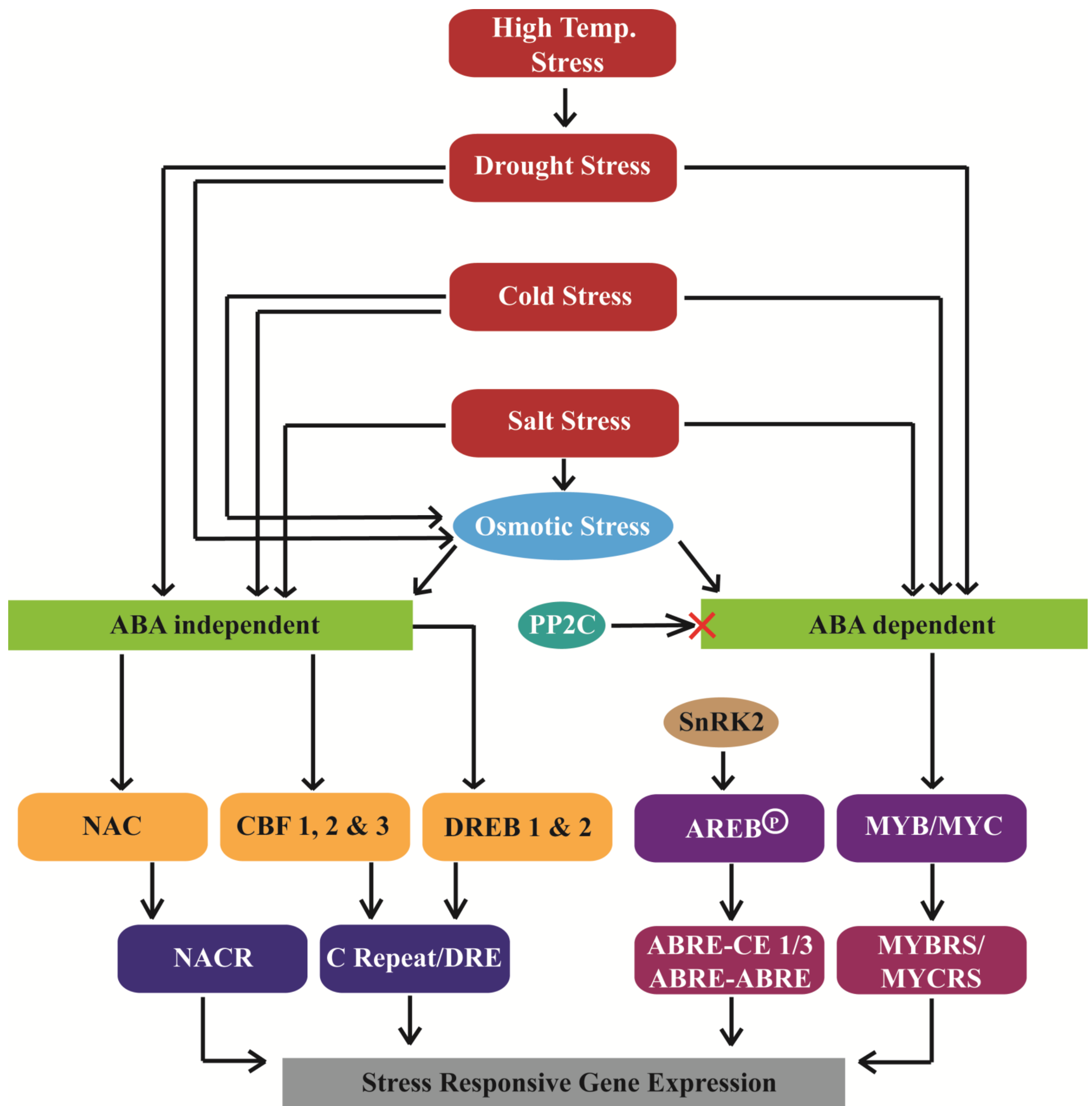


Figure 1-2 Overview of ABA-dependent and ABA-independent abiotic stress pathways.

responsive genes have ACGT as the core sequence, whereas pCREs concerned with biotic stress responsive regulation have TTGAC as the core sequence, also known as W-box (Zou et al., 2011).

The tetrameric ACGT sequence forms the core sequence of many other *cis* regulatory elements, such as light, UV, anaerobiosis, auxin, and salicylic acid, which perform their respective functions other than stress regulation as shown in the figure 1-3. The variants of ACGT element, reported in the PLACE database, due to the variation in the flanking sequence- ABREAZMRAB28 (GCC**ACGT**GGG), ABREMOTIFAOSOSEM (T**ACGT**GTC), ACGTATERD1 (**ACGT**), ABRE2HVA1 (CCT**ACGT**GGCGG, GCA**ACGT**GTC, CGC**ACGT**GTC), ABRE (**ACGT**GGC, CGC**ACGT**GTC, C**ACGT**G), **ACGT**GG, **ACGT**GT, CC**ACGT**, and GCC**ACGT**, function under different abiotic stress conditions.

Oliphant et al., (1989) assigned numbers to the position of nucleotide within the ACGT core element. The CG nucleotides of ACGT core represents the zero position and nucleotides towards 5' half represent the negative value and the sequence towards 3' half represent the positive value $C_{-3}C_{-2}A_{-1}C_{0}G_{0}T_{+1}G_{+2}G_{+3}$ (Figure 1-4) (Oliphant et al., 1989; Izawa et al., 1993). Therefore, based on the position of nucleotides at +2, the ACGT containing *cis* regulatory elements is classified into different boxes like A-box, C-box, G-box, or T-box. Among these boxes, C and/or G-boxes are the most preferred ones. The nucleotides at position 2 and 3 determine bZIP transcription factors binding specificity and affinity (Figure 1-4) (Izawa et al., 1993).

Earlier the G-box was identified in the promoter region of the small subunit of rubisco (Giuliano et al., 1988). Later it was identified in upstream region of other genes viz wheat *em* gene (Marcotte et al., 1989), *AtAdh* gene (Dolferus et al., 1994), *strictosidine synthase* gene (Ouwerkerk and Memelink 1999) and ABA-responsive genes. Many ABA inducible genes share the (C/T) ACGTGGC consensus, *cis*-acting ABRE in their promoter regions (Guiltinan et al., 1990). Liu et al., (1996) have identified that G-box binding factor binds *in vitro* to half G-box sequence (CCAAGTGG) where ACGT core element was missing. The neighbouring sequence of ACGT core element provides the difference which facilitates in the binding of appropriate transcription factors (Menkens et al., 1994; Uno et al., 2000; Pan et al., 2010).

The genome-wide *in silico* analysis by Gomez-Porras et al., (2007), showed the *cis*-regulatory sequence combinations, ABRE-ABRE pairs in the upstream region emerges as

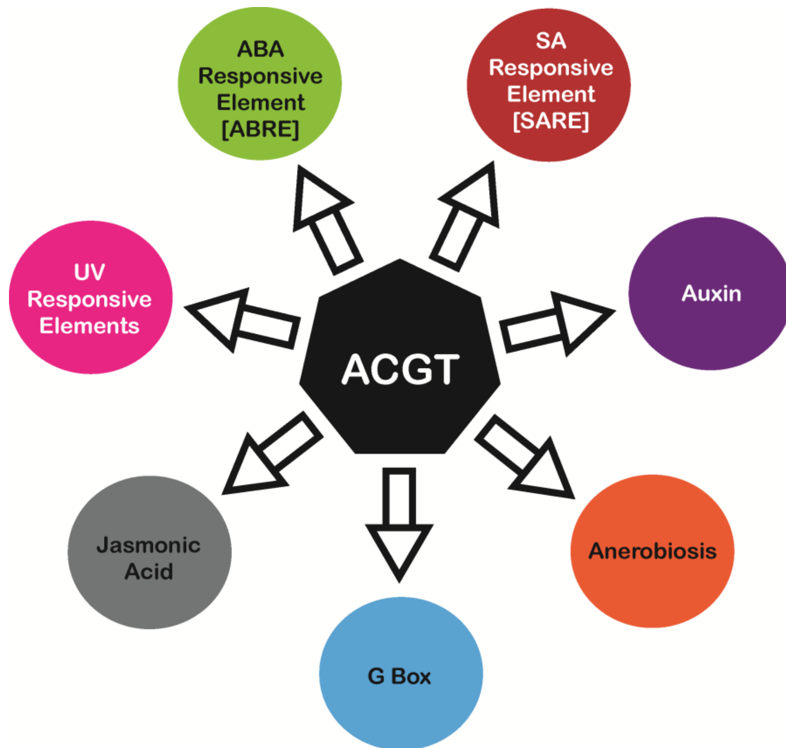


Figure 1-3 ACGT *cis* regulatory elements forms the core sequence of other *cis* regulatory elements.

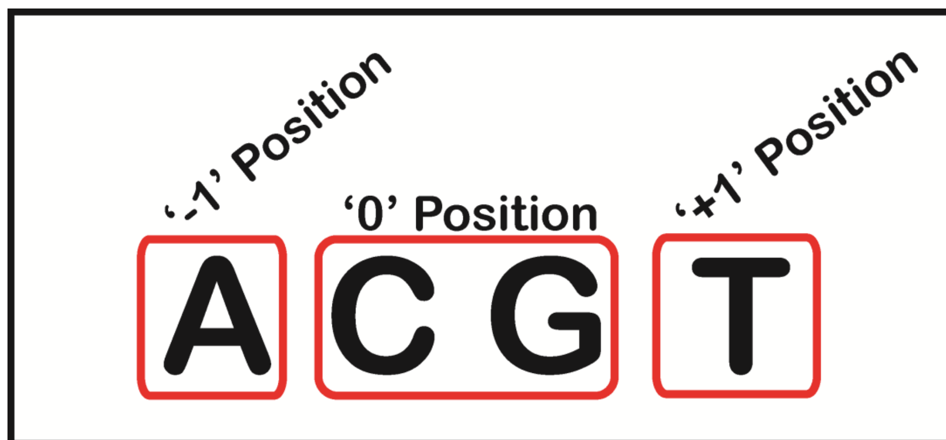


Figure 1-4 The numbering given to ACGT core and its flanking nucleotides by Oliphant et al., (1989).

a major combination in ABA-responsive genes in both *Arabidopsis* and rice genome (Gomez-Porras et al., 2007). A single copy of ABRE is not effective for the ABA-responsive genes (Hattori et al., 2002). ABA-dependent mediated regulation of gene expression involves presence of two ABREs or one ABRE and one coupling (CE) element, required for ABA responsiveness (Hattori et al., 2002). In wheat, *em1*a contains an ACGT box and *em2*a which is 16 bp upstream of *em1a* serve as a CE. An ACGT box and CE1 like sequence are present in *em1b*. The *rab16b* in rice contains an ACGT box and CE1 like sequence 140 bp upstream of this gene which was also induced in response to ABA. A study on barley gene promoters namely *HvA1* and *HvA22* reveals the arrangement of ABRE and suggested that the presence of ACGT element and CE1 and CE3 elements are necessary and sufficient for ABA induction. *HvA22* gene contains ACGT box and CE1, which were 20 bp apart (Hattori et al., 2002). The *rab16 A-D* (responsive to ABA) gene family promoter from rice was also found to be induced in response to ABA and osmotic stress. The analysis of upstream region of *rab16 A-D* gene showed two conserved regions – motif I (GTACGTCGC) having ACGT core sequence, and motif II (CGG/CCGGGCT). Two copies of motif II were observed in *rab16 A* while only one copy was present in *rab16 B-D*. Skriver et al., (1991) observed that motif I was important for conferring ABA-inducible gene expression. These results indicate that motif I is an essential part of the ABRE (Mundy and Chua 1988; Yamaguchi-Shinozaki et al., 1989). Similarly *rab16 A* promoter region of *em* gene in wheat also has two conserved regions, *em1a* (CACGTGGC) and *em1b* (ACACGTGCC) (Marcotte et al., 1989). The *em1a* and *em1b* are representatives of ABRE box.

The *rab -21* gene promoter was found to be induced in response to ABA and osmotic stress. The promoter of *rd29A* gene has been found to contain both ABRE as well as DRE/CRT elements which get induced in response to both ABA-dependent and ABA-independent pathway (Yamaguchi-Shinozaki and Shinozaki 1994). The promoter of *rd29B* gene has two AREB elements, which are induced in response to only ABA and activate the ABA-dependent pathway. The *rd22* gene promoter has MYC and MYB transcription factor recognition sites along with ABRE element and plays an important role in ABA signalling (Abe et al., 1997, 2003). The ABRE, CE3 and G-box *cis* regulatory elements were found to be widely distributed in the upstream regions of sub-family A of PP2C genes in the whole genome of *Arabidopsis* and rice (Xue et al., 2008). Another major finding is that the two ACGT *cis* elements separated by five nucleotides

are induced in response to SA and similar motifs separated by 25 bp are induced in response to ABA (Mehrotra et al., 2010). This indicates the importance of the spacer distance between two ACGT *cis* elements.

1.1.5 bZIP proteins and their regulation under ABA-dependent pathway

bZIP proteins are found to be present in *C. elegans*, yeast, animals and plants. The prevalence of bZIP is more in plants than other organisms and it is the eighth largest transcription factor family in *Arabidopsis* genome (Reechman et al., 2000). The number of bZIP genes present in *Arabidopsis* is 75, rice (*Oryza sativa)* - 84, sorghum (*Sorghum bicolour)* - 92, soybean (*Glycine max)* – 131, maize (*Zea mays)* – 125 and 55 genes in the grapevine (*Vitis vinifera*) genome (Wei et al., 2012). In plants, bZIP transcription factors play an important role in various biological processes such as light response (Scindler and Cashmore 1990), pathogen defence (Alves et al., 2013), abiotic stress signalling (Fujita et al., 2005), hormone signalling (Choi et al., 2000), energy metabolism (Baena-González et al., 2007), flowering development (Abe et al., 2005), senescence (Smykowski et al., 2010) and seedling maturation (Alonso et al., 2009).

The name bZIP is given to this transcription factor family due to the presence of a basic region followed by a stretch of leucine amino acids in their structure. The bZIP proteins have basic DNA binding region present towards the N-terminal and a leucine zipper region towards C-terminal and a variable region required for protein - protein interaction. The basic region is necessary for DNA sequence identification, binding and nuclear localization signal (Vander Krol and Chua 1991; Hakkoshima 2005). Some difference has been found in the leucine zipper portion of bZIP proteins in plant and animals. In animals, bZIP proteins belong to the CCAAT/enhancer-binding proteins (C/EBP) family. The animal bZIPs have a valine amino acid in the signature basic region at the 5th position. This valine residue is responsible for identification of A or G nucleotides present at -3 positions in the DNA sequence (Miller et al., 2003) and this helps in inhibiting the non-specific interaction. The bZIP family members- AP-1-like factors (YAP) and CAMP response element binding protein-2 (CREB2) in *Schizosaccharomyces pombe* has a hydrophobic amino acid at the 8th position in basic region which interacts with the AT-rich DNA binding site (Fujii et al., 2000).

In plants, the bZIP family was subdivided into 10 groups and were named A to I (Table 1-2). This subdivision was based upon the protein composition and DNA

recognition sequence. An additional subgroup of bZIP TFs named as S was found to be present in *Arabidopsis* (Jakoby et al., 2002; Nijhawan et al., 2007; Wei et al., 2012). The characteristic features of bZIP protein sub-groups are summarized in table 1-2.

A phylogenetic analysis of *bZIP* genes to identify homologues gene group was done by Guedes Corrêa et al., (2008) in algae, mosses, ferns, gymnosperms and angiosperms. It was concluded that bZIPs from *Homo sapiens*, *Caenorhabditis elegans*, *Drosophila melanogaster*, *Saccharomyces cerevisiae*, *Arabidopsis thaliana* and *Oryza sativa*, all have a common ancestor bZIP gene, which was found to be present in primitive eukaryote *Giardia lamblia*. This study highlights the existence of bZIP genes in the eukaryotic system even before the division of eukaryotes came into existence in different lineages. Groups B, C, D, E, F, G, I and J were identified in bryophytes and tracheophytes, Guedes Corrêa et al., (2008) study also mentions that there is a functional connection before the colonization to terrestrial plants. The above mentioned research also indicates that Group D and I were important for transition from early land plants to angiosperms, as their gene number increased as compared to the other groups of homologues (i.e. 5 to 10, and 4 to 11 genes in groups D and I, respectively). Group A bZIP genes were identified in spermatophytes and Groups K, L and S bZIP genes were exclusive found only in angiosperms. Group K and S bZIP members are related to Proto-B and Proto-C respectively, which emerged before green plant evolution. Guedes-Corrêa et al., (2008) analyzed Group H members and found them to be the most conserved group out of the ten groups of *bZIP* homologues in green plants. The members of group H- *HY5* and *HYH* are involved in regulation of light responses and anthocyanin biosynthesis.

The plant bZIP proteins are known as ABRE binding factors - (ABFs) target ABRE elements. The ABFs belong to group A, according to the classification done by Jakoby et al., (2002) as mentioned in Table 1-2. The nine bZIP proteins from *Arabidopsis*, being members of group-A bZIP family are classified as homologs of AREB/ABFs on the basis of presence of four conserved domains, in addition to the bZIP domain (Bensmihen et al., 2002).

The basic DNA binding domain of AREB/ABFs is found to be poised with twenty amino acids at N-terminal whereas leucine zipper domain present at C-terminal is composed of seven amino acids. A double hydrophilic leucine zipper helix structure is framed by seven amino acids which may interact with ABRE. The flanking sequence of ABRE has shown a degree of variability, for example in rice, a consensus ABRE with variable flanking sequence T/G/CACGTGG/TC has been observed (Hattori et al., 2002).

ABRE binding factor 1 (ABF1), a member of ABF family, interacts with ABRE sequences present in the promoter region of ABA responsive genes. A study was carried out on the interaction between a homodimer AtABF1 and ABRE sequence by using a comparative model followed by mutational analysis by an empirical force field based method FoldX (Sarkar and Lahiri 2013). The protein–DNA complex was observed even by changing the ABRE sequences using the protein design algorithm FoldX. With the help of high throughput free energy, binding of ABF1 to other core motifs like GCGT or AAGT was also studied to determine the binding specificity of ABF1. The results illustrated that variations occur around CACGTG, which in conformity with the data, stated that the flanking residues have a crucial role in protein -DNA recognition and the stability of the protein-DNA complex. Flanking sequences of the hexameric core can also differentiate between two or more members of the same ABF family (Izawa et al., 1993). Along with this, the role of the flanking sequences was also recognized in determination of specificity of the protein-DNA interaction (Sarkar and Lahiri 2013).

The bZIPs bind as dimers to the target DNA sequence, and bZIP dimers forms a Y-shaped structure to hold the DNA molecule via its DNA binding domain (Figure 1-5). While holding the DNA it forms a "L" shape, so that leucines form the interface in the bZIP dimer. The numbering code of the ACGT core and the flanking sequence, determine the protein binding affinity of bZIP proteins (Williams et al., 1992, Izawa et al 1993). Hence, this variation in the *cis*-element sequence plus ACGT core sequence and the existence of unique basic regions are able to discriminate the bZIP. The 5 residues of the basic region of bZIPs, come in contact with 12 base pair targeted sequence present in the major groove of DNA. These connections are further extended with the help of water molecules (Fujii et al., 2000). At a time, only five out of nine amino acids (N-X7-R/K) are able to make important contacts with the target sequence resulting in a variation, hence becoming the special feature of bZIPs. The homo- and hetero-dimerization property of bZIP allows them to carry out diverse biological functions. Dimerization process is due to the presence of luecine zipper (LZ), composed of repeats of seven leucine residues which are arranged around the alpha- helix turns. The arrangements of seven leucine residues are known as heptads. Deppmann et al., (2006) has given nomenclature to position of amino acids arranged in heptad from a to g (Figure 1-6).

The amino acids present at a, d, e and g positions are responsible for specific interaction (Figure 1-6) (Deppmann et al., 2004). According to amino acid composition, bZIP forms homo/hetero- dimers between two paralogs known as quasihomodimers.

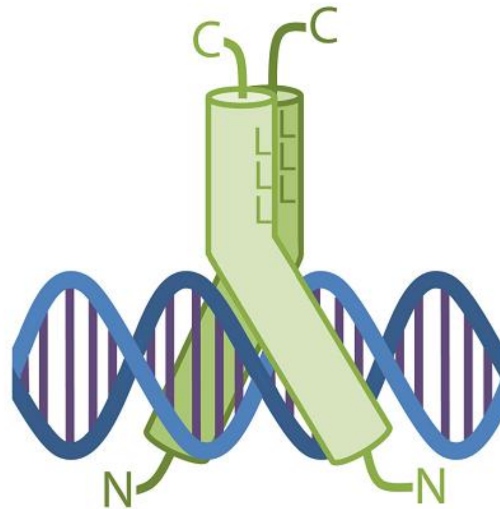


Figure 1-5 The two monomers of bZIP forms a Y-shape structure when it binds to its target DNA sequence perpendicularly. The bZIP monomers interact with major groove through their DNA-binding domains. C and N in figure symbolize the terminal ends and L represents the leucines forming the interface in the bZIP dimers (Adopted from Llolca et al., 2014 Frontiers of Plant Science)

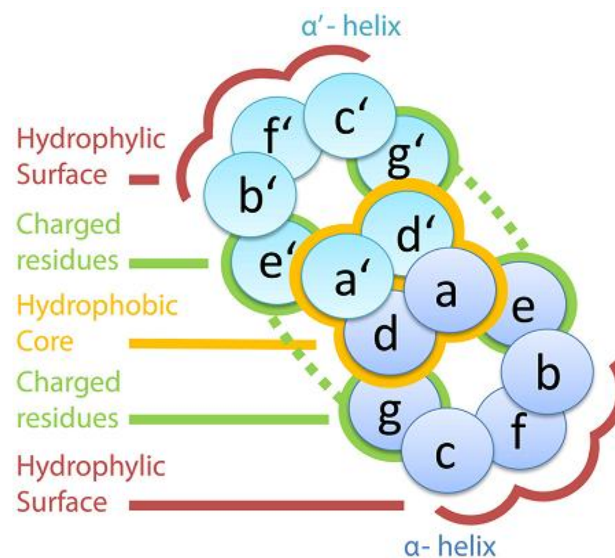


Figure 1-6 The amino acids in position a and d configure the hydrophobic core, which is indicated by yellow halo colour. The amino acids in position e and g depict the charged residues which generate electrostatic forces, represented by dashed green lines. The hydrophilic surface is represented by amino acids b, c and f (Adopted from Llolca et al., 2014 Frontiers of Plant Science)

Quasihomodimer formation occurs between group G, H or A bZIP family members. Heterodimers are formed between members of group E (bZIP34/ bZIP61) and I or S (bZIP51/ bZIP43) (Llorca et al., 2014). The presence of asparagine at a-a' position facilitates the homo-dimerization, while if this position is occupied by lysine or serine than hetero- dimerization is favoured. Charged or polar amino acids are present at the e and g positions and are responsible for stabilizing the helix. The charge of the amino acid residues determines the attractive or repulsive force. The dimerization specificity of bZIP monomers is regulated through post-translation modification mediated by phosphorylation/dephosphorylation activity of leucine zipper region.

The genome-wide transcriptome analysis done by Yoshida et al., (2010) showed that stress-responsive gene expression is extremely impaired in the *areb1 areb2 abf3* triple mutant in comparison to double mutant and wild type. Further, a novel ABRE-dependent AREB/ABF target gene expressed in response to osmotic stress was identified, including many LEA class genes and group-A PP2C genes and TFs (Yoshida et al. 2010). From this study, it can be concluded that AREB1/ABF2, AREB2/ABF4 and ABF3 are principal transcriptional activators that cooperatively regulate ABRE-dependent gene expression in ABA signalling under osmotic stress conditions.The AREB1/ABF2, AREB2/ABF4, ABF1 and ABF3 members of AREB/ABF subfamily are found to be mainly expressed in vegetative tissues under abiotic stress conditions (Choi et al., 2000; Uno et al., 2000; Kang et al., 2002; Fujita et al., 2005). Differential responses of (AREB1/ABF2, AREB2/ABF4, ABF1 and ABF3) members of AREB/ABF subfamily have been observed under abiotic stress conditions.

ABF1 gene mainly responds to cold stress (Kim 2006). In contrast, AREB1/ABF2, AREB2/ABF4 and ABF3 are found to be induced by both ABA and osmotic stresses such as dehydration and high salinity, not by cold stress (Uno et al., 2000, Kang et al., 2002, Kim et al., 2004b, Fujita et al., 2005). Yoshida et al., (2010) have observed that AREB1/ABF2, AREB2/ABF4 and ABF3 proteins form homo- and hetero-dimers and have largely overlapping functions.

ACGT *cis* elements are present in diverse group of promoters in plants. Studies have shown the over–representation of ABRE-ABRE, DRE-ABRE and MYB-ABRE pairs in the gene promoters in *Arabidopsis* (Gomez-Porras 2005; Gomez- Porras et al., 2007). From preliminary analysis, it was indicated that the two ACGT elements in close vicinity with difference in the spacing responded differentially to different elicitors. Many studies have shown the occurrence of ACGT elements in highly expressed genes

upstream of TATA box. This class of activator modules are present in different copy numbers and in different relative locations upstream of genes in plants. One such gene promoter located at chromosome number 5 (Accession No. AT5G59220) having putative ABA inducible ACGT *cis* element has been taken for the present study. This promoter has two ACGT elements which are separated by 30 nucleotides. In addition to this, it has another set of ACGT *cis* element separated by 5 nucleotides.

Hence, this class of *cis* regulatory elements were selected for the present study with the following objectives:

1. ***In-silico* analysis of *Arabidopsis thaliana* genome to identify stress responsive *cis* elements.**
2. **Cloning of the *PP2C*-like promoter (At5g59220) from *Arabidopsis thaliana* for investigation of its role in biotic and abiotic stress.**
3. **Characterization of *PP2C*-like promoter through deletion analysis.**
4. **Study of transient expression of promoter-reporter gene constructs under induced biotic and abiotic stress.**
5. **Development of stable transgenic plants carrying different deletion constructs and carrying out studies under different stress conditions.**

Table 1-2: Characteristic features of the bZIP group of proteins (Adopted from Jakoby et al., 2002 & Guedes Corrêa et al., 2008)

Group	Exon No.	Size (aa)	Basic domain in position	No. of Leucine repeats	Known interaction with other proteins*	Known binding sites	Additional features	Function
A	1-4	234 - 454	C-terminal	3-4	ABI5 with ABI3, In rice TRAB1 with OsVP1	C**ACGT**GG/Tc, CGCGTG For ABF1* & TRAB1	Conserved motifs containing phoshorylation sites	Involved in seed development germination, dessication/dehydration & abiotic stress
B	2,4	523 - 675	C-terminal to N-terminal	5	Not known	Not known	Putative transmembrane domain in C-terminus of Atb ZIP17 & AtbZIP 28. Proline-rich domain C-terminal of bZIP domain	Involved in oxidative stress signaling and unfolded protein response
C	6,7	294 - 411	Central to C-terminal	7-9	Opaque2 with PBF	GTG**AGT**CAT for barley BLZ1 & BLZ2 CC**ACGT**GG & TGACGTCA for CPRF2	N-terminal hydrophobic or acidic signature (activation domain), Ser/Thr cluster + acidic aa stretch. Putative phosphorylation	Involved in nitrogen/carbon balance
D	7,8,10-13	325 - 481	N-terminal	3, plus G at position 4	TGA3, 2 & 5 with NPRI1, TGA4 with AtEBP	TG**ACGt**/g for TGA1-6	None	Involved in pathogen defense, biotic responses & ion responses
E	4	--	C-terminal	6-7	Not known	Not known	N-terminal stretches of basic aa	-
F	1,2	157 - 260	Central	8, Q at position 4	Not known	Not known	None	-
G	--	--	C-terminal	5	GBF1, 2 & 3 heterodimerize pair wise	CC**ACGT**GG for GBF1*, GBF2* & 3	Proline- rich N-terminal activation domain	Involved in light responses & nitrogen/carbon balance
H	3,4	148, 169	C-terminal	5	HY5 with the N-terminus of COP1	AC**ACGT**GG for HY5	COP1 interaction domain	Majorly involved in light regulation
I	1-5	157 - 553	Central to C-termi	7	NtRSG with 14-3-3 proteins	TCC**AGCT**TGA , TCCAACTTGG A for tobacco	Conserved lysine in position -10 of the basic region	Involve in vascular gene expression

			nal			RSG GCTCCGTTG for tomato VSF-1		
S	1	145-186	Central	8-9	Snapdragon bZIP910 & 911 heterodimerize. Tobacco BZI, 2,3 & 4 heterodimerize with BZI-1 (group C)	TG**ACGTG** for Snapdragon bZIP910* & bZIP 911	Short N- & C-terminal extensions	Involved in unfolded protein response & energy homeostasis

*Interaction of bZIPs and other proteins in *Arabidopsis thaliana* and other plants, and heterodimerization between bZIP proteins. An asterisk (*) after the protein name indicates that binding sites selection experiments have been performed. The ACGT core or part of it is in bold when present. Abbreviation: aa- amino acid.

MATERIALS and METHODS

CHAPTER 2

2.0 Materials and methods

2.1 The sequence extraction from databases:

The recurring frequency of ACGT *cis*-elements was searched in *Arabidopsis thaliana* genome. To achieve this objective, *in-silico* analysis of the ACGT *cis* element sequence was done using TAIR (The *Arabidopsis* Information Resource, version 10) database. The *in-silico* analysis of two ACGT elements with and without spacer (ACGT $_{0 \leq N \leq 30}$ ACGT) sequence was carried out. The frequency of occurrence of ACGT element was analyzed across the genome of *Arabidopsis thaliana* by using NCBI BLASTn. The accession numbers of *Arabidopsis thaliana* chromosomes are:

- chromosome 1 NC_003070.9
- chromosome 2 NC_003071.7
- chromosome 3 NC_003074.8
- chromosome 4 NC_003075.7
- chromosome 5 NC_003076.8

BLAST was done for all the sequences (ACGT sequence with variable spacer length) used for analysis using the *Arabidopsis thaliana* genome as a subject (Table 2-1 and 2-1-1). Further, a random sequence was generated using the bioinformatics program SHUFFLE (http://www.bioinformatics.org/sms2/shuffle_dna.html) (Doelz 1990) which had the ACGT element or spacer along with the whole sequence, and this sequence was used as a control. The random sequence was further searched in the genome of *Arabidopsis thaliana* in order to analyze the frequency of occurrence.

Table No. 2-1 Sequences used for frequency analysis across the *Arabidopsis thaliana* genome

$ACGT_NACGT$	Sequence
(ACGT)$_{N1}$(ACGT)	ACGT_ACGT
(ACGT)$_{N2}$(ACGT)	ACGTG_ACGT
(ACGT)$_{N3}$(ACGT)	ACGTGG_ACGT
(ACGT)$_{N4}$(ACGT)	ACGTGGC_ACGT
(ACGT)$_{N5}$(ACGT)	ACGTGGCT_ACGT
(ACGT)$_{N6}$(ACGT)	ACGTGGCTA_ACGT
(ACGT)$_{N7}$(ACGT)	ACGTGGCTAT_ACGT
(ACGT)$_{N8}$(ACGT)	ACGTGGCTATG_ACGT
(ACGT)$_{N9}$(ACGT)	ACGTGGCTATGG_ACGT
(ACGT)$_{N10}$(ACGT)	ACGTGGCTATGGC_ACGT
(ACGT)$_{N11}$(ACGT)	ACGTGGCTATGGCG_ACGT
(ACGT)$_{N12}$(ACGT)	ACGTGGCTATGGCGG_ACGT
(ACGT)$_{N13}$(ACGT)	ACGTGGCTATGGCGGA_ACGT
(ACGT)$_{N14}$(ACGT)	ACGTGGCTATGGCGGAG_ACGT
(ACGT)$_{N15}$(ACGT)	ACGTGGCTATGGCGGAGC_ACGT
(ACGT)$_{N16}$(ACGT)	ACGTGGCTATGGCGGAGCA_ACGT
(ACGT)$_{N17}$(ACGT)	ACGTGGCTATGGCGGAGCAA_ACGT
(ACGT)$_{N18}$(ACGT)	ACGTGGCTATGGCGGAGCAAG_ACGT
(ACGT)$_{N19}$(ACGT)	ACGTGGCTATGGCGGAGCAAGA_ACGT
(ACGT)$_{N20}$(ACGT)	ACGTGGCTATGGCGGAGCAAGAT_ACGT
(ACGT)$_{N21}$(ACGT)	ACGTGGCTATGGCGGAGCAAGATT_ACGT
(ACGT)$_{N22}$(ACGT)	ACGTGGCTATGGCGGAGCAAGATTC_ACGT
(ACGT)$_{N23}$(ACGT)	ACGTGGCTATGGCGGAGCAAGATTCA_ACGT
(ACGT)$_{N24}$(ACGT)	ACGTGGCTATGGCGGAGCAAGATTCAC_ACGT
(ACGT)$_{N25}$(ACGT)	ACGTGGCTATGGCGGAGCAAGATTCACT_ACGT
(ACGT)$_{N30}$(ACGT)	ACGTAAGTGTTTCGTATCGCGATTTAGGAGAAGTACGT

Table 2-1-1 Sequences used for frequency analysis across the *Arabidopsis thaliana* genome

S.No	Sequence
$(ACGT)_2$	ACGTACGT
$(ACGT)_8$	ACGTACGTACGTACGTACGTACGTACGTACGT
$(ACGT)_{N1}(ACGT)$	ACGT_ACGT
$(ACGT)_{N2}(ACGT)$	ACGTG_ACGT
$(ACGT)_{N3}(ACGT)$	ACGTGG_ACGT
$(ACGT)_{N4}(ACGT)$	ACGTGGC_ACGT
$(ACGT)_{N5}(ACGT)$	ACGTGGCTAACGT
$(ACGT)_{N10}(ACGT)$	ACGTGGCTATGGCGACGT
$(ACGT)_{N25}(ACGT)$	ACGTGGCTATGGCGGAGCAAGATTCACTCACGT
$(ACGT)_{N30}(ACGT)$	ACGTAAGTGTTTCGTATCGCGATTTAGGAGAAGTACGT
$(ACGT)_{RN5}(ACGT)$	ACGT*GC*TAGACGT
$(ACGT)_{RN25}(ACGT)$	ACGTATATGAGATCGGCGCTTCACGGAGCACGT
$(ACGT)_{RN30}(ACGT)$	ACGT-TATGGGACTTTTAACCATAATAGGGGGTTG-ACGT
$(ACGT)_{N10}(ACGT)$ randomized	GCGGGCTATCGGTAGCAT
$(ACGT)_{N25}(ACGT)$randomized	TAAGGCTTAGCCACGCTTAGGGTGTGAGCACAC
$(ACGT)_{N30}(ACGT)$randomized	TTCGGAATTGGTAGAAGTCATTGGGATTACCCTGGATA

N stands for the spacer sequence. R stands for randomization of the sequence.

The spacer sequence was increased from 0 to 30.

RN5, RN10, RN25 - implies only spacer sequence was being randomized.

$(ACGT)_N$ (ACGT) randomized – complete sequence was randomized.

2.2 Analysis of transcription factor binding site

The transcription factor binding sites were searched using ACGT elements with or without spacer sequence conjugated to a 139 bp long minimal promoter sequence (MPS) – *Pmec* (minimal expression cassette). The sequence of this minimal promoter *Pmec* is as follows - TCACTATATATAGGAAGTTCATTTCATTTGGAATGGACACGTGTTGTCATTTCT CAACAATTACCAACAACAACAAACAACAAACAACATTATACAATTACTATTTA CAATTACATCTAGATAAACAATGGCTTCCTCC (Sawant et al., 2001).

These extended sequences were then used to explore transcription factor binding sites. For this JASPAR core database (Bryne et al., 2008) was used and the results obtained from this database were validated with CONSITE database (Sandelin et al., 2004). The following sequences were used for analysis of transcription factor binding sites.

- MPS
- ACGT
- (ACGT)2MPS
- (ACGT)8MPS
- (ACGT)N5(ACGT)MPS
- (ACGT)N25(ACGT)MPS
- (ACGT)N30(ACGT)MPS

2.3 Retrieval of promoter sequence from databases

The 1 kb sequence upstream of all identified chromosomal genes from the following genomes were retrieved using NCBI reference sequence database

- *Arabidopsis thaliana* (The Arabidopsis Genome Initiative v. 10, 2011)
- *Oryza sativa* (Rice) (International Rice Genome Sequencing Project, Build 4.0, 2009)
- *Glycine max* (Soybean) (US DOE Joint Genome Institute (JGI-PGF), v. 1.0, 2010)
- *Sorghum bicolour* (Sorghum) (Sorghum Consortium, v. 1.0, 2009)

The gene annotation information (Gene ID/ Arabidopsis TAIR ID and ATG site) from Gene bank files were used for the extraction of 1 kb promoter sequence in FASTA format for all the four species. The total frequency of occurrence of ACGT (N) ACGT with $0 \leq N \leq 30$ was searched in all extracted 1 kb regions. The four palindromes and non-palindrome sequences that were used as control elements have been listed in Table 2-2.

Table 2-2 Palindrome and non-palindrome tetramer sequences used as control elements

Palindrome	Non-palindrome
TAGC	AGCT
CGTA	TGCA
GCTA	CTAG
ATGC	GATC

2.4 The spacer sequence analysis

The spacer sequences, extracted from upstream 1 kb region of all promoters of *Arabidopsis thaliana,* were analyzed to identify nucleotide position in spacer sequence. The percentages of A, G, C and T at each position for all the spacer lengths ($0 \leq N \leq 30$) were calculated to identify the preferences at particular positions within the spacers sequence.

2.5 Ortholog analysis

In order to understand the mechanism of evolution of the ACGT (N) ACGT *cis* regulatory elements in the afore mentioned plant species, we analyzed its predominance for $0 \leq N \leq 30$ in all identified paralogs/orthologs among *Arabidopsis thaliana*, *Oryza sativa* (rice), *Glycine max* (soyabean) and *Sorghum bicolor* (sorghum). Gene names for rice were converted from MSU annotation to RAP annotation prior to analysis. Frequencies for co-occurring ACGT elements were analyzed from the extracted genes.

2.6 Functional analysis

Using published microarray data from the EBI Gene Expression Atlas for *Arabidopsis thaliana*, we studied whether genes containing multiple core ACGT *cis* elements were up /down regulated during developmental stages (embryo, seedling), by hormones (ABA, auxin, ethylene, gibberellin, jasmonic acid, salicylic acid), in different plant parts

(cambium, flower, leaf, root, pollen, seed, sperm cell, stem, vegetative apex, apical root meristem) and by environmental conditions (at baseline growth temperature, disease, drought, low water potential, at optimum photosynthetic temperature, salt, 20% inhibition from optimum photosynthetic temperature, 30% inhibition from optimum photosynthetic temperature).

In addition, we extended our functional analysis to rice by extracting genes up regulated by salt and drought stress in 7 days old seedling from publically available microarray data. By comparing genes regulated by a condition with genes containing multiple ACGT elements, we calculated the likelihood of occurrence by the following formula:

Likelihood of occurrence X/Y for a particular spacer length N,

Where, X= (A∩B)/ B and Y=P (A);

A= event that a given gene is regulated (up/down) by a particular condition

B = event that a given gene contains multiple ACGT elements separated by N base pairs (bp).

Further, we calculated the overall likelihood of occurrence for each condition (for N= 0–30), and conditions with likelihood of occurrence > 1.30 were subjected to further statistical analysis using the 8 control sequences described earlier.

2.7 Statistical analysis

The statistical analysis was performed to determine the significance of results. The frequency of ACGT (N) ACGT was assessed for significant peaks by box and whiskers plots, with 10% and 90% whiskers. The outliers were considered as potential peaks, especially if they were present across all species. The degree of correlation between the two monocot and dicot species was calculated by taking the frequencies for N consecutive spacer lengths at a time, beginning from 0 till 30 for each of the species, where N≥6. By assuming Gaussian distribution, the Pearson's correlation coefficient was calculated for each of the cases to determine significance. Consecutive spacer lengths of N with the highest degrees of correlation were interpreted as the most conserved spacer distances. To identify conditions regulating ACGT (N) ACGT containing promoters over the 8 control elements, a Grubb's outlier's test was performed on the likelihood of occurrences for each condition, assuming the data set was normally distributed. If the ACGT (N) ACGT likelihood emerged significantly higher than the controls by this test for a certain condition, it was interpreted to be specifically regulated by that condition.

2.8 Analysis of protein phosphatase 2C (*PP2C*)-like promoter sequence

From *in-silico* study of ACGT elements, protein phosphatase 2C (*PP2C*)-like promoter was identified, located at chromosome number 5 (Accession No. 59220) having putative ABA-inducible ACGT *cis* elements has been taken for the study. This promoter has unique genetic architecture which includes two ACGT elements separated by 30 nucleotides as shown in figure 2-1. In addition to this, it has another set of ACGT motifs separated by 5 nucleotides. To determine the complete set of *cis* regulatory elements in the promoter region under study, PLANT CARE (Lescot et al., 2002) (http://bioinformatics.psb.ugent.be/webtools/plantcare/html/) and PLACE (Higo et al., 1999) (http://www.dna.affrc.go.jp/PLACE/) databases were used.

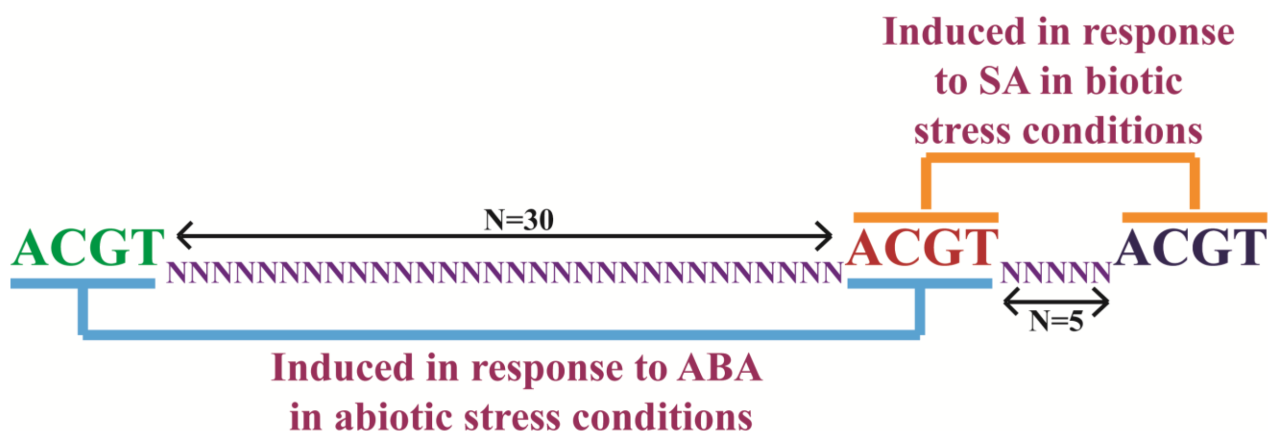


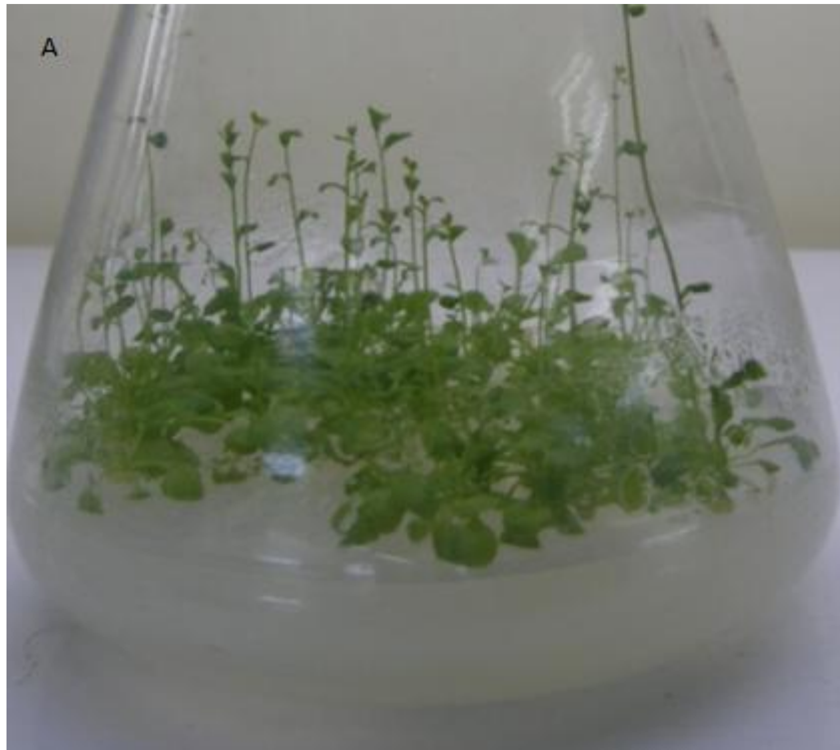
Figure 2-1 Genetic arrangement of ACGT regulatory motif in *PP2C*-like promoter. N represents- nucleotides

2.9 Materials

The PCR machine (Model Palm Cycler) was procured from Genetix. Thermo stable polymerase -*Taq* Polymerase was obtained from Bangalore Genei. The restriction enzymes were procured from the New England Biolabs Inc, USA and the T4 DNA ligase enzyme was obtained from the Thermo Fisher Scientific or Promega. Other chemicals were obtained from Himedia, SRL, Sigma, Qiagen. MilliQ grade water (Millipore Corp., Massachusetts, USA) was used in all molecular biology experiments. *Arabidopsis thaliana* (ecotype Columbia) and *Nicotiana tabacum* cv Xanthi growth conditions were standardized in the laboratory conditions. *Agrobacterium tumefaciens* strain GV3101 harbouring pMP90 a helper plasmid was obtained from NABI, Mohali. *Agrobacterium tumefaciens* strain GV3101 is resistant to rifampicin (rif) and gentamycin antibiotics. The resistant gene *rif is* present in GV3101 chromosome and gentamycin resistant gene is present in helper plasmid pMP90. *Agrobacterium tumefaciens* strain GV3101 was used for the development of transgenic *Arabidopsis thaliana* plants. Binary vector pBI101 (Jefferson et al., 1987) (procured from NABI, Mohali) was used for the construction of promoter-reporter constructs.

2.10 Standardization of growth conditions for *Arabidopsis thaliana.*

Arabidopsis thaliana (ecotype Columbia) seeds were grown in full Murashige and Skoog (MS) media. Seeds were sterilized with 0.1% of mercuric chloride ($HgCl_2$) for 30 seconds and then washed with autoclaved distilled water 3-4 times for 5 minutes. After washing, the seeds were inoculated on MS media. The inoculated flasks were incubated for 3 days at 4°C in dark for stratification. After 3 days, the flasks were transferred from refrigerator to culture room and were allowed to grow at 22±2°C under 16 hours light and 8 hours dark period (Figure 2-2 A and B).



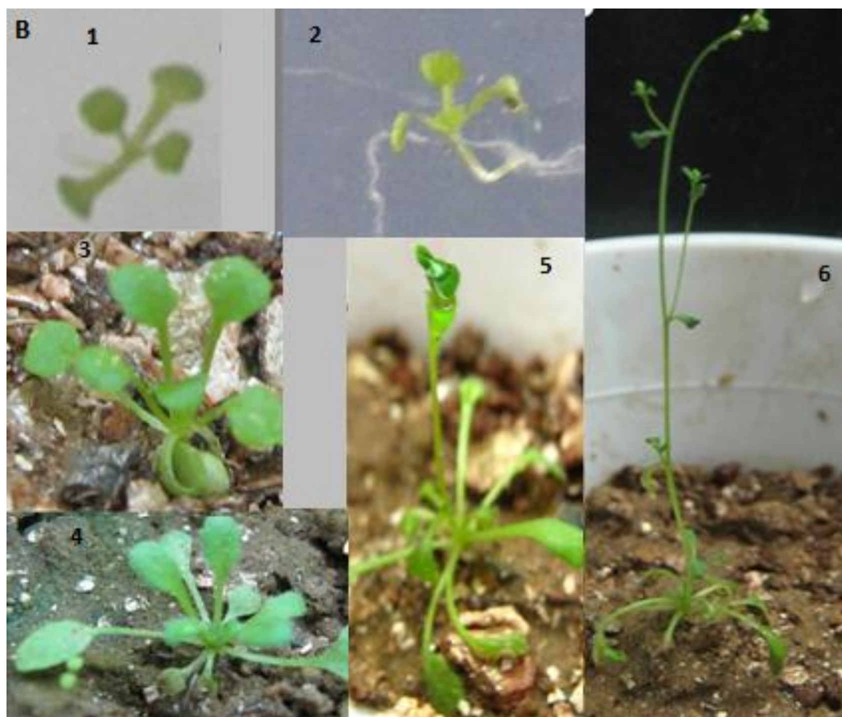


Figure 2-2(A) *In-vitro* grown *Arabidopsis thaliana* on MS media. Figure 2-2(B) Showing different stages of *Arabidopsis thaliana* in soil (1-6)

2.11 Genomic DNA isolation from *Arabidopsis thaliana* leaves using Qiagen DNA Easy plant Mini kit (Murray and Thompson 1980)

- *Arabidopsis thaliana* leaves were weighed (100 mg) and grounded to a fine powder in liquid nitrogen (liq. N_2) using a clean and pre-chilled mortar and pestle. The tissue powder was transferred to a micro-centrifuge tube (2 mL).
- A volume of 400 µL of buffer AP1 and 4µL of RNaseA (stock solution of 100 mg/mL) were added to the sample.
- The sample was vortexed vigorously to remove tissue clumps. The cells were lysed by incubating the sample at 65°C for 10-15 min. Meanwhile, the sample was mixed gently by inverting the tubes 2-3 times.
- After incubation, 130 µL of buffer AP2 was added to the lysate mixture and the mixture was incubated for 5 min on ice.
- The sample was centrifuged at 20,000g (14,000 rpm) for 5 min. The supernatant was collected and added to the QIA shredder Mini spin column. The column was centrifuged for 2 min at 20,000g (14,000 rpm).
- After centrifugation, the flow through was collected in micro-centrifuge without disturbing the pellet. 1.5 volumes of buffer AP3/E were added to flow through and mixed gently by inverting the tube.
- The mixture was added to the DNeasy Mini spin column and the column was centrifuged for 1 min at ≥6000g (≥8000 rpm). The flow through was discarded.
- DNeasy Mini spin column was transferred to a new 2 mL collection tube. The 500 µL of buffer AW was added and the column was centrifuge for 1 min at ≥6000g (≥8000 rpm). The flow through was discarded.
- Again 500 µL of buffer AW was added and the column was centrifuge for 2 min at 20,000g (14,000 rpm) to dry the membrane of the column.
- The DNeasy Mini spin column was transferred to a 1.5 mL micro-centrifuge tube and the sample was eluted using nucleases free water (50-100 µl). The column was incubated at room temperature for 5-15 min and centrifuged for 1 min at ≥6000g (≥8000 rpm) to elute the nucleic acid.
- After elution the genomic DNA was checked on 0.8% agarose gel (Figure 2-3)

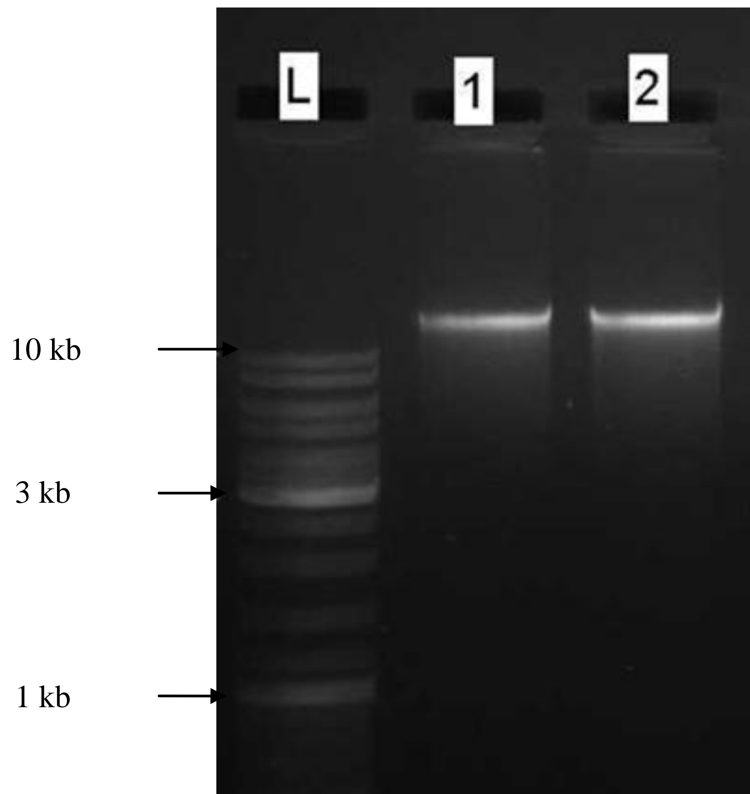


Figure 2-3 Gel image of genomic DNA isolated from *Arabidopsis thaliana* leaves, run on 0.8% agarose gel. Lane L has DNA ladder from 100 bp to 10 kb (#SM0331). Lane 1 and 2 has genomic DNA

2.12 Oligonucleotides designing and synthesis

The genomic DNA was isolated from the leaves of *Arabidopsis thaliana* and used as template (Figure 2-3) to amplify full length *PP2C*-like promoter (AT5G59220) and its various deletion promoter fragments. The deletion was introduced from 5' upstream end of the promoter and the primers were designed accordingly. Primers designed for *PP2C*-like promoter (AT5G59220) are listed in Table 2-3. The primers were synthesized commercially from EUROFINS at a scale of 10 nmole.

Table 2-3 List of primers used for amplification of *PP2C* like promoter and its deletion variants

S.No.	Sequence 5'-3'	Temperature (°C)
1	(F) FULL 5'AA**CTGCAG**AAGTATTCACGCACCAAGGT3'	58
2	(R) REVERSE' CG**GGATCC**ACAAACACACTCCATCAC3'	58
3	(S) SEQUENCING 5' TTTAGAGATCGCATGCGTGT3'	58
4	900 5' AA**CTGCAG**TGTCCTTGAACACACCAAAC3'	60
5	500 5' AA**CTGCAG**AAATGGTAAGGTAAATTTCCAC3'	58
6	400 5' AA**CTGCAG**CACTTTAGGTCTGAGTAGTGT3'	58
7	NRM 5' AA**CTGCAG**GCGATTTAGGAGAAGTACGT 3'	58

The bold sequence represents the restriction enzyme site *PstI* (CTGCAG) and *BamHI* (GGATCC). The 5 pairs of primers (F/R), (900/R), (500/R), (400/R) and (NRM/R) were used to amplify the full-length *PP2C*-like promoter and its deletion variant fragments, the PCR was carried out using Bangalore Genei *Taq* polymerase. PCR reaction was set in 50µL reaction mixture and thermo-cycler conditions are mentioned in figure 2-4.The amplified PCR products were checked on 1.5% agarose gel (Figure 2-5 A and B).

2.12.1 The composition of reaction mixture is as follows:

2mM dNTPs mix = 1 µL

Taq Polymerase = 1 µL (5U/µL)

Taq Polymerase buffer (10X) (15 mM $MgCl_2$) = 5 µL (1X) (1.5 mM $MgCl_2$)

Primer Forward = 3 µL (30 pmole)

Primer Reverse = 3 µL (30 pmole)

Template = 0.5 µL (0.5 ng)

Autoclaved Milli Q H_2O = 36.5µL

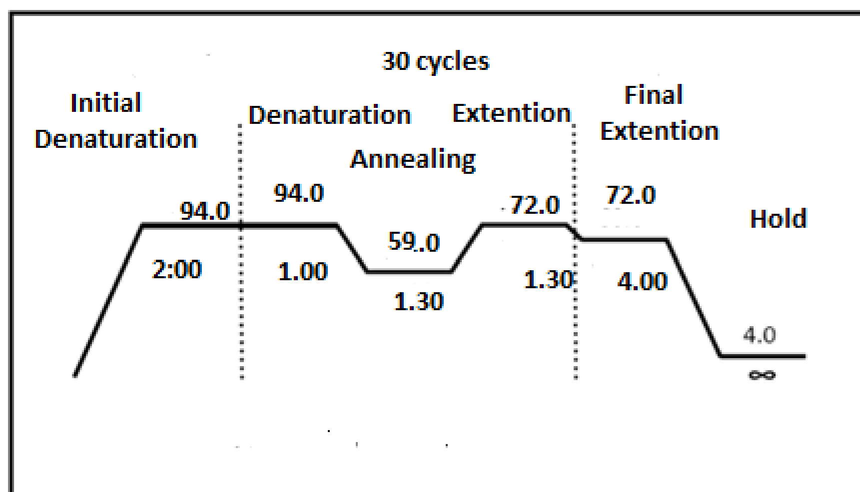


Figure 2-4 Thermo-cycler conditions which was followed to amplify full-length *PP2C*-like promoter and its deletion variants (900 bp, 500 bp, 400 bp and NRM)

2.13 Agarose gel electrophoresis

The amplified products were analyzed on agarose gel. The agarose gel electrophoresis was carried out in submarine horizontal agarose slab gel apparatus procured from Tarson and Bio-Rad (Sambrook et al., 1989).

The agarose gels were prepared and electrophoresis was done in 1X TAE buffer. Agarose powder was suspended in 1X TAE buffer, and was dissolved by heating in a microwave oven. The melted agarose gel was cooled down to approximately 50°C, and just before pouring in the casting tray, ethidium bromide (EtBr) (1µg/mL in TE buffer, pH 8.0) was added in the gel solution. After adding EtBr, the gel was poured in the electrophoresis boat. Wells were made by fixing the comb over the boat, prior to pouring of the molten agarose. After solidification the comb was removed and the gel was transferred to electrophoresis tank containing 1X TAE buffer. The DNA samples and molecular markers were mixed with tracking dye 1X (1/6 volume of DNA sample). Electrophoresis was carried out at 100v. The gel was then examined under transmitted UV light (302-312 nm) using a gel doc (Biorad). Double stranded DNA was analyzed on agarose gel. Gels of different strengths were prepared depending on the size of DNA to be analyzed. A table which illustrates appropriate gel strength for different sizes of DNA is given below:

Gel strength (w/v)	Size of DNA (bp)
4%	50-120
3%	120-120
2%	300-800
1.5%	800-1000
1.2%	1000-1200
1%	1200-2000
0.8%	2000 longer

Gel loading buffer (Tracking dye 6X)

EDTA (pH 8.0) = 1mM

Glycerol = 50% (v/v)

Bromophenol Blue = 0.4% (w/v)

Xylene Cyanol FF = 0.4% (w/v)

TAE buffer 50 X 1L pH 8.0

242 gm Tris Base

57.1 mL Glacial acetic acid

100 mL EDTA pH 8.0

TE Buffer 10 ml pH 8.0

100 mM Tris HCl (pH 8.0)

10 mM EDTA (pH 8.0)

The amplified products were cloned into pEGFP vector of 3.4 kb (Figure 2-6) to upstream region of reporter gene which is the enhanced version of GFP (eGFP) in the multiple cloning sites between *PstI* and *BamH*I restriction enzyme sites.

Figure 2-5 A Gel image of PCR amplification of *PP2C*-like promoter and its deletion variants (lane 2 to 7) and amplified DNA fragments were run on 1.5% agarose gel. Lane 1 has DNA ladder from 100 bp to 1000 bp (1 kb ladder)

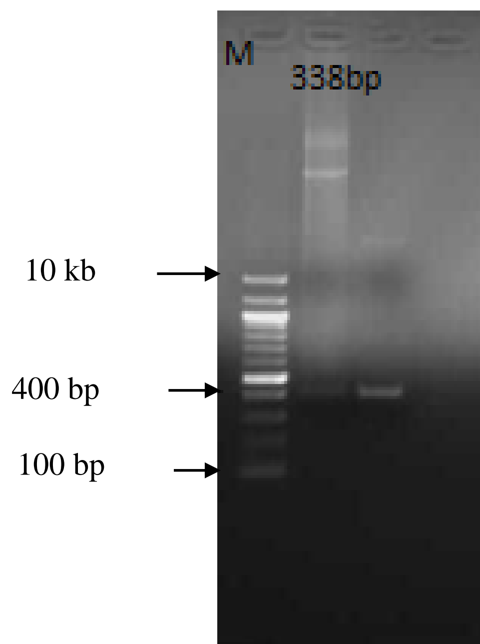


Figure 2-5 B Gel image of PCR amplification of *PP2C*- like promoter's deletion variant NRM in lane 2 & 3 (338 bp) and it was run on 1.5% agarose gel. Lane 1 has DNA ladder from 100 bp to10 kb (#SM0331)

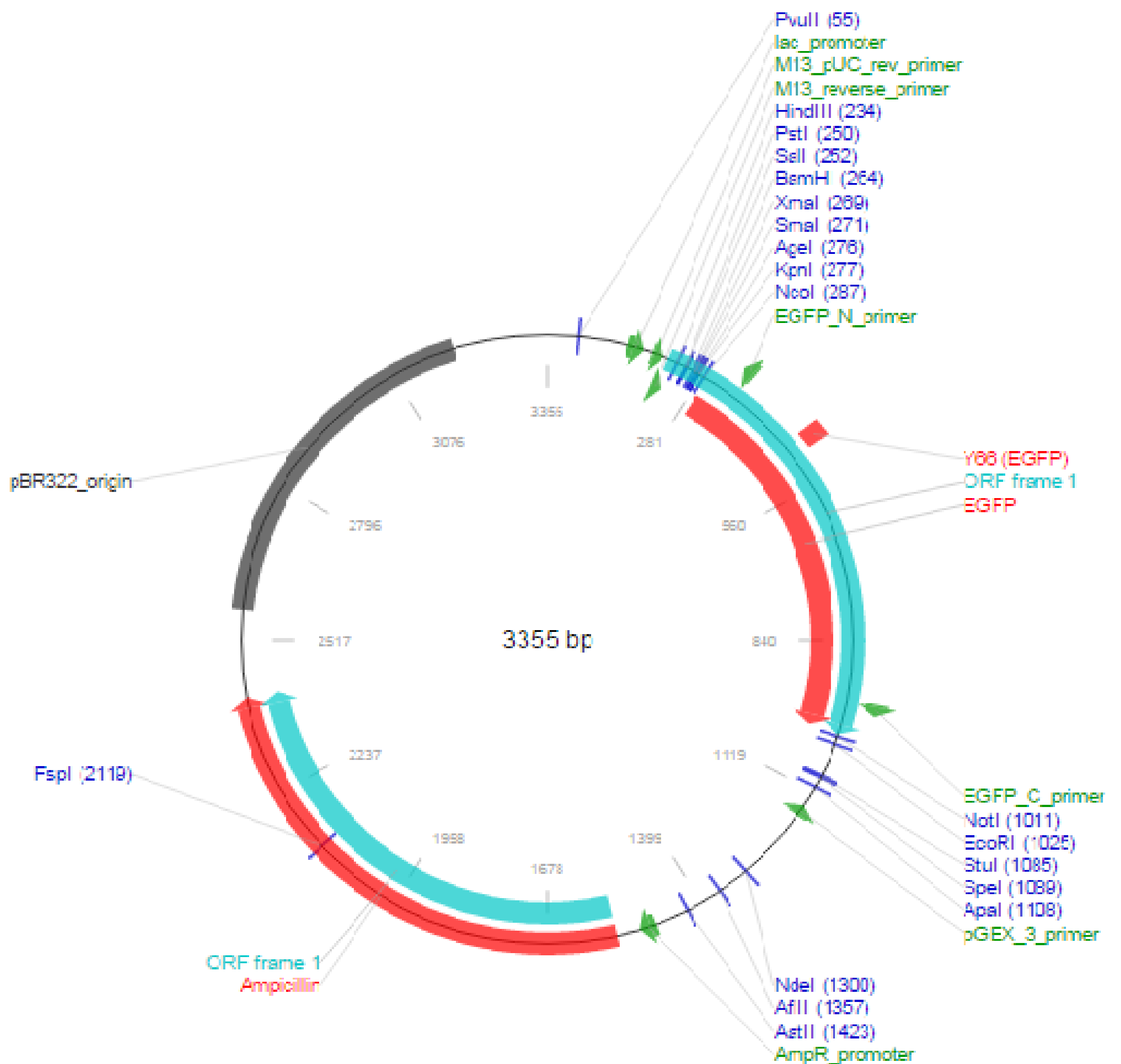


Figure 2-6 Vector map of pEGFP. The desired amplified products were cloned into pEGFP vector in restriction enzyme sites *PstI* (CTGCAG) and *BamHI* (GGATCC). This pEGFP vector was isolated from the *E.coli* by using alkaline lysis method as described by Brinboim and Dolly (1979) as described in Sambrook et al. (1989) (Figure 2-7)

2.14 Alkaline lysis method of isolation of plasmid

2.14.1Reagents used for plasmid isolation

- Solution I GET Buffer (Glucose, EDTA, Tris)
- 50mM Glucose
- 25mM TrisHCl (pH 8.0)
- 10mM EDTA (pH 8.0)
- Solution II Lysis Buffer
- NaOH 0.2N
- 10% SDS (Sodium dodecyl sulphate)
- Solution III (Sodium acetate 3M (pH 5.2)
- Tris saturated phenol
- Chloroform
- Isoamyl alcohol
- 70% Ethanol
- Isopropanol
- Nuclease free Milli-Q water
- TE buffer pH 8.0

2.14.2 Protocol for mini preparation of plasmid isolation described by Brinboim and Dolly (1979) as described in Sambrook et al. (1989)

- Single bacterial colony was inoculated in 5 mL Luria Bertani broth (LB) medium containing 100 µg/mL of ampicillin and incubated at 37°C for 12-16 hours with continuous shaking at 250 rpm in a shaker incubator (Mac). A volume of 1.5 mL of the overnight grown culture was transferred to a micro-centrifuge tube and the cells were centrifuged at 13000 rpm for 1 min at room temperature.
- The pellet was suspended in 100 µL of ice cold solution I (GET buffer - 0.05M glucose, 0.025 M Tris HCl, pH 8.0 and 0.01 M EDTA, pH 8.0) by vortexing.
- To the resuspended cells, 200 µL of freshly prepared solution II (alkaline SDS - 1% SDS and 0.2N NaOH) was added and the cells were lysed by gently inverting the tubes 2-5 times.
- The sample was incubated at room temperature for 5 min.

- After 5 min incubation, 150 µL of chilled solution III (3M sodium acetate pH 5.2) was added and samples were gently mixed by inverting the tubes 2-5 times. The sample was again incubated on ice for 15-30 min. The lysate was centrifuged at 13,000 rpm for 20 min at 4°C. The clear transparent supernatant was transferred into fresh micro-centrifuge tube. DNase free RNaseA (20 µg/mL) was added and the sample was incubated at 37°C for 1 hour.
- Plasmid Purification:-Equal volume of phenol: chloroform: isoamyl alchohol (25:24:1 v/v) was added to the supernatant obtained at previous step. The solutions were mixed by inverting the tubes to get a homogenous mixture. The sample was centrifuged at 10,000 rpm for 10 min.
- After centrifugation, the upper aqueous phase was removed carefully and this layer was transferred to fresh 1.5 mL micro-centrifuge tube. Equal volume of chloroform:isoamyl alcohol (24:1) was added and the tubes were mixed by inverting. The sample was centrifuged at 10,000 rpm for 10 min.
- The upper aqueous phase was removed carefully and this layer was transferred to fresh 1.5 mL micro-centrifuge tube.
- DNA precipitation was done after plasmid purification.
    i. The upper aqueous phase obtained after chloroform: isoamyl alcohol treatment was taken and equal volume of absolute isopropanol was added.
    ii. The samples were kept at -20°C for 15-30 min.
    iii. The samples were centrifuged at 13,000 rpm for 20 min at 4°C.
    iv. The pellet was washed with 70% ethanol by centrifuging at 10,000 rpm for 10 min at 4°C.
    v. The precipitated nucleic acid was dried and dissolved in 30 µL of TE buffer (10 mM pH 8.0).
    vi. The samples were stored at -20°C.
    vii. The plasmid DNA was checked on 0.8% agarose gel.

2.15 Restriction analysis of amplified PCR products and plasmid DNA

Endonuclease digestion of amplified PCR products and the plasmid was done by *PstI* and *BamHI* restriction enzymes with their respective buffers (Figure 2-7). A 50 µL reaction volume wasset up. The digestion mixture was incubated at 37°C for 3 hr and checked on 1% agarose gel.

2.15.1 Digestion of vector & amplified products

The composition of reaction mixture was as follows:
Plasmid = 1.25 µg
Autoclaved Milli-Q water = 24 µL
Buffer (10X) = 5 µL
Enzyme *BamHI* = 1 µL (20 U/µL)
Enzyme *PstI* =1 µL (20 U/µL)
Total = 50 µL

2.16 Elution of nucleic acid by using gel extraction kit (Qiagen)

To obtain pure vector and the amplified PCR products for cloning purpose, gel elution of the vector and amplified PCR products were done using QIAGEN Gel Extraction kit.

2.16.1 Gel Elution Protocol

- The DNA fragment was excised with a clean scalpel.
- The gel was weighed in pre-weighed eppendorf tube and 3 volumes of buffer QG was added to 1 volume of gel (100 mg = 100 µL x 3).
- The samples were incubated at 50°C (water bath) until a clear solution was obtained. One volume of isopropanol was added to the sample and mixed properly.
- The samples were kept at room temperature for 5 min.
- 750 µL of sample was applied to the QIA quick spin column and placed in a 2 mL collection tube and centrifuged at 13000 rpm for 1min.

  The flow through was discarded and the column was placed back in the same collection tube. The previous step was repeated until the sample was finished.
- 500 µL of QG buffer was added to QIA quick spin columns and centrifuged at 13000 rpm for 1min.

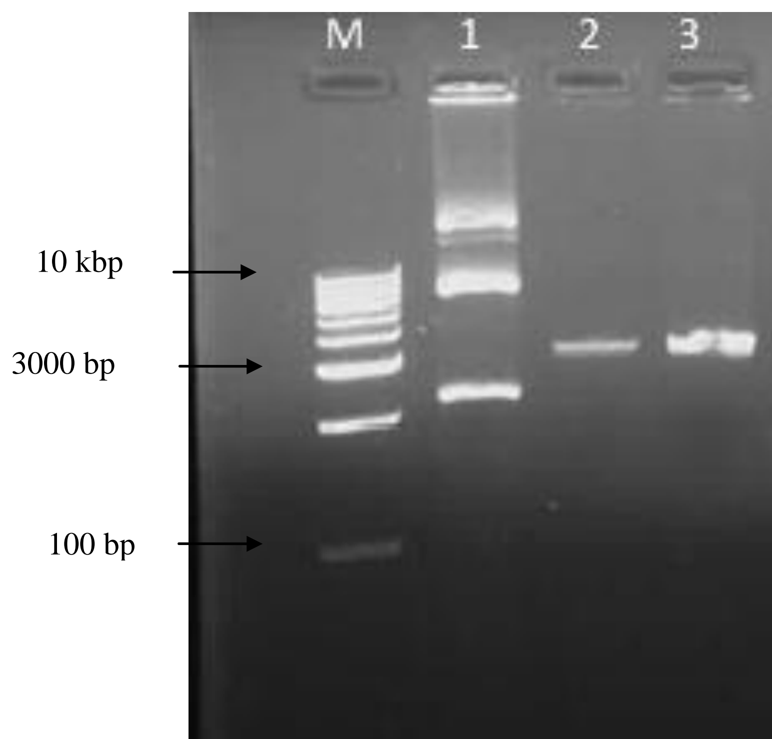


Figure 2-7 Gel image of isolated plasmid pEGFP by alkaline lysis method. Lane M has DNA ladder from 100 bp to 10,000 bp (#SM0331), Lane 1 has undigested pEGFP vector, and lane 2 and 3 have double digested pEGFP vector (3000 bp) with *PstI* and *BamHI*, respectively run on 0.8% agarose gel

- Flow through was discarded and 750 µL of PE buffer was added to the column.

- The QIA quick spin column was incubated at room temperature for 5 min and centrifuged at 13000 rpm for 1min.
- The flow through was discarded and an empty spin was given at 13000 rpm for 1min.
- The spin columns were transferred to fresh 1.5 mL micro-centrifuge tubes and 15 µL of nucleases free Milli-Q water was added to centre of the spin column. Tubes were incubated at room temperature for 5-10 min.
- After incubation, tubes were centrifuged at 13,000 rpm for 1 min. The samples were stored at -20°C.

2.17 Ligation of digested purified PCR products in the cloning vector

Double digested and purified PCR amplicons and vector (pEGFP) was quantified, and the inserts were cloned in MCS of pEGFP vector between *PstI* and *BamHI* enzyme site. 50 ng of digested vector and appropriate amount of insert in a molar ratio of 3:1 were mixed and ligated. The reaction mixture was set in 10 µL reaction volume. The ligation reaction was carried out at 16°C for 16 hours. The ligation mixture was used to transform the *E.coli* DH5α competent cells.

2.17.1 Ligation reaction

The composition of reaction mixture was as follows:
Vector = 50 µg
Insert = 3X in molar ratio
Buffer (10X) = 1 µL
Enzyme T4 DNA ligase = 1 µL 20 U/µL
Autoclaved Milli-Q water = 4 µL
Total Reaction Volume = 10 µL

2.18 Preparation of *E.coli* DH5α competent cells

2.18.1 Requirements

- 100 Mm $CaCl_2$ (50 mL) - autoclaved and pre-cooled to 4°C
- 60% glycerol solution- autoclaved – pre-cooled to 4°C
- Luria Bertani broth (LB)

5 mL LB media for primary (1°) culture - autoclaved (without antibiotic)

50 mL LB media for secondary (2°) culture (without antibiotic)

- Oakridge tubes- autoclaved and stored at 4°C before use
- Eppendorf 1.5 & 2 mL – autoclaved and stored at 4°C
- Autoclaved 1 mL and 200 µL tips
- LB agar + ampicillin (100 µg/mL)

2.18.2 Procedure

Preparation of primary (1°) culture: 5 mL of LB was inoculated with a single colony of *E.coli* DH5α culture and incubated at 37°C for 14-16 hours. Preparation of secondary culture: After 14-16 hours of 1° culture, 1% inoculum from 1° culture was taken, i.e. 250 µL for 25 mL medium, and secondary culture was inoculated and incubated at 37°C for 2-3 h, until the OD at 560 nm reached 0.3-0.4. It was centrifuged at 6000 rpm at 4°C for 8 min. The supernatant was discarded without disturbing the pellet. Ice cold 100 mM $CaCl_2$ was added to each tube. The pellet was dislodged completely and carefully. Once the resuspension was done properly, it was centrifuged at 6000 rpm, 4°C for 8 min. The supernatant was discarded very gently and the pellet was resuspended in 100 mM $CaCl_2$. 0.9 mL of 60% glycerol was added so as to make the final concentration 15%. It was mixed well by tapping. Mini cooler was used to keep the eppendorfs cool. Finally, 0.3 mL aliquots were transferred in sterile eppendorf tubes, labelled and frozen in liquid nitrogen and stored at -80°C.

2.18. 3 Transformation of *E.coli* DH5α competent cells

- The competent cells *E.coli* (DH5α) were thawed on ice. The ligation mixture was added to 100 µL of competent cells and incubated on ice for 30 min.
- A heat shock was given at 42°C (water bath) for 90 sec and the tubes were incubated on ice for 2-3 min.
- 800 µL of LB medium was added, mixed well and incubated at 37°C incubator for 45 min with shaking at 200 rpm.
- The cells were pelleted by spinning at 3000 rpm for 5 min. 750 µL of LB media was removed from the supernatant without disturbing the pellet.
- The 150 µL bacterial suspension was suspended in the remaining supernatant and was spread on LB agar plate with ampicillin. The plates were incubated overnight at

37°C. The bacterial colonies were checked for recombinant plasmid by isolation of plasmid and restriction analysis.

Colonies were observed and master plate of these colonies was prepared and plasmid isolation was done by alkaline lysis method. After plasmid isolation, positive clones were analysed by digestion and finally, recombinant plasmids were confirmed by sequencing (Figure 2-8 A and B). After confirmation of the recombinant, pEGFP vector was used for transient expression studies.

2.19 Transient expression studies using Biolistic system (Biorad Gene gun PDS-1000/He)

2.19.1 Standardization of growth conditions for *Nicotiana tabacum* cv Xanthi

Nicotiana tabacum cv Xanthi seeds were grown in full Murashige and Skoog (MS) media. Seeds were treated with 0.1% of mercuric chloride ($HgCl_2$) for 30 sec and then washed with autoclaved distilled water for 35 minutes, 6-7 times. After washing, the seeds were inoculated on MS media. The flasks were transferred to culture room and grown at 24^0C under 16 hours light and 8 hours dark conditions (Figure 2-9).

Figure 2- 9 *In vitro* grown *Nicotiana tabacum* cv Xanthi

Figure 2- 8 Electropherogram or chromatogram showing the sequence of positive clones of NRM

N5 (500base pair) with primer M13 reverse

Figure 2- 8 Electropherogram or chromatogram showing the sequence of positive clones of NRM

2.19.2 Transient expression studies using biolistic system (Biorad Gene gun PDS-1000/He)

The transient expression studies were carried out by microprojectile mediated delivery of DNA in tobacco leaf discs to study the expression of EGFP which was used as a reporter gene (*Nicotiana tabacum* cv Xanthi).

2.19.3 Preparation of recombinant plasmids using Qiagen Columns

- A single colony was isolated from a freshly streaked plate and inoculated in culture of 10 mL LB medium containing the ampicillin as an antibiotic and incubated for 12 hrs at 37°C with continuous shaking (approx. 300 rpm).
- The bacterial cells were harvested by centrifugation at 6000*g* for 15 min at 4°C.
- The bacterial pellet was resuspended in 4 mL of buffer P1. The bacterial suspension was resuspended completely by vortexing or pipetting until no cell clumps remained.
- 4 mL of buffer P2 was added and mixed thoroughly by inverting the sealed tube 4–6 times, and incubated at room temperature (15–25°C) for 5 min.
- 4 mL of chilled buffer P3 was added and mixed immediately and thoroughly by inverting 4–6 times, and incubated on ice for 15 min.
- The sample was centrifuged at ≥20,000*g* for 30 min at 4°C and the supernatant was removed containing plasmid DNA promptly. This step was repeated. The supernatant collected at this step was applied to column after column equilibration.
- A QIAGEN-tip 100 was equilibrated by buffer QBT and the column was allowed to empty by gravity flow.
- The supernatant obtained from previous step was applied to the QIAGEN-tip and allowed to enter the resin by gravity flow.
- The QIAGEN-tip was washed with 2 x 10 mL of buffer QC and allowed buffer QC to move through the QIAGEN-tip by gravity flow.
- Plasmid DNA was eluted with 5 mL of buffer QF.
- DNA was precipitated by adding 3.5 mL of isopropanol to the eluted DNA, mixed and centrifuged immediately at ≥15,000*g* for 30 min at 4°C and the supernatant was carefully decanted.
- The DNA pellet was washed with 2 mL of 70% ethanol and centrifuged at ≥15,000*g* for 10 min.

- The supernatant was decanted carefully without disturbing the pellet. The DNA pellet was air-dried for 5–10 min, and then suspended in 50 µL of nuclease free water. The DNA concentration was determined using UV spectrophotometer by taking absorbance reading at 260 nm.

2.19.4 Sterilization of Plant material (leaves)

Leaves were first washed with autoclaved RO water. These leaves were further sterilized with 0.1% $HgCl_2$ for 30 seconds followed by 3-4 times washing with RO water. After sterilization, midrib was removed and leaves were cut into small sections and incubated on MS agar plates at 25°C in 16 hours light and 8 hours dark light cycle for 12 hours before bombardment (Figure 2-10).

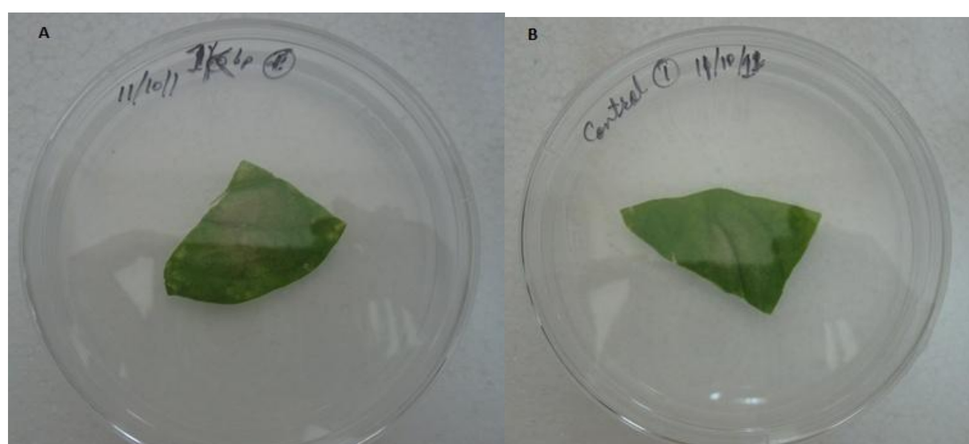


Figure 2-10 Preparation of plant material for bombardament.

2.19.5 Materials required for Biolistic transformation

A. Macrocarriers
B. Rupture disks (1100 psi)
C. Stopping screens
D. Gold micro carriers (1µm)

2.19.6 Preparation of gold micro carriers for bombardment (Sawant et al., 2000)

- 1.5 mg of gold particles was weighed in a 1.5 mL micro-centrifuge tube. Gold particles were baked at 192°C overnight in hot air oven and washed with 1 mL of 70% isopropanol (v/v).
- It was vortexed vigorously for 3–5 minutes.

- The gold particle suspension was allowed to stand for 5 minutes to allow the gold particles to settle down.
- After five minutes, isopropanol was removed without disturbing the settled gold particle.
- Gold particles were washed with autoclaved Milli-Q water, and particles were again allowed to settle down and then vortexed for 3 minutes. This step was repeated 3 times.
- After the third wash, 50% glycerol was added to bring the microparticle concentration to 60 mg/mL.

2.19.7 Coating of Washed Gold Microcarriers with DNA

- The microcarriers were prepared in 50% glycerol and were vortexed to resuspend and disrupt the agglomerated particles.
- Microcarriers were transferred to different microcentrifuge tubes. Continuous agitation of the microcarriers was needed for uniform DNA precipitation onto microcarriers. While vortexing vigorously, following were added in order:
  - 5 µL DNA (1 µg/µL)
  - 50 µL of 2.5 M $CaCl_2$
  - 20 µL of 0.1 M spermidine
- Continuous vortexing was done for 2–3 minutes.
- The microcarriers were allowed to settle for 1 minute.
- Pelleted microcarriers were centrifuged for 2 seconds. The liquid was removed.
- The suspension was washed with 70% ethanol. The liquid was removed and discarded. 30 µL of 100% ethanol was added. The solution was continuously vortexed at low speed for 2–3 seconds.
- The DNA coated gold particles were spread on macrocarriers.

2.19.8 Bombardment of DNA onto tobacco leaves

Macrocarriers carrying DNA coated gold microcarriers particles (10 µL) were fixed on macrocarrier holders and allowed to air dry for 3-5 min. The microprojectile apparatus was assembled. The tobacco leaf plate was placed inside the chamber of the Gene gun machine and vacuum was created in the chamber. Bombardment was carried out at 1100 lbs/cm^3 pressure using 1100 psi rupture disk (Bio-Rad).

2.19.9 Induction studies by using abscisic acid (ABA), jasmonic acid (JA) and salicylic acid (SA) elicitors

A. Bombarded leaves were placed in Hoagland media without any hormones and were incubated at 24°C in 16 hours light and 8 hours dark for 48 hours.
B. Bombarded leaves were kept floating on Hoagland media and were subjected to different hormonal treatments like Hoagland media containing 100 µM abscisic acid (ABA), 100 µM salicylic acid (SA) and 50µM jasmonic acid (JA) and incubated at 24°C in 16 hours light and 8 hours dark for 48 hours.

2.19.10 EGFP assay of bombarded leaves

- After 48 hours incubation the bombarded leaves were crushed in liquid nitrogen and suspended in pre-chilled 1 mL of extraction buffer (1M Tris-HCl, 0.5M EDTA, 0.5M EGTA, 10% Glycerol, 14.3 M β-mercaptoethanol, 0.5 M sodium fluoride, 1M sodium orthovandate). The whole procedure was performed at 4°C.
- Cellular debris was pelleted down by centrifugation at 13000 rpm for 10 min at 4°C. The supernatant was transferred to fresh tubes. 200 µL of the sample was added to 96 well titre plate.
- The excitation and emission wavelengths were 484 nm and 509 nm, respectively, which is the excitation and emission peak of EGFP, respectively. The EGFP concentration was calculated using the reported excitation coefficients 55000/cm/M (Reichel et al., 1996; Wu et al., 2011). Total soluble protein in leaf extract was quantified by using the Bradford assay (Bradford 1976).

2.20 Transgenic expression studies

2.20.1 Oligonucleotide designing and synthesis

The genomic DNA was isolated from the *Arabidopsis thaliana* leaves to amplify full length *PP2C*-like promoter (AT5G59220) and its deletion promoter fragments. The deletion was introduced from 5' upstream end of the promoter and primers were designed accordingly. The primers were synthesized commercially from EUROFINS at a scale of 10 nmole. Primers designed for *PP2C*-like promoter (AT5G59220) are listed in Table 2-4.

Table 2-4 List of primers used for amplification of *PP2C*-like promoter and its deletion constructs

S.No.	Sequence 5'-3'	Temperature °C
1	(F) Forward 5' **TGCTCTAGA**AAGTATTCACGCACCAAGGT 3'	58
3	900 bp 5' **TGCTCTAGA**TGTCCTTGAACACACCAAAC 3'	60
4	500 bp 5' **TGC TCTAGA**AAATGGTAAGGTA AATTTCCAC 3'	58
5	400 bp 5' **TGCTCTAGA**CACTTTAGGTCTGAGTAGTGT 3'	58
6	NRM 5' **TGCTCTAGA**GCGATTTAGGAGAAGTACGT 3'	58
7	(R)Reverse 5' **CGCGGATCC**TTTATATTAGCTTCTTTCACCAG3'	60

The bold sequence represents the restriction enzyme site for *XbaI* (TCTAGA) and *BamHI* (GGATCC) (Table 2-5). The five primer pairs (F/R), (900/R), (500/R), (400/R) and (NRM/R) were used to amplify th*e PP2C*-like promoter and its deletion variants (Figure 2-11). The PCR reaction mixture and thermo cycler conditions were same as mentioned in 2.11.1 and figure 2-4 respectively. The amplified PCR products were checked on 1% agarose gel (Figure 2-11).

The amplified fragments were then cloned using pGEMT easy vector. The promoter fragments were isolated, purified and cloned into pGEM easy plasmid vector (Promega Corp., Madison, WI, USA). Five different promoter fragments of full length *PP2C*-like promoter and the deletion promoter fragments cloned into pGEMT easy vector were released by *Xab*I and *BamH*I digestion and then cloned in pBI101 (promoter less vector as shown in figure 2-13) between the same restriction sites. Five vectors containing various promoter constructs of *PP2C*-like promoter were obtained and verified by restriction digestion.

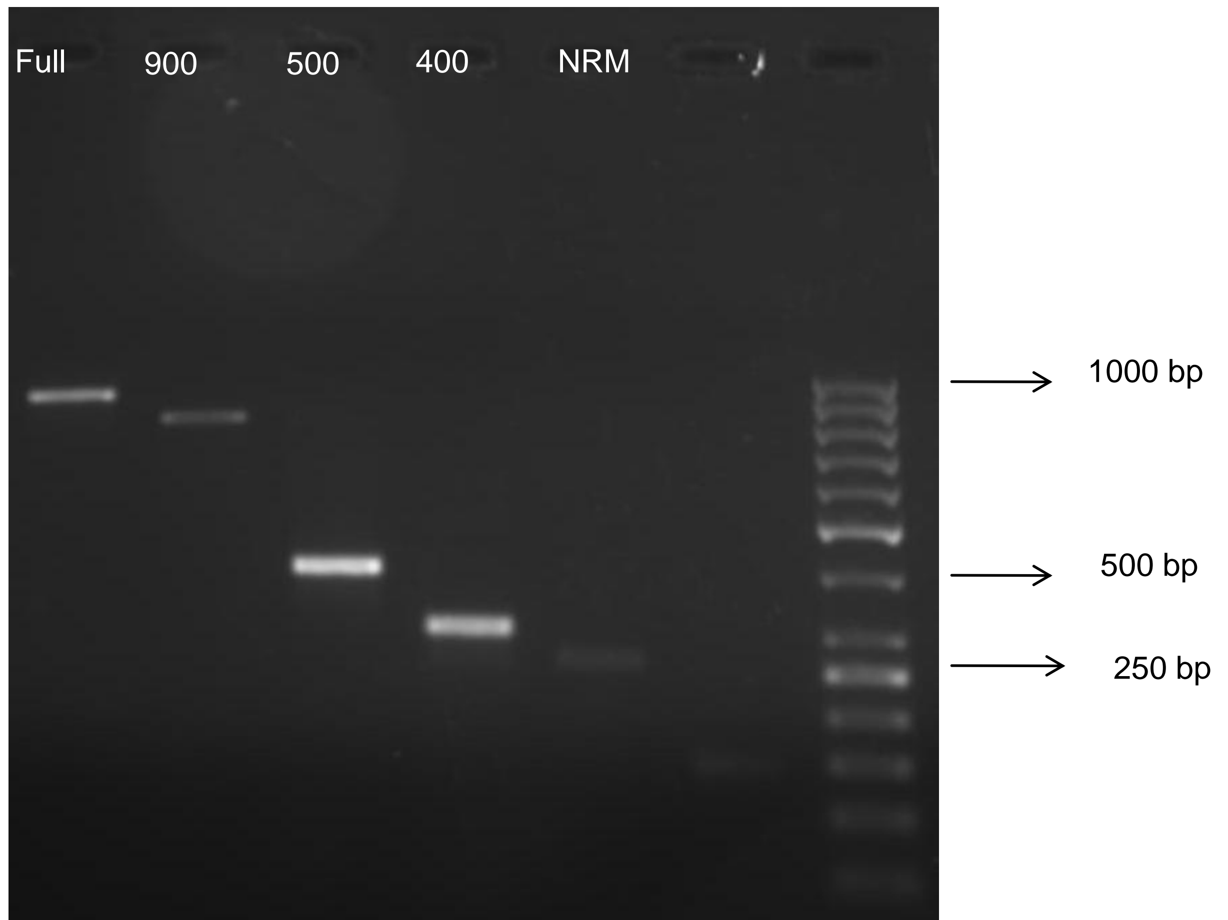


Figure 2-11 Gel image of PCR amplification of *PP2C*-like promoter and its deletion variants (lane 1 to 5) and amplified DNA fragments run on 2 % agarose gel. Lane 7 has DNA ladder from 50 bp to 1000 bp. Full – Full-length, 900 – 900 bp, 500- 500 bp, 400 – 400 bp and NRM are constructs of *PP2C*-like promoter

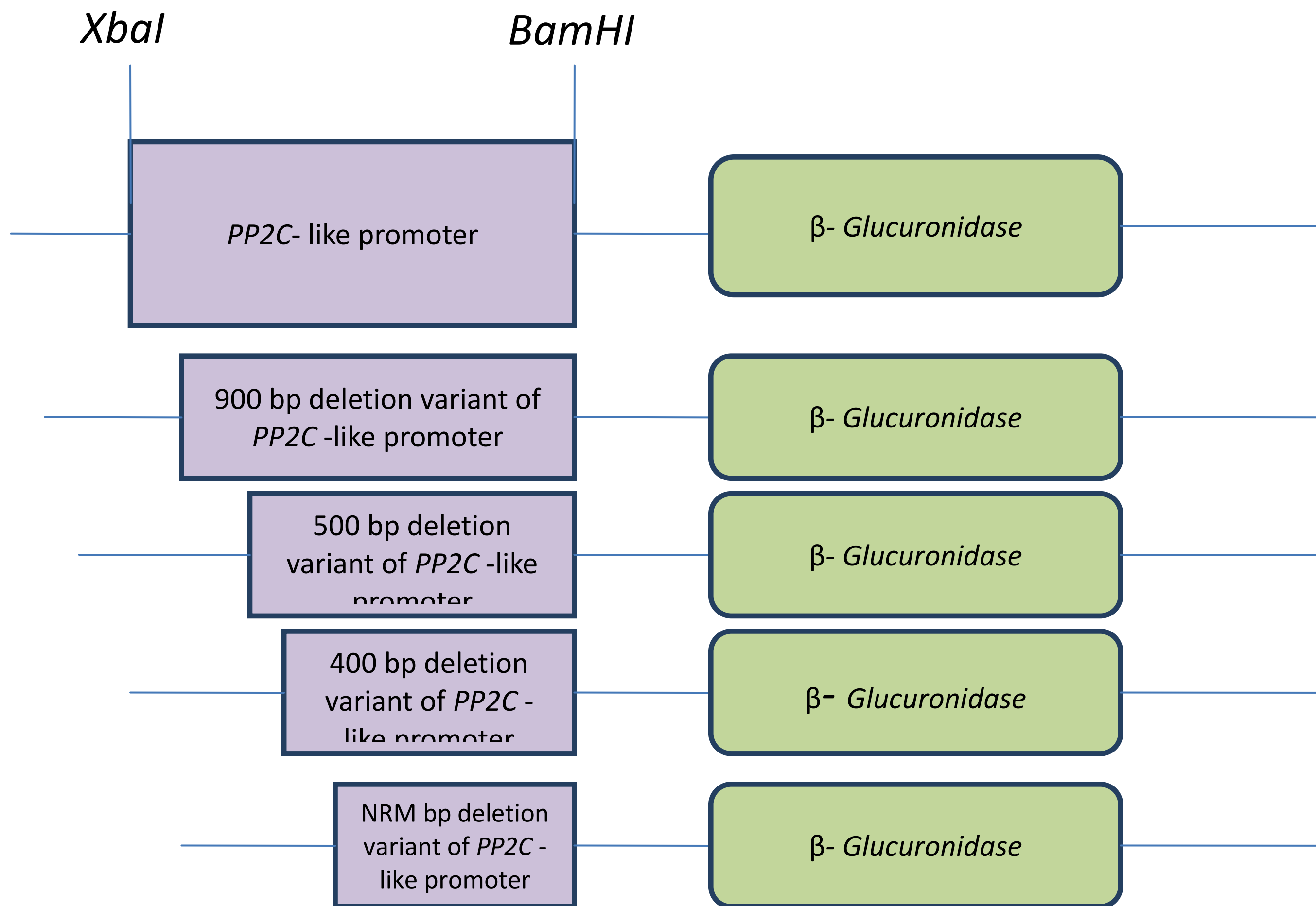
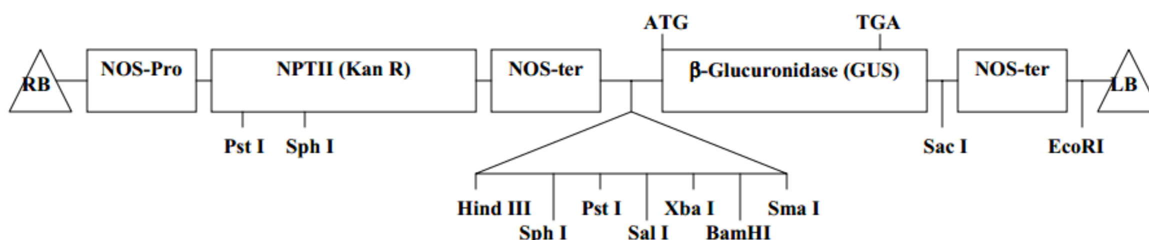


Figure 2-12 Promoter- reporter constructs were made using the amplified products of deletion variants of *PP2C* promoter by cloning in pBI101

pBI101 (12.2 kb)



AAG CTT GCA TGC CTG CAG GTC GAC TCT AGA GGA TCC CCG GGT GGT CAG TCC CTT ATG TTA
Hind III Sph I Pst I Sal I Xba I Bam HI Sma I

Figure 2-13 pBI101 vector was used for expression in plant system and promoter fragments were cloned between the *XabI* and *BamHI* endonucleases restriction sites

These amplified products were cloned into pBI101 vector (promoter less) of 12 kb, upstream to the reporter gene GUS in the multiple cloning sites between *XbaI* and *BamHI* restriction enzyme sites (Figure 2-12 and 2-13). The pBI101 vector used for cloning was isolated using alkaline lysis method as mentioned in section 2.14.2.

2.20.2 Restriction analysis of amplified PCR products and plasmid DNA

Restriction digestion of amplified PCR products and the plasmid was done by *Xba*I and *BamHI* restriction enzymes with their respective buffers (Figure 2-14 A and B). A 50 µL reaction was set up. The digestion mixture was incubated at 37°C for 3 hours.

2.20.2.1 Digestion of vector & amplified products

The composition of reaction mixture was as follows:

Plasmid = 1.25 µg

Autoclaved Milli-Q water = 24 µL

Buffer (10X) = 5 µL

Enzyme *BamHI*= 1 µL (20U/µL)

Enzyme *XbaI* =1 µL (20U/µL)

Total Reaction Volume = 50 µL

2.20.3 Elution of nucleic acid by using gel extraction kit (Qiagen) To obtain pure vector as well as amplified PCR products for cloning, gel elution of the vector and

amplified PCR products were done using QIAGEN gel extraction kit (as mentioned in 2.16.2).

2.20.4 Ligation of digested purified PCR products in the cloning vector

Double digested and purified amplicons and vector (pBI101) were quantified, and the insert was cloned in MCS of vector between *Xba*I and *BamH*I enzyme site. 50 ng of digested vector and appropriate amount of insert in a molar ratio of 3:1 were mixed and ligated. The reaction mixture volume was 10 µL. The ligation reaction was carried out at 16°C for 16 hour.

2.20.4.1 Ligation reaction

The composition of reaction mixture was as follows:

Vector = 50µg

Insert = 3x in molar ratio

Buffer (10X) = 1 µL

Enzyme T4 DNA ligase = 1 µL (20 U/µL)

Autoclaved Milli-Q water = 4 µL

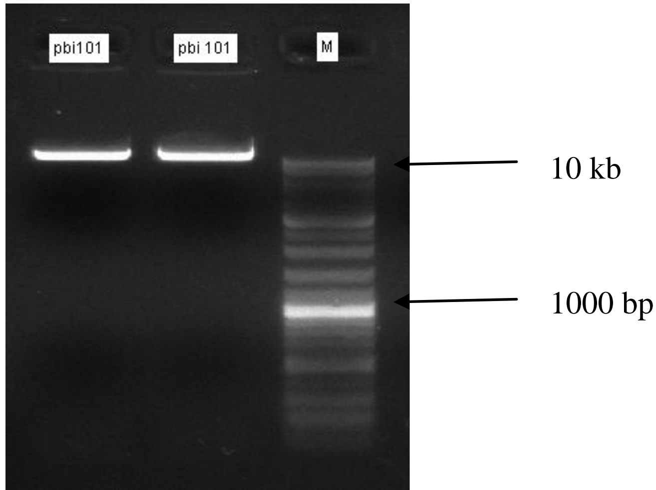
Total Reaction Volume =10 µL

2.20.5 Transformation of *Agrobacterium* GV3101

The ligation mixture was used to transform the *Agrobacterium* GV3101competent cells.

2.20.5.1 Preparation of *Agrobacterium* (GV3101) competent cells

One single colony of *Agrobacterium* (stored at -80°C) was streaked on LB agar plate containing gentamycin, rifampicin and kanamycin antibiotics and grown for 48 hrs at 28°C. A single isolated colony was picked from the plate and inoculated in 5 mL LB medium containing antibiotics and incubated at 28°C for 36 hrs at 200 rpm. 100 µL of 1° culture was inoculated into 200 mL fresh LB containing gentamycin and rifampicin antibiotics, and incubated at 28°C for 36 hrs at 200 rpm. The cultures were kept on ice and centrifuged at 4,000 rpm for 5 minutes at 4°C. Supernatant was discarded and the pellet was resuspended in 40 mL of 20 mM $CaCl_2$. It was again centrifuged at 4,000 rpm for 5 minute at 4°C. Finally, the pellet was resuspended in mixture of 20 mM $CaCl_2$ (4.5 mL) and 10% glycerol (0.5 mL). Volume of aliquotsof competent cells were stored in sterile micro-centrifuge tubes and stored at -80°C.



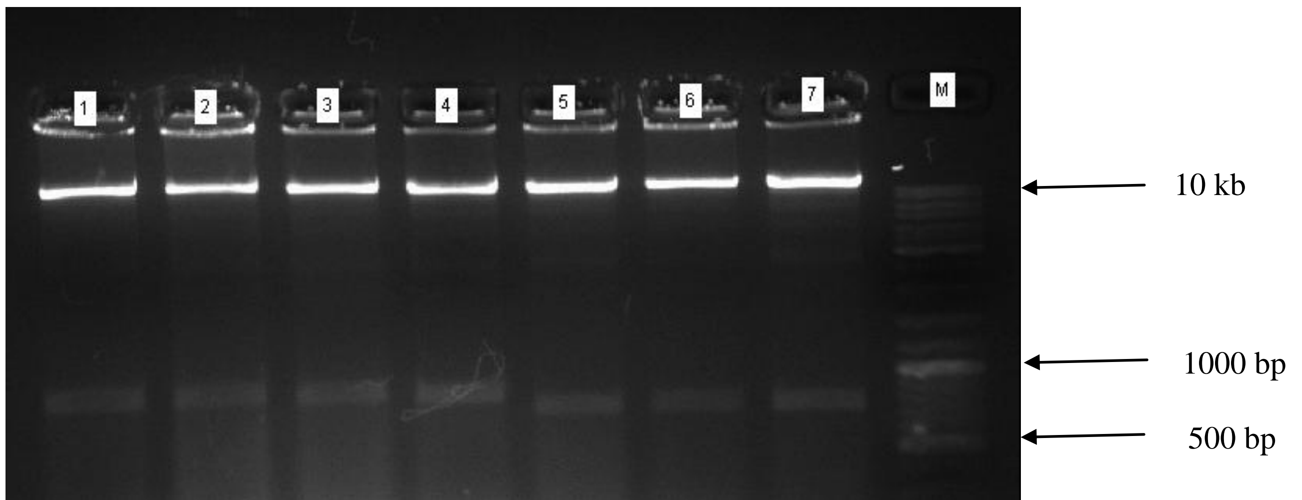


Figure 2-14 Gel image for restriction digestion analysis of clones with *BamHI* and *XabI* run on 0.8% agarose gel. Lane M has DNA ladder from 100 bp to 10,000 bp. Lane 1 to 4 has full length fragment (880 bp) and lane 5-7 900 bp construct.

2.20.5.2 Transformation of *Agrobacterium tumefaciens* GV3101 competent cells

The competent *Agrobacterium tumefaciens* strain (GV3101) was taken out from -80°C and thawed on ice. 10 µL of construct plasmid product was added to 100 µL of competent *Agrobacterium tumefaciens* cells mixed gently (5 times) by tapping the tube and incubated on ice for 45 minutes. It was frozen in liquid nitrogen for 3 minutes and then immediately transferred to 37°C for 15 minutes and then incubated on ice for 10 minutes. 600 µL of fresh Luria broth was added, gently mixed and incubated for 3 hour at 28°C, 220 rpm. After incubation, the tubes were centrifuged at 7000 rpm for 2 min and 500 µL supernatant was discarded and the remaining solution was mixed gently by pipetting and spread on the LB agar media containing gentamycin (50 µg/mL), rifampicin (25 µg/mL) and kanamycin (50 µg/mL). It was incubated for 48 hours at 28°C.

The colonies were picked up after transformation, plasmid isolation was done and clone was confirmed by restriction digestion (*XbaI* and *BamHI*), sequencing and PCR with forward and reverse primer set.

2.20.6 Transformation of *Arabidopsis thaliana* plants by floral dip method (Clough and Bent 1998; Zhang et al., 2006)

The plants were grown under lab conditions and primary bolts were clipped to allow emergence of multiple secondary bolts for greater number of floral buds per plant. As the secondary bolts started to flower, plants were transformed with recombinant plasmid constructs by floral dip method as described by Clough and Bent (1998). For this a single colony of *Agrobacterium tumefaciens* GV3101 strain culture carrying gene of interest was picked and inoculated in 5 mL LB medium supplemented with antibiotics rifampicin (50 µg/mL), kanamycin (50 µg/mL) and gentamycin (30 µg/mL). The primary culture was grown at 28°C, 220 rpm for nearly 24 hours. From this 1mL culture was inoculated in 500 mL of LB medium containing the same set of antibiotics and grown for nearly 24 hours at 28°C. The cells were harvested by centrifugation at 4000 rpm for 5 min at room temperature. The pellet was smoothly re-suspended in 5% sucrose solution to a final OD_{600} of 0.8 to 1. Prior to dipping the flowers, the detergent Silwet L-77 (Lehle seeds, USA) was added in *Agrobacterium* cell suspension prepared in 5% sucrose to a final concentration of 0.05% and mixed gently by swirling the container. This solution was termed as infiltration medium. The floral inflorescences were dipped in the infiltration medium for 10 to 15 sec with gentle swirling. A film of

liquid coating the plant was seen and excess liquid was dipped off. Dipped plants were placed in dark for 16 to 24 hours to maintain high humidity and then back in the rack and subjected to normal light/dark period. For a higher rate of transformation the same plants with newer flowers were again dipped in a fresh suspension of *Agrobacterium* with the same construct after one week. Plants were allowed to grow and moisture of soil was maintained with nutrient medium till all the siliques began to dry. At this satge, watering of plants was stopped and seeds were harvested, separated from other dried tissues and stored at room temperature in dry place.

2.20.7 Selection of putative transformants using an antibiotic marker

The transformants were selected using antibiotics kanamycin as selection marker on OS medium plate. Seeds were surface sterilized by treating them with 0.1% $HgCl_2$ for 30 secs and then seeds were thoroughly rinsed with autoclaved Mill-Q water 5-6 times. The seeds were suspended in distilled water and dispersed on OS medium containing kanamycin (50 µg/mL) (Figure 2-15 A and B). Extra water was removed with pipettes, plates were sealed using parafilm and kept at 4°C for stratification for 3 days before being placed in a growth room for germination and then grown for 6-8 days at 22°C under 24h light. Transformants were identified as kanamycin resistant seedlings that produced green leaves and well established roots in the selection medium. The independent transformants were screened on OS media containing (50 µg/mL) kanamycin. The transformed seedlings of green and expanded leaves were screened out from the yellow-coloured non-transformed seedlings and transferred to soil after 2–3 weeks (Figure 2-16 A and B).

Plastic pots were filled with autoclaved and nutrient saturated soilrite. Each viable seedling was transferred into pots filled with soilrite saturated with nutrient medium. The seedlings were transferred to soil and grown to maturity in growth chamber. Temperature in the growth chamber was 23±2°C. Light provided by cool-white fluorescent bulbs was 50-70lEm–s–l (constant) for seedlings in agar plates and~100lEm–2s–l (16 hours light/8 hours dark) for potted plants. To ascertain that plants were true transformants PCR was carried out with GUS/kanamycin gene specific primers using genomic DNA isolated from leaves of transgenic lines as template. The T2 plants were used for the RNA isolation, GUS assay and histochemical staining.

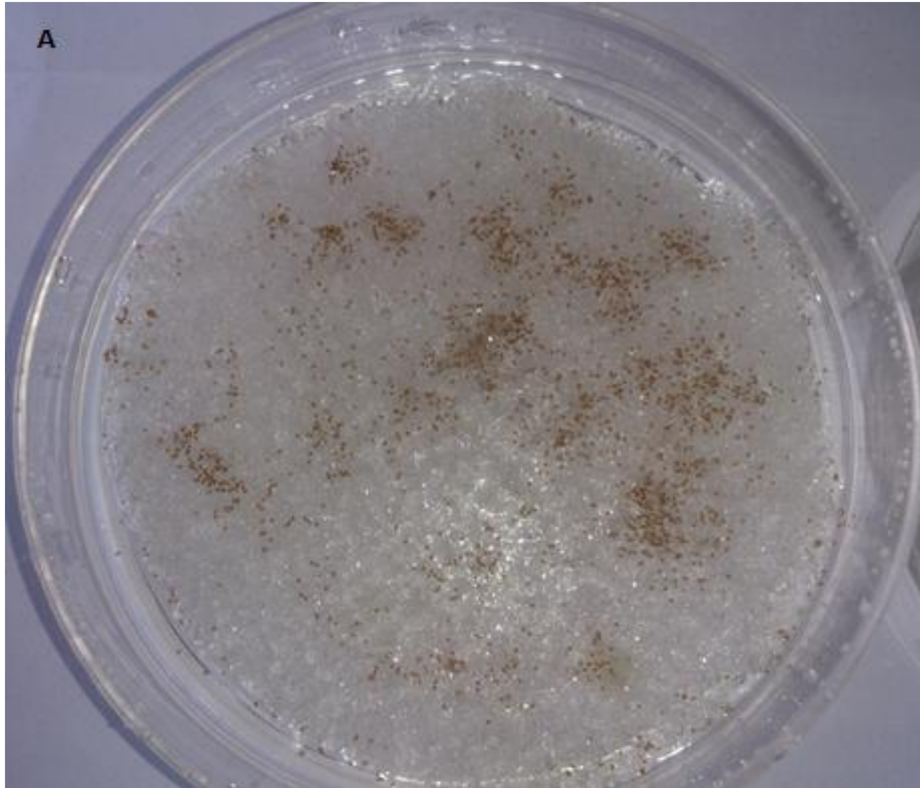


Figure 2- 15 A *Arabidopsis thaliana* seeds dispersed on OS medium.

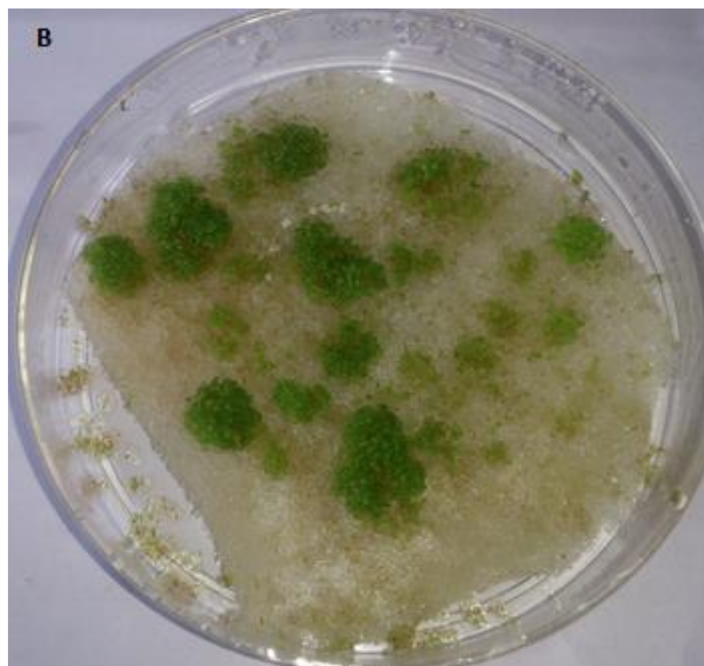
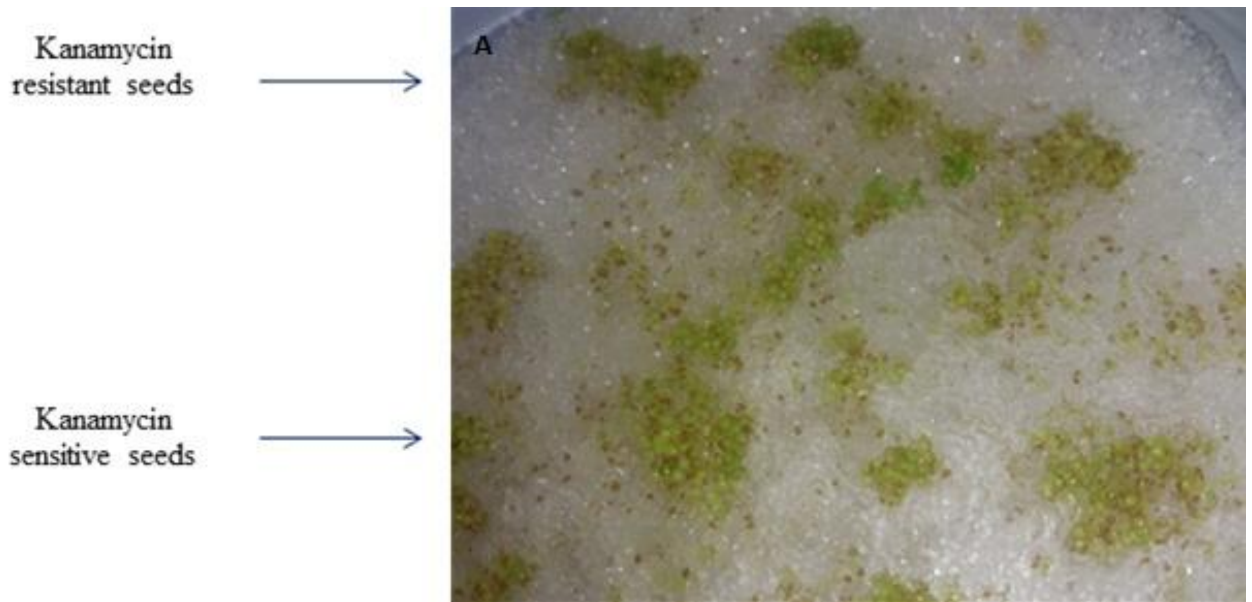


Figure 2-15 B Germination of *Arabidopsis thaliana* seeds on OS medium.



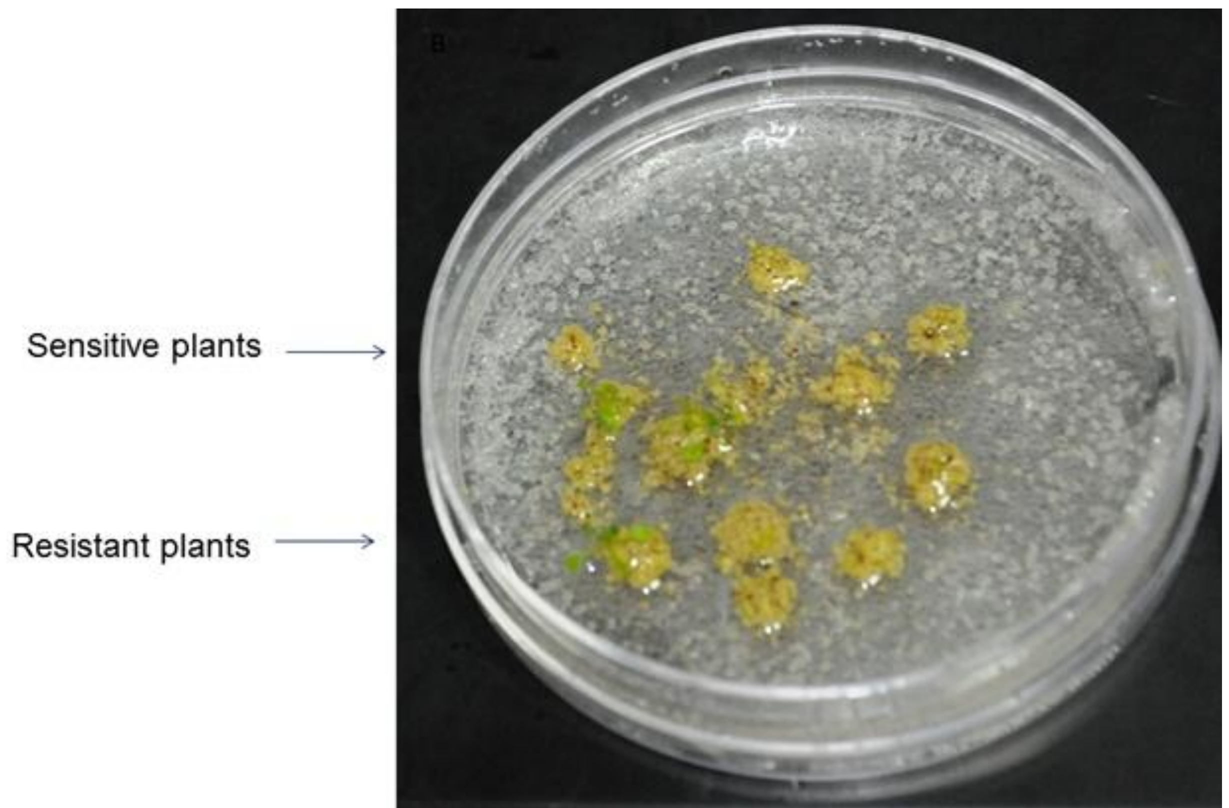


Figure 2-16 A and B Selection of transformed and non-transformed transgenic *Arabidopsis thaliana* seedlings in the presence of kanamycin.

2.20.8 GUS Histochemical Staining (Jefferson et al., 1987)

GUS histochemical staining of transgenic *Arabidopsis* plants containing *PP2C*-like promoter -GUS fusion constructs was carried out by following the method described by Jefferson et al., (1987). The GUS reaction mixture consisted of the solution containing sodium phosphate buffer (pH 7.0), Triton-X-100, potassium ferricyanide (K_3FeCN_6), potassium ferrocyanide (K_4FeCN_6), and 5-bromo-4-chloro-3-indolyl-β-D-glucuronide (X-gluc). GUS histochemical staining on transgenic plants was conducted for plants under ABA, MeJA and SA treatments, respectively. The image of blue-coloured plants was recorded with a Canon scanner.

The GUS-positive plant tissues were examined with a bright field microscope (Leica Q500MC, Cambridge, England). The reaction was stopped by adding 70% (v/v) ethanol and the pigments and chlorophylls were removed by ethanol treatment.

2.20.9 Protein extraction and Fluorimetric GUS-assay (Jefferson et al., 1987)

- GUS activity in solution was measured with the fluorometric substrate 4-methylumbelliferyl-β-D-glucuronide (MUG). Enzymatic action of GUS on MUG which is non-fluorescent substance gives fluorescent compound 4-methylumbelliferone (4MU) and glucuronic acid. MU production was measured fluorometrically at 365 nm excitation and 455 nm emission wavelengths.
- *Arabidopsis* plant leaf sample, approximately 10 to 15 mg were crushed in liquid nitrogen, suspended in 150 μL GUS extraction buffer (50 mM Na_2HPO_4 pH 7.0, 10 mM di-thiothreitol (DTT), 1 mM EDTA, 0.1% Sodium Lauryl Sarcosine (SLS), 0.1% Triton X 100) and transferred to 1.5 mL eppendorf tube and keep at 4°C.
- Cellular debris was pelleted by centrifuging the sample at 13000 rpm for 20 min at 4°C and supernatant were transferred into fresh tubes. 9 μL of extract was mixed with 1 μL of 10 mM GUS assay buffer (4-Methylumbelliferyl-β-D-glucuronide (MUG) suspended in GUS extraction buffer) in fresh tubes and mixed by pipetting. The tubes were covered to protect from light and incubated at 37°C for 60 min. After the incubation period reaction was stopped by adding 90 μL of 0.2 M sodium carbonate solution.

- The tubes were mixed properly by vortexing. Relative fluorescence of MU was recorded using spectrofluorometer (Microplate spectrofluorometer, SPECTRAmax GEMINI XS, Biocompare) at an excitation of 365 nm and emission of 455 nm. Compared the fluorescence with MUG standard curve and the amount of GUS protein content was calculated. Total soluble protein was quantified by Bradford method (Bradford 1976). A volume of 5 µL of extract was added in micro titer plate wells and then 250 µL of Bradford reagent was added in each well and kept for 5 min at room temperature. Absorbance was taken at 595 nm on spectrofluorometer reader. Compared the absorbance with BSA standard curve and amount of protein content was calculated. Each assay was repeated three times.

2.20.9 RNA isolation by using RNeasy plant mini kit

- The plant tissue was crushed in the liquid nitrogen, and it was ground thoroughly with a pre-chilled mortar and pestle. The tissue powder was transferred into RNase-free,pre-cooled 2 mL microcentrifuge tube. β-mercaptoethanol was added to buffer RLT before use.
- 450 µl of buffer RLT was added to tissue powder and sample tubes were vortexed vigorously.
- The lysate was transferred to a QIA shredder spin column and it was centrifuged for 2 min at full speed. The supernatant was transferred carefully to a new micro centrifuge tube without disturbing the cell-debris pellet in the collection tube. 0.5 volume of 100% ethanol was added to the cleared lysate and was mixed immediately by inverting the tubes 2-3 times.
- The sample was transferred (650 µl) to RNeasy spin column and column was centrifuged for 30 sec at 10,000 rpm. The flow-through was discarded.
- 700 µl of buffer RW1 was added to the RNeasy spin column and column was centrifuged for 30 sec at 10,000 rpm. The flow-through was discarded.
- 500 µl of buffer RPE was added to the RNeasy spin column and column was centrifuged for 30 sec at 10,000 rpm to wash the spin column membrane. The flow-through was discarded.

- Again 500 µl of buffer RPE was added to the RNeasy spin column and column was centrifuged for 2 min at 10,000 rpm to wash the spin column membrane. The flow through was discarded and a empty spin was given to RNeasy spin column at full speed for 1 min.
- The RNeasy spin column was placed in a new 1.5 mL micro centrifuge tube. RNA was eluted by adding 30 µl RNase-free water directly to the spin column membrane and centrifuged for 1 min at 10,000 rpm. The eluted product was stored at -80 °C (Figure 2-17).

2.20.11 DNase I Reaction Protocol (NEB M0303)

RNA sample was resuspended in 1X DNase I Reaction Buffer to a final volume of 100 µl. 2 units of DNase I was added, mixed thoroughly and incubated at 37°C for 10 minutes. 1 µl of 0.5 M EDTA was added (to a final concentration of 5 mM to protect RNA from being degraded during enzyme inactivation). Samples were heat inactivated at 75°C for 10 minutes. The sample was incubated on ice.

2.20.12 Quantification of RNA was done by using The Infinite® 200 PRO Nano Quant instrument

For RNA quantification, readings were taken at wavelengths of 260 nm and 280 nm. The reading at 260 nm allows calculation of the concentration of nucleic acid in the sample. The reading at 280 nm gives the amount of protein in the sample. Pure preparations of RNA have OD_{260}/OD_{280} values of 2.0. Nucleic acids have absorbance maxima at wavelength of 260 nm. Organic contaminants like phenol and other aromatic compounds and some reagents used in RNA extraction absorb light of a 230 nm wavelength. Samples with a low 260/230 (below about 1.8) have a significant presence of these organic contaminants that may interfere with RT-PCR lowering their efficiency. Nucleotides, RNA, ssDNA, and dsDNA all will absorb at 260 nm and contribute to the total absorbance. The ratio of absorbance at 260 nm and 280 nm was used to assess the purity of DNA and RNA. A ratio of ~1.8 is generally accepted as "pure" for DNA; a ratio of ~2.0 is generally accepted as "pure" for RNA. If the ratio is appreciably lower in either case, it may indicate the presence of protein, phenol or other contaminants that absorb strongly at or near 280 nm. After RNA quantification, cDNA synthesis and RT PCR preparation was done.

2.20.13 Protocol for First-Strand cDNA Synthesis (according to the manufacture Invitrogen instruction)

The following protocol has been optimized for generating first-strand cDNA for use in two-step qRT-PCR. The following procedure is designed to convert 0.1 pg to 5 µg of total RNA or 0.1 pg to 500 ng of poly(A)+ RNA into first-strand cDNA:

The RNA used for cDNA synthesis was 600 ng, for each sample. The each component was mixed and briefly centrifuged before use. The thermal cycler was preheated to 65°C. The following components were added in a 0.2 mL thin-walled PCR tube on ice:

- Up to 600 ng RNA
- Primer (50 µMoligo-dT)
- Annealing Buffer 1 µL
- RNase/DNase-free water to 8 µL

The samples were incubated in a thermal cycler at 65°C for 5 minutes followed by incubation of the samples immediately on ice for at least 1 minute. After the incubation the following were added to the sample tubes on ice:

- 2X First-Strand Reaction Mix 10 µL
- SuperScript® III/RNaseOUT™ Enzyme Mix 2 µL

The sample was vortexed briefly to mix the components and then briefly centrifuged. The sample was incubated at 50°C for 50 minutes. The reactions were terminated at 85°C for 5 minutes and were incubated on ice. The synthesized cDNA was stored at –20°C and it was used for PCR.

2.20.14 RT (real time) PCR reaction mixture

Fast SYBR® Green Master Mix (Applied biosystem) supplied in a 2X concentration, was used to perform real-time PCR using SYBR® Green I dye. The master mix contains:

- SYBR® Green I Dye
- AmpliTaq® Fast DNA Polymerase (Ultra Pure)
- Uracil-DNA Glycosylase (UDG)
- ROX™ dye Passive Reference

- dNTPs
- Optimized buffer components

Table 2-5 Following primers were used for the PCR reaction.

ATUBI (F) 5' CAC ACT CCA CTT GGT CTT GCG T 3'

ATUBI (R) 5' TGG TCT TTC CGG TGA GAG TCT TCA 3'

GUS (F) 5' CCG GGT GAA GGT TAT CTC TAT GAA C 3'

GUS (R) 5' CGA CGC GAA GCG GGT AGA TAT CA 3'

Mix the Fast SYBR® Green Master Mix thoroughly by swirling the bottle. The frozen cDNA samples and primers (Table 2-5) were placed on ice to thaw. After the samples were thawed, vortexd them, then centrifuge the tubes briefly. The reaction was carried out by using Applied Biosystems 7500 Fast Real-Time PCR System. The thermal cycling conditions using the default PCR thermal-cycling conditions are specified in the following table 2-6:

Table 2-6 Thermocycler conditions followed to amplify GUS gene

Step	Temperature°C	Duration	Cycles
AmpliTaq®Fast DNA Polymerase, UP Activation	**95**	**30 sec**	**Hold**
Denature	**95**	**30 sec**	**45**
Anneal/Extend	**60**	**1 min**	

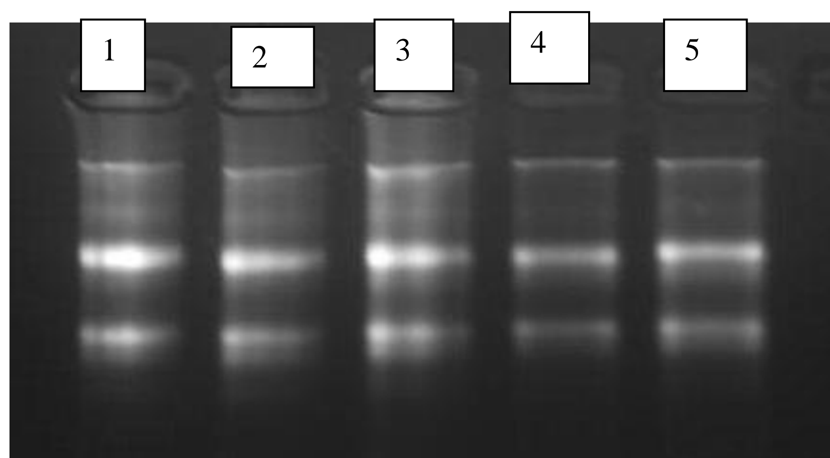


Figure 2-17 Gel image showing total RNA isolated from transgenic *Arabidopsis* leaves treated with ABA. Lanes 1-5 ethidium bromide stained gel showing RNA samples

RESULTS and DISCUSSION

CHAPTER-3

3.0 Results and Discussion

ACGT elements are involved in various biological processes viz auxin response, salicylic acid response, UV light response, ABA response, jasmonic acid response (Figure 1-3). Computational analysis of *PP2C* (AT5G59220) promoter revealed the occurrence of six copies of ACGT elements in the promoter region. Hence, an understanding of the pattern and distribution of ACGT elements is critical to this study. To achieve this, we analyzed four different plant genomes viz *Arabidopsis thaliana*, *Oryza sativa*, *Glycine max* and *Sorghum bicolour* to identify pattern and distribution of ACGT repeat elements across these four genomes (as mentioned in section 3.1 and 3.3)

3.1 Genome wide analysis of co- occurrence of ACGT elements in *Arabidopsis thaliana*

The frequency of occurrence of ACGT *cis* elements was analyzed in whole genome of *Arabidopsis thaliana*. For the analysis, two ACGT elements were searched with or without a spacer sequence ranging from 0 to 30 bp. When two ACGT elements were used in tandem and its frequency across the whole genome was looked for, its occurrence was 1885 times. It can be explained by the probability, that the occurrence of smaller sequence would be more prevalent in the genome. When occurrence of $(ACGT)_8$ element sequence was evaluated, its occurrence was found to be 200, which indicates that the frequency of occurrence of larger sequence decreases in the genome as expected. When the spacer distance between two ACGT elements was increased by one nucleotide at a time the frequency was found to be N1 (1605), N2 (110), N3 (16) and N4 (23) (Table 3-1). The probability of occurrence of two ACGT elements separated by a single nucleotide N1 was high as compared to that of N2, N3 and N4 spacer distance. No biological role has been assigned or reported in the literature to N1, N2, N3 and N4 spacer length so far, probably because this length of spacer is not sufficient enough to define the binding specificity of bZIP proteins (Table 3-1 and 3-1-1) (Pan et al., 2010). In addition, it has been shown that single ACGT *cis* element is not enough for the optimum ABA inducible gene expression (Gomez-Porras et al., 2007). For a functional promoter supporting elements known as coupling element (CE1 or CE3) are required viz. DRE, O2S, motif III, which form an ABA-responsive complex (ABRC) (Marcotte et al., 1989; Shen and Ho

1995; Shen et al., 1996; Singh 1998; Hobo et al., 1999). The studies have shown that an ABRE-ABRE pair also forms an ABRC, the second ABRE acting as the coupling element (Shen and Ho 1995; Gómez-Porraset al., 2007). Moreover Su et al., (1998) have studied one and four copies of 49 bp ABRC1 sequence, identified from barley *HVA22* promoter, which was fused with minimal rice *act1* (actin 1) promoter and showed three to eight fold inductions during stress conditions. Higher induction was obtained by using four ABRC1 repeats. *HVA1-like* promoter induces the enhanced gene expression of *OsLEA3-1* under drought conditions. *HVA1-like* promoter possesses multiple copies of ABRE and hence elevated *OsLEA3-1* gene expression leading to enhanced drought tolerance in transgenic rice under field conditions (Xiao et al., 2007).

The N7 and N13 spacer sequence between two ACGT elements have shown higher frequency of occurrence (77 for both) than N5 spacer sequence. in the present study major focus was on N5, N10, N25 and N30 spacer distances (Table 3-1 and Table 3-1-1). As per observation, gradual decrease in the frequency of occurrence was observed in the case of (ACGT) $_{N5}$ (ACGT) with a value of 62 (Table number 3-1 and Table number 3-1-1). This trend also continued with (ACGT) $_{N10}$ (ACGT) sequence where the occurrence further fell down to 39 from 62 (Table 3-1 and Table 3-1-1). In contrast to this, there was an unexpected sharp rise in the frequency of (ACGT) $_{N25}$ (ACGT) and (ACGT) $_{N30}$ (ACGT) to 72 and 86 respectively which doesn't follow the probabilistic distribution (Table 3-1). To understand this atypical observation, literature evidence supports the biological roles of (ACGT) $_{N5}$(ACGT) and (ACGT) $_{N25}$(ACGT) have been reported and are found to be induced in response to SA and ABA, respectively (Mehrotra et al., 2005 and Mehrotra and Mehrotra 2010). A similar study by Mehrotra et al., (2005) have reported the effect of copy number and the spacing between the two ACGT elements, using spacer sequences of length 5, 10 and 25 for the study. The highest expression of the GUS gene was observed when (ACGT)$_{N5}$(ACGT) or (ACGT)$_{N10}$(ACGT) was placed 50 bp upstream to *Pmec* (minimal expression cassette) series of the promoter followed by (ACGT)$_{N25}$(ACGT) placed 100 bp upstream to *Pmec* series (Mehrotra et al., 2005). Another study showed that certain distances were significantly enriched between two W boxes viz N1, N2, N3, N5, N7, N29 and N30 (Ciolkowski et al., 2008). The frequency of occurrence of N5, N25 and N30 was higher than the smaller spacer sequences. So, it was apparent from the data obtained that larger spacer sequences in some cases were more favored in comparison to smaller spacer sequences in *Arabidopsis thaliana* genome.

To understand the importance or preference of nucleotides in the spacer sequence between two ACGT elements, randomization of the spacer sequence was done to identify the significance of the spacer sequence. When the ACGT element along with the spacer sequence was randomized, a further decrease in frequency of randomized sequence was observed. In comparison to non-randomized N5 sequence, the randomized N5 (ACGT RN5 ACGT) sequence value decreased to 23 (Table 3-1). On comparing the randomized sequences N5, N25 and N30, the frequency of N25 and N30 increased to 32 and 34 respectively (Table 3-1). When the whole sequence including ACGT element along with spacer sequence was randomized, the frequency of ACGT N5 ACGT increased to 137 (Table 3-1 and 3-1-1). In other cases like RN10, RN25 and RN30, its occurrence decreased to 10, 18 and 18, respectively. This indicates that the preference for nucleotides or arrangement of nucleotides in spacer sequence length could plays a significant role in providing proper binding efficiency for the binding of transcription factors (Table 3-2). G box and C box have ACGT as their core sequence. The G box or C box and their hybrids have been observed in the promoter region of ABA-responsive genes. The two ACGT elements in close proximity cannot serve as flanking sequences for each other. According to Foster et al., (1994), in case of ACGT elements the flanking sequences also play an important role in defining the binding efficiency and specificity of bZIP transcription factors. The nucleotides present at +2/-2, +3/-3 and +4/-4 positions define the specificity and affinity for binding of bZIP class of transcription factors (Izawa et al., 1993; Foster et al., 1994). Like for G-box (presence of G at +2/-2 position), preferred nucleotides at +3/-3 are A/C whereas on +4/-4 are G/T. Similarly for C box, the preferred nucleotides at +3/-3 are C/A. In another study by Suzuki et al., (2005), five classes of stress responsive gene promoters were considered, namely ABA regulated (ABR), ABA dependent (ABD), VP1 (viviparous1) dependent (VPD), VP1 and ABA dependent (VAA), VP1 or ABA dependent (VOA). The distribution of flanking sequence of ACGT motifs was studied in all these gene promoters. It has been reported that out of 64 variants, 23 variants were differently distributed among these 5 stress responsive promoters. GAC as flanking sequence of ACGT was observed in all classes especially in VAA. Like this, another variant GCCACGT was reported in VAA, VPD and in two classes of VP1 dependent. Similarly, AACACGT and CCCACGT were reported in ABA dependent and ABA regulated genes, respectively. It has been reported that flanking sequence is also responsible for differential distribution in stress responsive promoters (Suzuki et al., 2005).

Table 3-1 Frequency of occurrence of the selected spacer sequence between two ACGT elements

Cis element	Sequence	Total
$(ACGT)_2$	ACGTACGT	1885
$(ACGT)_8$	ACGTACGTACGTACGTACGTACGTACGTACGT	200
$(ACGT)_{N1}(ACGT)$	ACGTGACGT	1605
$(ACGT)_{N2}(ACGT)$	ACGTGGACGT	110
$(ACGT)_{N3}(ACGT)$	ACGTGGCACGT	16
$(ACGT)_{N4}(ACGT)$	ACGTGGCTACGT	23
$(ACGT)_{N5}(ACGT)$	ACGTGGCTAACGT	72
$(ACGT)_{N10}(ACGT)$	ACGTGGCTATGGCGACGT	39
$(ACGT)_{N25}(ACGT)$	ACGTGGCTATGGCGGAGCAAGATTCACTCACGT	62
$(ACGT)_{N30}(ACGT)$	ACGTAAGTGTTTCGTATCGCGATTTAGGAGAAGTACGT	86
$(ACGT)_{RN5}(ACGT)$	ACGT—GCTAG—ACGT	23
$(ACGT)_{RN10}(ACGT)$	ACGT—TGGGGCCGAT—ACGT	14
$(ACGT)_{RN25}(ACGT)$	ACGT—AGACACGTTGGGGGAACTTACTGCC—ACGT	21
$(ACGT)_{RN25}(ACGT)$	ACGT—ATATGAGATCGGCGCTTCACGGAGC—ACGT	32
$(ACGT)_{N5}(ACGT)$ randomized	GGAATCCTTGGCA	137
$(ACGT)_{N10}(ACGT)$ randomized	GCGGGCTATCGGTAGCAT	10
$(ACGT)_{N25}(ACGT)$ randomized	TAAGGCTTAGCCACGCTTAGGGTGTGAGCACAC	18
$(ACGT)_{N30}(ACGT)$ randomized	TTCGGAATTGGTAGAAGTCATTGGGATTACCCTGGATA	18
$(ACGT)_{RN30}(ACGT)$	ACGT—TATGGGACTTTTAACCATAATAGGGGGTTG—ACGT	34

Table 3-1-1 Frequency of occurrence of spacer sequence between two ACGT elements from 0 to 30 bp

Cis element	Sequence	Total
(ACGT) $_{N1}$(ACGT)	ACGT_ACGT	1605
(ACGT) $_{N2}$(ACGT)	ACGTG_ACGT	110
(ACGT) $_{N3}$(ACGT)	ACGTGG_ACGT	16
(ACGT) $_{N4}$(ACGT)	ACGTGGC_ACGT	23
(ACGT) $_{N5}$(ACGT)	**ACGTGGCT_ACGT**	**72**
(ACGT) $_{N6}$(ACGT)	ACGTGGCTA_ACGT	44
(ACGT) $_{N7}$(ACGT)	ACGTGGCTAT_ACGT	77
(ACGT) $_{N8}$(ACGT)	ACGTGGCTATG_ACGT	64
(ACGT) $_{N9}$(ACGT)	ACGTGGCTATGG_ACGT	32
(ACGT)$_{N10}$(ACGT)	**ACGTGGCTATGGC_ACGT**	**39**
(ACGT) $_{N11}$(ACGT)	ACGTGGCTATGGCG_ACGT	38
(ACGT) $_{N12}$(ACGT)	ACGTGGCTATGGCGG_ACGT	56
(ACGT) $_{N13}$(ACGT)	ACGTGGCTATGGCGGA_ACGT	77
(ACGT) $_{N14}$(ACGT)	ACGTGGCTATGGCGGAG_ACGT	53
(ACGT) $_{N15}$(ACGT)	ACGTGGCTATGGCGGAGC_ACGT	56
(ACGT) $_{N16}$(ACGT)	ACGTGGCTATGGCGGAGCA_ACGT	60
(ACGT) $_{N17}$(ACGT)	ACGTGGCTATGGCGGAGCAA_ACGT	42
(ACGT) $_{N18}$(ACGT)	ACGTGGCTATGGCGGAGCAAG_ACGT	49
(ACGT) $_{N19}$(ACGT)	ACGTGGCTATGGCGGAGCAAGA_ACGT	44
(ACGT) $_{N20}$(ACGT)	ACGTGGCTATGGCGGAGCAAGAT_ACGT	37
(ACGT) $_{N21}$(ACGT)	ACGTGGCTATGGCGGAGCAAGATT_ACGT	40
(ACGT) $_{N22}$(ACGT)	ACGTGGCTATGGCGGAGCAAGATTC_ACGT	53
(ACGT) $_{N23}$(ACGT)	ACGTGGCTATGGCGGAGCAAGATTCA_ACGT	55
(ACGT) $_{N24}$(ACGT)	ACGTGGCTATGGCGGAGCAAGATTCAC_ACGT	53
(ACGT)$_{N25}$(ACGT)	**ACGTGGCTATGGCGGAGCAAGATTCACT_ACGT**	**62**
(ACGT) $_{N30}$(ACGT)	**ACGTAAGTGTTTCGTATCGCGATTTAGGAGAAGT ACGT**	**86**

This could be one of the reasons behind the variation between the regulations or it could be a reason behind the differential regulation like ABA regulated or ABA dependent. Similar result is also obtained with the cold stress up regulated genes. Here, CCGACGT sequence was found to be predominant in up-regulated cold stress responsive genes and also in ABA and VP1 regulated genes. These results show that the differences in the ABA signalling is due to the difference in the binding specificity of the different members of bZIP class, which is determined by the flanking sequence of ACGT core sequence (Suzuki et al., 2005).

3.2 Increase in spacer length between two ACGT elements increases transcription factor binding sites

The binding specificity of transcription factor is based on the distance between the *cis* elements, spacer sequence and length specificity. Spacing between two *cis* elements is important for the promoter's activity. The gene expression can be affected by the increase or decrease in spacing between two *cis* elements (Gilmartin and Chua 1990). An optimum spacer sequence between *cis* elements is required for the proper binding of the transcription factors (Pan et al., 2010). GATA factor binds to (A/T) GATA (A/G) sequence with varying degrees of affinity. GATA factor binding specificity has been found to depend on the nucleotide sequences surrounding the assigned consensus sequence (A/T) GATA (A/G) (Ko et al., 1993; Merika and Orkin 1993; Jeong and Shih 2003).

Computational methods were applied to identify transcription binding sites. In order to identify various families of transcription factors associated with the desired sequences, *in-silico* studies were done using JASPAR (Bryne et al., 2008) and CONSITE (Sandelin et al., 2004) databases. For the present study, *Pmec* minimal promoter sequence was fused with the ACGT elements with or without spacer sequence (Sawant et al., 2001). The $(ACGT)_2$ element in tandem was used as a control sequence. According to the data obtained from computational analysis, no reported transcription factor binds to the $(ACGT)_2$ sequence element whereas different families of transcription factors were observed to bind to the minimal promoter (Table 3-2). No transcription factor was shown to bind to the single ACGT element according to the JASPAR (Bryne et al., 2008) and CONSITE (Sandelin et al., 2004) databases. When the minimal promoter sequence *Pmec* was juxtaposed with the single ACGT element and $(ACGT)_2$ without the spacer sequence, the binding sites for transcription factor increased to 36 in both the cases. The minimal promoter sequence *Pmec* was also juxtaposed with the ACGT-ACGT element separated by – N5, N10, N25 and N30 spacer length and the transcription factor binding sites were analyzed (Table 3-2). The transcription factor binding site were - 39, 40 and 44 for N5, N25 and N30 respectively and comes down, for N10 to 38. Observation from this data indicates that spacer sequence equally contributes along with the core ACGT element, in determining the biological functions. From the genome wide analysis, we noticed that the larger size element is preferred over the smaller element.

Table 3-2 Transcription factors binding site increases with the increase in the spacer length between the ACGT elements

Transcription factors	MPS	(ACGT)	(ACGT)$_2$ MPS	(ACGT)$_8$ MPS	(ACGT) $_{N5}$ (ACGT) MPS	(ACGT) $_{N10}$ (ACGT) MPS	(ACGT) $_{N25}$ (ACGT) MPS	(ACGT) $_{N30}$ (ACGT) MPS
MODEL NAME	**Frequency**							
ARR10	**0**	**0**	**0**	**0**	**0**	**0**	**1**	**2**
AGL3	**2**	**0**	**2**	**2**	**2**	**2**	**2**	**1**
ATHB-5	**1**	**0**	**1**	**2**	**1**	**1**	**1**	**1**
bZIP910	**0**	**0**	**0**	**0**	**1**	**1**	**1**	**1**
Dof3	**1**	**0**	**1**	**1**	**1**	**1**	**2**	**3**
EmBP-1	**2**	**0**	**2**	**1**	**2**	**2**	**2**	**2**
Gamyb	**5**	**0**	**5**	**5**	**5**	**5**	**5**	**5**
HAT5	**2**	**0**	**2**	**2**	**2**	**2**	**2**	**2**
HMG-1	**6**	**0**	**6**	**6**	**6**	**6**	**6**	**6**
HMG-I/Y	**6**	**0**	**6**	**6**	**6**	**6**	**6**	**6**
id1	**5**	**0**	**5**	**5**	**5**	**5**	**5**	**7**
myb.Ph3	**1**	**0**	**1**	**1**	**2**	**1**	**1**	**3**
PEND	**1**	**0**	**1**	**1**	**1**	**1**	**1**	**1**
Squamosa	**2**	**0**	**3**	**3**	**3**	**3**	**3**	**3**
TGA1A	**1**	**0**	**1**	**1**	**2**	**2**	**2**	**1**
	35	0	36	36	39	38	40	44

Mehrotra et al., (2010) explains that promoters are differentially regulated by the spacing between two copies of the ACGT elements. Another study states that copy number of ACGT elements and the position of this element from the TATA box, also affects the gene expression (Mehrotra et al., 2005). Roy Choudhury and Sen Gupta (2009) have conducted a study to identify the interaction between the promoter-elements of abiotic stress-inducible genes from cereals and nuclear proteins from tobacco using EMSA and RNA blot. For this study, synthetic promoters were designed containing either of the four tandem copies of ABREs from wheat (4XABREs:35s) (CGAAAGCTTGCC**ACGT**GGCGCC**ACGT**GGCGCC**ACGT**GGCGCC**ACGT**GGCA CCCTCCTT) or two copies of ABRCs (2XABRCs:35s) (GTCAGAAAGCTTGCC**ACGT**GGCTGCAGTGCCATTGCCACCGGATCAGCC**AC GT**GGCTCCAGTGCCATTGCCACCGGATGACGCACAATCCCAC) from barley. Each was fused with -70 bp *CaMV 35S* promoter (AGGCTAGGATCCTCTAGATTCCCTCTCCAAATGAAATGAACTTCCTTATATAG AGGAAGGGTCTTGCGAAGGATAGTGGGATTGTGCGTCAT). CAAT box and TATA box (70 LS) were also included in the construct. The authors report enhanced GUS expression in transgenic plants bearing constructs 2XABRCs:35s as compared to 4XABRES:35s with no GUS expression in wild type plants under abiotic stress conditions. EMSA studies depicted more intense band for the 2XABRCs construct than 4XABRES (Roy Choudhury and Sen Gupta 2009). This was due to the presence of the coupling element in 2XABRCs which acts together with ABRE to further support the promoter activity and binding of transcription factors. This observation shows that the binding of transcription factors to two ACGT elements separated by a particular length is required for the proper functioning of promoter. Flanking sequence along with the core sequence is also important for binding of other transcription factors. The binding specificity of WRKY proteins and Dof proteins also depends on the flanking sequence present in the neighborhood of core TGAC and AAAG $_{N7}$ CTTT *cis* regulatory elements respectively (Cilwolswki et al., 2008; Mehrotra et al., 2014b).

In order to further study the distribution pattern and evolution of co-occurrence of ACGT elements across the plant kingdom in promoter region, two monocots – *Sorghum bicolour* (Sorghum) and *Oryza sativa* (Rice), and two dicots –*Arabidopsis thaliana* and *Glycine max* (Soybean) were used to study.

3.3 Comparative analysis of co-occurring ACGT elements along with spacer distance in gene promoter region of *Sorghum bicolour, Oryza sativa*, *Arabidopsis thaliana* and *Glycine max*

In preliminary study ACGT element with or without spacer sequence was searched in whole genome of *Arabidopsis thaliana* (mentioned in section 3.1). ACGT (N) ACGT (N≤30) spacer frequency was analysed and compared among the *Arabidopsis thaliana, Sorghum bicolour, Oryza sativa* and *Glycine max* promoter regions to identify common peaks and dips for different spacer lengths. However we refined the search and restricted it to promoter region. The upstream 1 kb region from transcription start site (TSS) was extracted for 33,323 genes in *Arabidopsis thaliana*, 49,841 genes in soybean, 30,294 genes in rice and 32,886 genes in sorghum.

The peak at "0" spacer length between two ACGT sequences - ACGTACGT i.e. in tandem had shown the highest frequency of occurrence in promoter region and it was common across the four plant species (Figure 3-1 a and b). Other than this, no common peak was observed for other (N 1≤30) spacer sequences in these four plant species. The maximum frequency of spacer lengths observed in dicot and monocot species were as follows: in *Arabidopsis thaliana* N= 0, 2, 24, soybean N=0, 14, 28 (Figure 3-1 a), rice N= 0,7,26 and sorghum N= 0, 2, 4 (Figure 3-1 b). The 24 bp spacer sequence had the maximum frequency in promoters of *Arabidopsis thaliana*. This data again supports the fact that certain larger spacer sequence occurs more as compared to the smaller spacer sequences (Figure 3-2). Although in rice, spacer sequence of 7 nucleotides had shown very high frequency of occurrence but no corresponding function has been observed to be associated with this spacer sequence in the literature.

According to whisker's plot following frequencies (N=0, 2, 24) for *Arabidopsis thaliana*, (N=0, 14, 28) for soybean, (N= 0, 7, 26) for rice and (N= 0, 2, 4) for sorghum lay outside 90th percentile. Local regions of conservation were obtained in monocot and dicot plant species, based on variation in frequencies. *Arabidopsis thaliana* and soybean have shown high correlation in frequencies for spacer length (N) = 6-11 (r 0.974; t=8.599; p=0.0010; N=6) as shown in figure 3-1 a and c. Similarly in monocot plant species, rice and sorghum have shown high correlation in frequencies for spacer distance (N) = 0–6 (r = 0.984; t = 12.42; p = .0001; N = 7) and (N) = 18–26 (r = 0.934; t = 6.921; p =0.0002; N = 9) as shown in figure 3-1 b and c. From this observation, no significant co-relation (r < 0.8) was obtained for spacer length between these four plant genomes.

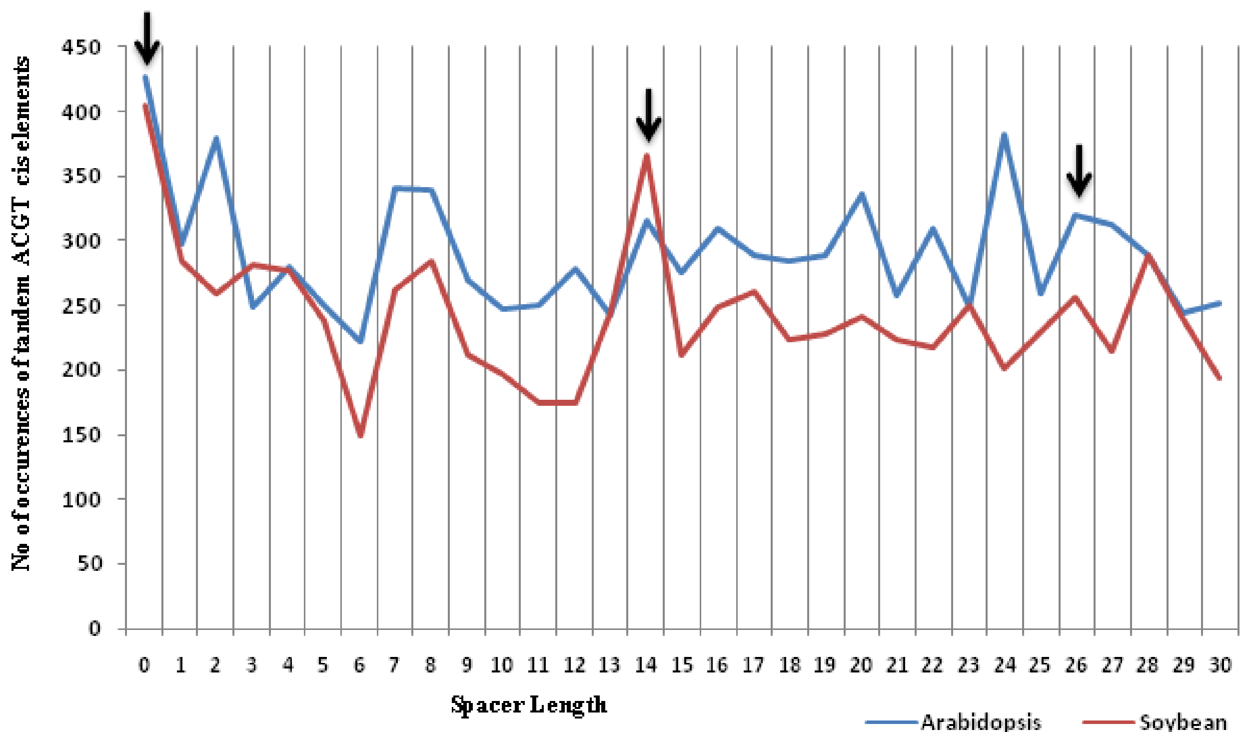


Figure 3-1 a Frequency of occurrence of co-occurring ACGT elements with variable spacer lengths N≤30in all promoters in two dicotyledonous plants - *Arabidopsis* (Blue) and Soybean (Red). High degree of directed evolution was observed in both species as suggested by similar trends in the range of 4-21 spacer lengths. Significant peaks (shown by arrows) are seen at 0, 14 and 26

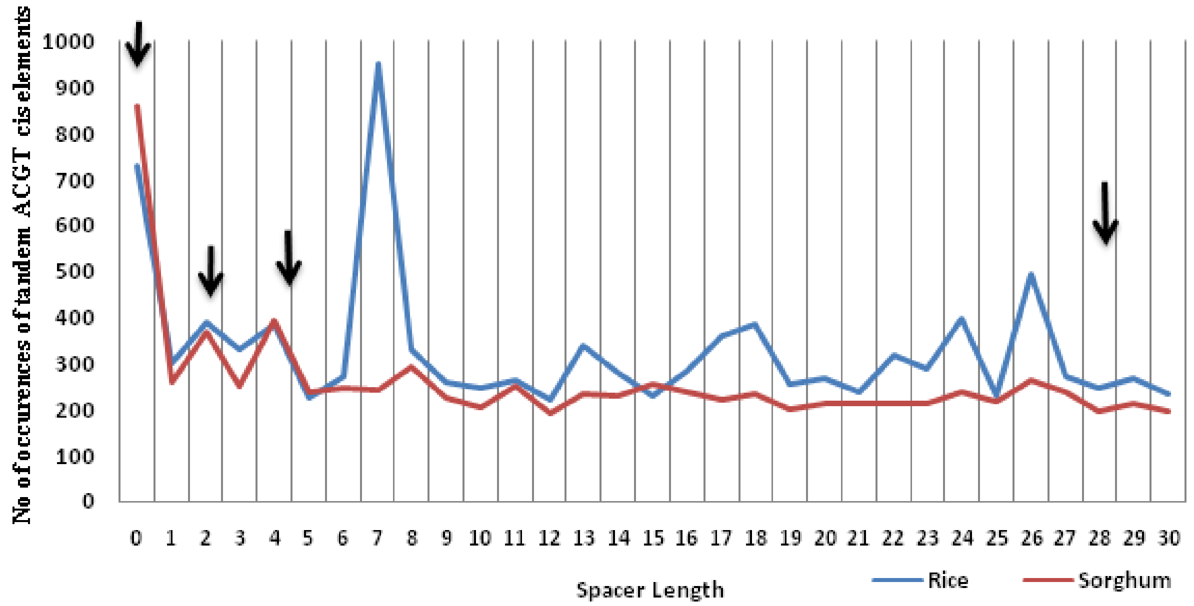


Figure 3-1 b Number of co-occurring ACGT elements with variable spacer lengths between them in all promoters of two monocotyledonous plants - Rice (blue) and Sorghum (Red). Although a conserved evolutionary pattern is not visible, common peaks are observed for spacer lengths 0, 2, 4 and 26 (shown as arrows). High frequency for Rice spacer distance 7 suggests conservation for that spacer distance

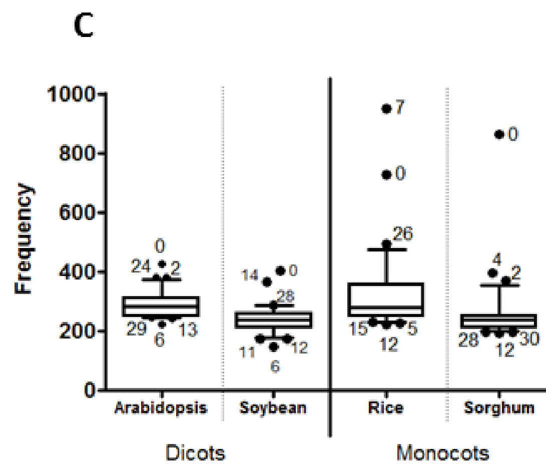


Figure 3-1 c 10–90% box and whiskers plot

A peak for larger spacer size was been observed in *Arabidopsis thaliana*, soybean and rice whereas in sorghum peak was observed for smaller spacer sequence. No spacer length was found to be (regularly occurrence) common across these plant species, though high correlation indicated some conserved areas.

Uniformity in the frequency dips was observed among monocot and dicot species. According to the box and whisker's plot (10% and 90%), peak at 0 is common in all the four genomes while the lowest frequency peaks were observed for spacer length N=6 and N= 11- 13 in the dicot plant species, whereas for spacer sequence N=12 in monocot plant species (Figure 3- 1 c). The theory of a class specific mode of regulation function among monocots and dicots is supported by the pattern observed through differences in regions of high correlation in monocots and dicots. Preference for certain spacer lengths over others could be due to the stable helical phasing for proper assembly of transcription activating complex while other spacer sequences cause sterical hindrance for binding of transcription factors. A study done by Lopez-Ochoa et al., (2007) found that the spacer sequence of 5 bp, 10 bp, 15 bp and 20 bp was inserted between the I-box and G-box to determine the efficiency of conserved modular array 5 (CMA5), a minimal light-regulatory unit of *rbcS* gene promoters. No GUS activity was detected for spacer sequence 15 bp and 20 bp between I-box and G-box in comparison to the 5 bp and 10 bp spacer sequences. Lopez-Ochoa et al., (2007) analyzed that I- and G-boxes cannot be separated by more than 23–25 bases completely as it leads to complete loss of the GUS activity. The results show that helical phasing and distance constraint between the I- and G-box elements are essential for the assembly of gene-transcription activating complex (Lopez-Ochoa et al., 2007).

The sequences of nucleotides present in the spacer sequences were also analyzed. The most preferred nucleotide was G at the first position and C was most commonly present at the end of all the spacer sequences (Figure 3-2). Other than these positions, no other nucleotide was favoured at any other position in spacer sequence. In contrast, a high degree of conservation of nucleotides for spacer sequence N= 24 bp was observed. "N" in figure 3-2, represents that none of the nucleotides have met the threshold requirement**.** Apart from these (first and last) positions, 17 nucleotide positions were found to be conserved only in spacer length N= 24 (highlighted in blue colour in Figure 3-2) and not in any other spacer length. The predominance of G at +2 and C at -2 positions was also

Spacer Length	Consensus Spacer Sequence (Position-wise)																													
	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30
0																														
1	N																													
2	N	C																												
3	G	N	C																											
4	G	N	C	C																										
5	G	N	N	N	C																									
6	G	N	N	N	N	C																								
7	G	N	C	N	N	N	C																							
8	G	C	A	C	G	N	N	C																						
9	G	G	C	N	N	N	N	N	C																					
10	G	N	N	N	N	N	N	N	N	C																				
11	N	N	N	N	N	N	N	G	N	N	N																			
12	G	N	N	N	N	N	C	T	G	N	N	N																		
13	G	N	N	N	N	N	N	N	N	N	N	N	C																	
14	C	N	N	N	N	N	N	N	C	N	N	N	N	C																
15	G	N	N	N	N	N	N	N	N	N	N	N	N	A	C															
16	G	N	N	C	N	N	N	N	N	N	N	N	N	N	N	C														
17	G	N	N	N	N	N	N	N	N	N	N	N	N	N	N	N	C													
18	N	N	N	C	G	N	T	N	N	N	N	N	N	N	N	N	N	C												
19	G	N	N	C	N	N	N	N	N	N	N	N	N	N	N	N	N	N	C											
20	G	C	C	N	N	N	N	N	N	N	N	N	N	C	N	C	N	N	C	C										
21	N	C	N	C	N	N	N	N	N	N	N	N	N	N	N	N	N	N	N	A	C									
22	G	G	N	N	N	N	N	N	N	N	N	A	N	N	N	N	N	N	N	N	C	C								
23	N	N	N	C	N	C	N	N	N	N	N	A	N	N	N	N	N	A	N	N	N	N	C							
24	N	N	N	C	C	C	G	N	G	A	T	T	N	T	C	C	C	S	S	C	A	A	N	N						
25	G	N	N	N	N	N	G	N	N	N	N	N	N	N	T	N	N	N	C	N	N	N	N	N	C					
26	G	N	C	N	N	N	N	G	N	N	N	N	N	N	N	N	N	N	N	N	N	A	N	N	C	C				
27	G	N	N	N	N	N	N	N	N	N	N	N	N	N	N	N	N	N	N	N	N	N	N	N	G	N	C			
28	G	N	C	N	N	N	N	N	N	N	N	N	N	N	N	N	N	N	N	N	N	N	N	N	N	C	N	C		
29	G	N	N	N	N	N	N	N	N	N	N	N	N	N	N	N	N	N	N	N	N	N	N	N	N	N	N	C	C	
30	G	N	C	N	N	N	N	N	N	T	A	N	N	N	N	N	N	A	N	N	N	N	N	N	N	N	N	N	N	C

Figure 3-2 The spacer sequences of varying length from 0 to 30 between the two ACGT elements. The frequency of occurrence of 24 bp spacer sequence is found to be more (shown in blue colour) in *Arabidopsis thaliana*. Guanine in red and cytosine in green colours are showing first and last position of nucleotides in spacer sequence respectively. Single letter IUPAC DNA codes were used for all spacer lengths.

A- Adenine, T- Thymine, G- Guanine, C-cytosine, S-G/C, Here, N-none of the nucleotides met the threshold requirements

evident from the spacer sequence analysis results. Earlier reports by Ishige et al., (1999) validate this data, who evaluated the regulatory function of G-box flanking sequences and also concluded that a G-box (GC**CACGTG**CC) confers high-level constitutive expression in seed, root, leaf, and axillary bud, almost all parts of flower bud and pollen of transgenic tobacco, carrot and rice. This data is also validated by the fact that bZIP transcription factor family members have shown enhanced binding towards ACGT core elements followed by G and C nucleotide (Izawa et al., 1993).

3.4 Comparative study of the frequency of co-occurring ACGT elements and random tetramer sequences along with the spacer length

Distribution preference of co-occurring ACGT element was compared with 4 palindrome and 4 non-palindrome tetramer sequences. CTAG, AGCT, GATC, GCTA, ATGC, CGTA, TAGC and TGCA sequences were used as control sequences and for these sequences the frequency of occurrence with or without spacer sequence ($N= 0 \leq 30$) were analyzed in the gene promoters of *Arabidopsis*, soybean, rice and sorghum. The frequency of occurrence for these 4 palindrome and non-palindrome random sequences were also compared among the four plant species (Figure 3-3). Among these control sequences, the numbers of ATGC, GATC and TGCA sequences were observed to be predominantly high with respect to ACGT sequence. From the analysis it was revealed that ACGT was not the most preferred tetramer sequence in all gene promoters. The frequency of occurrence of ACGT lies around the mean of frequencies of the control tetramer sequences, except ACGT elements which had a very low occurrence in sorghum (Figure 3-3).

3. 5 Analysis of pattern of ACGT elements in the orthologous gene groups

A frequency analysis for ACGT elements in promoters of all orthologous gene groups reported between *Arabidopsis* (12013 genes), rice (11453 genes) and sorghum (10575 genes) were performed. Similar trends in spacer distance frequency were noticed in the orthologous genes in *Arabidopsis*, rice and sorghum (Figure 3-4 A), with significant correlations across all spacer lengths for *Arabidopsis* and rice ($r = 0.3945$, $p < 0.05$, $n = 31$) (Figure 3-4 B), *Arabidopsis* and sorghum ($r = 0.4177$, $p < 0.05$, $n = 31$)

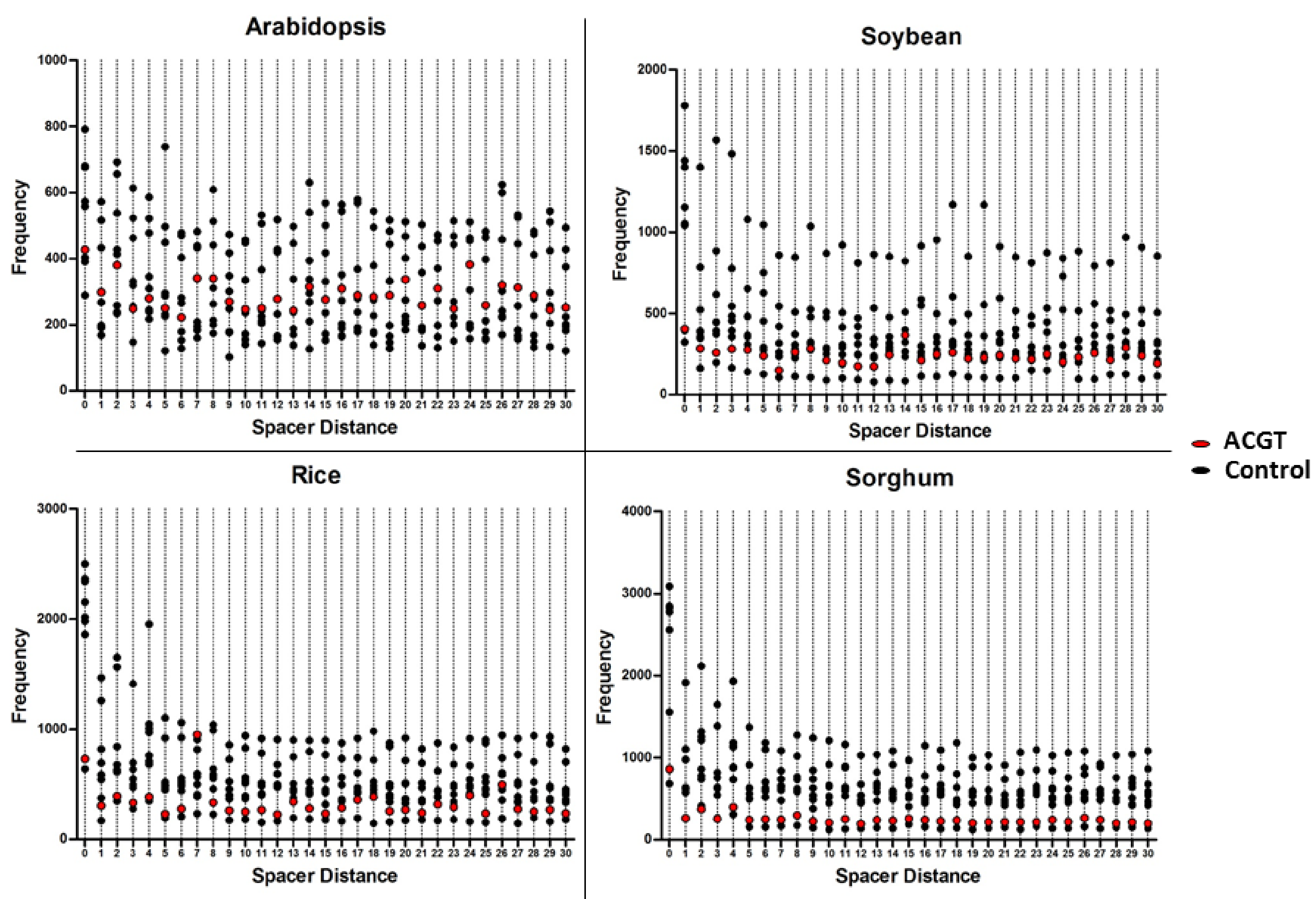


Figure 3-3 The frequency of co-occurring ACGT elements was compared with the palindromic (GATC and TGCA) and non-palindromic (ATGC) tetramer sequences in all four genomes. These random sequences were used as control

(Figure 3-4 C), and rice and sorghum ($r = 0.4495$, $p < 0.5$, $n = 31$) (Figure 3-4 D).

Biological relevance of ACGT (N) ACGT and random elements were analysed in *Arabidopsis* in genes containing ACGT and random *cis* elements along with spacer lengths for conditions. The overall likelihood of occurrence was calculated for each condition, with a likelihood of 1 being that of random chance. Conditions which showed value greater than 1.3 were selected for further analysis. Conditions which were used for ACGT (N) ACGT elements were as follows: up/down regulated during developmental stages (embryo, seedling), by hormones (ABA, auxin, ethylene, gibberellin, jasmonic acid, salicylic acid), in different plant parts (cambium, flower, leaf, root, pollen, seed, sperm cell, stem,vegetative apex, apical root meristem) and by environmental conditions (at baseline growth temperature, disease, drought, low water potential, at optimum photosynthetic temperature, salt, 20% inhibition from optimum photosynthetic temperature, 30% inhibition from optimum photosynthetic temperature). Genes containing random control elements were also analyzed in salt (all, up and down), drought (all, up and down) and jasmonic acid (all, up and down). Biological functional analysis of rice genes containing ACGT (N) ACGT and random control elements was done for salt and drought up-regulation conditions. Microarray data analysis showed the biological significance of ACGT (N) ACGT elements and observed their predominance in spite of lower frequencies in comparison to random elements in all promoters. The likelihood for each of these conditions was compared with that of the same condition for the previously described 8 random control sequences (Figure 3-5 A). Dot plots for each of the conditions and a Grubbs' test for outliers indicated a significant effect for ACGT containing promoters to be up-regulated by salt (Mean = 0.9791; $Z = 2.21$: $p < 0.5$; $n = 9$) and drought (Mean = 0.9539; $Z = 2.21$: $p < 0.5$; $n = 9$) (Figure 3-5 B). An individual spacer distance -wise split up revealed considerable fluctuation in the likelihoods for each spacer length, and despite no clear pattern emerging over all functions (Figure 3-5 A), suggested potential spacer length specific gene regulation. Similar analysis was performed by using microarray data for *Oryza sativa* for two functions – salt and drought up-regulation. Both conditions showed likelihood of occurrences greater than 1.20, and the Grubbs' test for outliers emerged to be significant for both salt (Mean = 0.9611; $Z = 2.21$: $p < 0.5$; $n = 9$) and drought (Mean =0.8453; $Z = 2.21$; $p < 0.5$; $n = 9$) up-regulation

(Figure 3-5 C). Interestingly, similar fluctuating spacer length-wise patterns were observed for this dataset.

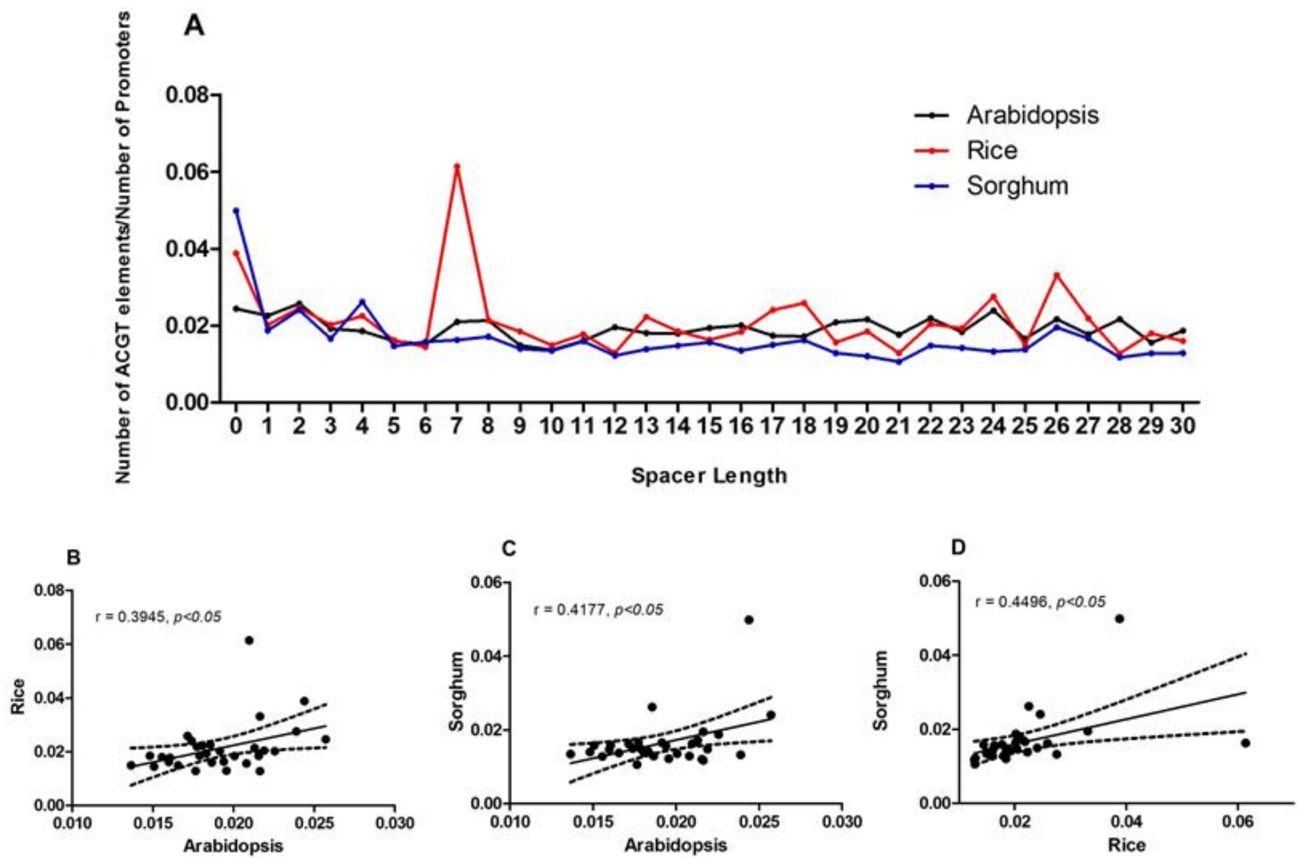


Figure 3-4 Patterns of co-occurring ACGT elements in orthologous genes in *Arabidopsis thaliana*, rice and sorghum

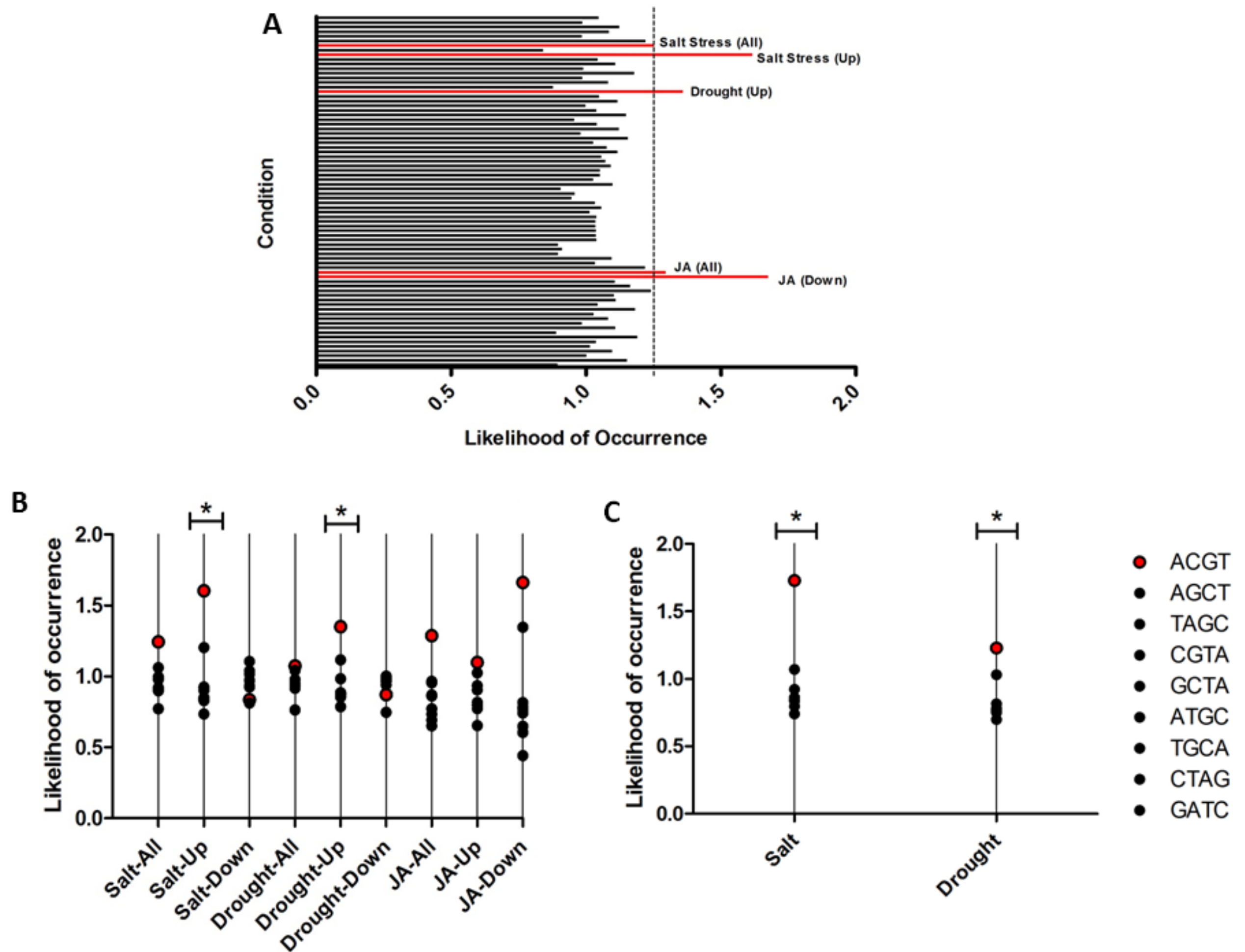


Figure 3-5 (A) The frequency of ACGT (N) ACGT is relatively abundant in stress responsive promoters. The microarray data shows the up & down regulation of genes in the presence of salt, drought, jasmonic acid. (B) Dot plot of ACGT and control random elements in *Arabidopsis thaliana* have shown that the up-regulation of salt, drought and jasmonic acid gene promoters with ACGT elements are more preferred in comparison to control sequences. (C) Dot plot for rice have shown similar result for ACGT elements in response to salt and drought stress

Biological functional analysis inferred that co-occurring ACGT elements are involved in gene regulation in response to salt and drought stress conditions in both *Arabidopsis* and rice. ACGT (N) ACGT was over-represented in the upstream region of genes which were up-regulated in response to salt and drought stress. This information is supported by up-regulation of bZIP transcription factors in response to drought and salinity stress conditions in plants system. Further, although regulation by jasmonic acid could not satisfy the criteria for the Grubs' outlier tests, the likelihood of occurrence of genes regulated by jasmonic acid was higher than the cutoff of 1.30 (Figure 3-5 A). With jasmonic acid's conventional involvement in mediating stress responses in plants this observation is extremely interesting in light of our findings. The presence of ACGT element is exclusively more in stress responsive gene promoters in comparison to random sequence. Further, it is evident from the microarray data of salt and drought responsive gene promoters that they are more likely to co-occur within a range of N5 to N30 (N= spacer sequence length) (Figure 3-5 B and C).

From the genome-wide studies of Gomez-Porras et al., (2007) for divergent occurrence of ABRE and CE3 in *Arabidopsis thaliana* and rice was done, it is evident that the occurrence of CE3 was more prevalent in rice and other monocot genomes, and absent in the dicot i.e. *Arabidopsis thaliana.* It also showed the predominance of ACGT-ACGT element pair in *Arabidopsis thaliana*. CE3 elements have shown preference for the minus strand (Ross and Shen 2006).The common distribution pattern of ACGT elements obtained from the present study across the four plant genomes reveals that it has emerged from a common ancestral gene. This result also correspond to reports which state that the spacer sequence of light responsive regulatory elements differs but it is found to be conserved in orthologous gene promoters from other plant species (Lopez-Ochoa et al., 2007). Maruyama et al., (2012) analysed the 5' upstream region (-1 kb) of 10,000 genes in *Arabidopsis,* rice and soybean. The frequency of occurrence of ABRE element was found to be more in (5') -51 to -100 bp promoter region of *Arabidopsis,* rice and soybean. According to the microarray data of this study, ABRE element was found to be one of the most conserved regulatory elements in the promoter region of dehydration inducible and AREB genes in *Arabidopsis thaliana,* rice and soybean. The orthologous genes of AREB family have been found to be up regulated in rice and soybean during dehydration stress.

The important observation made by Maruyama et al., (2012) was that ABRE dependent signalling pathway was established in all land plants before dividing in two major groups viz., monocots and dicots. In another study, Tran and Mochida (2010) have predicted and identified abiotic stress responsive transcription factors which are found to be involved in abiotic stress signalling in soybean. From studies it has been observed that most transcription factors are found to be conserved among orthologous, paralogous, and co-regulated genes (Tran and Mochida 2010). ABRE element sequence is found to be conserved in the ABA responsive gene promoters in both the monocot and dicot plant species. Following ABRE consensus sequences CGTACGTGTC, GACGTG and CCACGTGG were observed in the *Oryza sativa*, *Zea mays* (maize) and *Arabidopsis thaliana* respectively (Rabbani et al., 2003; Yazaki et al., 2003). Another study by Zhang et al., (2005) and Won et al., (2009) that *cis* regulatory elements were found to be highly conserved among orthologous or paralogous genes and even in co-regulated genes. The evolution of ACGT element across the plant genome is found to be evolved in parallel fashion.

3.6 Characterization of protein phosphates 2C- like promoter (*PP2C*)

According to literature, two ACGT elements in close proximity is a feature of stress responsive promoter (Mehrotra et al., 2012, 2013). From *in -silico* analysis, a *PP2C*-like promoter (accession number AT5G59220) was identified in *Arabidopsis thaliana*. The *PP2C*-like promoter having a unique genetic arrangement of ACGT elements (Figure 3-6 a). The *PP2C*-like promoter has two ACGT elements separated by 30 nucleotides is a possible feature of stress response and two ACGT elements separated by 5 nucleotides is found to be induced in response to SA (Figure 3-6 a and b) (Mehrotra et al., 2005; Mehrotra and Mehrotra 2010). The *PP2C*- like promoter harbors multiple identified *cis* regulatory elements viz light regulatory elements, stress responsive *cis* elements and plant hormones elements. These putative regulatory elements analyzed by PLACE and PlantCARE databases, propose that *PP2C*- like promoter may respond to a variety of inducers like ABA, JA, SA signals (Prestridge 1991; Higo et al., 1999) (Table 3-3). Therefore it can be predicted that *PP2C*- like promoter could be an inducible promoter, regulated by abiotic factors and hormones. Four homologue sequences of the pathogenesis- and salt-related *cis* element GT1GMSCAM4 (GAAAAA) were detected in

full-length *PP2C-* like promoter. Two homologue sequences of WBOXATNPR1 (TTGAC) were also noticed in full-length *PP2C-* like promoter. In addition, light-responsive elements such as GT1 (GRWAAW) and GATABOX (GATA) were also

5'AAGTATTCACGCACCAAGGTTATATTTGTAGTGACATATTCTACAATTATCAC
ATTTTTCTCTTATGTTTCGTAGTCGCAGATGGTCAATTTTTTCTATAATAATTTG
TCCTTGAACACACCAAACTTTAGAAACGATGATATATACCGTATTGTCACGCT
CACAATGAAACAAACGCGATGAATCGTCATCACCAGCTAAAAGCCTAAAACA
CCATCTTAGTTTTCACTCAGATAAAAAGATTATTTGTTTCCAACCTTTCTATTG
AATTGATTAGCAGTGATG**ACGT**AATTAGTGATAGTTTATAGTAAAACAAATG
GAAGTGGTAATAAATTTACACAACAAAATATGGTAAGAATCTATAAAATAAG
AGGTTAAGAGATCTCATGTTATATTAAATGATTGAAAGAAAAACAAACTATTG
GTTGATTTCCATATGTAATAGTAAGTTGTGATGAAAGTGA**TGACGT**AATTAG
TTGTATTTATAGTAAAACAAATTAAAATGGTAAGGTAAATTTCCACAACAAAA
CTTGGTAAAAATCTTAAAAAAAAAAAAAGAGGTTTAGAGATCGCATGCGTGT
CATCAAAGGTTCTTTTTCACTTTAGGTCTGAGTAGTGTTAGACTTTGATTGGTG
C**ACGT**AAGTGTTTCGTATCGCGATTTAGGAGAAGT**ACGT**TTTAC**ACGT**GG
ACACAATCAACGGTCAAGATTTCGTCGTCCAGATAGAGGAGCGAT**ACGT**CAC
GCCATTCAACAATCTCCTCTTCTTCATTCCTTCATTTTGATTTTGAGTTTTGATC
TGCCCGTTCAAAAGTCTCGGTCATCTGCCCGTAAA**TATAAAGATGATTATATT**
TATTTATATCTTCTGGTGAAAGAAGCTAATATAAAGCTTCCATGGCTAATCTT
GTTTAAGCTTCTCTTCTTCTTCTCTCTCCTGTGTCTCGTTCACTAGTTTTTTTCG
GGGGAGAGTGATGGAGTGTGTTTG 3'

Figure 3-6 a Full length sequence of *PP2C-* like promoter (AT5G59220) from *Arabidopsis thaliana*. Bold letters indicate ACGT elements. The elements in oval boxes show the unique arrangement of ACGT elements in *PP2C-* like promoter

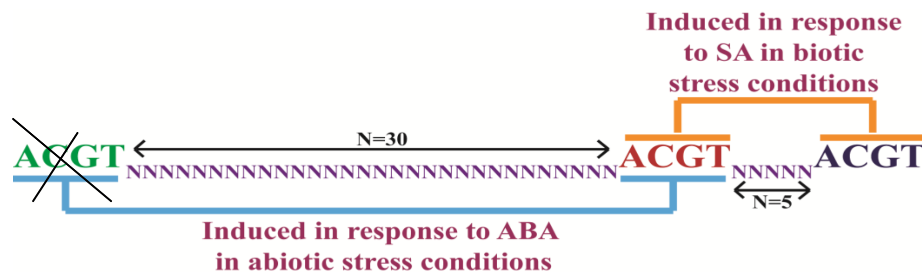


Figure 3-6 b Design of NRM construct by deletion of ACGT element, which is one of the deletion variants of *PP2C*-like promoter. The deleted element has been highlighted in yellow colour and N= 30 and 5 is underlined in figure 3-6 a.

N represents – nucleotides

Table 3-3 The *cis* elements present in *PP2C-* like promoter and in its deletion variants viz. 900 bp, 500 bp, 400 bp and NRM.

Cis element	Signal sequence	Copy number				
		1kb	900 bp	500 bp	400 bp	NRM
-10PEHVPSBD	TATTCT	1	N	N	N	N
-300ELEMENT	TGHAAARK	1	1	1	N	N
2SSEEDPROTBANAPA	CAAACAC	1	1	1	1	1
AACACOREOSGLUB1	AACAAAC	2	2	N	N	N
ABRELATERD1	**ACGT**G	3	3	3	3	2
ABRERATCAL	M**ACGY**GB	2	2	2	2	2
ACGTATERD1	**ACGT**	6	6	4	4	3
ANAERO1CONSENSUS	AAACAAA	5	5	N	N	N
ANAERO3CONSENSUS	TCATCAC	4	4	N	N	N
ARFAT	TGTCTC	1	1	1	1	1
ARR1AT	NGATT	15	15	9	8	7
ASF1MOTIFCAMV	TG**ACG**	4	4	1	1	1
AUXRETGA1GMGH3	TG**ACGT**AA	2	2	N	N	N
BIHD1OS	TGTCA	3	2	1	N	N
BOXIINTPATPB	ATAGAA	2	1	N	N	N
CAATBOX1	CAAT	11	9	3	3	2
CACGTGMOTIF	C**ACGT**G	1	1	1	1	1
CACTFTPPCA1	YACT	18	16	8	8	5
CANBNNAPA	CNAACAC	2	2	2	2	1
CATATGGMSAUR	CATATG	2	2	N	N	N
CCA1ATLHCB1	AAMAATCT	2	2	2	1	1
CCAATBOX1	CCAAT	2	2	1	1	N
CEREGLUBOX2PSLEGA	TGAAAACT	1	1	N	N	N
CGACGOSAMY3	CGACG	1	1	1	1	1
CPBCSPOR	TATTAG	1	1	1	1	1
CURECORECR	GTAC	1	1	1	1	1
DOFCOREZM	AAAG	15	15	9	6	4
DPBFCOREDCDC3	ACACNNG	2	2	2	2	2
DRE1COREZMRAB17	ACCGAGA	1	1	1	1	1
EBOXBNNAPA	CANNTG	5	4	2	2	2
EECCRCAH1	GANTTNC	1	1	N	N	N
ELRECOREPCRP1	TTGACC	2	1	1	1	1
GATABOX	GATA	8	7	4	4	4
GT1CONSENSUS	GRWAAW	12	9	5	1	1
GT1CORE	GGTTAA	1	1	N	N	N
GT1GMSCAM4	GAAAAA	5	3	2	1	1
GTGANTG10	GTGA	16	13	5	4	4
HEXMOTIFTAH3H4	**ACGT**CA	3	3	1	1	1
IBOXCORE	GATAA	2	1	N	N	N
INRNTPSADB	YTCANTYY	3	3	3	2	2
IRO2OS	C**ACGT**GG	1	1	1	1	1

LTRE1HVBLT49	CCGAAA	1	1	1	1	1
MARTBOX	TTWTWTTWTT	4	4	4	N	N
MYB2CONSENSUSAT	YAACKG	1	1	1	1	1
MYBCORE	CNGTTR	1	1	1	1	1
MYBCOREATCYCB1	AACGG	2	2	2	2	2
MYBPLANT	MACCWAMC	1	1	N	N	N
MYBPZM	CCWACC	1	1	N	N	N
MYCCONSENSUSAT	CANNTG	5	4	2	2	2
NODCON1GM	AAAGAT	2	2	1	1	1
NODCON2GM	CTCTT	6	5	3	2	2
NTBBF1ARROLB	ACTTTA	2	2	1	N	N
OSE1ROOTNODULE	AAAGAT	2	2	1	1	1
OSE2ROOTNODULE	CTCTT	6	5	3	2	2
POLASIG1	AATAAA	2	2	1	1	1
POLASIG2	AATTAAA	1	1	N	N	N
POLASIG3	AATAAT	2	1	N	N	N
POLLEN1LELAT52	AGAAA	5	3	N	N	N
PRECONSCRHSP70A	SCGAYNR$(N)_{15}$ HD	2	1	N	N	N
RAV1AAT	CAACA	3	3	2	1	1
REALPHALGLHCB21	AACCAA	1	1	N	N	N
RGATAOS	CAGAAGATA	1	1	1	1	1
RHERPATEXPA7	KCACGW	1	1	1	1	N
ROOTMOTIFTAPOX1	ATATT	7	5	3	3	3
S1FBOXSORPS1L21	ATGGTA	2	2	1	N	N
SEBFCONSSTPR10A	YTGTCWC	1	1	N	N	N
SEF1MOTIF	ATATTTAWW	1	1	1	1	1
SEF4MOTIFGM7S	RTTTTTR	1	1	1	N	N
SP8BFIBSP8BIB	TACTATT	1	1	N	N	N
SURECOREATSULTR11	GAGAC	2	2	2	2	2
TAAAGSTKST1	TAAAG	4	4	3	3	2
TATABOX1	CTATAAATAC	1	1	N	N	N
TATABOX2	TATAAAT	2	2	1	1	1
TATABOX5	TTATTT	3	3	1	1	1
TBOXATGAPB	ACTTTG	1	1	1	1	1
TGACGTVMAMY	TG**ACGT**	3	3	1	1	1
WBOXATNPR1	TTGAC	2	1	1	1	1
WBOXNTERF3	TGACY	3	2	2	2	2
WRKY71OS	TGAC	10	8	4	3	3

Not present is denoted as N

present. Only two homologue sequences of ABRERATCAL (MACGYGB) element, a calcium responsive *cis* element were identified in full-length *PP2C* like promoter. The *PP2C* –like promoter was characterized by making a series of 5' deletion constructs viz. 900 bp, 500 bp, 400 bp and NRM. The *cis* elements were analyzed in *PP2C* -like promoter and in its deletion constructs viz. 900 bp, 500 bp, 400 bp and NRM by using PLACE and PlantCARE data base (Table 3-3) (Higo et al., 1999; Lescot et al., 2002). In

our study, inducers like ABA, JA and SA were used to illustrate the promoter activity in the presence and absence of these elicitors.

3.7 Transient expression studies of *PP2C* -like promoter and its deletion constructs

The full length *PP2C*- like promoter and its 5' series of deletion constructs 900 bp, 500 bp, 400 bp and NRM were amplified from *Arabidopsis thaliana* genome and cloned into pEGFP vector. These promoter-reporter constructs were bombarded on to the leaves of *Nicotiana tabacum* cv. Xanthi. The promoter reporter constructs bombarded leaves were incubated in Hoagland medium for 48 hours. The *PP2C*- like promoter activity was assessed by measuring the expression of EGFP reporter gene present downstream to *PP2C*- like promoter and its deletion variants - 900 bp, 500 bp, 400 bp and NRM. The non-recombinant pEGFP vector was used as a control in this experiment. The expression of EGFP in the control was low in un-induced and induced conditions, in comparison to the other promoter reporter constructs of *PP2C*- like promoter. The expression of the EGFP, observed in case of the control was due to the presence of *lacZ* promoter (Figure 3-7). The expression of reporter gene is higher in the full length *PP2C*- like promoter as well as in deletion variants of *PP2C* like promoter as compared to the control i.e. pEGFP vector (Figure 3-7). As mentioned previously, full length of *PP2C*- like promoter is having 79 *cis* elements and hence, it was expected to have more expression compared to other deletion constructs but results were not as expected, and several deviations from the expected results were observed. The difference in expression of reporter gene was observed due to the deletion of some of the *cis* elements (for example -10 PEHVPSBD in 900 bp, Table 3-3) while in others, the variation in copy number could be the reason for difference in expression (like for ABRE) table 3-3. The activity of *PP2C*- like promoter was high in 400 bp and NRM construct while comparatively low expression of EGFP was observed in full length, 900 bp and 500 bp constructs of *PP2C*- like promoter, as shown in figure 3-7. This is due to the presence of some negatively regulating *cis* elements in the full length, 900 bp and 500 bp constructs which were responsible for the low expression of reporter gene. When the copy numbers of these elements were reduced in 400 bp and NRM constructs, the higher expression of reporter gene was observed.

3.8 AACA element negatively regulates the gene expression of *PP2C*- like promoter

Under un-induced conditions the expression of EGFP is higher in 400 bp and NRM whereas in full length, 900 bp and 500 bp constructs of *PP2C*- like promoter the

expression of EGFP was low. The disparity in expression of the promoter reporter construct could be due to the presence of AACA elements in the full length, 900 and 500 bp constructs. AACA element is a tissue specific *cis* element and is known to express only in the endosperm (Wu et al., 2000). Wu et al., (2000) observed that deletion of AACA element leads to decrease in the seed storage *Glu b-1* promoter activity. The presence of AACA element in promoter of genes in any other tissue, acts as negative *cis* element for other elements (Yoshihara et al., 1996). The AACA element has positive effect in endosperm. The AACA element being present in higher copy number, in full length, 900 bp and 500 bp could be responsible for reduced expression of EGFP reporter gene. This negative effect of the *cis* element AACA is observed in *PP2C*- like promoter. The occurrence of AACA element is more in full length *PP2C* like promoter, deletion promoter – reporter constructs of *PP2C* like promoter viz 900 bp and 500 bp. AACA element is present in *PP2C*- like promoter in 9 copies. The copy number falls down to one in case of deletion constructs like 400 bp and NRM construct and hence it shows high activity in terms of expression of the reporter gene (the position of the AACA element is mentioned in Table 3-4).

Table 3-4 The copy number and the position of the AACA motif in context to TATA box

Constructs	**No. of copies of AACA element**	**Position upstream to TATA box**
Full length *PP2C* like promoter	9	-100,-322,-353,-432,-504,-533,-635,-675.-730
900 base pair	9	-100,-322,-353,-432,-504,-533,-635,-675.-730
500 base pair	2	-100,-322
400 base pair	1	-100
NRM	1	-100

3.9 Expression of *PP2C*- like promoter in the presence of inducers

The induction was observed in each case of *PP2C*- like promoter constructs in comparison to the control incubated in Hoagland media supplemented with ABA, JA and SA, separately. The control pEGFP vector does not show any induction response in the presence of ABA, JA and SA, individually as shown in figure 3-8.

The elements which are responsible for response towards ABA have ACGT as a core sequence. These elements are like ACGTERD1 and ABRELATERD1 (as mentioned in Table 3-3). The six copies of ACGTATERD1 elements are present in full length and 900 bp of *PP2C*- like promoter while four copies of ACGTATERD1 motif are present in both 500 and 400 bp constructs. The ACGT elements in close vicinity lead to increase in the expression of the reporter gene. The difference in the expression of promoter reporter constructs observed in the presence of ABA is due to the presence of AACA, which act as negative element in full length, 900 bp and 500 bp constructs (Figure 3-8). The AACA element causes decrease in overall expression of the reporter gene in constructs like full length, 900 bp and 500 bp deletion construct of *PP2C* like promoter. The maximum expression was observed in 400 bp construct of *PP2C* like promoter.

The expression of NRM construct is lower than 500 bp and 400 bp constructs even though AACA copy number falls down in both the constructs. The NRM construct showed expression closer to full length and 900 bp constructs (Figure 3-8). The genetic architecture of NRM construct has been disturbed by removing one ACGT motif out of three ACGT motifs present in close proximity (Figure 3-6 a and b). The central ACGT elemnt is separated by 30 nucleotides from ACGT element towards 5' side and is separated by five nucleotides towards 3' side by another ACGT element (shown above in Figure 3-6 a). In NRM construct, it retains only the arrangement where two ACGT elements are separated by 5 bp (Figure 3-6 a). The induction of NRM deletion construct was observed to be comparatively lower than 400bp construct in the presence of ABA (Figure 3-8). Hence this shows that two ACGT elements separated by 30 bp, are responsible for induction by ABA.

In the induction studies, in the presence of JA, the highest level of expression is observed in the case of 500 bp cassette and very low expression observed in 400 bp cassette (Figure 3-8). The full length, 900 bp and NRM constructs of *PP2C* like promoter have not shown similar levels of EGFP expression. The deletion of 100 bp from 500 bp construct led to deletion of *cis* elements and decrease in the copy number of the CAAT box. The CAAT box region is also necessary for the maximum response to JA. The

wounding caused by the bombardment also effects the expression of EGFP (Kim et al., 1993).

In the presence of SA the full length, 500 bp, 400 bp and NRM constructs (Figure 3-8) showed more expression than 900 bp construct. Minimal elements required for the induction by SA is present in full length, 900 bp and 500 bp. The 400 bp and NRM constructs have shown more expression, which have 1 copy of AACA element. The 400 bp and NRM has the basic structure i.e. two ACGT element separated by 5 bp. This leads to the increase in the expression of EGFP. The W box (TTGAC/G) is present in 2 copy number in full length promoter-reporter construct and the copy number of W box decrease to 1 in rest of the constructs this also effects the expression of EGFP. Whereas the copy number of AACA also decrease so in the presence of SA 400 bp and NRM promoter-reporter construct have shown maximum expression.

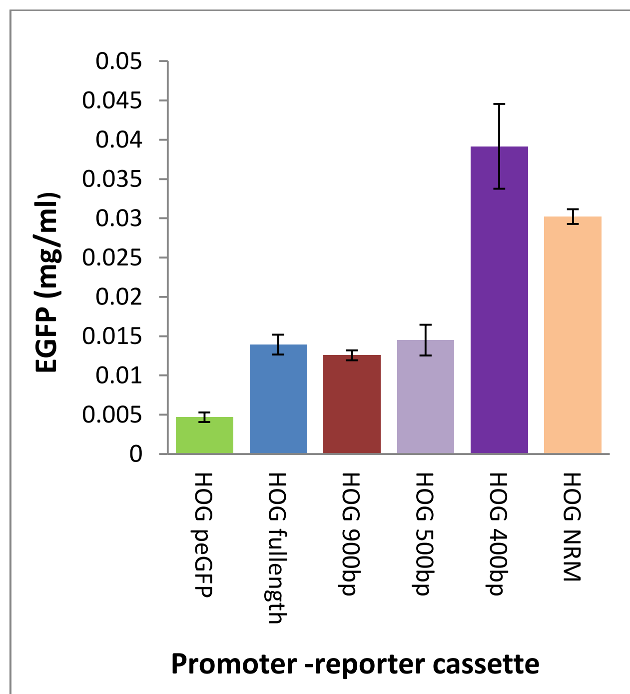


Figure 3-7 Expression of the *PP2C* like promoter under un-induced conditions (with different constructs). The vertical bar indicates standard deviation. (HOG-Hoagland media)

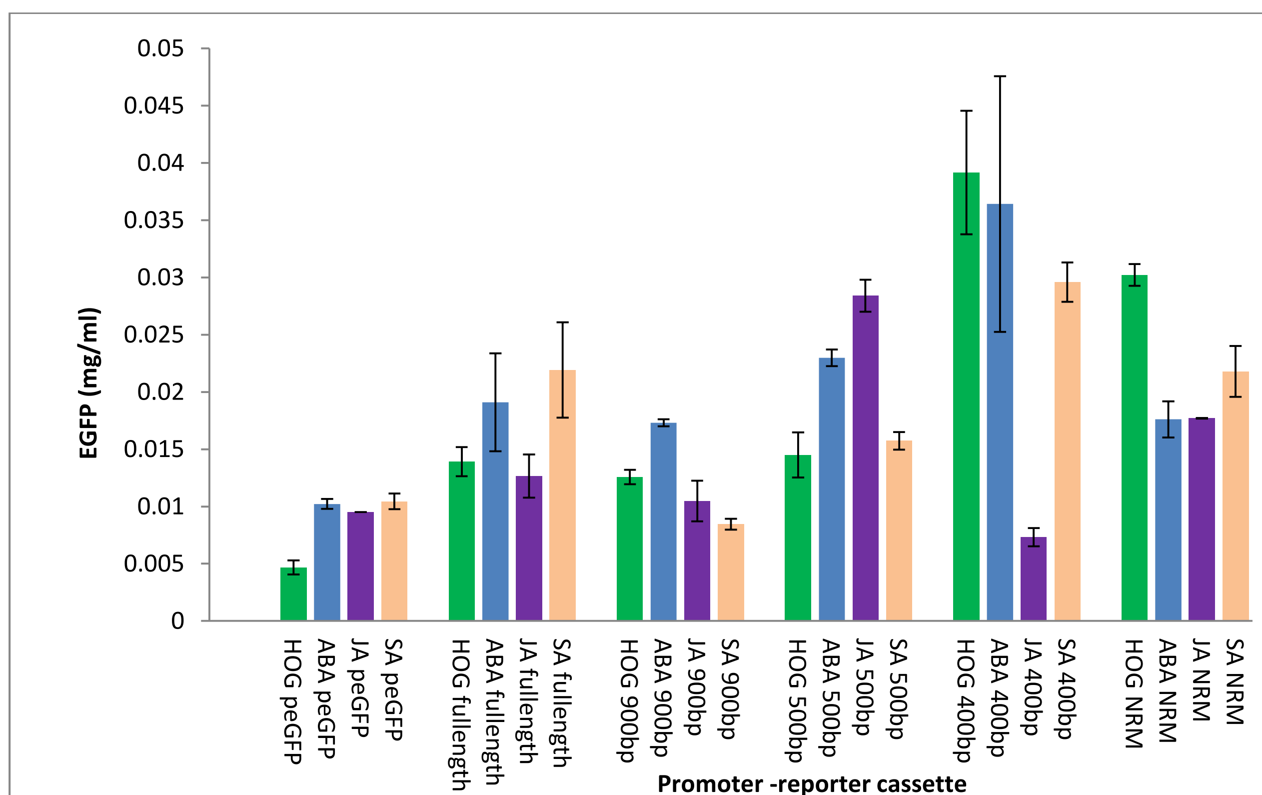


Figure 3-8 Expression of the *PP2C* like promoter (with different constructs) under un-induced and induced conditions (green – Hoagland media (H.M), blue – H.M+ABA, purple– H.M+JA, orange– H.M+SA). The vertical bar indicates standard deviation

3.10 Generation of transgenic *Arabidopsis thaliana* plants harboring *PP2C*-like promoter

The full-length and deletion fragments of *PP2C*- like promoter were amplified from *Arabidopsis* genomic DNA using the *PP2C*-like promoter specific primer sets (Table 2-4). The amplified fragments were first cloned into pGEMT easy vector. Five different promoter fragments were released by *XabI* and *BamHI* digestion (Figure 3-9). Further these five fragments were cloned in pBI101 (promoter less vector) in the same restriction sites.

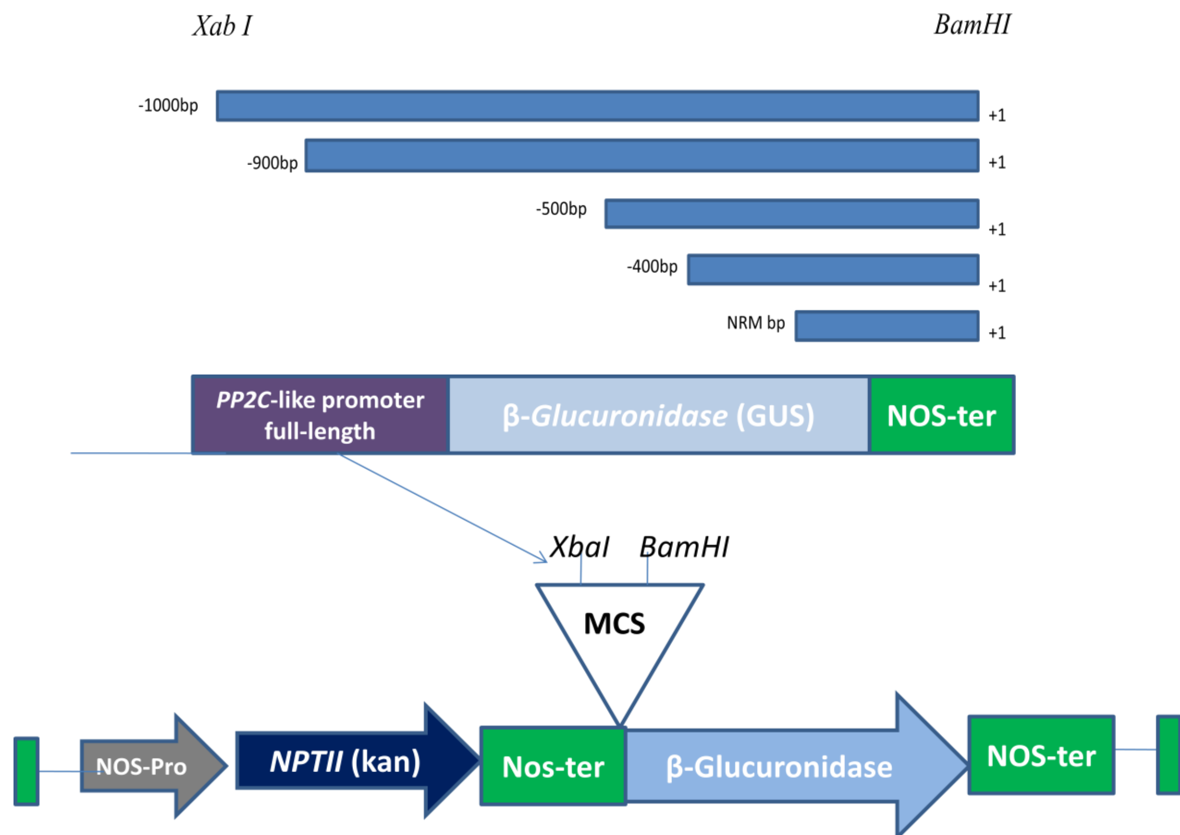


Figure 3-9 A diagrammatic representation of full-length construct in pBI101

Five promoter-reporter constructs were transformed in *Agrobacterium*. The *Arabidopsis* plants, were transformed through floral dip method to obtain transgenic plants. Transgenic *Arabidopsis* plant lines containing full-length promoter, 900 bp, 500 bp, 400 bp and NRM deletions were obtained. The GUS activity in transformed *Arabidopsis* plants were assayed by a fluorometric quantification method and measured in nmol 4-methylumbelliferone (4MU)/mg/min proteins. The expression of GUS gene was analysed through RT-PCR. The total RNA was isolated from both treated and untreated transformed *Arabidopsis* plants leaves.

3.11 Fluorometric analysis of GUS activity under un-induced condition

Fluorometric assay was performed under normal conditions, to estimate the GUS activity in the transgenic *Arabidopsis* lines harboring full-length and the different deletion promoter-reporter constructs of *PP2C*-like promoter. Results obtained have been shown in figure 3-10. The GUS activity was high in the case of full length promoter and 900 bp deletion promoter constructs as compared to other three deletion promoter constructs under normal conditions. Results of the GUS assay showed that deletion of -500 bp and -640 bp regions from full length *PP2C*-like promoter, considerably affected the quantitative behavior of *PP2C*-like promoter (Figure 3-10).

3.12 ABA-induced expression analysis of GUS gene and GUS activity of promoter(s) in *Arabidopsis* plants

Five promoters – reporter cassettes of *PP2C*-like promoter viz full length, 900 bp, 500 bp, 400 bp and NRM were differently expressed upon exogenous application of ABA (100µM). The RNA was isolated from the leaves of 24 and 48 hour treated plants with ABA. Fluorometric analysis of GUS activity was done on the plants treated with ABA (100µM) for 12, 24, 36 and 48 hours, respectively. In leaves, the full-length and 900 bp deletion constructs of *PP2C*-like promoter responded significantly to ABA, whereas other deletion constructs of *PP2C*-like promoter showed relatively very less GUS gene expression (Figure 3-11 A and B). The 900 bp deletion construct has shown enhanced *ß-glucuronidase* gene expression than full length promoter segment after 48 hour treatment. Rest other three deletion construct responded very weakly, i.e. almost no expression was

observed to ABA after 24 and 48 hours. The decrease in copy number of CAAT elements could be one of the possible reasons for decrease in expression of GUS gene observed in the three deletions constructs of *PP2C*-like promoter, which is required for efficient promoter activity

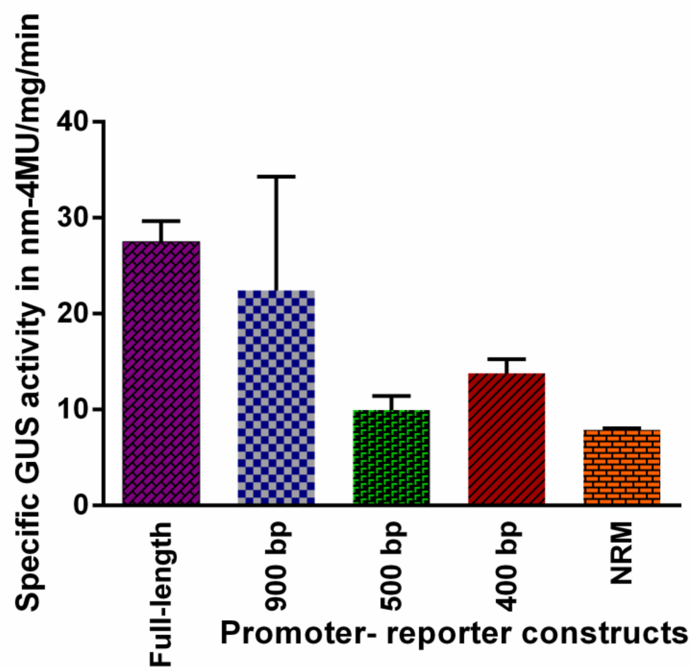


Figure 3-10 Flurometric assay of GUS activity in transgenic lines of *Arabidopsis* plants harboring full length and deletion constructs viz 900 bp, 500 bp, 400 bp and NRM of *PP2C*-like promoter under normal conditions. The vertical bar indicates standard deviation

Moreover, the deletion constructs also have shown a decrease in copy number of enhancer-like *cis*-elements i.e. A/T rich region. In the *β-phaseolin* gene promoter, it has been observed that the A/T rich sequences possibly act as general enhancers of expression in the developing embryo (Bustos et al., 1989) and light-regulated photosynthesis gene because of its positive role in gene expression (Lam et al., 1990). For initiation of transcription, CAAT (enhancer *cis* element) and TATA sites were used for synthesizing inducible promoters (Roy Choudhury and Sen Gupta 2009). The activity of *OsAct2* (actin 2 promoter from rice) could be defined by the presence of a negative regulator (+96 and +274 within the intron) which represses *OsAct2* expression. Conversely, positive regulators i.e. the two CAAT-boxes found to be present (between −448 and −445 and positions −635 and −632) which enhances *OsAct2* expression (de los Reyes BG et al., 2015).

GUS protein activity was assayed and the results are shown in figure 3-11 C, D, E, F. At 12th hour, the full-length promoter construct was induced in response to ABA (Figure 3-11 C). Whereas, an increase in GUS activity was observed for 900 bp promoter-reporter construct after 24 hours, with no change in the GUS activity for full length promoter-reporter construct (Figure 3-11 D). After 36 hours of treatment with ABA, higher GUS activity for full length promoter and 900 bp promoter deletions constructs was found (Figure 3-11 E). However, any change in the GUS activity of the other three promoter- reporter cassettes, i.e. 500 bp, 400 bp and NRM; was observed only at 12 hour and were found unchanged at 24, 36 and 48 hours (Figure 3-11 C, D, E, F). It was remarkably notable that the 900 bp construct responded positively to ABA amongst all deletion promoter cassettes. After 48 hour of treatment the full length promoter and 900 bp deletion cassettes expressed a decrease in GUS activity (Figure 3-11 F). The rest three deletion cassettes showed negligible GUS activity. From this, it can be inferred that the full length and 900 bp deletion construct of *PP2C*-like promoter is an ABA inducible promoter. This implied that an ABR complex may form appropriately in full length and 900 bp deletions construct. This also demonstrates that the deleted or

decreases in copy number of elements are required for gene expression as well as for quantitative behavior of GUS protein (Table 3-1).

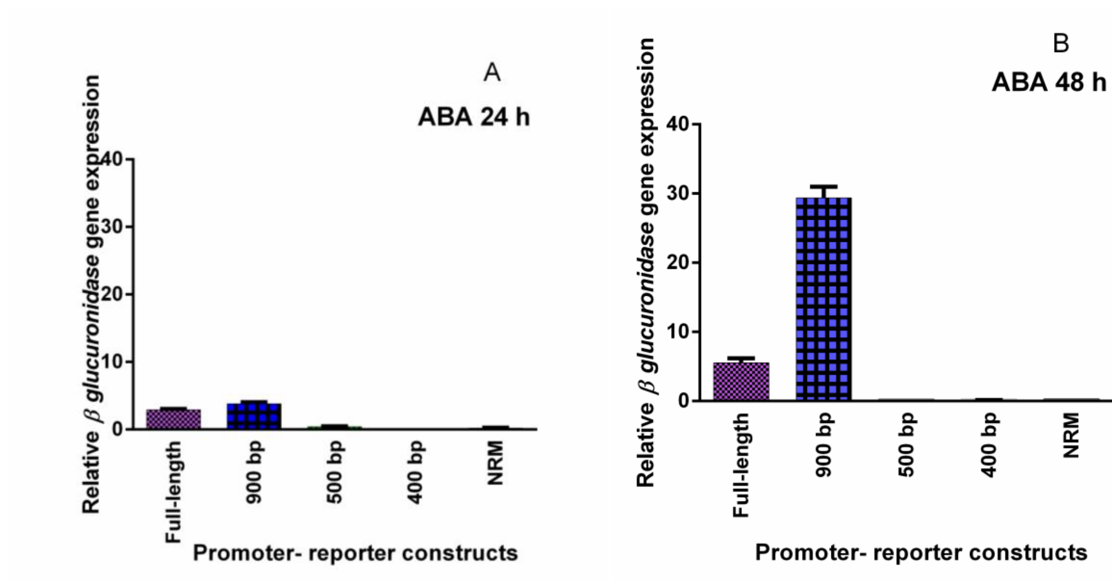


Figure 3-11 A and B Relative GUS gene expression in response to exogenous ABA - full length promoter, 900 bp, 500 bp, 400 bp and NRM deletion constructs of *PP2C*-like promoter. The vertical bar indicates standard deviation

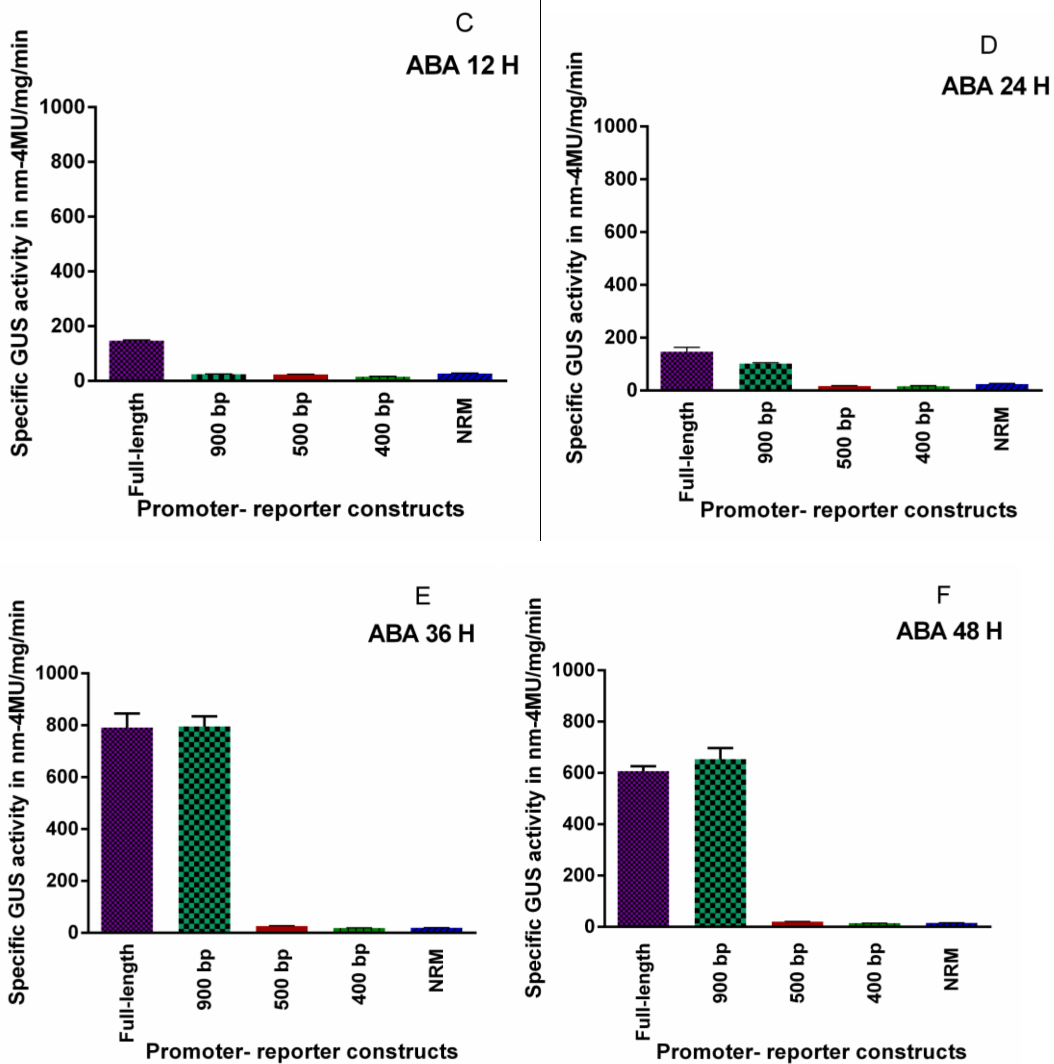


Figure 3-11 C, D, E and F Fluorometric assay of GUS activity in transgenic lines of *Arabidopsis* plants harboring full length *PP2C*-like promoter construct and *PP2C*-like promoter deletion constructs viz 900 bp, 500 bp, 400 bp and NRM under ABA treatment. The vertical bar indicates standard deviation

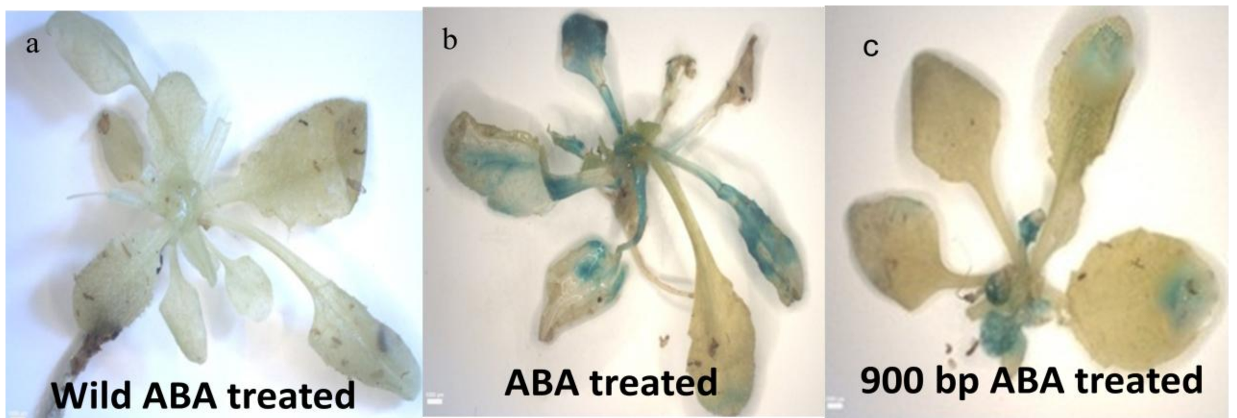


Figure 3-12 Histochemical analysis (a) Wild type ABA treated, (b) Full length ABA treated, (c) 900 bp ABA treated

It can be observed in histochemical staining results as well. The GUS stain was visible in ABA treated leaves in case of full length and 900 bp deletion constructs of *PP2C*-like promoter (Figure 3-12 b and c), whereas GUS stain was not visible for other three deletion constructs. No GUS activity was detected in wild-type *Arabidopsis* plants (Figure 3-12 a). In the case of the longest deletion construct NRM, ACGT $_{N30}$ ACGT was deleted (Figure 3-6 a). It could be possible that the loss in expression was due to decrease in copy number of ACGT elements and their coupling elements. This indicates that possibly appropriate ABR complex was not forming in the smaller deletion constructs like -500 bp, -400 bp and NRM. The 500 bp and 400 bp promoter-reporter constructs have ACGT elements, and then also it did not respond to ABA in case of leaves, possibly because of the reported decrease in the number of *cis*-regulatory elements in this region. A study by Hong and Hawang (2009) found that -593 bp deletion construct to be insensitive to ABA treatment, although this construct had ABA responsive bZIP and MYB binding sites. The results suggest that *PP2C*-like promoter deletion of −100 bp i.e. 900 bp promoter-reporter cassettes was sufficient to drive GUS gene expression.

According to PLACE and PlantCARE databases, ABA responsive element (ABRE), drought responsive elements (DRE) or MYC and MYB recognition sites are present in the *PP2C*-like promoter sequence (Higo et al., 1999; Lescot et al., 2002). A single copy of DRE1COREZMRAB17 (ACCGAGA) is present in the *PP2C*-like promoter and in all its deletion cassettes (Table 3-3). ABRE-ABRE or ABRE-CE pair is required for ABA response. Some *cis* elements occur within a specific distance from one another, forming dyad elements such as is in the case of ABRE elements (Gomez porraz et al., 2007). The GUS activity falls due to the deletion from −1.9 kb or to −1.0 kb the promoter length of *gbss1* (granule-bound starch synthase 1from wheat). It is proposed that the enhancer elements and *cis*-acting elements located in −1.9 kb region of the promoter were deleted, which responsible for *gbss1* transcription could (Kluth et al., 2002). *Cis* regulatory elements are involved in a variety of regulatory networks, affecting each other's role and their respective position varies in different promoters (Wray 2003).

Protein-protein interactions also play a role in affecting the gene expression. For instance, SCOF-1 protein (soybean cold inducible factor-1) interacts with SGBF-1 (soybean G-box binding bZIP transcription factor) in response to cold stress (Kim et al., 2001).

Variation in the intensity of GUS gene expression was noticed in different transgenic lines tested. Moon and Callahan (2004) have observed variations in the intensity of the GUS among individual transgenic plants. The levels of the transgene GUS RNA was found to be very low in the transgenic plants carrying shortest promoter constructs of *ACO-GUS* whereas a slight increase was detected in the -610 promoter-reporter construct bearing transgenic line. The transgenic plants with the three (-2919 to −2141, -1319 to -901 and the first 403 bp) longer promoter-reporter constructs have shown a significant increase in the amount of mRNA (Moon and Callahan 2004). Some of the possible reasons for this difference could be attributed to different transgene integration and variation in transgene copy number.

3.13 MeJA -induced expression analysis of GUS gene and GUS activity of promoter(s) in *Arabidopsis* plants

MeJA is a plant hormone which regulates tendril (root) coiling, flowering, seed and fruit maturation and biotic stress. Any change in the MeJA levels affects the flowering time, flower morphology and the number of open flowers. MeJA induces ethylene-forming enzyme activity, which increases the amount of ethylene required for fruit maturation. In addition, increased levels of MeJA also inhibit plant root growth. The *ß-glucuronidase* gene expression level was found to be high in full length promoter and 900 bp deletion constructs in leaves in response to MeJA at 24th hours of treatment (Figure 3-13 A and B). After 48th hour of treatment, *ß-glucuronidase* gene expression decreases in full length and 900 bp deletion constructs (Figure 3-13 A and B). Other deletion cassettes, 500 bp, 400 bp and NRM resulted in low *ß-glucuronidase* gene expression in comparison to the full length and 900 bp construct at 24th and 48th hour treatment. This might be due to the removal of some transcription factor binding sites, which may be responsible for MeJA induced gene expression (Figure 3-13 A and B). For -500 bp and -400 bp promoter-reporter cassettes, no change was observed after 24th and 48th hour. Although, after 48th hour treatment, NRM deletion construct showed increase in the GUS activity, this could be due to the induction of late responsive nature of this promoter-reporter constructs.

From the fluorometric analysis, high GUS activity in response to MeJA was observed in leaves from 900 bp promoter constructs after 12th hour treatment (Figure 3-13 C). Further decrease in the GUS activity was observed for all the promoter-reporter constructs of *PP2C*-like promoter, which could be due to the saturation of GUS activity after 24th, 36th and 48th h treatment with MeJA (Figure 3-13 D, E and F).

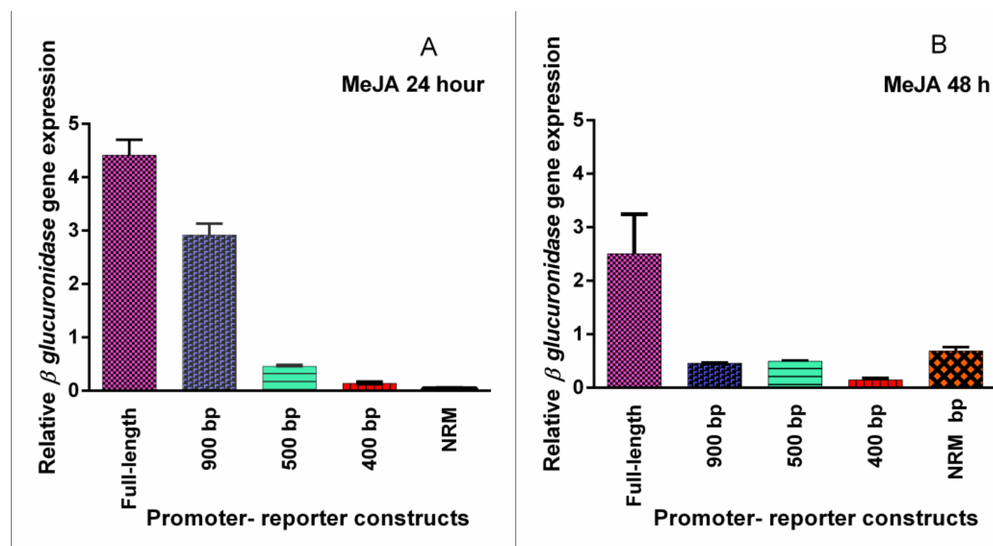


Figure 3-13 A and B Relative GUS gene expression of five *PP2C*-like promoter constructs full promoter, 900 bp deletion, 500 bp, 400 bp and NRM deletion constructs, 24th and 48th hour after MeJA treatment. The vertical bar indicates standard deviation

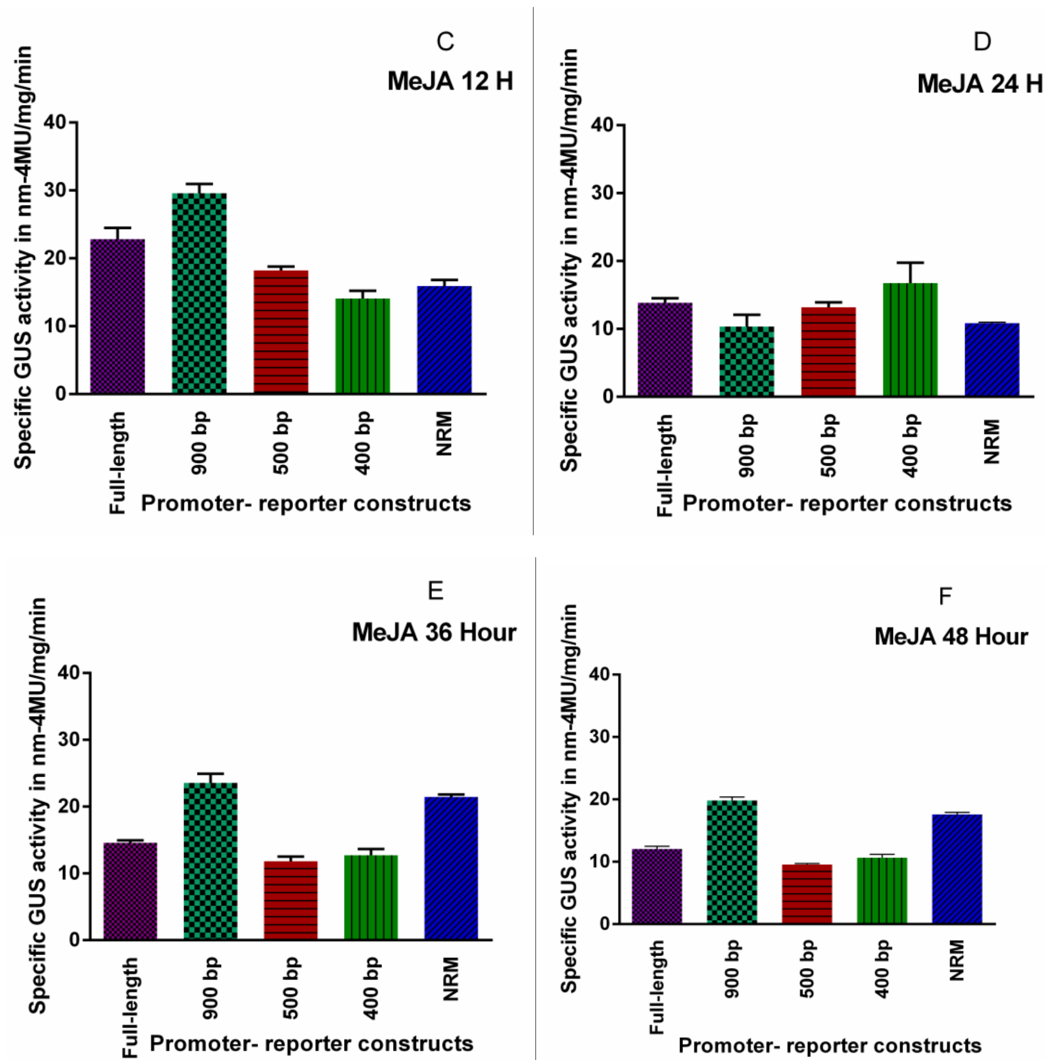


Figure 3-13 C, D, E and F Fluorometric assay of GUS activity in transgenic lines of *Arabidopsis* plants harboring full length *PP2C*-like promoter and *PP2C*-like promoter deletion constructs viz 900 bp, 500 bp, 400 bp and NRM under MeJA treatment. The vertical bar indicates standard deviation

The variation in activity may be due to the decrease in enhancer *cis*-acting elements viz CAAT elements, A/T rich region. In another study, it was observed that, −1210 to −886 bp was found to be sufficient for MeJA-induced GUS activity in *OsPMCa*$^{2+}$*ATPase* promoter, whereas the −519 bp deletion from *OsPMCa*$^{2+}$*ATPase* promoter showed decreases MeJA-responsive promoter activity in leaves (Kamrul-Huda et al., 2013). The GCC-box- jasmonic acid-responsive element was not observed in the *PP2C*-like promoter region. This is also one of the reasons for weak induction shown by the *PP2C*-like promoter's constructs in response to MeJA.

3.14 SA-induced expression analysis of GUS gene and GUS activity of promoter(s) in *Arabidopsis* plants

Low level of *β-glucuronidase* gene was observed in leaves in response to SA with all the promoter constructs after 24 hours of treatment. The 900 bp promoter constructs induced at 48th hour gave high GUS expression in leaves, while in case of other constructs slight increase in gene expression was observed (Figure 3-14 A and B).

Fluorometric analysis showed increase in GUS activity for full length promoter construct after 12 hour of treatment (Figure 3-14 C). For 24 h, 36 h and 48 h, no change in GUS activity was observed (Figure 3-14 D, E and F). In comparison to un-induced condition, there was no change in GUS activity was observed (Figure 3-10). This indicates that *PP2C*-like promoter was not inducible in response to SA. Decrease in activity was observed when two ACGT elements separated by N5 and placed 100 nucleotides away from *Pmec* (Mehrotra et al., 2005). This was observed in all the cassettes of *PP2C*-like promoter. A TTGAC element was observed within the *OsPMCa*$^{2+}$*ATPase* promoter region located at −1261 bp still no induction was observed in response to SA (Karmula-Huada et al., 2013). In our case, the promoter constructs containing these *cis* elements were activated by SA treatment. In the deletion constructs, W-box and TGACG copy number decreases, resulting in reduced activity. *Cis* elements essential for activation in response to ABA, JA and SA, are detected mostly

in full length and 900 bp constructs. The change in mRNA and protein levels do not correlate well mainly due to gene expression regulation at different levels, viz transcriptional, post-transcriptional, translational and post-translational level. Hence, it can be proposed that due to multiple levels of gene regulation in some subsets, there is a weak correlation between mRNA and protein levels (Walling et al., 1986).

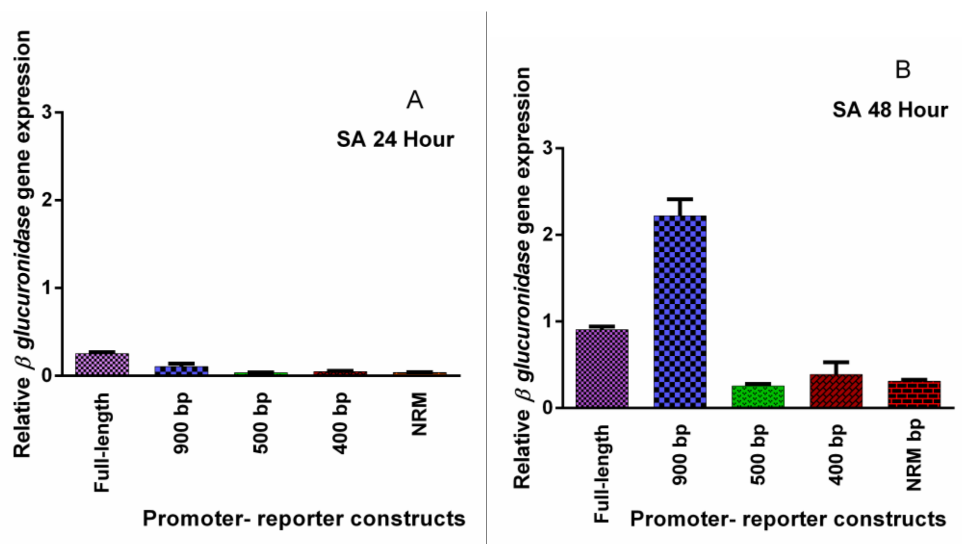


Figure 3-14 A & B Relative GUS gene expression of five *PP2C*-like promoter constructs including full promoter, 900 bp deletion, 500 bp, 400 bp and NRM deletion constructs, at 24th and 48th hour after SA treatment. The vertical bar indicates standard deviation

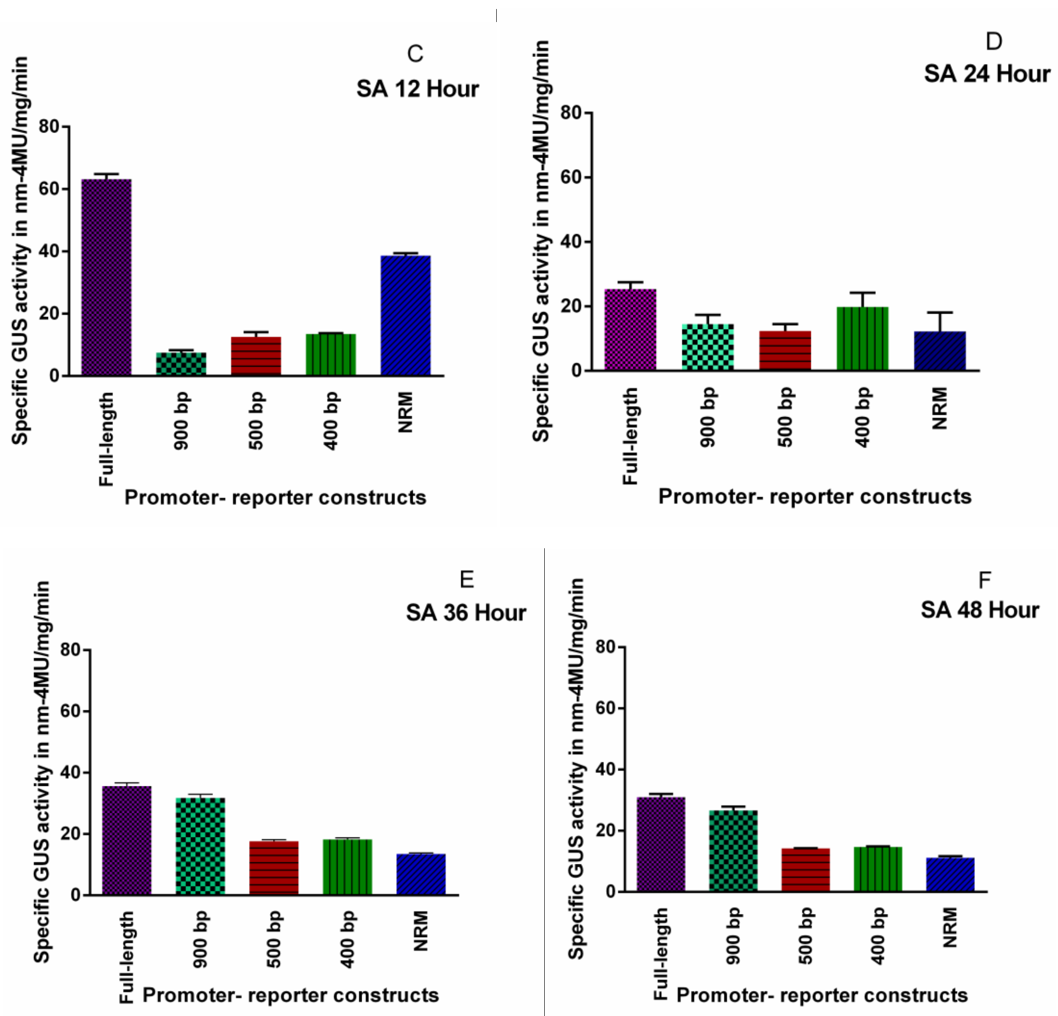


Figure 3-14 C, D, E and F Fluorometric assay of GUS activity in transgenic lines of *Arabidopsis* plants harboring full length *PP2C*-like promoter and *PP2C*-like promoter deletion constructs viz 900 bp, 500 bp, 400 bp and NRM under SA treatment. The vertical bar indicates standard deviation

SUMMARY

CHAPTER 4

4.0 Summary

The interaction between *cis* elements and cognate transcription factors is responsible for specific and differential transcriptional activity. Availability of high-throughput expression profiles and genome-wide analysis datasets along with various bioinformatics tools have made it possible to study the promoter region of co-expressed genes and identify the role of *cis* elements. ABRE elements are found to be over-represented in promoter region of ABA or abiotic stress responsive genes in *Arabidopsis thaliana*. Zhang et al., (2005) have used *cis* regulatory element based targeted gene finding method to identify ABA or cold, salinity and drought responsive genes in *Arabidopsis thaliana* genome. The over-representation of G-box have been observed in the promoter region of light responsive genes, ABA responsive genes, abiotic stress responsive genes, in intra-genomic conserved non-coding sequence, circardian and diurnal regulation of gene expression in *Arabidopsis* (Giuliano et al., 1988; Freeling et al., 2007; Freeling and Subramaniam 2009; Mockler et al., 2007; Michael et al., 2007; Priest HD et al., 2009).

The information available about *cis* elements can be useful in promoter engineering. CaMV35S is preferentially used to drive constitutive expression (Benfey and Chua, 1990), in both - stable and transient transformation. However, over-expression of gene under CaMV35S regulation is showing co-suppression or gene silencing of transgene (Elmayan and Vaucheret 1996; Mishiba et al., 2005) and long-term developmental effects of expressing the transgene in plants. For say those transgenic plants which are constitutively over-expressing DREB/CBFs genes have shown some undesirable developmental features in plants, such as stunted growth and delay in flowering (Morran et al., 2011) whereas, the expression of DREB factors under the

regulation of drought inducible promoter have shown normal development process with improved drought tolerance in wheat and barley transgenics.

Plant genetic engineering is facing problems such as expression of multiple trans-gene, inducibility and homology dependent gene silencing. The homology dependent gene silencing is caused due to a homology between transgene or promoter with endogenous gene or promoter, repeated use of same promoter also causes gene silencing and use of two homologous transgene. To work out with these problems there is a need to design or identify promoters to achieve inducibility, tissue-specificity and high expression. To design a promoter, study and optimization of *cis* elements is required. The promoter strength is based on the arrangement of *cis* elements, copy number of *cis* -regulatory elements and different combination of *cis* elements affects the promoter activity (Rushton et al., 2002). A -450 bp synthetic promoter has been designed known as *Pcec* (complete expression cassette) made up of -138 bp minimal expressions cassette- *Pmec* and -312 bp transcription activation module (Sawant et al., 2001).

Uni or bi directional synthetic promoters can be synthesized with less homology with the endogenous genes by using two strategies as explained by Venter and Botha (2010). A core promoter is fused to a stretch of a specific *cis* element inducible for a stimulus or fused to stretch of multiple repeat of a specific *cis* element. Bidirectional promoter is having additional core promoter fused on the other side of *cis* elements and allows the transcription of two genes. Bidirectional promoter synthesis also solves gene stacking problem in transgenic plants. The availability whole genome sequencing of Arabidopsis and rice has enabled the identification of large sets of promoter sequences. Although many wild type promoters have been identified, they are not able to achieve all the challenges. Further strategies involving genetic transformation with multiple transgenes expressed in an inducible manner could help to improve stress tolerance (Chaturvedi et al., 2006).

Different promoters have been designed from plant, virus and bacterial origin to develop transgenic plants (Venter and Botha 2010). The constitutive expression of multiple genes leads to severe effect on the transgenic plants. In support to this it has been observed that the constitutive expression of *OsbZIP71* by CaMV35S causes delay in the flowering time (Liu et al., 2014). Further in the case of DREB/CBF genes under the regulation of CaMV35S promoter have shown improved tolerance to salt, drought and cold stress along with multiple growth and developmental abnormalities (Jaglo-Ottosen 1998). This problem was solved by the use of inducible promoter. Liu et al.,

(2014) have used *rd29A* gene promoter from Arabidopsis which is an abiotic inducible promoter instead of CaMV35S promoter to regulate *OsbZIP71*. This result implies that *PP2C*- like promoter can be used as inducible promoter. Enhanced drought resistance was observed due to the constitutive expression of *LeNCED1* gene under the control of *CaMV-35S* in transgenic tomato plants and petunia. Some disadvantages were also noticed like stunted growth, delayed germination, and seed dormancy, photobleaching in transgenic tomato plants. Improvement in drought resistance with the expression of *NCED* gene in petuntia under the control of an inducible promoter *rd29A*, is an effective genetic engineering strategy with no side effects, resultant due to constitutive expression of transgene (Estrada-Melo et al., 2015).

The study reported in this thesis is aimed at examining the fundamental aspects of gene regulation by *cis* elements. The broad approach and findings described in this thesis are as follows:

1. The large spacer sequences between the two ACGT elements are preferred in some cases in *Arabidopsis thaliana*. From the frequency analysis, it has been found that the spacer length N-7 between two ACGT elements occurred more in rice. A high degree of nucleotide conservation was also observed for N-24 spacer length and not for any other spacer length in *Arabidopsis*.

2. ACGT element is more likely to be present in the upstream region of the salt and drought responsive genes. From our study, it was inferred that the two ACGT elements present in close vicinity in promoter region are the feature of stress responsive genes. Co-occurrence of ACGT element across the four genomes gives an idea that ACGT *cis* regulatory elements shows parallel evolution across different plant genomes.

3. The transient study reveals that AACA element negatively regulates the expression of the *PP2C*-like promoter. The presence of the AACA element in *PP2C*-like promoter decreases the expression of the reporter gene EGFP in the presence or absence of the inducers – ABA, JA and SA.

4. In the present study, we have verified that the full-length *PP2*C-like promoter from *Arabidopsis* is a induced by ABA, SA and MeJA

5. We tried to identify an efficient promoter region for stress induced expression activity by making a series of deletions in *PP2C*-like promoter. The full length and 900

bp deletion constructs of *PP2C*-like promoter drives high levels of transgene expression in response to ABA. These promoter constructs will be useful for the development of stress tolerant transgenic plants.

Future Perspective

The constitutive promoters which are widely used in development of transgenic plant suffer from the drawback of being highly energy demanding as they are always active even in favourable conditions. The demand for high energy by constitutive promoters leads to economic imbalance in terms of energy which may significantly affect the overall plant productivity. The solution can be the use of stress inducible promoters which are intended to activate precisely under abiotic and biotic stress conditions.

The *PP2C*-like promoter (At5g59220) which is one of the stress inducible promoters in *Arabidopsis thaliana,* has three ACGT elements in close vicinity. Based on *in-silico* information, two ACGT elements separated by 30 bp were found to be induced in response to ABA. This study involves the assessment of the role of *cis* element responsible for stress induction in the above promoter.

The aim of *cis* engineering of plants is to achieve multiple stress tolerance with increased productivity. There is a need to understand the cross-talk between different *cis* elements and transcription factors involved in different stress pathways. This understanding of the transcriptional cross-talk between pathways will help in designing better stress tolerant plants.

References

References

- Abe H, Urao T, Ito T, Seki M, Shinozaki K, Yamaguchi- Shinozaki K. (2003). Arabidopsis AtMYC2 (bHLH) and AtMYB2 (MYB) function as transcriptional activators in abscisic acid signaling. Plant Cell 15:63-78.
- Abe H, Yamaguchi-Shinozaki K, Urao T, Iwasaki T, Hosokawa D, Shinozaki K. (1997). Role of Arabidopsis MYC and MYB homologus in drought- and abscisic acid- regulated gene expression. Plant Cell 9:1859-1869.
- Abe M, Kobayashi Y, Yamamoto S, Daimon Y, Yamaguchi A, Ikeda Y, Ichinoki H, Notaguchi M, Goto K, Araki T. (2005). FD, a bZIP protein mediating signals from the floral pathway integrator FT at the shoot apex. Science 309:1052-1056.
- Alonso R, Onate-Sanchez L, Weltmeier F, Ehlert A, Diaz I, Dietrich K, Vicente-Carbajosa J, Droge-Laser W. (2009). A pivotal role of the basic leucine zipper transcription factor bZIP53 in the regulation of *Arabidopsis* seed maturation gene expression based on hetero-dimerization and protein complex formation. Plant Cell 21:1747-1761.
- Alves M, Dadalto S, Gonçalves A, De Souza G, Barros V, Fietto L. (2013). Plant bZIP transcription factors responsive to pathogens: a review. Int J Mol Sci 14:7815-7828.
- Anderberg RJ, Walker-Simmons MK. (1992). Isolation of a wheat cDNA clone for an abscisic acid-inducible transcript with homology to protein kinase. Proc Natl Acad Sci USA 89:10183-10187.
- Asseng S, Ewert F, Martre P, Rötter RP, Lobell DB, Cammarano D, Kimball BA, Ottman MJ,Wall GW, White JW, et al. (2014). Rising temperatures reduce global wheat production. Nature Climate Change 5:143-147.
- Baena-González E, Rolland F, Thevelein JM, Sheen J. (2007). A central integrator of transcription networks in plant stress and energy signalling. Nature 448:938-942.
- Barbosa EGG, Leite JP, Marin SRR, Marinho JP, Fátima Corrêa Carvalho J, Fuganti-Pagliarini R, Farias BRJ, Neumaier N, Marcelino-Guimaraes CF, Neves de- Oliveira CM,Yamaguchi-Shinozaki K, Nakashima K, Maruyama K, Kanamori N, Fujita Y, Yoshida T, Nepomuceno LA. (2013). Over expression of the ABA-dependent *AREB1* transcription factor from *Arabidopsis thaliana* improves soybean tolerance to water deficit. Plant Mol Biol Rep 31:719-730.
- Becker PB, Horz W. (2002). ATP-dependent nucleosome remodeling. Annu Rev Biochem 71:247-273.
- Benfy PN, Chua N-H. (1990). The cauliflower mosaic virus 35S promoter: combinatorial regulation of transcription in plants. Science 250: 959-966.

- Bensmihen S, Rippa S, Lambert G, Jublot D, Pautot V, Granier F, Giraudat J, Parcy F. (2002). The homologous ABI5 and EEL transcription factors function antagonistically to fine-tune gene expression during late embryogenesis. Plant Cell 14:1391-403.
- Ben-Tabou DS, Davidson EH. (2007). Gene regulation: gene control network in development. Annu Rev Biophys Biomol Struct 36:191-212.
- Bhattacharjee S, Renganaath K, Mehrotra R, Mehrotra S. (2013). Combinatorial control of gene expression. Biomed Research International Article Id 407263, 11 pages (doi.org/10.1155/2013/407263).
- Bradford MM. (1976). A rapid and sensitive method for the quantitation of microgram quantities of protein utilizing the principle of protein-dye binding. Anal Biochem 7:248-254.
- Bray EA. (1994). Alterations in gene expression in response to water deficit. In Stress-Induced Gene Expression in Plants, A.S. Basra, ed (Amsterdam: Harwood Academic), pp. 1–23.
- Brinboim HC, Dolly J. (1979). A rapid alkaline extraction procedure for screening recombinant plasmid DNA. Nucleic Acids Res 7: 1513-1523.
- Bryne JC, Valen E, Tang MH, Marstrand T, Winther O, da Piedade I, Krogh A, Lenhard B, Sandelin A. (2008). JASPAR, the open access database of transcription factor-binding profiles: new content and tools in the 2008 update. Nucleic Acids Res 36:D102–D106.
- Bulger M, Groudine M. (2011). Functional and mechanistic diversity of distal transcription enhancers. Cell 144:327-339.
- Busk PK, Pages M. (1998). Regulation of abscisic acid-induced transcription. Plant Mol Biol 37:425-435.
- Bustos MM, Guiltlnan MJ, Jordano J, Begum D, Kalkan FA, Hall TC. (1989). Regulation of *β-glucuronidase* expression in transgenic tobacco plants by an AT-rich, *cis*-acting sequence found upstream of a French bean *β-phaseolin* gene. Plant Cell 1:839-853.
- Cairns RB. (2009). The logic of chromatin architecture and remodeling at promoters. Nature 461:193-198.
- Chaturvedi CP, Sawant SV, Kiran K, Mehrotra R, Lodhi N, Ansari SA, Tuli R. (2006). Analysis of polarity in the expression from a multifactorial bidirectional promoter designed for high-level expression of transgenes in plants. J Biotechnol 123:1-12.
- Choi H, Hong J, Ha J, Kang J, Kim SY. (2000). ABFs, a family of ABA-responsive element binding factors. J Biol Chem 275:1723-1730.
- Clough SJ, Bent A. (1998). Floral dip: a simplified method for Agrobacterium-mediated transformation of *Arabidopsis thaliana*. Plant J 16:735–743.
- Davuluri RV, Sun H, Palaniswamy SK, Matthews N, Molina C, Kurtz M, Grotewold E. (2003). AGRIS: Arabidopsis gene regulatory information server, an information resource of Arabidopsis *cis*-regulatory elements and transcription factors. BMC Bioinformatics 4:25.

- de los Reyes BG, Mohanty B, Yun JS, Park M-R and Lee D-Y. (2015). Upstream regulatory architecture of rice genes: summarizing the baseline towards genus-wide comparative analysis of regulatory networks and allele mining. Rice J 8: 14.
- de Pater S, Greco V, Pham K, Memelink J, Kijne J. (1996). Characterization of a zinc-dependent transcriptional activator from *Arabidopsis*. Nucleic Acids Res 24:4624-4631.
- de Torres Zabala M, Bennett MH, Truman WH, Grant MR. (2009). Antagonism between salicylic and abscisic acid reflects early host–pathogen conflict and moulds plant defense responses. Plant J 59:375-386.
- Deppmann CD, Acharya A, Rishi V, Wobbes B, Smeekens S, Taparowsky EJ, et al. (2004). Dimerization specificity of all 67 B-ZIP motifs in *Arabidopsis thaliana*: a comparison to *Homo sapiens* B-ZIP motifs. Nucleic Acids Res 32:3435-3445.
- Deppmann CD, Alvania RS, Taparowsky EJ. (2006). Cross-species annotation of basic leucine zipper factor interactions: insight into the evolution of closed interaction networks. *Mol Biol Evol* 23:1480-1492.
- Doelz L. BioCompanion, Biocomputing Essentials Series, Dr. Ing. U. Doelz, Basel, Switzerland, 1990.
- Dolferus R, Jacobs M, Peacock WJ, Dennis ES. (1994). Differential interactions of promoter elements in stress responses of the Arabidopsis *Adh* gene. Plant Physiol 105:1075–1087.
- Elmayan T, Vaucheret H. (1996). A strongly-expressed 35S-driven transgene undergoes post-transcriptional silencing in all tobacco transformants irrespective of the copy number. Plant J 9:787–797.
- Estrada-Melo CA, Ma C, Michael S Reid SM, Jiang ZC. (2015). Overexpression of an ABA biosynthesis gene using a stress-inducible promoter enhances drought resistance in petunia. Horticulture Research. doi:10.1038/hortres.2015.13
- Eulgem T, Rushton PJ, Robatzek S, Somssich IE. (2000). The WRKY superfamily of plant transcription factors. *Trends Plant Sci* 5:199-206.
- Finkelstein RR, Gampala SS, Rock CD. (2002). Abscisic acid signaling in seeds and seedlings. Plant Cell 14:S15-S45.
- Finkelstein RR, Lynch TJ. (2000). The Arabidopsis abscisic acid response gene ABI5 encodes a basic leucine zipper transcription factor. Plant Cell 12:599-610.
- Folter S, Angenent GC. (2006). Trans meets *cis* in MADS science. Trends Plant Sci 11:224-231.
- Foster R, Izawa T, Chua NH. (1994). Plant bZIP proteins gather at ACGT elements. FASEB J 8:192-200.
- Freeling M, Rapaka L, Lyons E, Pedersen B, Thomas CB. (2007). G-Boxes, Bigfoot genes, and environmental response: characterization of intra-genomic conserved noncoding sequences in Arabidopsis. Plant Cell 19:1441–1457
- Freeling M, Subramaniam S. (2009). Conserved noncoding sequences (CNSs) in higher plants. Curr Opin Plant Biol 12:126–132

- Fujii H, Zhu JK. (2009). Arabidopsis mutant deficient in 3 abscisic acid-activated protein kinases reveals critical roles in growth, reproduction, and stress. Proc Natl Acad Sci USA 106:8380-8385.
- Fujii HM, Verslues PE, Zhu JK. (2007). Identification of two protein kinases required for abscisic acid regulation of seed germination, root growth and gene expression in Arabidopsis. Plant Cell 19:485-494.
- Fujii Y, Shimizu T, Toda T, Yanagida M, Hakoshima T. (2000). Structural basis for the diversity of DNA recognition by bZIP transcription factors. Nat Struct Biol 7:889-893.
- Fujita M, Fujita Y, Noutoshi Y, Takahashi F, Narusaka Y, Yamaguchi-Shinozaki K, Shinozaki K. (2006). Crosstalk between abiotic and biotic stress responses: a current view from the points of convergence in the stress signaling networks. Curr Opin Plant Biol 9:436-442.
- Fujita Y, Fujita M, Satoh R, Maruyama K, Parvez MM, Seki M, Hiratsu K, Ohme-Takagi M, Shinozaki K, Yamaguchi-Shinozaki K. (2005). AREB1 is a transcription activator of novel ABRE-dependent ABA signaling that enhances drought stress tolerance in Arabidopsis. Plant Cell 17:3470-88.
- Fujita Y, Nakashima K, Yoshida T, Katagiri T, Kidokoro S, Kanamori N, Umezawa T, Fujita M, Maruyama K, Ishiyama K, Kobayashi M, Nakasone S, Yamada K, Ito T, Shinozaki K, Yamaguchi-Shinozaki K. (2009). Three SnRK2 protein kinases are the main positive regulators of abscisic acid signaling in response to water stress in Arabidopsis. Plant Cell Physiol 50:2123-2132.
- Furihata T, Maruyama K, Fujita Y, Umezawa T, Yoshida R, Shinozaki K, Yoshida R, Shinozaki K, Yamaguchi-Shinozaki K. (2006). Abscisic acid-dependent multisite phosphorylation regulates the activity of a transcription activator AREB1. Proc Natl Acad Sci USA 103:1988-1993.
- Gilmartin PM, Chua NH. (1990). Spacing between GT-1 binding sites within a light responsive element is critical for transcriptional activity. Plant Cell 2:447-455.
- Giraudat J, Parcy F, Bertauche N, Gosti F, Leung J, Morris PC, Bouvier-Durand M, Vartanian N. (1994). Current advances in abscisic acid action and signalling. Plant Mol Biol 5:1557-1577.
- Giuliano G, Pichersky E, Malik VS, Timko MP, Scolnik PA, Cashmore AR. (1988). An evolutionarily conserved protein binding sequence upstream of a plant light-regulated gene. Proc Natl Acad Sci USA 85:7089-7093.
- Glycine Max, US DOE Joint Genome Institute (JGI-PGF), Ref Seq Assembly ID: GCF_000004515.1. http://www.ncbi.nlm.nih.gov/assembly/310798/.
- Gomez-Porras JL, Riano-Pachon DM, Dreyer I, Mayer JE, Mueller-Roeber B. (2007). Genome-wide analysis of ABA-responsive elements ABRE and CE3 reveals divergent patterns in Arabidopsis and rice. BMC Genomics 8:260-.
- Gómez-Porras JL: *In silico* identification of genes regulated by abscisic acid in *Arabidopsis thaliana (L.) Heynh.* Volume Dr. rer. nat. PhD thesis. University of Potsdam. Faculty of Mathematics and Natural Sciences; 2005::141.

- Gorantla M, Babu PR, Lachagari VBR, Reddy AMM, Wusirika R, Bennetzen JL, Reddy AR. (2007). Identification of stress responsive genes in an indica rice (Oryza sativa L.) using ESTs generated from drought stressed seedlings. J Exp Bot 58:253-265.
- Guedes Corrêa GL, Riaño-Pachón MD, Schrago GC, dos Santos VR, Mueller-Roeber B, et al. (2008) The Role of bZIP Transcription Factors in Green Plant Evolution: Adaptive Features Emerging from Four Founder Genes. PLoS ONE 3: e2944.
- Guiltinan MJ, Marcotte WR, Quatrano RS. (1990). A plant leucine zipper protein that recognizes an abscisic acid response element. Science 250:267-270.
- Guo AY, Chen X, Gao G, Zhang H, Zhu QH, Liu XC, Zhong YF, Gu X, He K, Luo J. (2008). Plant TFDB: a comprehensive plant transcription factor database. Nucleic Acids Res 36:D966-D969.
- 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.
- Hernandez-Garcia MC, Finer JJ. (2014). Identification and validation of promoters and *cis* acting regulatory elements. Plant Sci 217:109-119.
- Hey SJ, Byrne E, Halford NG. (2010). The interface between metabolic and stress signaling. Ann Botany 105:197–203.
- Higo K, Ugawa Y, Iwamoto M, Korenaga T. (1999). Plant cis-Acting Regulatory DNA Elements (PLACE) Database. Nucleic Acids Res 27:297-300.
- Hirayama T, Shinozaki K. (2007). Perception and transduction of abscisic acid signals: keys to the function of the versatile plant hormone ABA. Trends Plant Sci 12:343-351.
- Hobo T, Asada M, Kowyama Y, Hattori T. (1999). ACGT-containing abscisic acid response element (ABRE) and coupling element 3 (CE3) are functionally equivalent. Plant J 19:679-689.
- Hong JK, Hwang BK. (2009). The promoter of the pepper pathogen-induced membrane protein gene CaPIMP1 mediates environmental stress responses in plants. Planta 229:249-259.
- Ishige F, Takaichi M, Foster R, Chua NH, Oeda K. (1999). A G-box motif (GCCACGTGCC) tetramer confers high-level constitutive expression in dicot and monocot plants. Plant J 18:443-448.
- Ishiguro S, Nakamura K. (1994). Characterization of a cDNA encoding a novel DNA-binding protein, SPF1, that recognizes SP8 sequences in the 59 upstream regions of genes coding for sporamin and b-amylase from sweet potato. Mol Gen Genet 244:563-571.
- Izawa T, Foster R, Chua NH. (1993). Plant bZIP protein DNA binding specificity. J Mol Blot 230:131-140.
- Jaglo-Ottosen KR, Gilmour SJ, Zarka DG, Schabenberger O, Thomashow MF. (1998). Arabidopsis CBF1 overexpression induces COR genes and enhances freezing tolerance. Science 280:104–106.

- Jakoby M, Weisshaar B, Dröge-Laser W, Vicente-Carbajosa J, Tiedemann J, Kroj T, Parcy F. (2002). bZIP transcription factors in Arabidopsis. Trends Plant Sci 7:106-111.
- Jefferson RA, Kavanagh TA, Bevan MW. (1987). GUS fusions: *β-glucuronidase* as a sensitive and versatile gene fusion marker in higher plants. EMBO J 6:3901-3907.
- Jefferson RA. (1987). Assaying chimeric genes in plants: The GUS gene fusion system. Plant Mol Biol Rep 5: 387-405.
- Jeong MJ, Shih MC. (2003). Interaction of a GATA factor with cis-acting elements involved in light regulation of nuclear genes encoding chloroplast glyceraldehyde-3-phosphate dehydrogenase in Arabidopsis. Biochem Bioph Res Co 300:555-562.
- Johnson R, Wagner R, Verhey SD, Walker-Simmons MK. (2002). The ABA-responsive kinase PKABA1 interacts with a seed-specific ABA response element binding factor, TaABF, and phosphorylates TaABF peptide sequences. Plant Physiol 130:837-846.
- Kamrul-Huda MK, Akhter-Banu MS, Pathi MK, and Tuteja N. (2013). Reproductive Organ and Vascular Specific Promoter of the Rice Plasma Membrane Ca^{2+}ATPase Mediates Environmental Stress Responses in Plants. PLoS One 8:1-15.
- Kang J, Choi H, Im M, Kim SY. (2002). Arabidopsis basic leucine zipper proteins that mediate stress-responsive abscisic acid signaling. Plant Cell 14:343-357.
- Kim JC, Lee SH, Cheong YH, Yoo CM, Lee SI, Chun HJ, Yun DJ, Hong JC, Lee SY, Lim CO, Cho MJ. (2001). A novel cold-inducible zinc finger protein from soybean, SCOF-1, enhances cold tolerance in transgenic plants. Plant J 25:247-259.
- Kim JM, To TK, Nishioka T, Seki M. (2010). Chromatin regulation functions in plant abiotic stress responses. Plant Cell Environ 33:604-610.
- Kim SR, Kim Y, An G. (1993). Identification of methyl-jasmonate and salicylic acid response elements from the nopaline synthase (*nos*) promoter. Plant Physiol 103:97-103.
- Kim S, Kang J, Cho D, Park JH, Kim SY. (2004). ABF2, an ABRE-binding bZIP factor, is an essential component of glucose signaling and it's over expression affects multiple stress tolerance. Plant J 40:75-87.
- Kleinjan DA, Van -Heyningenn V. (2005). Long range control of gene expression: emerging mechanisms and disruption in disease. Am J Hum Genet 76:8-32.
- Kluth A, Sprunck S, Becker D, Lörz H, Lütticke S. (2002). 5′ deletion of a *gbss1* promoter region from wheat leads to changes in tissue and developmental specificities. Plant Mol Biol 49: 665-678.
- Ko LJ, Engel JD. (1993). DNA-binding specificities of the GATA transcription factor family. Mol Cell Biol 13:4011-4022.
- Kobayashi Y, Yamamoto S. Minami H, Kagaya Y, Hattori T. (2004). Differential activation of the rice sucrose nonfermenting1-related protein kinase2 family by hyperosmotic stress and abscisic acid. Plant Cell 16:1163-1177.

- Komatsu K, Suzuki N, Kuwamura M, Nishikawa Y, Nakatani M, Ohtawa H, Takezawa D, Seki M, Tanaka M, Taji T, Hayashi T, Sakata Y. (2013). Group A PP2Cs evolved in land plants as key regulators of intrinsic desiccation tolerance. Nat Commun 4:2219-.
- Kulik A, Wawer I, Krzywinska E, Bucholc M, Dobrowolska G. (2011). SnRK2 protein kinases-key regulators of plant response to abiotic stresses. OMICS 15:859-872.
- Lam E, Kano-Murakami Y, Gilmartin P, Niner B, Chua N-H. (1990). A metal-dependent DNA-binding protein interacts with a constitutive element of a light-responsive promoter. Plant Cell 2:857-866.
- Lee TI, Young RA. (2000). Transcription of eukaryotic protein-coding genes. Annu Rev Genet 34:77-137.
- Lescot M, Déhais P, Moreau Y, De Moor B, Rouzé P, Rombauts S. (2002). Plant CARE: 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.
- Levine M, Tjian R. (2003).Transcription regulation and animal diversity. Nature 424:147-151.
- Li B, Carey M, Workman JL. (2007). The role of chromatin during transcription. Cell 128:707-719.
- Liu C, Mao B, Ou S, Wang W, Liu L, Wu Y, Chu C, Wang X. (2014). OsbZIP71, a bZIP transcription factor, confers salinity and drought tolerance in rice. Plant Mol Biol 84:19-36.
- Liu ZB, Hagen G, Guilfoyle, TJ. (1997) A G-box-binding protein from soybean binds to the E1 auxin-response element in the soybean *GH3* promoter and contains a proline-rich repression domain. Plant Physiol 115: 397–407.
- Llorca MC, Potschin M, Zentgraf U. (2014). bZIPs and WRKYs: two large transcription factor families executing two different functional strategies. Front Plant Sci 5:169-.
- Lopez-Ochoa L, Acevedo-Hernandez G, Martinez-Hernandez A, Argiiello-Astorga G, Herrera- Estrella L. (2007). Structural relationships between diverse *cis*-acting elements are critical for the functional properties of a *rbcS* minimal light regulatory unit. J Exp Bot 58:4397-4406.
- Luan S. (2003). Protein phosphatases in plants. Annu Rev Plant Biol 54:63-92.
- Ma S, Bohnert HJ. (2007). Integration of Arabidopsis thaliana stress-related transcript profiles, promoter structures and cell-specific expression. Genome Biol 8:R49.
- Ma S, Gong Q, Bohnert HJ. (2006). Dissecting salt stress pathways. J Exp Bot 57:1097-1107.
- Ma Y, Szostkiewicz I, Korte A, Moes D, Yang Y, Christmann A, Grill E. (2009). Regulators of PP2C phosphatase activity function as abscisic acid sensors. Science 324:1064-1068.
- Marcotte WR Jr, Russell SH, Quatrano RS. (1989). Abscisic acid-responsive sequences from the *em* gene of wheat. Plant Cell 1:969-976.

- Maruyama K, Todaka D, Mizoi J, Yoshida T, Kidokoro S, Matsukura S, Takasaki H, Sakurai T, Yamamoto YY, Yoshiwara MK, Sakakibara H, Shinozaki K, Shinozaki-Yamaguchi K. (2012). Identification of *cis*-acting promoter elements in cold and dehydration-induced transcriptional pathways in *Arabidopsis*, rice and soybean. DNA Res 19:37-49.
- Mauch-Mani B, Mauch F. (2005). The role of abscisic acid in plant–pathogen interactions. Curr Opin Plant Biol 8:409-414.
- Mehrotra R, Sethi S, Zutshi I, Bhalothia P, Mehrotra S. (2013). Patterns and evolution of ACGT repeat *cis*-element landscape across four plant genomes. BMC Genomics 14:203-210.
- Mehrotra R, Bhalothiaa P, Bansali B, Basantani KM, Bharti V and Mehrotra S. (2014). Abscisic acid and abiotic stress tolerance – Different tiers of regulation. J Plant Physiol 171:486-496.
- Mehrotra R, Jain V, Shekhar C, Mehrotra S. (2014b). Genome wide analysis of *Arabidopsis thaliana* reveals high frequency of $AAAG_{N7}CTTT$ motif. Meta Gene 2: 606–615.
- Mehrotra R, Kiran K, Chaturvedi CP, Ansari SA, Lodhi N, Sawant S, Tuli R. (2005). Effect of copy number and spacing of the ACGT and GT *cis* elements on transient expression of minimal promoter in plants. J Genet 84:183-187.
- Mehrotra R, Mehrotra S. (2010). Promoter activation by ACGT in Response to Salicylic and abscisic acids is differentially regulated by the spacing between two copies of the motif. J Plant Physiol 167:1214-1218.
- Mehrotra R, Yadav A, Bhalothia P, Karan R, S Mehrotra S. (2012). Evidence for directed evolution of larger size motif in *Arabidopsis thaliana* genome. Scientific World J (2012). 2012: 983528. doi: 10.1100/2012/983528.
- Melotto M, Underwood W, Koczan J, Nomura K, He SY. (2006). Plant stomata function in innate immunity against bacterial invasion. Cell 126:969-980.
- Menkens AE, Cashmore AR. (1994). Isolation and characterization of a fourth Arabidopsis thaliana G-box-binding factor, which has similarities to Fos oncoprotein. Proc Natl Acad Sci USA 91, 2522–2526. doi:10.1073/pnas.91.7.2522.
- Merika M, Orkin SH. (1993). DNA-binding specificity of GATA family transcription factors. Mol Cell Biol 13:3999-4010.
- Michael TP, Mockler TC, Breton G, McEntee C, Byer A, Trout JD, Hazen SP, Shen R, Priest HD, Sullivan CM, et al. (2008). Network discovery pipeline elucidates conserved time of day specific *cis*-regulatory modules. PLoS Genet 4 p. e14
- Mikołajczyk M, Awotunde OS, Muszyńska G, Klessig DF, Dobrowolska G. (2000). Osmotic stress induces rapid activation of a salicylic acid-induced protein kinase and a homolog of protein kinase ASK1 in tobacco cells. Plant Cell 12:165-178.
- Miller M, Shuman JD, Sebastian T, Dauter Z, Johnson PF. (2003). Structural basis for DNA recognition by the basic region leucine zipper transcription factor CCAAT/enhancer-binding protein alpha. J Biol Chem 278:15178-15184.

- Mishiba K, Nishihara M, Nakatsuka T, Abe Y, Hirano H, Yokoi T, Kikuchi A, Yamamura S. (2005). Consistent transcriptional silencing of 35S-driven transgenes in gentian. Plant J 44:541-556.
- Mockler TC, Michael TP, Priest HD, Shen R, Sullivan CM, Givan SA, McEntee C, Kay SA, Chory J. (2007). The DIURNAL project: DIURNAL and circadian expression profiling, model-based pattern matching, and promoter analysis. Cold Spring Harb Symp Quant Biol, 72:353–363.
- Molina C, Grotewold E. (2005). Genome wide analysis of Arabidopsis core promoters. BMC Genomics 6:25.
- Monks DE, Aghoram K, Courtney PD, Dewald DB, Dewey RE. (2001). Hyperosmotic stress induces the rapid phosphorylation of a soybean phosphatidyl inositol transfer protein homolog through activation of the protein kinases SPK1 and SPK2. Plant Cell 13:1205-1219.
- Moon H, Callahan MA. (2004). Developmental regulation of peach *ACC o*xidase promoter–GUS fusions in transgenic tomato fruits J Exp Bot 55:1519–1528.
- Morran S, Eini O, Pyvovarenko T, Parent B, Singh R, Ismagul A, Eliby S, Shirley N, Langridge P, Lopato S. (2011). Improvement of stress tolerance of wheat and barley by modulation of expression of DREB/CBF factors. Plant Biotechnol J 9:230-49.
- MSU Rice Genome Annotation Project. http://rice.plantbiology.msu.edu/annotation_pseudo_apk.shtml.
- Muller F, Demeny MA, Tora L. (2007). New problems in RNA polymerase II transcription initiation: matching the diversity of core promoters with a variety of promoter recognition factors. J Biol Chem 282:14685-14689.
- Mundy J, Chua NH. (1988). Abscisic acid and water-stress induce the expression of a novel rice gene. EMBO J 7:2279-2286.
- Murashige T, Skoog F. (1962). A revised medium for rapid growth and bioassays with tobacco tissue cultures. Plant Physiol 15:473- 497.
- Murray MG, Thompson WF. (1980). Thompson rapid isolation of high molecular weight plant DNA. Nucleic Acids Res 8:4321-4326.
- Naika M, Shameer K, Mathew KO, Gowda R, Sowdhamini R. (2013). STIFDB2: An Updated Version of Plant Stress-Responsive TranscrIption Factor DataBase with Additional Stress Signals, Stress-Responsive Transcription Factor Binding Sites and Stress-Responsive Genes in Arabidopsis and Rice.
- Nakashima K, Fujita Y, Kanamori N, Katagiri T, Umezawa T, Kidokoro S, Maruyama K, Yoshida T, Ishiyama K, Kobayashi M, Shinozaki K, Yamaguchi-Shinozaki K. (2009a). Three Arabidopsis SnRK2 protein kinases, SRK2D/SnRK2.2, SRK2E/SnRK2.6/OST1 and SRK2I/SnRK2.3, involved in ABA signaling are essential for the control of seed development and dormancy. Plant Cell Physiol 50:1345-1363.

- Nakashima K, Ito Y, Yamaguchi-Shinozaki K. (2009b). Transcriptional regulatory networks in response to abiotic stresses in Arabidopsis and grasses. Plant Physiol 149:88-95. doi: 10.1104/pp.108.129791
- Nakashima K, Yamaguchi-Shinozaki K. (2013). ABA signaling in stress-response and seed development. Plant Cell Rep 32:959-970.
- Narlikar GJ, Fan HY, Kingston RE. (2002). Cooperation between complexes that regulate chromatin structure and transcription. Cell 108:475-487.
- Narusaka Y, Nakashima K, Shinwari ZK, 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 34:137-148.
- Neill JS, Desikan R, Clarke A, Hurst DR, Hancock TJ. (2002). Hydrogen peroxide and nitric oxide as signalling molecules in plants. J Exp Bot 53:1237-1247.
- Nijhawan A, Jain M, Tyagi AK, Khurana JP. (2007). Genomic survey and gene expression analysis of the basic leucine zipper transcription factor family in rice. Plant Physiol 146:333-350.
- Nishimura N, Hitomi K, Arvai AS, Rambo RP, Hitomi C, Cutler SR, Schroeder JI, Getzoff ED. (2009a). Structural mechanism of abscisic acid binding and signaling by dimeric PYR1. Science 326:1373-1379.
- Nishimura N, Yoshida T, Kitahata N, Asami T, Shinozaki K, Hirayama T. (2007). ABA Hypersensitive Germination1 encodes a protein phosphatase 2C, an essential component of abscisic acid signaling in Arabidopsis seed. Plant J 50:935-49.
- Oh SJ, Song SI, Kim YS, Jang HJ, Kim SY, Kim M, Kim YK, Nahm BH, Kim JK. (2005). Arabidopsis CBF3/DREB1A and ABF3 in transgenic rice increased tolerance to abiotic stress without stunting growth. Plant Physiol 138:341-351.
- Oliphant AR, Brandle CJ, Struhl K. (1989). Defining the sequence specificity of DNA-binding proteins by selecting binding sites from random-sequence oligonucleotides: analysis of yeast GCN4 protein. Mol Cell Biol 9:2944-2949.
- Ong CT, Corces VG. (2011). Enhancer function: new insights into the regulation of tissue-specific gene expression. Nature Rev Genet 12:283-293.
- *Oryza sativa* Japonica Group, International Rice Genome Sequencing Project, Build 4.0, Ref Seq Assembly ID: GCF_000005425.2. http://www.ncbi.nlm.nih.gov/assembly/313038/.
- Ouwerkerk PBF, Memelink J. (1999). A G-box element from the *Catharanthus roseus strictosidine synthase* (*Str*) gene promoter confers seed-specific expression in transgenic tobacco plants. Mol General Genet 261, 635-643.
- Pan Y, Tsai CJ, Ma B, Nussinov R. (2010). Mechanisms of transcription factor selectivity. Trends Genet 26:75-83.
- Park SY, Fung P, Nishimura N, Jensen DR, Hiroaki F, Zhao Y, Lumba S, Santiago J, Rodrigues A, Chow TF, Alfred SE, Bonetta D, Finkelstein R, Provart NJ, Desveaux D, Rodriguez PL, McCourt P, Zhu JK, Schroeder JI, Volkman

BF, Cutler SR. (2009). Abscisic acid inhibits PP2Cs via the PYR/PYL family of ABA binding START proteins. Science 324:1068-1071.

- Polizel AM, Medri ME, Nakashima K, Yamanaka N, Farias JR, De Oliveira MC, Marin SR, Abdelnoor RV, Marcelino-Guimarães FC, Fuganti R, Rodrigues FA, Stolf-Moreira R, Beneventi MA, Rolla AA, Neumaier N, Yamaguchi-Shinozaki K, Carvalho JF, Nepomuceno AL. (2011). Molecular, anatomical and physiological properties of a genetically modified soybean line transformed with *rd29A*:AtDREB1A for the improvement of drought tolerance. Genet Mol Res 10:3641-3656.
- Prestridge DS. (1991). SIGNAL SCAN: A computer program that scans DNA sequences for eukaryotic transcriptional elements. Comput Appl Biosci7:203-206.
- Priest HD, Filichkin SA, Mockler TC. (2009). *Cis* regulatory elements in plant cell signaling. Curr Opin Plant Biol 12:643-649
- Rabbani MA, Maruyama K, Abe H, Khan MA, Katsura K, ItoY, Yoshiwara K, Seki M, Shinozaki K, Yamaguchi-Shinozaki K. (2003). Monitoring expression profiles of rice gene under cold, drought, high-salinity stresses and abscisic acid application using cDNA microarray and RNA gel blot analyses. Plant Physiol 133:1755–1767.
- Reese JC. (2003). Basal transcription factors. Curr Opin Genet Dev 13:114-118.
- Reichel, Mathur J, Eckes P, Langenkemper K, Koncz C, Schell J, Reiss B, Maas C. (1996). Enhanced green fluorescence by the expression of an *Aequorea victoria* green fluorescent protein mutant in mono and dicoty -ledonous plant Cells. Proc Natl Acad Sci USA 93:5888-5893.
- Riaño-Pachón DM, Ruzicic S, Dreyer I, Mueller-Roeber B. (2007). PlnTFDB: an integrative plant transcription factor database. BMC Bioinformatics 8:42.
- Riechmann JL, Heard J, Martin G, Reuber L, Jiang C, Keddie J, Adam L, Pineda O, Ratcliffe OJ, Samaha RR, Creelman R, Pilgrim M, Broun P, Zhang JZ,Ghandehari D, Sherman BK, Yu G. (2000). *Arabidopsis* transcription factors: genome-wide comparative analysis among eukaryotes. Science 290:2105-2110.
- Romero, Fuertes, Benito, Malpica, Leyva, Paz A. (1998). More than 80 R2R3 MYB regulatory genes in the genome of *Arabidopsis thaliana*. Plant J 14:273-284.
- Ross C, Shen QJ. (2006). Computational prediction and experimental verification of *HVA1*-like abscisic acid responsive promoters in rice (Oryza sativa). Plant Mol Biol 62:233-246.
- Roychoudhury A, Sengupta DN. (2009).The promoter-elements of some abiotic stress-inducible genes from cereals interact with a nuclear protein from tobacco. Biologia Plantarum 53:583-587.
- Rubio S, Rodrigues A, Saez A, Dizon MB, Galle A, Kim TH, Santiago J, Flexas J, Schroeder JI, Rodriguez PL. (2009). Triple loss of function of protein phosphatases type 2C leads to partial constitutive response to endogenous abscisic acid. Plant Physiol 150:1345-1355.

- Rushton PJ, Macdonald H, Huttly AK, Lazarus CM, Hooley R. (1995). Members of a new family of DNA-binding proteins bind to a conserved *cis*-element in the promoters of *α-Amy 2* genes. Plant Mol Biol 29(4):691-702.
- Rushton PJ, Reinstadler A, Lipka V, Lippok B, I.E. Somssich IE. (2002). Synthetic plant promoters containing defined regulatory elements provide novel insights into pathogen- and wound-induced signaling. Plant Cell, 14:749–762.
- Rushton PJ, Torres JT, Parniske M, Wernert P, Hahlbrock K, Somssich IE. (1996). Interaction of elicitor-induced DNA-binding proteins with elicitor response elements in the promoters of parsley *PR1* genes. EMBO J 15:5690-5700.
- Saavedra X, Modrego A, Rodriguez D, Gonzalez-Garcia MP, Sanz L, Nicolas G, Lorenzo O. (2010). The nuclear interactor PYL8/RCAR3 of *Fagus sylvatica* FsPP2C1 is a positive regulator of abscisic acid signalling in seeds and stress. Plant Physiol 152:133-150.
- Sablowski RW, Moyano E., Culianez-Macia FA, Schuch W, Martin C, Bevan M. (1994). A flower-specific Myb protein activates transcription of phenylpropanoid biosynthetic genes. *EMBO J* 13(1):128-137.
- Saha A, Wittmeyer J, Cairns BR. (2006). Chromatin remodeling: the industrial revolution of DNA around histones. Nature Rev Mol Cell Biol 7:437-447.
- Sambrook J, Fritsch EF, Maniatis Y. (1989). Molecular cloning: a laboratory manual, 2nd edn. Cold Spring Harbor Laboratory, Cold Spring Harbor.
- Sandelin, Wasserman WW, Lenhard B. (2004). ConSite: web-based prediction of regulatory elements using cross species comparison. Nucleic Acids Res 32:W249–W252.
- Santiago J, Dupeux F, Round A, Antoni R, Park S, Jamin M, Cutler SR, Roddriguez PL, Marquez JA. (2009a). The abscisic acid receptor PYR1 in complex with abscisic acid. Nature 462:665-668.
- Santiago J, Rodrigues A, Saez A, Rubio S, Antoni R, Dupeux F, Park SY, Márquez JA, Cutler SR, Rodriguez PL. (2009b). Modulation of drought resistance by the abscisic acid receptor PYL5 through inhibition of clade A PP2Cs. Plant J 60:575-578.
- Sarkar KA, Lahiri A. (2013). Specificity determinants for the abscisic acid response element. FEBS Open Bio 3:101-105.
- Sawant S, Kiran K, Mehrotra R, Chaturvedi CP, Ansari SA, Singh P, Lodhi N, Tuli R. (2005). A variety of synergistic and antagonistic interactions mediated by *cis*-acting DNA motifs regulate gene expression in plant cells andmodulate stability of the transcription complex formed on a basal promoter. J Exp Bot 56:2345-2353.
- Sawant S, Singh PK, Madanala R, Tuli R. (2001). Designing of an artificial expression cassette for the high-level expression of transgenes in plants. Theoretical Applied Genet 102:635-644.
- Sawant S, Singh PK, Tuli R. (2000). Pretreatment of microprojectiles to improve delivery of DNA in plant transformation. Biotechniques 29:246–248.
- Sazegari S, Niazi A, Ahmadi FS. (2015). A study on the regulatory network with promoter analysis for *Arabidopsis DREB* genes. Bioinformation 28;11:101-106.

- Schweighofer A, Hirt H, Meskiene I. (2004). Plant PP2C phosphatases: emerging functions in stress signaling. Trends Plant Sci 9:236-243.
- Seki M, Narusaka M, Ishida J, Nanjo T, Fujita M, Oono Y, Kamiya A, Nakajima M, Enju A, Sakurai T, Satou M, Akiyama K, Taji T, Yamaguchi-Shinozaki K,Carninci P, Kawai J, Hayashizaki Y, Shinozaki K. (2002). Monitoring the expression profiles of 7000 Arabidopsis genes under drought, cold and high-salinity stresses using a full-length cDNA microarray. Plant J 31:279-92.
- Shaik R, Ramakrishna W. (2013). Genes and co-expression modules common to drought and bacterial stress responses in Arabidopsis and rice. PLoS ONE 8:e77261.
- Shaik R, Ramakrishna W. (2014). Machine learning approaches distinguish multiple stress conditions using stress-responsive genes and identify candidate genes for broad resistance in rice. Plant Physiol 164:481–495.
- Shen QJ, Ho TH. (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.
- Shen QJ, Zhang P, Ho TH. (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, Seki M. (2003). Regulatory network of gene expression in the drought and cold stress responses. Curr Opin Plant Biol 6:410-417.
- Shinozaki K, Yamaguchi-Shinozaki K. (2007). Gene networks involved in drought stress response and tolerance. J Exp Bot 58:221-227.
- Sikorski TW, Buratowski S. (2009). The basal initiation machinery: beyond the general transcription factors. Curr Opin Cell Biol 21:344-351.
- Singh A, Giri J, Kapoor S, Tyagi AK, Pandey GK. (2010). Protein phosphatase complement in rice: genome-wide identification and transcriptional analysis under abiotic stress conditions and reproductive development. BMC Genomics 11:435.
- Singh KB. (1998). Transcriptional regulation in plants: the importance of combinatorial control. Plant Physiol 118:1111-1120.
- Skriver K, Olsen FL, Roger J, Mundy J. (1991). *Cis*-acting DNA elements responsive to gibberellin and its antagonist abscisic acid. Proc Natl Acad Sci USA 88:7266-7270.
- Smykowski A, Zimmermann P, Zentgraf U. (2010). G-box binding factor1 reduces CATALASE2 expression and regulates the onset of leaf senescence in *Arabidopsis*. Plant Physiol 153:1321-1331.
- Sorghum Consortium, v. 1.0, RefSeq Assembly ID: GCF_000003195.2. http://www.ncbi.nlm.nih.gov/assembly/GCF_000003195.2/.
- Soufi A, Garcia MF, Jaroszewicz A, Osman N, Pellegrini M, Zaret KS. 2015 Pioneer transcription factors target partial DNA motifs on nucleosomes to initiate reprogramming. Cell 161(3):555-68.

- Srivastava R, Rai KM, Srivastava M, Kumar V, Pandey B, Singh SP, Bag SK, Singh BD, Tuli R, Sawant SV. (2014). Distinct role of core promoter architecture in regulation of light-mediated responses in plant genes. Mol Plant 7:626-641.
- Suzuki M, Ketterling MG, McCarty DR. (2005). Quantitative statistical analysis of *cis*-regulatory sequences in ABA/VP1- and CBF/DREB1-regulated genes of Arabidopsis. Plant Physiol 139:437-40.
- The Arabidopsis Information Resource (TAIR), v.10, Ref Seq Assembly ID: GCF_000001735.3. http://www.ncbi.nlm.nih.gov/assembly/GCF_000001735.3/.
- The Gene Expression Atlas. http://www.ebi.ac.uk/gxa/.
- Thomas MC, Chiang CM. (2006). The general transcription machinery and general cofactors. Crit Rev Biochem Mol Biol 41:105-178.
- Tougane K, Komatsu K, Bhyan SB, Sakata Y, Ishizaki K, Yamato KT, Kohchi T, Takezawa D. (2010). Evolutionarily conserved regulatory mechanisms of abscisic acid signaling in land plants: Characterization of abscisic acid insensitive 1-like type 2C protein phosphatase in the liverwort Marchantia polymorpha. Plant Physiol 152:1529-1543.
- Tran LS, Mochida K. (2010). Identification and prediction of abiotic stress responsive transcription factors involved in abiotic stress signaling in soybean. Plant Signal Behav 5: 255-257.
- Tran LS, Nakashima K, Sakuma Y, Simpson SD, Fujita Y, Maruyama K, Fujita M, Seki M, Shinozaki K, Yamaguchi-Shinozaki K. (2004). Isolation and functional analysis of Arabidopsis stress-inducible NAC transcription factors that bind to a drought-responsive *cis*-element in the *early responsive to dehydration stress 1* promoter. Plant Cell 16:2481-2498.
- Umezawa T, Nakashima K, Miyakawa T, Kuromori T, Tanokura M, Shinozaki K, Yamaguchi-Shinozaki K. (2010). Molecular basis of the core regulatory network in ABA responses: sensing, signaling and transport. Plant Cell Physiol 51:1821–39.
- Umezawa T, Sugiyama N, Takahashi F, Anderson JC, Ishihama Y, Peck SC, Shinozaki K. (2013). Genetics and phosphoproteomics reveal a protein phosphorylation network in the abscisic acid signaling pathway in Arabidopsis thaliana. Sci Signal 6:rs8.
- Umezawa T, Sugiyama N, Mizoguchi M, Hayashi S, Myouga F, Yamaguchi-Shinozaki K, Ishihama Y, Hirayama T, Shinozaki K. (2009). Type 2C protein phosphatases directly regulate abscisic acid-activated protein kinases in Arabidopsis. Proc Natl Acad Sci USA 106:17588-17593.
- Uno Y, Furihata T, Abe H, Yoshida R, Shinozaki K, Yamaguchi-Shinozaki K. (2000). Arabidopsis basic leucine zipper transcription factors involved in an abscisic acid-dependent signal transduction pathway under drought and high-salinity conditions. Proc Natl Acad Sci USA 97:11632-11637.
- Vander Krol AR, Chua N-H. (1991). The basic domain of part bZIP proteins facilitates import of a reporter protein into plant nuclei. Plant Cell 3:667-675.

- Venter M, Botha FC. (2010). Synthetic promoter engineering. Plant Developmental Biology-Biotechnological Perspectives 2:393-414.
- Vogel JT, Zarka DG, Van Buskirk HA, Fowler SG, Thomashow MF. (2005). Roles of the CBF2 and ZAT12 transcription factors in configuring the low temperature transcriptome of Arabidopsis. Plant J 41:195-211.
- Walley JW, Coughlan S, Hudson ME, Covington MF, Kaspi R., Banu G, et al., (2007). Mechanical stress induces biotic and abiotic stress responses via a novel *cis*-element. PLoS Genet 3:e172.
- Walling L, Drews GN, Goldberg RB. (1986). Transcriptional and post – transcriptional regulation of soyabean seed protein mRNA levels. Proc Natl Acad Sci USA 83:2123-2127
- Wei K, Chen J, Wang Y, Chen Y, Chen S, Lin Y, et al. (2012). Genome-wide analysis of bZIP-encoding genes in maize. DNA Res 19:463-476.
- Weiner JJ, Peterson FC, Volkman BF, Cutler SR. (2010). Structural and functional insights into core ABA signaling. Curr Opin Plant Biol 13:495-502.
- Williams ME, Foster R, Chua NH. (1992). Sequences flanking the hexameric G-box core CACGTG affect the specificity of protein binding. Plant Cell 4:485-496.
- Wittkopp JP, Kalay G. (2012). Cis-regulatory elements: molecular mechanisms and evolutionary processes underlying divergence. Nat Rev Genet 13:59-69.
- Won SK, Lee YJ, Lee HY, Heo YK, Cho M, Cho HT. (2009). *Cis*-element- and transcriptome-based screening of root hair-specific genes and their functional characterization in Arabidopsis. Plant Physiol 150:1459-1473.
- Wray GA, Hahn MW, Abouheif E, Balhoff JP, Pizer M, Rockman MV, Romano LA. (2003). The evolution of transcriptional regulation in eukaryotes. Mol Biol Evol 20: 1377-419.
- Wu C, Washida H, Onodera Y, Harada K, Takaiwa F. (2000). Quantitative nature of the Prolamin-box, ACGT and AACA motifs in a rice glutelin gene promoter: minimal *cis*-element requirements for endosperm-specific gene expression. Plant J 23:415-421.
- Wu JJ, Liu WY, Sun MX. (2011). Improved and high throughput quantitative measurements of weak GFP expression in transgenic plant materials. Plant Cell Rep 30:1253-1260.
- Xiao B, Huang Y, Tang N, Xiong L. (2007). Over-expression of a LEA gene in rice improves drought resistance under the field conditions. Theor Appl Genet 115(1):35–46.
- Xiong L, Schumaker KS, Zhu JK. (2002). Cell signaling during cold, drought and salt stress. Plant Cell 14:165-183.
- Xue T, Wang D, Zhang S, Ehlting J, Ni F, Jakab S, Zheng S, Zhong Y. (2008). Genome-wide and expression analysis of protein phosphatase 2C in rice and Arabidopsis. BMC Genomics 9:550.
- Yamaguchi-Shinozaki K, Mundy J, Chua NH. (1989). Four tightly linked rab genes are differentially expressed in rice. Plant Mo1 Biol 14:29-39.

- Yamaguchi-Shinozaki K, Shinozaki K. (1994). A novel *cis* acting element in an Arabidopsis gene is involved in responsiveness to drought, low-temperature, or high-salt stress. Plant Cell 6:251-264.
- Yamaguchi-Shinozaki K, Shinozaki K. (2005). Organization of *cis* acting regulatory elements in osmotic- and cold-stress-responsive promoters. Trends Plant Sci10:88-94.
- Yamaguchi-Shinozaki K, Shinozaki K. (2006). Transcriptional regulatory networks in cellular responses and tolerance to dehydration and cold stresses. Annu Rev Plant Biol 57:781-803.
- Yazaki J, Kishimoto N, Nagata Y, et al., (2003). Genomics approach to abscisic acid and gibberellin-responsive genes in rice. DNA Res 10:249-261.
- Yoon HW, Kim MC, Shin PG, Kim JS, Kim CY, Lee SY, et al. (1997). Differential expression of two functional serine/threonine protein kinases from soybean that have an unusual acidic domain at the carboxy terminus. Mol Gen Genet 255:359-371.
- Yoshida R, Hobo T, Ichimura K, Mizoguchi T, Takahashi F, Aronso J, Ecker JR, Shinozaki K. (2002). ABA-activated SnRK2 protein kinase is required for dehydration stress signaling in Arabidopsis. Plant Cell Physiol 43:1473-1483.
- Yoshida R, Umezawa T, Mizoguchi T, Takahashi S, Takahashi F, Schinozaki K. (2006). The regulatory domain of SRK2E/OST1/SnRK2.6 interacts with ABI1 and integrates ABA and osmotic stress signals controlling stomatal closure in Arabidopsis. J Biol Chem 281:5310-5318.
- Yoshida T, Fujita Y, Sayama H, Kidokoro S, Maruyama K, Mizoi J, Shinozaki K, Yamaguchi-Shinozaki K. (2010). AREB1, AREB2 and ABF3 are master transcription factors that cooperatively regulate ABRE-dependent ABA signaling involved in drought stress tolerance and require ABA for full activation. Plant J 61:672-685.
- Yoshihara T, Washida H, Takaiwa F. (1996). A 45-bp proximal region containing AACA and GCN4 motif is sufficient to confer endosperm-specific expression of the rice storage protein glutelin gene, GluA-3. FEBS Lett 383:2133-2138.
- Yoshiharu Y, Yamamoto YY, Yoshioka Y, Hyakumachi M, Obokata J. (2011). Characteristics of Core Promoter Types with respect to Gene Structure and Expression in *Arabidopsis thaliana*. DNA Res doi: 10.1093/dnares/dsr020
- Zhang W, Ruan J, Ho TH, You Y, Yu T, Quatrano RS. (2005). Cis-regulatory element based targeted gene finding: genome-wide identification of abscisic acid- and abiotic stress responsive genes in *Arabidopsis thaliana*. Bioinformatics 21:3074-81.
- Zhu JK. (2002). Salt and drought stress signal transduction in plants. Annu Rev Plant Biol 53:247-273.
- Zou C, Sun K, Mackaluso JD, Sedddon E, Jin R, Thomashow MF, Shiu S et al. (2011). *Cis*-regulatory code of stress-responsive transcription in *Arabidopsis thaliana*. Proc Natl Acad Sci USA 108:14992-14997.

- Zou H, Zhang X, Zhao JR, Yang Q, Wu Z, Wang F, Huang C. (2006). Cloning and characterization of maize ZmSPK1, a homologue to nonfermenting1-related protein kinase 2. Afr J Biotechnol 5:490-496.

Biography of Dr. Rajesh Mehrotra

Dr. Rajesh Mehrotra, is working as an Associate Professor in Department of Biological Sciences, Birla Institute of Technology and Science Pilani, Pilani campus, Rajsthan. Currently, he is head of the department of Biological Sciences. He obtained his master's degree in Agricultural Sciences from Banaras Hindu University with specialization in Genetics and Plant breeding. He obtained his Ph.D in the area of Plant molecular biology and Biotechnology from Banaras Hindu University, Varanasi (U.P.) while working at National Botanical Research Institute (NBRI), Lucknow. He was post doctoral fellow in the department of Gene Mechanism, Graduate School of Bio studies, Kyoto University, Japan. He worked on finding the mitotic function of topoisomerase II using genetical tools and time lapse microscopy of living yeast cell. His broad area of research interests are: promoter designing, host-pathogen interactions, cell division and lipase catalyzed biofuel production from microorganisms. Dr. Rajesh Mehrotra has handled four research projects, funded by Department of Science and Technology (DST), Aditya Birla Group of companies (ABG), Department of Biotechnology (DBT) and University Grants commission (UGC Major). He has published thirty research articles in journal of international repute and two chapters in edited books and has two patents to his credits. He is guiding three students for Ph.D and guided more than fifty undergraduate and post-graduate students for their research studies. He is a recipient of DST travel grant 2014. He was a visiting foreign scientist to Okinawa Institute of Science and Technology from July-Dec 2013 and June -July 2012, visiting foreign scientist Louisiana State University from Feb 2011 to August 2011, visiting foreign scientist Kyoto University, from May-June 2009, Dec 2008 – January 2009, June –July 2008 sponsored by Japanese Science and Technology Ministry. He has given several invited talks both nationally and internationally.

Biography of Purva

Mrs. Purva completed her M.Sc. Biotechnology from Gyan Vihar College University of Rajasthan, Jaipur and gets enrolled in the Ph.D programme of BITS, Pilani in January, 2009. She received financial assistance from various funding agencies such as BITS, Pilani, Department of Science and Technology (DST) and University Grant Commission-Basic Scientific Research (UGC-BSR) during Ph.D. She is recipient of DBT travel grant to attend international conference at Rome, Italy. She has also been involved in teaching various courses of Department of Biological sciences, Birla Institute of Technology and Science, Pilani. She has published research articles in renowned International journals and presented papers in various National and International conferences.

List of Publications by the Author

Rajesh Mehrotra, Purva Bhalothia, Prashali Bansal, Mahesh K Basantani, Vandana Bharti, Sandhya Mehrotra. (2014). Abscisic acid and abiotic stress tolerance - different tiers of regulation. J Plant Physiol 171:486-496.

Rajesh Mehrotra, Ipshita Zutsi, Sachin Sethi, Purva Bhalothia, Sandhya Mehrotra. (2013). Patterns and evolution of ACGT repeat *cis* element landscape across four plant genomes. BMC Genomics 14:203-213.

Purva Bhalothia, Anshu Alok, Sandhya Mehrotra, Rajesh Mehrotra. (2013). AACA element negatively interferes with expression of Protein Phosphatase 2C like gene promoter. American J Plant Sci 4:549-554.

Purva Bhalothia, Sandhya Mehrotra and Rajesh Mehrotra. (2013). Cloning and characterization of protein phosphatase 2C like promoter from *Arabidopsis thaliana*. International journal of Biosci Biochem Bioinfor 3:103-105.

Rajesh Mehrotra, Amit Yadav, Purva Bhalotia, Ratna Karan and Sandhya Mehrotra. (2012). Evidence for directed evolution of larger size motif in *Arabidopsis thaliana* genome. Sientific World J doi:10.1100/2012/983528.

Rajesh Mehrotra, Gauri Gupta, Riccha Sethi, Purva Bhalothia, Narayan Kumar, Sandhya Mehrotra. (2011). Designer promoter: an artwork of *cis* engineering. Plant Mol Biol 75:527-536.

Akhil Bajaj, Purva Bhalothia, Prabhat Nath Jha, Rajesh Mehrotra. (2010). Biodiesel production though lipase catalysed transesterification: An overview. Journal of Molecular Catalysis B: Enzymatic 62:9-14.

List of Presentations /Conferences /Work Shops /Attended

Full papers in Conference Proceedings

Purva Bhalothia, Anshu Alok, Sandhya Mehrotra, Rajesh Mehrotra.Cloning and characterization of protein phosphatase 2C like promoter from *Arabidopsis thaliana*. 3rd international conference on Bioscience, Biochemistry and Bioinformatics. February 24th-25th, 2013, Rome Italy.

Papers presented in Conferences, Seminars, Workshops, Symposia Conference Proceedings:

Vipul Dwivedi, Sachin Saxena, Purva Bhalothia, Sanddhya Mehrotra and Rajesh Mehrotra. (2015). Genome wide analysis of W-box in *Arabidopsis thaliana* reveals skewed distribution of $TGAC_{N6}TGAC$ sequence in chromosome 1. 5th Annual International Conference on Advances in Biotechnology (BIOTECH 2015) IIT, March 13th -15th, Kanpur.

Purva Bhalothia, Anshu Alok, Bhaskar Sharma, Sandhya Mehrotra and Rajesh Mehrotra. (2014). Cloning and characterization of protein phosphatase 2C (*PP2C*) like promoter from *Arabidopsis thaliana*. International Conference and XI convention of The Biotech Research Society, India on Emerging Trends in Biotechnology (ICETB-2014). The convention centre, of Environmental Sciences, Jawaharlal Nehru University and The Biotech Research Society, November 6th -9th November New Delhi.

Purva Bhalothia, Anshu Alok, Sidharth Tiwari and Rajesh Mehrotra. (2012). AACA element negatively regulates expression of protein phosphatase 2C (*PP2C*) like promoter in *Arabidopsis thaliana*. International conference on industrial biotechnology (IXthconvention of the biotechnology research society, India) 21st-23rd November, Patiala, Punjab

Rajesh Mehrotra, Ipshita Zutshi, Sachin Sethi, Purva Bhalothia and Sandhya Mehrotra. (2012). Patterns and evolution of ACGT repeat *cis* element landscape across four plant gemones. Next generation Plant Science 26th -28th September, Cologne, Germany.

Purva Bhalothia, Prasali Banshal and Rajesh Mehrotra. (2011). Cloning and characterization of Protein phosphatase 2C-like promoter from *Arabidopsis thaliana*. International conference on New Horizons in Biotechnology 21st -24th November, Trivandrum, Kerala.

Rajesh Mehrotra, Amit Yadav and Purva Lohan. (2011). Evidence for directed evolution of larger size motif in *Arabidopsis thaliana*. The 18th Penn State Plant Biology Symposium 18th -21st May Pennsylvania State University, PA, USA.

Rajesh Mehrotra, Sandhya Mehrotra, Akhil Bajaj, Purva Lohan. (2009). Promoter activation by ACGT in response to salicylic acid and abscisic acids is differentially regulated by the spacing between two copies of the motif: in the international conference on emerging trends in Biotechnology (ETBT-2009) and 6th annual convention on Biotechnology, INDIA (BRSI) Banaras Hindu University 4th -6th December Varanasi, Uttar Pradesh.

Journal of Plant Physiology 171 (2014) 486–496

Contents lists available at ScienceDirect

Journal of Plant Physiology

journal homepage: www.elsevier.com/locate/jplph

Molecular Biology

Abscisic acid and abiotic stress tolerance – Different tiers of regulation

Rajesh Mehrotra[a,b,1], Purva Bhalothia[a,1], Prashali Bansal[a,c], Mahesh Kumar Basantani[d], Vandana Bharti[e], Sandhya Mehrotra[a,*]

[a] Department of Biological Sciences, Birla Institute of Technology & Sciences, Pilani, Rajasthan 333031, India
[b] G_o Unit, Okinawa Institute of Science and Technology, 1919-1, Onnason, Okinawa, Japan
[c] Cancer Science Institute, National University of Singapore, Singapore, Singapore
[d] Division of Endocrinology, University of Pittsburgh, 200 Lothrop Street, BST E1140, Pittsburgh, PA 15261, USA
[e] Department of Biotechnology, St. Columba's College, Vinoba Bhave University, Hazaribagh, India

ARTICLE INFO

Article history:
Received 10 August 2013
Received in revised form
12 December 2013
Accepted 13 December 2013

Keywords:
ABA signaling
Abiotic stress
Abscisic acid
ACGT *cis* element
Protein phosphatase 2C

SUMMARY

Abiotic stresses affect plant growth, metabolism and sustainability in a significant way and hinder plant productivity. Plants combat these stresses in myriad ways. The analysis of the mechanisms underlying abiotic stress tolerance has led to the identification of a highly complex, yet tightly regulated signal transduction pathway consisting of phosphatases, kinases, transcription factors and other regulatory elements. It is becoming increasingly clear that also epigenetic processes cooperate in a concerted manner with ABA-mediated gene expression in combating stress conditions. Dynamic stress-induced mechanisms, involving changes in the apoplastic pool of ABA, are transmitted by a chain of phosphatases and kinases, resulting in the expression of stress inducible genes. Processes involving DNA methylation and chromatin modification as well as post transcriptional, post translational and epigenetic control mechanisms, forming multiple tiers of regulation, regulate this gene expression. With recent advances in transgenic technology, it has now become possible to engineer plants expressing stress-inducible genes under the control of an inducible promoter, enhancing their ability to withstand adverse conditions. This review briefly discusses the synthesis of ABA, components of the ABA signal transduction pathway and the plants' responses at the genetic and epigenetic levels. It further focuses on the role of RNAs in regulating stress responses and various approaches to develop stress-tolerant transgenic plants.
© 2013 Elsevier GmbH. All rights reserved.

Introduction

Plants, though continuously exposed and challenged by environmental cues, are capable of surviving by means of adaptations at the molecular, cellular and chromatin level of organization (Kim et al., 2010). Drought, salinity and temperature variations are among the major abiotic stresses, which hamper plant growth and productivity and often cause a series of morphological, physiological and biochemical changes. One such example can be seen in plants adapted to cold conditions, which show up-regulation of genes involved in the synthesis of unsaturated fatty acids, while those exposed to drought show a decrease in the number and surface area of leaves and an increase in root length (Shinozaki et al., 2003). Abscisic acid (ABA) is a widely studied phytohormone, and its role in ameliorating abiotic stress in plants is well established. In plants, ABA is of paramount significance as it plays an important role in mediating host responses to both biotic and abiotic stresses. It not only helps in seed maturation and seed dormancy but also gives desiccation tolerance to the cells during dehydration stress and is hence aptly called as a stress hormone.

ABA: evolutionary significance, synthesis and stress regulatory mechanisms

ABA: ubiquitous presence across different lineages

ABA is present in cyanobacteria, algae, bryophytes, fungi, lichens and higher plants. Out of eleven species of cyanobacteria tested for ABA synthesis under varied stress conditions, four species showed an increase in the level of ABA when subjected to salt stress (reviewed in Wolfram, 2010). Even though the majority of bacterial species does not exhibit ABA synthesis, a study has reported the presence of ABA in endophytic bacteria, which are present in the roots of *Helianthus annuus* under water stress conditions (Forchetti et al., 2007). In the cyanobacterium *Anacystis nidulans*, exogenous ABA has been shown to increase growth (Ahmad et al., 1978; Wolfram, 2010). ABA is believed to have protective functions

Abbreviations: PP2C, protein phosphatase 2C; SNF, sucrose non-fermenting; SnRK2, SNF1-related protein kinase 2.

* Corresponding author. Tel.: +91 1596244183; fax: +91 1596245073.
E-mail address: sandhyamehrotrabits@gmail.com (S. Mehrotra).
[1] Both these authors have contributed equally to the manuscript.

0176-1617/$ – see front matter © 2013 Elsevier GmbH. All rights reserved.
http://dx.doi.org/10.1016/j.jplph.2013.12.007

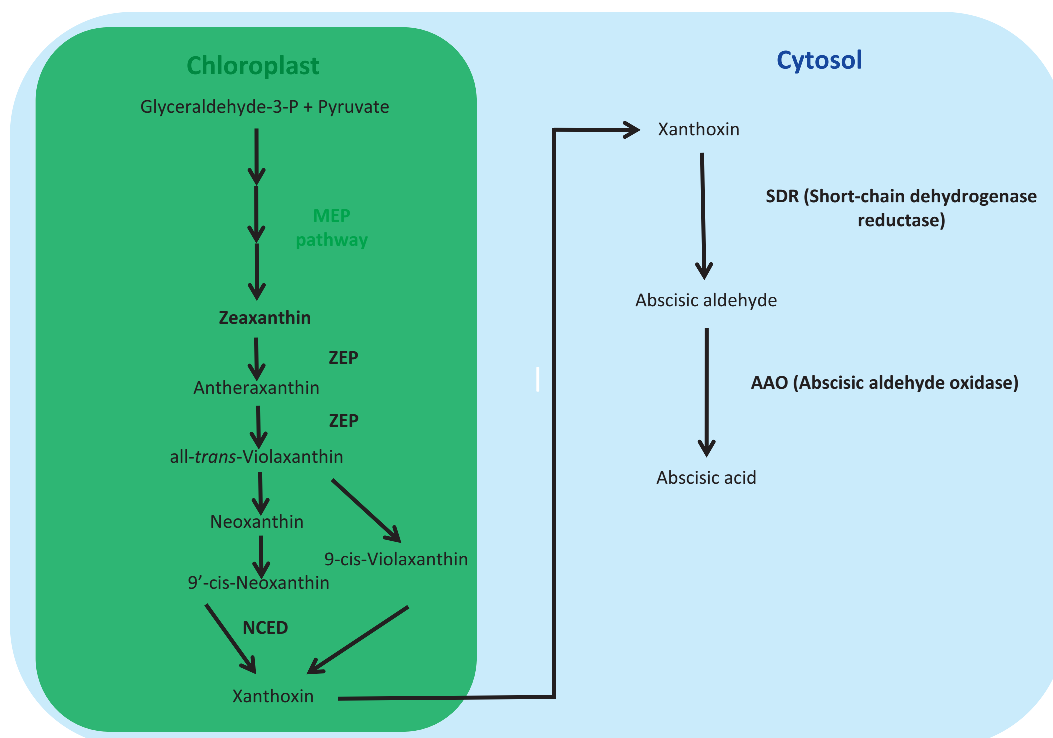


Fig. 1. Cellular synthesis of abscisic acid: ABA biosynthesis occurs partly in chloroplasts, and partly in the cytoplasm. The preliminary steps involve generation of zeathanxin from carotenoids via the MEP pathway. ABA synthesis starts from the breakdown of zeaxanthin to all-*trans* violaxanthin under the catalysis of zeaxanthin epoxidase (ZEP). The all-*trans* violaxanthin, an intermediate product, is catalyzed to xanthoxin in the presence of 9-*cis* epoxycarotenoid dioxygenase (NCED) in the chloroplast which is transported to the cytoplasm. In the cytoplasm, xanthoxin is first converted to abscisic aldehyde by short chain dehydrogenase reductase (SDR) and finally abscisic aldehyde is converted to ABA by the action of abscisic aldehyde oxidase (AAO).

in lower plants as it shields *Chlamydomonas reinhardtii* cells from photoinhibition and oxidative damage caused by salt and osmotic stress (Saradhi et al., 2000). ABA has also been found to regulate opening and closing of stomata in mosses and sporophytes of hornworts elucidating yet another function. More than 100 species of algae have also been reported for the presence of ABA though different algae respond differently to the applied stress (reviewed by Wolfram, 2010). With the evolution of liverworts, mosses and ferns, the endogenous levels of ABA have been observed to rise. However, the quantity of ABA has been shown to be very low in lower classes of plants.

ABA biosynthesis

ABA biosynthesis occurs via the mevalonic-acid independent pathway in plastids ABA, the 15-carbon isoprenoid plant hormone, is synthesized in plastids via the mevalonic acid-independent pathway called the 2-C-methyl-d-erythritol-4-phosphate (MEP) pathway. The 15 carbon atoms of ABA are derived from the cleavage of C_{40} carotenoids originating from the MEP pathway (Nambara and Marion-Poll, 2005). The biosynthesis of ABA starts in the plastids from the C_{40} carotenoid zeaxanthin and ends in the cytosol with the formation of abscisic aldehyde, which in turn, is oxidized into ABA (Seo and Koshiba, 2002). The mechanism of ABA biosynthesis is shown in Fig. 1.

In the plastid: from zeaxanthin to xanthoxin

The first step of the ABA-biosynthetic pathway is the conversion of zeaxanthin to all-*trans*-violaxanthin, catalyzed by zeaxanthin epoxidase (ZEP). Antheraxanthin is formed as an intermediate in this reaction. Thereafter, the conversion of all-*trans*-violaxanthin to 9-*cis*-violaxanthin or 9′-*cis*-neoxanthin occurs. The enzyme(s) involved in this reaction are not known (Seiler et al., 2011). The oxidative cleavage of 9-*cis*-violaxanthin and/or 9-*cis*-neoxanthin, catalyzed by 9-*cis*-epoxy carotenoid dioxygenase (*NCED*), leads to the formation of a C_{15} product, xanthoxin, and a C_{25} metabolite. This reaction is considered to be the rate-limiting step, and *NCED* is the key enzyme in ABA biosynthesis (Tan et al., 1997).

In the cytosol: from xanthoxin to ABA

Xanthoxin is transferred to the cytosol where it is converted to ABA via two enzymatic reactions. In the first step, xanthoxin is converted to abscisic aldehyde by an enzyme belonging to short-chain dehydrogenase/reductase (SDR) family. The gene responsible for this has been identified in *Arabidopsis thaliana*, named *AtABA2* (Gonzalez-Guzman et al., 2002). The final step of ABA biosynthesis is the oxidation of abscisic aldehyde to ABA. This reaction is catalyzed by an abscisic aldehyde oxidase (AAO).

ABA transport

ABA transport is both passive and mediated by ABA transporters

The level of endogenous ABA is maintained by interaction between anabolic and catabolic pathways. The catabolic products like ABA glucosyl ester and phaseic acid are stored in the vacuole or apoplast pool. Under drought stress, ABA is released from its conjugate form, which is facilitated by β-glucosidase, and is transported to guard cells where its accumulation leads to stomatal closure. One such enzyme, AtBG1, a type of β-glucosidase, has been reported in *A. thaliana* (Lee et al., 2006a).

It is not clear if specific transporters facilitate the movement of ABA between cells. It has been shown that ABA transport to the outside of cells in response to pH changes can occur without a specific transporter (Seo and Koshiba, 2011). Similarly, at the site of action, ABA uptake into cells could be mediated by diffusion or by

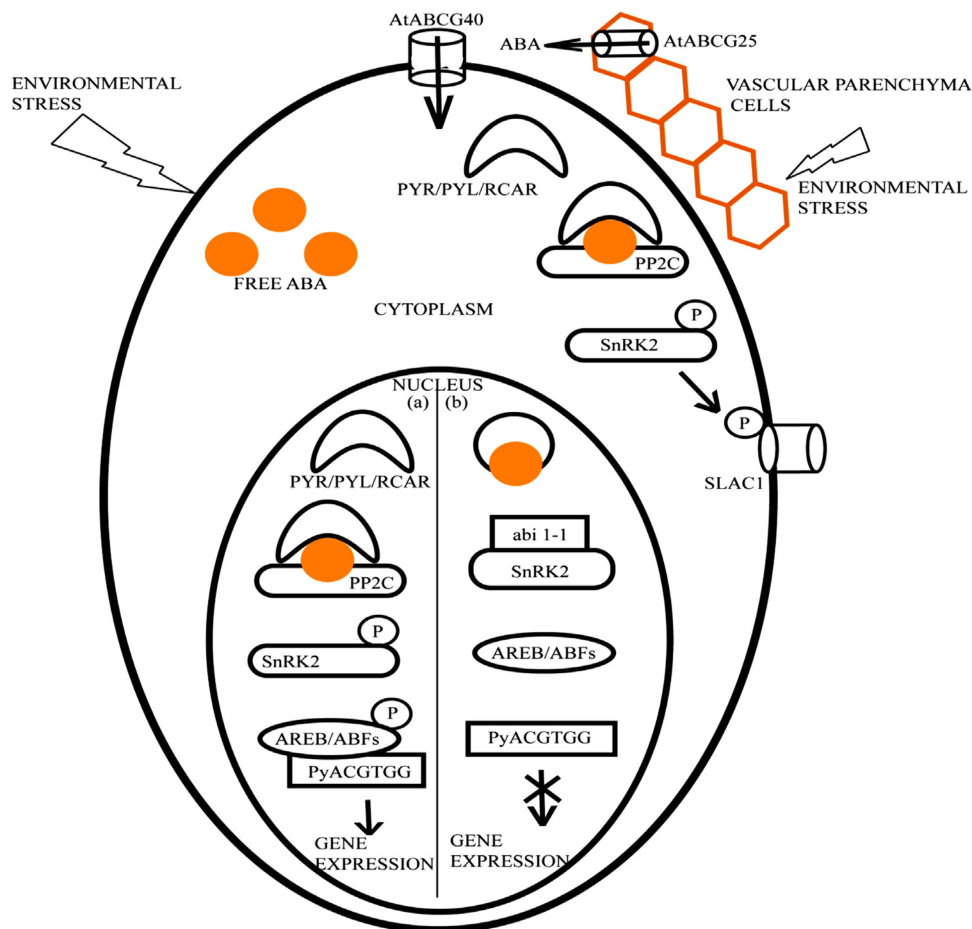


Fig. 2. Brief outline of ABA signaling and its transport "as shown in *Arabidopsis thaliana*": The components involved in ABA signaling include – receptor PYR/PYL/RCAR, a negative regulator PP2C, a positive regulator SnRK2s and transcription factors (AREB/ABFs). The ABA signaling occurs in both, cytoplasm and nucleus. Under abiotic stress conditions (and developmental conditions) the synthesis of ABA increases in vascular parenchyma cells. ABA is transported via AtBCG25 (from parenchyma cells) and AtBCG40 transporters (to target cells). In the presence of ABA, a complex is formed between ABA, receptor and PP2C (ABA-PYR/PYL/RCAR-PP2C) which prevents PP2C in carrying out its inhibitory activity. SnRK2 gets phosphorylated and further activates bZIP transcription factors such as AREB/ABFs by phosphorylation which subsequently induces the expression of target genes. In ABA insensitive condition, ABA insensitive 1-1 (Abi1-1) interacts with SnRK2, hence inhibiting its interaction with ABA-PYR/PYL/RCAR complex and thereby preventing the expression of the downstream target genes.

specific transporters. Therefore, both carrier-mediated and passive ABA transport mechanisms are suggested to exist.

Studies in *A. thaliana* have revealed several ABA transporter genes in vascular tissues belonging to the ATP-binding cassette (ABC) family. One such gene is *AtABCG25*, which codes for an ABA transporting protein (Kuromori and Shinozaki, 2010). It has been suggested that *AtABCG25* mediates ABA export from the inside to the outside of the cell. On the other hand, another transporter *AtABCG40*, expressed in guard cells, has been reported to be responsible for importing ABA from the outside to the inside of cells. Both these genes are reported to be involved in ABA signaling (Kuromori and Shinozaki, 2010).

ABA signaling

ABA plays an important role in various aspects of plant stress responses, embryo maturation, seed dormancy, germination, root branching, stomatal closure, *etc*. A great many studies have attempted to shed light on the mechanisms and core components of ABA signaling (reviewed in Cutler et al., 2010; Umezawa et al., 2010; Liu, 2012). There are three major components of ABA signaling: pyrabactin resistance (PYR)/PYR1-like (PYL)/regulatory component of ABA receptor (RCAR), protein phosphatase 2C (PP2C; a negative regulator) and (sucrose non-fermenting) SNF1-related protein kinase 2 (SnRK2: a positive regulator). These are assembled as a double negative regulatory system and form a signaling complex known as 'ABA signalosome' (Fig. 2).

PP2C is a global negative regulator of ABA signaling

Phosphorylation and de-phosphorylation, mediated by kinases and phosphatases, respectively, help in the transmission of signals from external to internal environment of the cell, and thereby bring about physiological changes.

In *A. thaliana*, 76 genes encode for PP2Cs and are divided into 10 groups (A-J). PP2Cs belong to the PPM family of phosphatases (Mg^{2+}-dependent phosphatases). PP2Cs like ABA insensitive 1 (ABI1), ABI2 (homolog to ABI1), AHG3/PP2CA, ABA-Hypersensitive Germination1 (AHG1), hypersensitive to ABA1 (HAB1) and HAB2 are members of the group A (Saez et al., 2004; Schweighofer et al., 2004; Yoshida et al., 2006). The members of the group A have definite roles in different tissues and organs, and demonstrate tissue-specific expression patterns.

ABA-Insenstive1 (ABI1) and ABI2 were identified in a genetic screen for ABA-insensitive *A. thaliana* mutants. The *abi1-1* and *abi2-1* mutants show ABA insensitivity in various tissues and developmental stages, suggesting that PP2C acts as a negative regulator of ABA signaling (Leung et al., 1997; Rodriguez et al., 1998; Umezawa et al., 2010). The ABA signaling regulators HAB1 and

HAB2 were isolated based on sequence similarity to ABI1 (Saez et al., 2004). AHG1 and AHG3/AtPP2CA were cloned from genetic screens of *A. thaliana* and a yeast complementation test. It has been shown that all of the loss-of-function mutants of group A PP2Cs exhibit significant hypersensitivity, which establishes them as the major negative regulators of ABA signaling (Umezawa et al., 2010). The negative regulatory role of PP2C in ABA signaling is now well founded, and this has been shown in various other plant species. This is further corroborated by the fact that double or triple PP2C knockout mutants show increased ABA hypersensitivity.

Gene expression data and genetic analysis has indicated a dominant role of HAB1, ABI1, ABI2 and PP2CA in ABA signaling pathways in both seeds and vegetative tissues. Recessive *Hab1-1* mutants show enhanced ABA responsive gene expression, enhanced ABA-mediated stomatal closure and ABA-hypersensitivity leading to seed germination and growth inhibition. Conversely, HAB1 reduces ABA sensitivity upon constitutive expression. These results are consistent with the fact that HAB1 negatively regulates ABA signaling (Saez et al., 2004).

Negative regulation of MAPK (MAPK4 and MAPK6) by *A. thaliana* PP2C, AP2C1 (Schweighofer et al., 2004) also potentially links PP2Cs with cold and drought stress responses. A PP2C encoding gene MP2C has been indicated to be responsible for inactivation of stress-activated MAPK (SAMK) pathway, by accumulation of MP2C phosphatase which inhibits further signal transmission (Meskiene et al., 1998) hence confirming the role of PP2Cs as global negative regulators of ABA signaling.

SnRK2 is a positive regulator of ABA signaling

The identification of PP2C indicated that protein phosphorylation events should be important components of the ABA signaling pathway. Many protein kinases were actually found to be involved in ABA signaling. The first SnRK2 cDNA clone (*PKABA1*) was isolated from an ABA-treated wheat embryo cDNA library (Anderberg and Walker-Simmons, 1992). In fava bean, the SnRK2 family was identified as ABA-Activated Protein Kinase (AAPK) (Li et al., 2000) and *A. thaliana* SRK2E/Open STomata1 (OST1)/SnRK2.6 (Yoshida et al., 2002, 2006).

SnRKs are serine/threonine protein kinases grouped under SNF1 family, which also includes yeast's SNF1 kinases and AMPKs from mammals. These are positive regulators of ABA mediated signaling, responsible for the phosphorylation of bZIP transcription factors known to bind *cis*-regulatory motif having ACGT as a core sequence. SnRKs can be classified into 3 sub-divisions: SnRK1, SnRK2 and SnRK3 (Kulik et al., 2011).

The SnRK2 family is the key regulator of plant response to abiotic stress. SnRK2s are monomeric serine/threonine protein kinases. SnRK2s are divided into 3 subgroups based on the affinity toward ABA *viz*. group I, II and III. There are 10 SnRK2 members in *A. thaliana*, SRK2A-SRKJ or SnRK2.1-SnRK2.10. Group I kinases do not respond to ABA, group II does not, or weakly, respond to ABA, and group III strongly responds to ABA. Experimental data suggest functional overlap between group II and group III SnRK2s inactivating bZIP type transcription factor ABA responsive element binding factor/ABA binding factor (AREB/ABF). Group III is believed to be the key regulator in ABA-dependent pathway of gene expression, whereas group II may be involved by a parallel pathway which means that group II SnRK2s could be variants of group III kinases that are not activated or weakly activated by ABA. Phylogenetic studies have also revealed group II as an intermediate during transition of group I to group III (for review refer Hrabak et al., 2003; Kulik et al., 2011).

Phosphorylation at the post-translational level and gene response to changes in the environment regulate SnRK2 expression. Members of SnRK2 family are found in algae (Ballester et al., 2008), monocots like rice, sorghum (Kobayashi et al., 2004; Li et al., 2010), maize (Huai et al., 2008) and wheat (Mao et al., 2010; Zhang et al., 2010), and dicots like tobacco (Kelner et al., 2004), soybean, *A. thaliana* (Monks et al., 2001; Boudsocq et al., 2004) and fava bean (Li et al., 2000). Enhanced expression of SnRK2 genes in maize occurs in response to NaCl and cold while heat suppresses it (Huai et al., 2008). Induced expression of SnRK2 genes in soybean occurs in response to hyperosmolarity, and the regulation of expression of some SnRK2s in response to dark-light cycle has also been reported (Monks et al., 2001).

SnRK2 kinases encode a protein of about 40 kDa. The N-terminal of the amino acid sequence is a highly conserved kinase domain, and the C-terminal is a variable regulatory domain with stretches of glutamate (Glu) or aspartate (Asp) which functions in protein interactions. Domain I, a subdomain of C-terminal is independently activated by osmotic stress whereas domain II is ABA dependent and is responsible for activation by ABA and interaction with ABI1 (Yoshida et al., 2010). Group I has a glutamate rich region in domain II, whereas group II/III have aspartate rich regions in this domain, which accounts for their differential response toward ABA activation (Mizoguchi et al., 2010; Yoshida et al., 2010; Fujita et al., 2013).

Three members of SnRK2 subgroup III namely SnRK2D/SnRK2.2, SRK2E/OST1/SnRK2.6 and SRK2I/SnRK2.3 are major regulators of plant response to ABA. SRK2E/OST1/SnRK2.6 is expressed in guard cells leading to stomatal closure in the presence of ABA. Other SnRK2s like SRK2D/SnRK2.2 and SRK2I/SnRK2.3 are expressed in seeds and vegetative tissues. Although partial functional redundancy has been shown for all of them but their importance in regulating water stress as well as ABA-dependent developmental processes cannot be neglected. A triple mutant in *A. thaliana* not only had disrupted stomatal closure but also had down regulated ABA and water stress induced genes resulting in ABA insensitivity and reduced drought tolerance (Yoshida et al., 2002, 2006; Fujita et al., 2009; Fujii and Zhu, 2009; Nakashima et al., 2009a). This mutant also showed up-regulation of jasmonic acid (JA) dependent genes, genes regulating flowering and those related to photosynthesis and tetrapyrrole synthesis. Researchers have also reported down regulation of almost all genes encoding dehydration-responsive late embryogenesis abundant (LEA) proteins and group A PP2Cs under water stress conditions in this mutant indicating the essentiality of these three genes in controlling seed development and dormancy (Nakashima et al., 2009a).

SRK2E/OST1/SnRK2.6 majorly regulates ABA dependent stomatal closure in water deficit conditions. Its mutant is unable to withstand low humidity and shows a wilted phenotype indicating its crucial role in combating environmental stress (Yoshida et al., 2002). It also regulates reactive oxygen species (ROS) production in the cell wall through phosphorylation of the two plasma membrane NADPH oxidases (*RbohD* and *RbohF*) in an ABA-dependent manner leading to stomatal closure (Kwak et al., 2003). Apart from this, OST1 has also been shown to positively regulate stress response genes like *RD29B* and *RD22* (Yoshida et al., 2002).

ABA and osmotic stress activate SRK2E/OST1 mediated stomatal closure by two independent pathways namely ABA and OS-dependent pathways as shown by experiments on ABA-insensitive or ABA-deficient mutants. C-terminal domain II is believed to be responsible for ABA-dependent activation of OST1 while low humidity or OS-mediated (ABA-independent) activation transduces stress signals to induce stomatal closure. Thus, both pathways are required to respond to sudden water stress by causing complete stomatal closure (Yoshida et al., 2006).

SRK2E/OST1 in *Arabidopsis* also interacts with ABI1/ABI2, where both have distinct roles in controlling its activity. Experimental results have indicated two roles of ABI1 in ABA dependent pathway. It acts as a negative regulator of SRK2E in ABA signaling and as a positive regulator in activation of SRK2E in low humidity.

Interaction of SRK2E with ABI1 and activation by ABA requires the presence of domain II, whereas in its absence, SRK2E is activated by low humidity, partially complementing the wilted phenotype of *srk2e*. A proposed explanation is the interaction of domain II with a downstream factor triggered by low humidity leading to stomatal closure. It has also been reported that ABI-1 protein acts as a dominant negative inhibitor, which blocks interactions of downstream proteins with domain II of SRK2E (Yoshida et al., 2006).

SnRK2s and downstream effectors of ABA action

SnRK2s are central components of ABA signaling pathways. The *SnrK2.2/2.3/2.6* triple mutant plants are nearly insensitive to ABA, suggesting that most of the molecular actions of ABA are triggered by the SnRK2s-mediated phosphorylation of substrate protein. Only a few substrate proteins of SnRK2s are known. In an effort to identify additional substrates of SnRK2s, Wang et al. (2013) used quantitative phosphoproteomics to compare the global changes in phosphopeptides in wild type (WT) and *snrK2.2/2.3/2.6* triple mutant seedlings in response to treatment. They reported that ABA could increase the phosphorylation of 166 peptides and decrease the phosphorylation of 117 in WT seedlings. They further reported that with *snrK2.2/2.3/2.6* triple mutant, 84 of the 166 peptides, representing 58 proteins, could not be phosphorylated. These results shed new light on the role of the SnRK2 protein kinases and on the downstream effectors of ABA action.

ABA-PYR/PYL/RCAR receptor

ABA receptors encode STAR domain/Bet V allergen super family proteins. Genome wide studies of *A. thaliana* have revealed presence of 14 highly conserved PYR/PYL/RCAR protein encoding genes out of which 9 are major in interactions with ABI1 (Ma et al., 2009; Park et al., 2009; Hubbard et al., 2010; Nishimura et al., 2010).

It has been reported that the PYR/PYL/RCAR receptor, which is a direct ABA receptor, interacts with PP2C in the presence of ABA, which inhibits phosphatase activity of PP2C and acts as negative regulator. These receptors play an important role and are observed in the case of quadruple mutants *pyr1*, *pyl1*, *pyl2* and *pyl4*, which are insensitive toward ABA. On the other hand, over-expression of RCAR1/PYL1, PYL5/RCAR8 or PYL8/RCAR3 gives more tolerance toward drought stress in *A. thaliana*. Even in case of ABI1 and ABI2, HAB1 activity is inhibited by stereo-specific RCAR1, RCAR3, RCAR8, RCAR11 and RCAR12 receptors in the presence of ABA, but not observed in the absence of receptors (Ma et al., 2009; Park et al., 2009; Nishimura et al., 2010).

PYR/PYL/RCAR interaction with ABI1 can be both ABA-dependent as in case of PYR1 and PYL1-4 or constitutive as shown in case of PYL5-12. Similar results have been obtained for *A. thaliana* (Nishimura et al., 2010). Bimolecular fluorescence complementation (BiFC) and co-immunoprecipitation assays have established PYL5 as an HAB1 antagonist. PYL5 inhibits HAB1, ABI1 and ABI2 functions in an ABA dependent manner activating ABA signaling cascade (Ma et al., 2009; Park et al., 2009).

Resolution of ABA binding with PYR1/PYL1/PYL2 has revealed the presence of a proline cap and a leucine lock. ABA binding leads to closing of the cap, limiting the enzyme within the cavity, and preventing solvent contact while the lock rearranges sealing the lid over ABA. The conformational changes brought about by this binding create an oval site for PP2C interaction further stabilizing the structure. This shields the catalytic site of PP2C leading to its inactivation by PYR/PYL/RCAR (Ma et al., 2009; Nishimura et al., 2009; Park et al., 2009). The crucial role of this cap and lock can be indicated by reduced ABA sensitivity in the mutant. Also disruption of ABA-PYR1 interactions by mutating the residues, which make direct contact with ABA highlights the importance of ABA-induced conformational (Nishimura et al., 2009) changes .

Antoni et al. (2013) showed that ABA-dependent inhibition of PP2Cs by PYR/PYLs is required for the perception of moisture gradient. The *pyl8* mutant showed reduced sensitivity to ABA mediated root growth inhibition. They discovered that this was due to lack of PYL8-mediated inhibition of several group A PP2Cs. It has been demonstrated *in vivo* that PYL8 interacted with at least five PP2Cs, namely HAB1, HAB2, ABI1, ABI2 and PP2CA/AHG3. Consistent with the above observations are the facts that ABA-hypersensitive *pp2c* quadruple mutant showed enhanced hydrotropism, whereas an ABA-insensitive sextuple *pyr/pyl* mutant showed reduced hydrotropic response.

Changes at gene expression level

Onset of drought and salt stress induces the accumulation of ABA, which then acts as a stress signal. ABA regulates expression of many stress-responsive genes grouped as ABA responsive genes. *Cis*-elements such as ABA-responsive element (ABRE) are present up stream of the ABA-dependent stress responsive genes, which have a core sequence PyACGTGGC along with one or two coupling elements (CE) also found upstream of stress-responsive genes. Genes having ABRE sequence belong to G box element. This sequence is target of many members of the bZIP transcription factor family known as ABA-responsive element binding protein (AREB). A major finding is the importance of distance between the ACGT *cis*-elements. Two ACGT *cis* elements separated by five nucleotides have been found to be induced in response to salicylic acid (SA), and similarly motifs separated by twenty-five base pairs are induced in response to ABA (Shen et al., 2004; Mehrotra and Mehrotra, 2010).

In accordance with the basic rules, the probability of occurrence of two ACGT motifs separated by a single nucleotide (ACGT$_{N1}$ACGT) has been found to be more as compared to that of $_{N2}$, $_{N3}$, $_{N4}$ and $_{N5}$ separating nucleotides in *A. thaliana* genome. Statistically, the occurrence of ACGT$_{N1}$ACGT is 1605 but no biological role has been assigned or reported to $_{N1}$, $_{N2}$, $_{N3}$ and $_{N4}$ so far. Only $_{N5}$ and $_{N25}$ have been found to be associated with biotic and abiotic stress responses (Mehrotra and Mehrotra, 2010; Mehrotra et al., 2012, 2013). A study has also shown that flanking sequences play an equally important role in defining the binding efficiency of bZIP transcription factors (Foster et al., 1994). Nucleotides at +2/−2, +3/−3 and +4/−4 positions define the specificity and affinity for binding of bZIP class of transcription factors. Presence of G at +2/−2 positions (G-box), presence of A/C at +3/−3 (C-box) and presence of G/T at +4/−4 are preferred over other nucleotides. Hence it could be reasonably concluded that two ACGT motifs in close proximity cannot serve as flanking sequences for each other (Foster et al., 1994; Mehrotra et al., 2012).

A similar study (Suzuki et al., 2005) considered 5 classes of stress-responsive gene promoters namely ABA-regulated (ABR), ABA-dependent (ABD), VP1-dependent (VPD), VP1 and ABA-dependent (VAA), VP1 or ABA-dependent (VOA). In all these classes, distribution of flanking sequences of ACGT motifs was studied. It has been reported that out of 64 variants, 23 variants were distributed among these 5 stress-responsive promoters. GAC as a flanking sequence of ACGT has been observed in all classes majorly in VAA. Another variant GCCACGT has been dominantly reported in VAA, VPD and in two sub-classes of VP1-dependent class of promoters. Similarly, AACACGT and CCCACGT were reported in ABA-dependent and ABA-regulated gene promoters, respectively. Genes regulated by cold also exhibit the same whereas CCGACGT sequence was found to be predominating in up-regulated cold genes along with ABA and VP1-regulated genes (Suzuki et al., 2005).

Different members of the bZIP class of transcription factors bind to the ACGT core specifically. This binding also depends upon the sequences flanking the core region. Flanking sequences are also

responsible for differential distribution of stress-responsive promoters across the genome, which could be one of the reasons behind the often-observed differences in ABA regulation and/or dependency (Riechmann et al., 2000; Mehrotra and Mehrotra, 2010). The flanking sequences might influence the binding specificity of different members of bZIP class to the ACGT core sequence (Foster et al., 1994; Suzuki et al., 2005).

Transcription factors involved in stress signaling

The interaction between specific transcription factors and their *cis*-elements, causes expression of stress inducible genes. Major transcription factor (TF) families, which are involved in the regulation of abiotic stress responses, are bZIP, MYB, MYC, NAC, ERF and DREB/CBF (C repeat binding factor).

The ABRE-binding (AREB) proteins or ABRE-binding factors (ABFs) encode bZIP transcription factors among which AREB1/ABF2, AREB2/ABF4 and ABF3 are induced by dehydration, high salinity or ABA treatment in vegetative tissues, and are involved in enhanced drought stress tolerance (Riechmann et al., 2000; Yoshida et al., 2010).

AREB1 requires ABA for full activation, and its activity is regulated by ABA-dependent multi-site phosphorylation of the conserved domains by SnRK2. SnRK2 protein kinases phosphorylate serine/threonine residues at R-X-X-S/T sites in the conserved regions of AREB/ABF proteins causing its inactivation which is further promoted by ABA-mediated inactivation of the PP2Cs, known to negatively regulate protein kinases (Nakashima et al., 2009a).

GmbZIP1, a member of AREB subfamily has been identified (Gao et al., 2011) in soybean cv.Tiefeng8. Its expression was found to increase in the presence of ABA, drought, high salt and low temperature. Stomatal closure under stress was induced because of the over expression of GmbZIP1 in transgenic plants, leading to enhanced tolerance to abiotic stresses.

Members of bZIP family, like ABP9, have been found to be associated with enhanced photosynthetic capacity of plants in drought and heat stresses (Zhang et al., 2008). Another bZIP transcription factor, OsABF1, reported (Hossain et al., 2010) in roots of the seedlings and shoots of rice has been found to be involved in abiotic stress responses and ABA signaling. In tomato, a bZIP transcription factor SIAREB1 participates in abiotic stress by regulating oxidative-stress related proteins, LEA proteins and lipid transfer proteins (LTPs) (Orellana et al., 2010).

MYC and MYB also play an important role in ABA signaling by activating some stress-inducible genes like *RD22* (Abe et al., 2003). MYB in *A. thaliana* namely *AtMYB60*, *AtMYB44* and *AtMYB15* have been found to be involved in stomatal closure, and protects against drought and salt stresses.

NAC family studies reveal that cDNAs encoding NAC-like proteins namely ANAC019, ANAC055, ANAC072 bind to ERD1 (early response to dehydration 1) present in the promoter region of stress inducible genes. Enhanced expression of these transcription factors ANAC019, ANAC055, ANAC072 and SNAC1, OsNAC6/SNAC2, ONAC045 leads to tolerance against drought stress in *A. thaliana* and rice vegetative tissues, respectively (Nakashima et al., 2009b). In wheat, *TaNAC69* has been found to be involved in the regulation of drought stress responses (Xu et al., 2011).

AtNAP (a NAC family transcription factor in *Arabidopsis*) and senescence associated gene113 (*SAG113*) (a PP2C family member localized in Golgi apparatus) have been reported to get activated in the presence of ABA. Knock out studies have shown that expression of *SAG113* is dependent on AtNAP (Zhang and Gan, 2012). In the presence of ABA, binding of AtNAP to 9 bp core sequence (CACGTAAGT) present in the promoter of *SAG113* leads to stomatal closure, creating a dehydration condition, which ultimately leads to leaf senescence (Zhang and Gan, 2012).

The central concept involved in ABA signaling is the inhibition of phosphatase activity of PP2C by ABA bound PYR/PYL/RCAR receptors. This leads to the subsequent activation of SnRK2 and phosphorylation of downstream targets such as AREB/ARF basic domain leucine zipper (bZIP), slow (S) anion channels and others, leading to stress induced response in plants (Ma et al., 2009; Park et al., 2009). Activation of ABA-responsive genes requires the ABA responsive element (ABRE). Three members of AREB/ABF family viz., AREB1, AREB2 and AREB3 have been identified to be expressed in response to ABA and water stress. AREBs require ABA for their full activation and show certain level of redundancy (Yoshida et al., 2010). The AREB/ABF encode bZIP transcription factors and belong to group A subfamily (Riechmann et al., 2000). Several late embryogenesis abundant (LEA) genes, group A PP2C genes and transcription factors as downstream targets of AREB/ABF have also been revealed (Yoshida et al., 2010). By generating a triple mutant of *areb1 areb2 abf3*, their role in stress tolerance and regulation of ABRE dependent gene expression in ABA signaling was elucidated. The triple mutant not only shows down regulation of *LEA* genes but also group A PP2Cs such as *AHG1*, *AHG3*, *HAI1*, *HAI2* and *HAI3* (Schweighofer et al., 2004; Yoshida et al., 2010).

The *srk2dei* (same as SnrK2.2/2.3/2.6) triple mutants also exhibit reduced phosphorylation of bZIP transcription factors including ABI5 which indicates that through phosphorylation of ABI5, SRK2D/E/I also plays a role in controlling genetic expression in ABA signaling pathways (Nakashima et al., 2009a).

Epigenetic changes in response to abiotic stress

Chromatin remodeling plays a major role in the regulation of gene expression under stressful conditions and the expression of stress-responsive genes requires them to be free from the confines of the nucleosome organization (Kim et al., 2010). Such histone modifications are mainly carried out by enzymes such as HAT (histone acetyl transferases), HDAC (histone deacetylases), HMT (histone methyltransferases) and HDM (histone demethylases).

Acetylation (ac) and/or methylation (me) of specific lysine residues in histone proteins have been reported in mammals, yeast and plants and H3K9Me3, H3K27me3 and H3K9ac are used as a marker. 28 such sites have also been reported in abiotic stress related genes in *A. thaliana* with the aid of techniques such as mass spectroscopy and chromatin immunoprecipitation (ChIP) (Zhang et al., 2007; Kim et al., 2008, 2010). In case of dehydration stress in *A. thaliana*, changes in nucleosome occupancy have been observed and the levels of methylation or acetylation of histone H3 have also been found to vary (Kim et al., 2008).

Not only genes but also promoters can be constrained by chromatin architecture. PvAlf, a putative seed specific *phas* activator (encodes phaseolin, seed storage protein of *Phaseolus vulgaris*) potentiates the *phas* promoter by repositioning nucleosomes on three TATA boxes initiating remodeling, and ABA further recruits the machinery leading to gene expression. A PvAlf factor has also been reported to assign chromatin switches (SWI/SNF) or in an alternative pathway acetylate or phosphorylate histone tail by histone kinases for transcriptional activation (Li et al., 1999). These findings not only elucidate diverse physiological roles of ABA but also provide a link between ABA and chromatin remodeling.

Histone modifications by acetylation and de-acetylation

Histone acetylation causes relaxation of the chromatin structure. Twenty-six lysine sites that undergo acetylation have been reported which are 9, 14, 18, 23 in H3; lysine 5, 8, 12, 16, 20 in H4; lysine 5 and 9 in H2A; and lysine 5, 12, 15 and 20 in H2B (Turner, 2002).

HATs have been classified into four groups on the basis of sequence homology: GNAT-MYST family, p300/CBP family, TAF_{11}250-related family and the nuclear receptor co-activator family. Apart from modifying histones, HDACs and HATs have also been reported to target non-histone proteins like p53, retinoblastoma protein (pRb), tubulin, and c-Myc (reviewed by Pawlak and Deckert, 2007).

Salt stress causes an increase in transcriptional expression through histone acetylation (Waterborg et al., 1989) and drought stress leads to acetylation of H3K23 and H3K27 in a gene specific manner (Kim et al., 2008). High salinity, cold and ABA stress also increase phosphorylation of H3Ser10, phosphoacetylation of H3Ser10 and Lys14 and acetylation of H4 in *A. thaliana* T87 and tobacco BY-2 cell lines (Sokol et al., 2007).

AtGCN5, a HAT protein in *A. thaliana*, which responds to environmental changes such as cold (Stockinger et al., 2001), is dephosphorylated by AtPP2C-6-6 (a PP2C protein) *in vitro* (Servet et al., 2008). However, *atpp2c-6-6* mutants do not show reduction of acetylation levels of H3K14 and H3K27 but the same is not true for *atgcn5* mutants. Also, reduced expression of several ABA stress inducible genes like *RAB18*, *RD29A* (responsive to dehydration 29A), *RD29B* and *COR15A* under high-salinity conditions was observed in *atpp2c-6-6* mutants, whereas in *atgcn5* mutants *RAB18* and *RD29B* genes were up-regulated in unstressed conditions with no effect on genes under high-salinity stress. These results suggest that although AtPP2C-6-6 plays a role in induction of stress-inducible genes, this interaction might not involve AtGCN5.

In vitro studies have shown that GCN5 protein interacts with transcriptional adaptor proteins Ada2a and Ada2b as well (Stockinger et al., 2001). Interaction of GCN5 and ADA with CBF1, which transcribes downstream cold-responsive genes by recruiting ADA/SAGA-like complexes, regulates cold tolerance by mediating chromatin remodeling in target genes. During cold stress, low expression of the cold-induced *COR* genes was observed in mutant for ADA and SAGA histone acetyltransferase in *A. thaliana* (Vlachonasios et al., 2003).

HDACs deacetylate lysine residues resulting in gene inactivation. In eukaryotes, they have been divided into three groups in *A. thaliana* namely RPD3/HDA1, HD2 and SIR2 families. RPD3/HDA1 family is further divided into classes I, II, III (Fu et al., 2007) and IV in rice (Sridha and Wu, 2006). A plant-specific HD2-type HDAC in *A. thaliana* (AtHD2) (Song et al., 2005) has been found to be repressed by ABA and its over expression can modulate ABA stress-responsive genes by regulating stomatal aperture. Over expression of AtERF7 (which represses transcription via its interaction with histone deacetylase HDA19) in transgenic *A. thaliana* plants shows reduced ABA sensitivity of guard cells and elevated water loss through transpiration. These results show that histone deacetylation also regulates guard cell mediated stomatal closure and that ABA regulated stomatal response mediated by chromatin remodeling plays an important role in abiotic stress tolerance.

Another HDAC known as HDA6 in *A. thaliana* has been found to be involved in RNA-directed DNA methylation (Aufsatz et al., 2002) and histone deacetylation in response to stress (Zhou et al., 2005). Low expression of ABA, *DREB2A*, *RD29A* and salt stress marker genes have been observed in those plants which show low expression of *HDA6* during stresses (Chen et al., 2010). Enhanced expression of *HDA19* caused by stress, affecting chromatin modifications by histone deacetylation have also been reported. Studies have revealed that acetylation level of H3K9 is affected in *hda19* mutant plants at seedling stage, which greatly affects gene expression (Zhou et al., 2010).

HOS15 of *A. thaliana* (high expression of osmotically responsive-genes 15) through histone H4 deacetylation represses abiotic stress response genes by encoding a protein complex WD40. *HOS15* mutants have been found to be hyper sensitive to low temperatures (Zhu et al., 2008).

Experimental data has provided evidence for differential roles of different family members of HDACs in regulating processes controlled by plant hormones. Interplay of HDACs and HATs may regulate stress response genes by integrating signaling pathways for ethylene, jasmonic acid, salicylic acid and abscissic acid (Demetriou et al., 2009).

HDACs, HvA22, HvADC2 in barley and HDT701 and HDT702 in rice are induced by ABA and jasmonic acid respectively suggesting functional similarities among monocots for HD2 genes. But on the same hand, repression of HDT genes in rice and AtHD2 in *A. thaliana* (Sridha and Wu, 2006) upon ABA treatment and differential responses of barley genes to ABA may indicate functional differences for HD2 members not only among monocots but between dicots and monocots as well (Demetriou et al., 2009).

Methylation and stress

DNA cytosine methylation represses gene promoters and gene transcription by asymmetric (mCpHpH) and symmetric (mCpGandmCpHpG) methylation. De novo methylation is catalyzed by methyl transferase domains rearranged methylase1 and 2 (DRM1 and DRM2), while enzymes like MET1 (DNMT-1 like enzymes) and plant-specific enzyme chromomethylase3 (CMT3) maintain symmetric CG and CHG methylation respectively (Henderson, 2007; Kim et al., 2010). MET1 and CMT3 may also catalyze new methylation, and maintenance of symmetric methylation may also require DRM1 and DRM2.

Importance of occurrence of methylation was reported in mutant methyl transferase *Dnmt1* as cells, which failed to maintain methylation process lost their viability (reviewed by Madlung and Comai, 2004). Mutant *MET1*, a homologue of *Dnmt1* results in a severe pleiotropic phenotype as observed in plants (Finnegan and Dennis, 1993, Finnegan et al., 1996). Also methylation has been reported to be crucial in retrotransposon suppression (Hirochika et al., 2000). Transposons, which are repressed by methylation, can be activated through demethylation induced by environmental factors. Similar findings have been observed in *Antirrhinum majus*, where cold stress mediated hypomethylation led to transposition of *Tam3* transposon (Hashida et al., 2006).

Histone H3Lys4 tri methylation and histone H3Lys9 and Lys27 di methylation have been correlated with gene activation and silencing respectively (Chinnusamy and Zhu, 2009). Tri methylation of H3K27, a negative histone modification marker was found to decrease in response to cold stress in *A. thaliana* (Kwon et al., 2009).

Methylation of histone H3 has been found to be regulated by mono ubiquitination by H2B mediated by *hub* gene, which positively regulates ABA-induced seed dormancy (Liu et al., 2007). *H2B* deubiquitination by nuclear ubiquitin protease SUP32/UBP26 directs CpNpG and CpNpN methylation leading to gene inactivation. The genome wide map of H3K4 methylation pattern of chromatin during drought stress in *A. thaliana* has been studied (Dijk et al., 2010). In this study, genes with high levels of expression were found to contain nucleosome free regions located upstream of the transcription start site.

H3K9 undergoes either acetylation or mono-di or tri-methylation. These epigenetic modifications have been reported to be involved in the regulation of gene transcription and chromatin organization, with varied impact. H3K9 acetylation has been consistently found to be associated with high gene expression/activation while dimethylation of H3K9 is reported to be linked with low gene expression. The balancing mechanism, which occurs between methylation and acetylation of H3K9makes transcription insensitive to variation disruption which affects the

process. A study shows that H3K9 acetylation effect decreases in the presence of H3K27 tri methylation (Zhou et al., 2010).

Chromatin remodeling factors

Chromatin modifying factors remodel or reposition the nucleosome by sliding histone octet in an ATP dependent process. Chromatin modifying factors include switch SWI/SNF complexes, which consist of ATPase SWI/SNF2, a pair of SWI3, SNF5 and SWP73 polypeptides. Plant genomes contain these ATPases of SWI2/SNF2 subgroup chromatin remodeling ATPases, which are called BRAHMA (BRM), SPLAYED (SYD) and SNF5 and FASCIATA (FAS).

Four SWI3-like proteins in *A. thaliana* have been identified namely SWI3A, SWI3B, SWI3C and SWI3D (Sarnowski et al., 2002; Zhou et al., 2003). SWI/SNF remodeling complexes have been found to be involved in hormonal response to abiotic stress conditions (Saez et al., 2008).

Interaction of SWI3 with a functional catalytic domain of HAB1 has been established in the plant cell nucleus by using BiFC and co-immunoprecipitation assays. BiFC assays of tobacco nuclei have also shown interaction of SWI3B with PP2CA, ABI1 and ABI2. Analysis of mutants of *swi3b* alleles, *swi3b-3* and *swi3b-4* showed that SWI3B positively regulates ABA in seeds and vegetative tissues, and also plays a key role in plant growth and development. It has been postulated that SWI3B regulates only a subset of ABA-inducible genes, and shows redundancy for other ABA-responsive genes (Saez et al., 2008).

ABA-insensitive phenotype of *Hab1-1swi3b-3* mutants, but not of *hab1-1* mutants, showed that SWI3B functions downstream of HAB1 in ABA signaling pathways. ChIP assays suggested repression of ABA-induced transcription by HAB1 through direct chromatin interaction, which can be released by subsequent ABA treatment. HAB1 has been found to be present near RAB18 and RD29B promoters, and its presence can be eliminated by ABA treatments. This suggests that SWI3B might anchor HAB1 to a putative SWI/SNF complex targeted to some ABA-responsive promoters and HAB1 regulates its activity by dephosphorylating a component essential for proper functioning of the complex. Given the opposite roles of HAB1 and SWI3B in ABA-signaling pathways, it is also possible that HAB1 by negatively regulating SWI3B modulates its role as a positive regulator (Saez et al., 2008).

In pea, *PsSNF5* gene encoding a homologous SNF5 protein of the SWI/SNF complex was found to be up-regulated by ABA and drought stress conditions. This shows a possible role of PsSNF5 in ABA signaling pathways through changes in chromatin structure (Rio et al., 2007).

In *A. thaliana*, AtCHR12, a SNF2/Brahma-type chromatin remodeling factor has been found to arrest growth of normally active primary bud and primary stem temporarily when subjected to drought and heat stress. Also the growth arrest response was found to be proportional to degree of severity of stress applied (Mlynárová et al., 2007). Han et al. (2012) described the role of *A. thaliana* BRM SWI/SNF2 chromatin remodeling complex components in direct transcriptional repression of ABI5 during post germination development.

The chromo domain helicase-DNA-binding (CHD3) class of SWI/SNF2 has been found to be associated with regulating stress-response genes in plants. *PICLKE* (PKL), which codes for *A. thaliana* CHD3 chromatin remodeling factor together with HDACs, has been implicated to play an important role in ABA signaling (Ogas et al., 1999) and also in methylation of H3K27 (Zhang et al., 2007).

Mutations in DDM1 (decrease in DNA methylation1), an *Arabidopsis* SWI2/SNF2 subfamily member results in methylation depletion by building heterochromatin structures, which are targeted by methyl transferases. However, role of DDM1 in DNA hemimethylation right after DNA replication by recruiting methyl transferases has also been established (Jeddeloh et al., 1999). This establishes a link between chromatin remodeling factors and DNA methylation.

RNAs and abiotic stress

Abiotic stress regulation also occurs at post-transcriptional and post-translational levels. The former involves processing of pre mRNA which starts with splicing of introns and joining of exons. In *A. thaliana*, a gene *STA1*, which codes for nuclear pre-mRNA splicing factor, has been identified under cold stress (Lee et al., 2006b). *Sta1-1* mutant was found to have defective splicing mechanisms as it could not splice introns from *COR15A* pre-mRNA. This unprocessed pre-mRNA showed hypersensitivity toward cold, salt stress and ABA.

Some glycine rich proteins known as RNA binding proteins (RBP) have also been identified to have an important role to play in mRNA splicing. GR-RBP over expressing transgenic plants has been observed to develop more resistance toward cold stress (Kim et al., 2007; Zhu et al., 2007). Up regulation of these proteins was also reported under environmental stress. A similar protein oligouridylate binding protein (UBP1) has been identified in *A. thaliana*. Interaction of UBP1 and RBPs results in formation of UBP associated proteins (UBA) (Lambermon et al., 2000). UBA1 leads to increased accumulation of processed mRNA in protoplasts, which suggests a role of UBA proteins in stabilizing mRNA in cytoplasm as well (Lambermon et al., 2002). A homolog of UBA1 from *Vicia faba*, ABA-activated protein kinase (AAPK)-interacting protein 1 (AKIP1), has been identified, which upon activation by AAPK through ABA dependent phosphorylation stabilizes dehydrin gene mRNA which is involved in cell protection (Li et al., 2000, 2002).

Abiotic stresses like temperature have also been found to be regulated by specific small RNAs. In *A. thaliana*, 24-nt *SRO5-P5CDH* nat-siRNA cleaves mRNA and thereby down regulates the expression of *P5CDH* leading to increased proline accumulation and hence increased salt tolerance (Borsani et al., 2005). Microarray analysis has also revealed that abiotic stresses, particularly ABA play an important role in RNA-mediated DNA methylation. However, further studies are required to elucidate the exact mechanism. *A. thaliana RDR2* has been reported to be involved in transcriptional gene silencing (Chan et al., 2004). Epigenetic changes through chromatin remodeling or DNA methylation can also be induced by non-coding RNAs (Bond and Finnegan, 2007).

Abiotic stresses including ABA have been reported to enhance transcription of miR402, which regulates Demeter-like protein 3 (DML3), a DNA glycosylase domain protein encoded by At4g34060. It has been hypothesized that miR402 under ABA induction reduces the transcript level of DML3 resulting in gene repression through methylation (Sunkar and Zhu, 2004; Chinnusamy et al., 2008).

Discussion

ABA seems to be ubiquitously distributed from cyanobacteria to plants. In many groups, where 9-*cis*-neoxanthin is missing, ABA is likely to be synthesized directly via farnesyl diphosphate. Several recent reports suggest that algal ABA may have protective role against oxidative stress (Yoshida et al., 2003), although a clear role has not been described. The possibility that the complete biochemical pathway to form and metabolize ABA was invented in bacteria, cyanobacteria and algae cannot be ruled out, although it may not have been adding to fitness advantage (Wolfram, 2010). During the course of evolution, when organisms such as submerged living liverworts started to occupy terrestrial habitats, ABA must have gained an important role as a stress hormone. The ability to tolerate drought has been one of the critical steps in the evolution

of land plants. Liu et al. (2008) has shown that *Selaginella* exhibits ABA responsive mechanisms with a subset of core components. This means that ABA signaling pathway was established in bryophyte. Bryophytes occupy an intermediate position between aquatic and land plants and the establishment of ABA signaling mechanism was of great advantage in achieving drought tolerance.

The central principle involved in ABA signaling is the inhibition of phosphatase activity of PP2C by ABA bound PYR/PYL/RCAR receptors. PYR/PYL/RCAR-PP2C-SnRK2 pathway is highly redundant, suggesting that combination of various members of the complex have distinctive roles in ABA signaling among different cells, tissues and organs. To elucidate the finer details of ABA signal transduction, further functional analyses of each combination are required. Many direct substrates or interacting proteins of PP2C or SnRK2 have been discovered and many others need to be unraveled. To fully understand the ABA signaling in plants, screening of SnRK2 substrates is still a major challenge. Global changes in phosphopeptides in WT and SnrK2.2/2.3/2.6 triple mutant seedlings in response to ABA treatment have been analyzed by Wang et al. (2013) with the use of qualitative phosphoproteomics. It was observed that ABA could increase the phorphorylation of 166 peptides and decrease the phosphorylation of 117 peptides in WT seedlings, whereas 84 of the 166 peptides representing 58 proteins could not be phosphorylated by the same. These results shed new light on the role of the SnRK2 protein kinases and on the downstream effectors of ABA action.

Once a plant detects the onset of stress, TFs characteristically respond by inducing the expression of a cascade of downstream targets. However, their activation is in part also dependent on their chromatin structure, which is largely determined by epigenetic means (Boyko and Kovalchuk, 2008; Bilichak et al., 2012). There exists a very clear time lag between transcriptional activation, DNA methylation and histone modification. Song et al. (2012) have shown in soybean that a promoter was demethylated within 1 h of the imposition of salt stress. However, there was no evidence for the TF's activation before 3 h. Similar time lags have been noted for the building of H3K4me3 within the coding regions of *A. thaliana* drought related genes *rd29A* and *rap2.4* during stress (Kim et al., 2008) suggesting the events of transcriptional activation, DNA methylation and histone modification are not simultaneous.

Future perspectives

A good number of genes have been documented which encode for enzymes conferring tolerance to abiotic stress. Transgenics of rice, oat, *etc*., have been engineered to express osmo-protectants such as glycine betaine, proline, polyamines and several other genes encoding enzymes involved in abiotic stress responses (Amudha and Balasubramani, 2011). Constitutive expression of group3 LEA proteins has resulted in improved growth characteristics and drought tolerance. HVA1 (group3 LEA) gene from barley (Xu et al., 1996) a PMA1959 (group1 LEA) from wheat (Cheng et al., 2002) have been reported to confer drought and salt stress tolerance.

To eliminate adverse effects of ROS generated under environmental stress, detoxifying genes have been incorporated in plants. These genes express anti-oxidant proteins like glutathione peroxidase, superoxide dismutase, ascorbate peroxidase and glutathione reductase, *etc*. that scavenge the ROS thereby enhancing tolerance to oxidative stress. Over-expression of superoxide dismutase (SOD), particularly in mitochondria and chloroplast, has been reported to have a profound effect on the relative tolerance of transgenic plants to stress conditions. Apart from this, transformation of plants with TFs like MYC, MYB, bZIP has also been proved to show a dramatic improvement in the performance of plants (Amudha and Balasubramani, 2011).

Early approaches were targeted on stress responsive genes for creation of transgenic plants. But continuous expression of stress inducible genes is a highly energy intensive process which casts a heavy burden on the plant machinery subsequently compromising the quality and quantity of the yield. This can be overcome by designing an inducible promoter where expression is under the control of an inducer. With improvements in sequencing techniques, advent of whole genome sequencing and construction of databases like PLACE, TRANSFAC, PlantCARE, JASPER, *etc*., the task of promoter designing and analysis has become easier. Strategies have been developed to design tissue specific promoters, which direct gene expression in particular tissues only (Mehrotra et al., 2011).

Synthetic promoters have found great application in plant biotechnology particularly in studies related to abiotic stress responses. A minimal expression cassette (Pmec) and transcriptional activation module (TAM) were designed which contain many *cis* motifs responding to biotic and abiotic stresses and growth and development (Sawant et al., 1999). These motifs show not only self-activation but also synergistic activation. Using this concept, a synthetic bidirectional expression module was made by arranging the positions of Pmec and TAM, Pmec being placed on the 5′ and 3′ end of TAM (Chaturvedi et al., 2006).

More recently, synthetic promoters have been produced for stringent regulation. However, the challenge of homology based gene silencing (HBGS) is yet to be met. On similar lines, bidirectional promoters have been developed which not only overcome the risk posed by synthetic promoters but also reduce the quantity of foreign DNA being introduced in the organism (Mehrotra et al., 2011).

Conclusion

Advancements in systemic computational and molecular biology have led to a deeper understanding of the underlying mechanisms of the involvement of ABA in stress tolerance; though a lot is yet to be discovered. The physiological changes that lead to adaptive responses are far more advanced in plants as compared to animals. An understanding of ABA signal transduction pathways and epigenetic modifications can be further exploited to produce better transgenic plants that can withstand adverse climatic conditions. Genetic regulation through *cis* engineering can be achieved by designing promoters targeted at tissue specific expression through which plant dynamics can be further explored.

Acknowledgements

This work is supported by the project of Dr. Sandhya Mehrotra and Dr. Rajesh Mehrotra funded by University Grants Commission, New Delhi, Government of India. The authors are grateful to the administration of Birla Institute of Technology and Sciences, Pilani, Rajasthan, for providing logistic support. The authors are also thankful to UGC/BSR for providing financial support to Purva Bhalothia.

References

Abe H, Urao T, Ito T, Seki M, Shinozaki K, Shinozaki KY. *Arabidopsis* AtM.Y.C2 (bHLH) and AtMYB2 (MYB) function as transcriptional activators in abscisic acid signaling. Plant Cell 2003;15:63–78.

Ahmad MR, Saxena PM, Amla DV, Shyam R. Effect of abscisic acid on the blue green algae *Anacystis nidulans* and *Nostoc muscorum*. Beitr Biol Pflanzen 1978;54:33–40.

Amudha J, Balasubramani G. Recent molecular advances to combat abiotic stress tolerance in crop plants. Biotechnol Mol Biol Rev 2011;6:31–58.

Anderberg RJ, Walker-Simmons MK. Isolation of a wheat cDNA clone for an abscisic acid-inducible transcript with homology to protein kinase. Proc Natl Acad Sci U S A 1992;89:10183–90.

Antoni R, Gonzalez-Guzman M, Rodriguez L, Peirats-Llobet M, Pizzio AG, Fernandez AM, et al. Pyrabactin resistance1-likeplays an important role for the regulation of abscisic acid signaling in root. Plant Physiol 2013;161:931–40.

Aufsatz W, Mette MF, Van der Winden J, Matzke M, Matzke AJ. HDA6, a putative histone deacetylase needed to enhance DNA methylation induced by double-stranded RNA. EMBO J 2002;21:6832–40.

Ballester DG, Pollock SV, Pootakham W, Grossman AR. The central role of a SnRK2 kinase in sulfur deprivation responses. Plant Physiol 2008;147:216–20.

Bilichak A, Ilnystkyy Y, Hollunder J, Kovalchuk I. The progeny of *Arabidopsis thaliana* plants exposed to salt exhibit changes in DNA methylation, histone modifications and gene expression. PLoS ONE 2012:7e30515.

Bond DM, Finnegan EJ. Passing the message on: inheritance of epigenetic traits. Trends Plant Sci 2007;12:211–20.

Borsani O, Zhu J, Verslues PE, Sunkar R, Zhu JK. Endogenous siRNAs derived from a pair of natural *cis*-antisense transcripts regulate salt tolerance in *Arabidopsis*. Cell 2005;123:1279–80.

Boudsocq M, Barbier-Brygoo H, Lauriere C. Identification of nine SNF1-related protein kinase 2 activated by hyperosmotic and saline stresses in *Arabidopsis thaliana*. J Biol Chem 2004;279:41758–60.

Boyko A, Kovalchuk I. Epigenetic control of plant stress response. Environ Mol Mutagen 2008;49:61–72.

Chan SW, Zilberman D, Xie Z, Johansen LK, Carrington JC, Jacobsen SE. RNA silencing genes control de novo DNA methylation. Science 2004;303:13–36.

Chaturvedi CP, Sawant SV, Kiran K, Mehrotra R, Lodhi N, Ansari SA, et al. Analysis of polarity in the expression from a multifactorial bidirectional promoter designed for high-level expression of transgenes in plants. J Biotechnol 2006;123:1–12.

Chen LT, Luo M, Wang YY, Wu KQ. Involvement of *Arabidopsis* histone deacetylase HDA6 in ABA and salt stress response. J Exp Bot 2010;61:3345–50.

Cheng Z, Targolli J, Huang X, Wu R. Wheat LEA genes, PMA80 and PMA1959, enhance dehydration tolerance of transgenic rice (*Oryza sativa* L.). Mol Breed 2002;10:71–82.

Chinnusamy V, Zhu JK. Epigenetic regulation of stress responses in plants. Curr Opin Plant Biol 2009;12:133–40.

Chinnusamy V, Gong Z, Zhu JK. Abscisic acid-mediated epigenetic processes in plant development and stress responses. J Integr Plant Biol 2008;50:1187–90.

Cutler SR, Rodriguez PL, Finkelstein RR, Abrams SR. Abscisic acid: emergence of a core signaling network. Annu Rev Plant Biol 2010;61:651–60.

Demetriou K, Kapazoglou A, Tondelli A, Francia E, Stanca MA, Bladenopoulos K, et al. Epigenetic chromatin modifiers in barley: I. Cloning, mapping and expression analysis of the plant specific *HD2* family of histone deacetylases from barley, during seed development and after hormonal treatment. Physiol Plant 2009;136:358–60.

Dijk van K, Ding Y, Malkaram S, Riethoven JJM, Liu R, Yang J, et al. Dynamic changes in genome wide histone H3 lysine 4 methylation patterns in response to dehydration stress in *Arabidopsis thaliana*. BMC Plant Biol 2010;10:238–40.

Finnegan EJ, Dennis ES. Isolation and identification by sequence homology of a putative cytosine methyltransferase from *Arabidopsis thaliana*. Nucleic Acids Res 1993;21:2383–90.

Finnegan EJ, Peacock WJ, Dennis ES. Reduced DNA methylation in *Arabidopsis* results in abnormal plant development. Proc Natl Acad Sci USA 1996;93:8449–50.

Forchetti G, Masciarelli O, Alemano S, Alvarez D, Abdala G. Endophytic bacteria in sunflower (*Helianthus annuus* L.): isolation, characterization and production of jasmonates and abscisic acid in culture medium. Appl Microbiol Biotechnol 2007;76:1145–50.

Foster R, Izawa T, Chua NH. Plant bZIP proteins gather at ACGT elements. FASEB J 1994;8:192–200.

Fu W, Wu K, Duan J. Sequence and expression analysis of histone deacetylases in rice. Biochim Biophys Acta 2007;356:843–50.

Fujii H, Zhu JK. *Arabidopsis* mutant deficient in 3 abscisic acid-activated protein kinases reveals critical roles in growth, reproduction and stress. Proc Natl Acad Sci USA 2009;19:8380–90.

Fujita Y, Nakashima K, Yoshida T, Katagiri T, Kidokoro S, Kanamori N, et al. Three SnRK2 protein kinases are the main positive regulators of abscisic acid signaling in response to water stress in *Arabidopsis*. Plant Cell Physiol 2009;50:2123–30.

Fujita Y, Yoshida T, Shinozaki KY. Pivotal role of the AREB/ABF-SnRK2 pathway in ABRE-mediated transcription in response to osmotic stress in plants. Physiol Plant 2013;147:15–27.

Gao SQ, Chen M, Xu ZS, Zhao CP, Li L, Xu H, et al. The soybean GmbZIP1 transcription factor enhances multiple abiotic stress tolerances in transgenic plants. Plant Mol Biol 2011;75:537–40.

Gonzalez-Guzman M, Apostolova N, Belles JM, Barrero JM, Piqueras P, Ponce MR, et al. The short-chain alcohol dehydrogenase ABA2 catalyzes the conversion of xanthoxin to abscisic aldehyde. Plant Cell 2002;14:1833–40.

Han SK, Sang Y, Rodrigues A, Wu MF, Rodriguez PL, Wagner D. The SWI2/SNF2 chromatin remodeling ATPase BRAHMA represses abscisic acid responses in the absence of the stress stimulus in *Arabidopsis*. Plant Cell 2012;24:4892–900.

Hashida SN, Uchiyama T, Martin C, Kishima Y, Sano Y, Mikami T. The temperature-dependent change in methylation of the *Antirrhinum* transposon Tam3 is controlled by the activity of its transposase. Plant Cell 2006;18:104–10.

Henderson IR, Jacobsen SE. Epigenetic inheritance in plants. Nature 2007;447:418–20.

Hirochika H, Okamoto H, Kakutani T. Silencing of retrotransposons in *Arabidopsis* and reactivation by the *ddm1* mutation. Plant Cell 2000;12:357–60.

Hossain MA, Lee Y, Cho JI, Ahn CH, Lee SK, Jeon JS, et al. The bZIP transcription factor OsABF1 is an ABA responsive element binding factor that enhances abiotic stress signaling in rice. Plant Mol Biol 2010;72:557–60.

Hrabak EM, Chan CWM, Gribskov M, Harper JF, Choi JH, Halford N, et al. *Arabidopsis* CDPK-SnRK superfamily of protein kinases. Plant Physiol 2003;132:666–70.

Huai J, Wang M, He J, Zheng J, Dong Z, Lv H, et al. Cloning and characterization of the SnRK2 gene family from *Zea mays*. Plant Cell Rep 2008;12:1861–70.

Hubbard KE, Nishimura N, Hitomi K, Getzoff ED, Schroeder JI. Early abscisic acid signal transduction mechanisms: newly discovered components and newly emerging questions. Genes Dev 2010;24:1695–700.

Jeddeloh JA, Stokes TL, Richards EJ. Maintenance of genomic methylation requires a SWI2/SNF2-like protein. Nat Genet 1999;22:94–7.

Kelner A, Pekala A, Kaczanowski S, Muszynska G, Hardie DG, Dobrowolska G. Biochemical characterization of the *Nicotiana tabacum* 42-kD protein kinase activated by osmotic stress. Plant Physiol 2004;136:3255–60.

Kim JM, To TK, Ishida J, Morosawa T, Kawashima M, Matsui A, et al. Alterations of lysine modifications on the histone H3 N-tail under drought stress conditions in *Arabidopsis thaliana*. Plant Cell Physiol 2008;49:1580–90.

Kim JM, To TK, Nishioka T, Seki M. Chromatin regulation function in plant abiotic stress responses. Plant Cell Environ 2010;33:604–10.

Kim JY, Park SJ, Jang B, Jung C-H, Ahn SJ, Goh CH, et al. Functional characterization of a glycine-rich RNA-binding protein 2 in *Arabidopsis thaliana* under abiotic stress conditions. Plant J 2007;50:439–40.

Kobayashi Y, Yamamoto S, Minami H, Kagaya Hattori T. Differential activation of the rice sucrose non fermenting1-related protein kinase 2 family by hyper-osmotic stress and abscisic acid. Plant Cell 2004;16:1163–70.

Kulik A, Wawer I, Krzywinska E, Bucholc M, Dobrowolska G. SnRK2 protein kinases-key regulators of plant response to abiotic stresses. OMICS: J Integr Biol 2011;15:859–60.

Kuromori T, Shinozaki K. ABA transport factors found in *Arabidopsis* ABC transporters. Plant Signal Behav 2010;5:1124–30.

Kwak JM, Mori IC, Pei ZM, Leonhardt N, Torres MA, Dangl JL, et al. NADPH oxidase *AtrbohD* and *AtrbohF* genes function in ROS-dependent ABA signaling in *Arabidopsis*. EMBO J 2003;22:2623–30.

Kwon CS, Lee D, Choi G, Chung W. Histone occupancy-dependent and -independent removal of H3K27 trimethylation at cold-responsive genes in *Arabidopsis*. Plant J 2009;60:112–20.

Lambermon MHL, Simpson GG, Kirk DAW, Hemmings-Mieszczak M, Klahre U, Filipowicz W. UBP1, a novel hn RNP-like protein that functions at multiple steps of higher plant nuclear pre-mRNA maturation. EMBO J 2000;19:1638–40.

Lambermon MH, Fu Y, Wieczorek Kirk DA, Dupasquier M, Filipowicz W, Lorković ZJ. UBA1 and UBA2, two proteins that interact with UBP1, a multifunctional effector of pre-mRNA maturation in plants. Mol Cell Biol 2002;22:4346–50.

Lee BH, Kapoor A, Zhu J, Zhu JK. STABILIZED1, a stress-upregulated nuclear protein, is required for pre mRNA splicing, mRNA turnover, and stress tolerance in *Arabidopsis*. Plant Cell 2006a;18:1736–40.

Lee KH, Piao HL, Kim HY, Choi SM, Jiang F, Hartung W, et al. Activation of glucosidase via stress-induced polymerization rapidly increases active pools of abscisic acid. Cell 2006b;126:1109–10.

Leung J, Merlot S, Giraudat J. The *Arabidopsis* abscisic acid-insensitive2 (*ABI2*) and *ABI1* genes encode homologous protein phosphatases 2C involved in abscisic acid signal transduction. Plant Cell 1997;9:759–60.

Li G, Bishop KJ, Chandrashekharan MB, Hall TC. β-Phaseolin gene activation is a two-step process: PvALF-mediated chromatin modifications followed by abscisic acid-mediated gene activation. Proc Natl Acad Sci USA 1999;96:7104–10.

Li J, Wang XQ, Watson MB, Assmann SM. Regulation of abscisic acid-induced stomatal closure and anion channels by guard cell AAPK kinase. Science 2000;287:300–10.

Li J, Kinoshita T, Pandey S, Ng CK, Gygi SP, Shimazaki K, et al. Modulation of an RNA-binding protein by abscisic-acid-activated protein kinase. Nature 2002;418:793–800.

Li LB, Zhang YR, Liu KC, Ni ZF, Fang ZJ, Sun QX, et al. Identification and bioinformatics analysis of SnRK2 and CIPK family genes in sorghum. Agric Sci China 2010;9:19–30.

Liu MS, Chien CT, Lin TP. Constitutive components and induced gene expression are involved in the desiccation tolerance of *Selaginella tamariscina*. Plant Cell Physiol 2008;49:653–60.

Liu Y. Roles of mitogen-activated protein kinase cascades in ABA signaling. Plant Cell Rep 2012;31:1–12.

Liu Y, Koornneef M, Soppe WJJ. The absence of histone H2B monoubiquitination in the *Arabidopsis* hub1 (rdo4) mutant reveals *a role for chromatin* remodeling in seed dormancy. Plant Cell 2007;19:433–40.

Ma Y, Szostkiewicz I, Korte A, Moes D, Yang Y, Christmann A, et al. Regulators of PP2C phosphatase activity function as abscisic acid sensors. Science 2009;324:1064–70.

Madlung A, Comai L. The effect of stress on genome regulation and structure. Ann Bot 2004;94:481–90.

Mao X, Zhang H, Tian S, Chang X, Jing R. TaSnRK2.4, an SNF1-type serine/threonine protein kinase of wheat (*Tricitum aestivum* L.), confers enhanced multistress tolerance in *Arabidopsis*. J Exp Bot 2010;61:683–90.

Mehrotra R, Mehrotra S. Promoter activation by ACGT in response to salicylic and abscisic acids is differentially regulated by the spacing between two copies of the motif. J Plant Physiol 2010;167:1214–20.

Mehrotra R, Gupta G, Sethi R, Bhalothia P, Kumar N, Mehrotra S. Designer promoter: an artwork of *cis* engineering. Plant Mol Biol 2011;75:527–30.

Mehrotra R, Sethi S, Zutshi I, Bhalothia P, Mehrotra S. Patterns and evolution of ACGT repeat *cis*-element landscape across four plant genomes. BMC Genomics 2013;14:203–10.

Mehrotra R, Yadav A, Bhalotia P, Karan R, Mehrotra S. Evidence for directed evolution of larger size motif in *Arabidopsis thaliana* genome. Sci World J 2012., http://dx.doi.org/10.1100/2012/983528.

Meskiene I, Boegre L, Glaser W, Balog J, Brandstoetter M, Zwerger TK, et al. MP2C, a plant protein phosphatase 2C, functions as a negative regulator of mitogen-activated protein kinase pathways in yeast and plants. Proc Natl Acad Sci USA 1998;95:1938–40.

Mizoguchi M, Umezawa T, Nakashima K, Kidokoro S, Takasaki H, Fujita Y, et al. Two closely related subclass II SnRK2 protein kinases cooperatively regulate drought-inducible gene expression. Plant Cell Physiol 2010;51:842–50.

Mlynárová L, Nap JP, Bisseling T. The SWI/SNF chromatin-remodeling gene *AtCHR12* mediates temporary growth arrest in *Arabidopsis thaliana* upon perceiving environmental stress. Plant J 2007;51:874–80.

Monks DE, Aghoram K, Courtney PD, DeWald DB, Dewey RE. Hyperosmotic stress induces the rapid phosphorylation of a soybean phosphatidylinositol transfer protein homolog through activation of the protein kinases SPK1 and SPK2. Plant Cell 2001;13:1205–10.

Nakashima K, Fujita Y, Kanamori N, Katagiri T, Umezawa T, Kidokoro S, et al. Three *Arabidopsis* SnRK2 protein kinases, SRK2D/SnRK2.2, SRK2E/SnRK2. 6/OST1 and SRK2I/SnRK2. 3, involved in ABA signaling are essential for the control of seed development and dormancy. Plant Cell Physiol 2009a;50:1345–50.

Nakashima K, Ito Y, Yamaguchi-Shinozaki K. Transcriptional regulatory networks in response to abiotic stresses in *Arabidopsis* and grasses. Plant Physiol 2009b;149:88–95.

Nambara E, Marion-Poll A. Abscisic acid biosynthesis and catabolism. Annu Rev Plant Biol 2005;56:165–70.

Nishimura N, Hitomi K, Arvai AS, Rambo RP, Hitomi C, Cutler SR, et al. Structural mechanism of abscisic acid binding and signaling by dimeric PYR1. Science 2009;326:1373–80.

Nishimura N, Sarkeshik A, Nito K, Park S, Wang A, Carvalho P, et al. PYR/PYL/RCAR family members are major *in-vivo* ABI1 protein phosphatase 2C-interacting proteins in *Arabidopsis*. Plant J 2010;61:290–300.

Ogas J, Kaufmann S, Henderson J, Somerville C. PICKLE is a CHD3 chromatin-remodeling factor that regulates the transition from embryonic to vegetative development in *Arabidopsis*. Proc Natl Acad Sci U S A 1999;96:13839–40.

Orellana S, Yañez M, Espinoza A, Verdugo I, González E, Lara SR, et al. The transcription factor SlAREB1 confers drought, salt stress tolerance and regulates biotic and abiotic stress-related genes in tomato. Plant Cell Environ 2010;33:2191–200.

Park SY, Fung P, Nishimura N, Jensen DR, Hiroaki F, Zhao Y, et al. Abscisic acid inhibits PP2Cs via the PYR/PYL family of ABA binding START proteins. Science 2009;324:1068–70.

Pawlak S, Deckert J. Histone modifications under environmental stress. Biol Lett 2007;44:65–73.

Riechmann JL, Heard J, Martin G, Reuber L, Jiang C, Keddie J, et al. *Arabidopsis* transcription factors: genome-wide comparative analysis among eukaryotes. Science 2000;290:2105–10.

Rio G, Gagete AP, Castillo J, Berbel A, Franco L, Rodrigo I. Abscisic acid and desiccation-dependent expression of a novel putative *SNF5*-type chromatin-remodeling gene in *Pisum sativum*. Plant Physiol Biochem 2007;45:427–30.

Rodriguez PL, Benning G, Grill E. ABI2, a second protein phosphatase 2C involved in abscisic acid signal transduction in *Arabidopsis*. FEBS Lett 1998;421:185–90.

Saez A, Apostolova N, Gonzalez-Guzman M, Gonzalez-Garcia MP, Nicolas C, Lorenzo O, et al. Gain of function and loss of function phenotypes of the protein phosphatase 2C HAB1 reveal its role as a negative regulator of abscisic acid signaling. Plant J 2004;37:354–60.

Saez A, Rodrigues A, Santiago J, Rubio S, Rodriguez PL. HAB1-SWI3B interaction reveals a link between abscisic acid signaling and putative SWI/SNF chromatin-remodeling complexes in *Arabidopsis*. Plant Cell 2008;20:2972–80.

Saradhi PP, Suzuki I, Katoh A, Sakamoto A, Sharmila P, Shi DJ, et al. Protection against the photo-induced inactivation of the photosystem II complex by abscisic acid. Plant Cell Environ 2000;23:711–20.

Sarnowski TJ, Swiezewski S, Pawlikowska K, Kaczanowski S, Jerzmanowski A. AtSWI3B, an *Arabidopsis* homolog of SWI3, a core subunit of yeast Swi/Snf chromatin remodeling complex, interacts with FCA, a regulator of flowering time. Nucleic Acids Res 2002;30:3412–20.

Sawant SV, Singh PK, Gupta SK, Madanala R, Tuli R. Conserved nucleotide sequence in highly expressed genes in plants. J Genet 1999;78:1–8.

Schweighofer A, Hirt H, Meskiene I. Plant PP2C phosphatases: emerging functions in stress signaling. Trends Plant Sci 2004;9:236–40.

Seiler C, Harshavardhan VT, Rajesh K, Reddy PS, Strickert M, Rolletschek H, et al. ABA biosynthesis and degradation contributing to ABA homeostasis during barley seed development under control and terminal drought-stress conditions. J Exp Bot 2011;62:2615–20.

Seo M, Koshiba T. Complex regulation of ABA biosynthesis in plants. Trends Plant Sci 2002;7:41–8.

Seo M, Koshiba T. Transport of ABA from the site of biosynthesis to the site of action. J Plant Res 2011;124:501–10.

Servet C, Benhamed M, Latrasse D, Kim W, Delarue M, Zhou DX. Characterization of a phosphatase 2C protein as an interacting partner of the histone acetyltransferase GCN5 in *Arabidopsis*. Biochim Biophys Acta 2008;1779:376–80.

Shen QJ, Casaretto JA, Zhang P, Ho THD. Functional definition of ABA-response complexes: the promoter units necessary and sufficient for ABA induction of gene expression in barley (*Hordeum vulgare* L.). Plant Mol Biol 2004;54:111–20.

Shinozaki K, Yamaguchi-Shinozaki K, Seki M. Regulatory network of gene expression in the drought and cold stress responses. Curr Opin Plant Biol 2003;6:410–20.

Sokol A, Kwiatkowska A, Jerzmanowski A, Prymakowska-Bosak M. Up-regulation of stress-inducible genes in tobacco and *Arabidopsis* cells in response to abiotic stresses and ABA treatment correlates with dynamic changes in histone H3 and H4 modifications. Planta 2007;227:245–50.

Song CP, Agarwal M, Ohta M, Guo Y, Halfter U, Wang P, et al. Role of an *Arabidopsis* AP2/EREBP-type transcription al repressor in abscisic acid and drought stress responses. Plant Cell 2005;17:2384–90.

Song Y, Ji D, Li S, Wang P, Li Q, Xiang F. The dynamic changes of DNA methylation and histone modifications of salt responsive transcription factor genes in Soybean. PLoS ONE 2012;7:1–11.

Sridha S, Wu K. Identification of *AtHD2C* as a novel regulator of abscisic acid responses in *Arabidopsis*. Plant J 2006;46:124–30.

Stockinger EJ, Mao Y, Regier MK, Triezenberg SJ, Thomashow MF. Transcriptional adaptor and histone acetyl-transferase proteins in *Arabidopsis* and their interactions with CBF1, a transcriptional activator involved in cold-regulated gene expression. Nucleic Acids Res 2001;29:1524–30.

Sunkar R, Zhu JK. Novel and stress-regulated micro RNAs and other small RNAs from *Arabidopsis*. Plant Cell 2004;16:2001–10.

Suzuki M, Ketterling MG, McCarty DR. Quantitative statistical analysis of *cis*-regulatory sequences in ABA/VP1- and CBF/DREB1-regulated genes of *Arabidopsis*. Plant Physiol 2005;139:437–40.

Tan BC, Schwartz SH, Zeevaart JA, McCarty DR. Genetic control of abscisic acid biosynthesis in maize. Proc Natl Acad Sci U S A 1997;94:12235–40.

Turner BM. Cellular memory and the histone code. Cell 2002;111:285–90.

Umezawa T, Nakashima K, Miyakawa T, Kuromori T, Tanokura M, Shinozaki K, et al. Molecular basis of the core regulatory network in ABA responses: sensing, signaling and transport. Plant Cell Physiol 2010;51:1821–30.

Vlachonasios KE, Thomashow MF, Triezenberg SJ. Disruption mutations of *ADA2b* and *GCN5* transcriptional adaptor genes dramatically affect *Arabidopsis* growth, development, and gene expression. Plant Cell 2003;15:626–30.

Wang P, Xue L, Batelli G, Lee S, Hou Y, Van Oosten MJ, et al. Quantitative phosphoproteomics identifies SnRK2 protein kinase substrates and reveals the effectors of abscisic acid action. Proc Natl Acad Sci U S A 2013;110:11205–10.

Waterborg JH, Harrington RE, Winicov I. Differential histone acetylation in alfalfa (*Medicago sativa*) due to growth in NaCl: responses in salt stressed and salt tolerant callus cultures. Plant Physiol 1989;90:237–40.

Wolfram H. The evolution of abscisic acid (ABA) and ABA function in lower plants, fungi and lichen. Funct Plant Biol 2010;37:806–10.

Xu D, Duan X, Wang B, Hong B, Ho THD, Wu R. Expression of a late embryogenesis abundant protein gene, HVA1, from barley confers tolerance to water deficit and salt stress in transgenic rice. Plant Physiol 1996;110:249–50.

Xu GP, Way HM, Richardson T, Drenth J, Joyce PA, McIntyre CL. Overexpression of *TaNAC69* leads to enhanced transcript levels of stress up-regulated genes and dehydration tolerance in bread wheat. Mol Plant 2011;4:697–700.

Yoshida K, Iigarashi E, Mukai M, Hirata K, Miyamoto K. Induction of tolerance to oxidative stress in the green alga, *Chlamydomonas reinhardtii*, by abscisic acid. Plant Cell Environ 2003;26:451–60.

Yoshida R, Hobo T, Ichimura K, Mizoguchi T, Takahashi F, Aronso J, et al. ABA-activated SnRK2 protein kinase is required for dehydration stress signaling in *Arabidopsis*. Plant Cell Physiol 2002;43:1473–80.

Yoshida R, Umezawa T, Mizoguchi T, Takahashi S, Takahashi F, Shinozaki K. The regulatory domain of SRK2E/OST1/SnRK2.6 interacts with ABI1 and integrates ABA and osmotic stress signals controlling stomatal closure in *Arabidopsis*. J Biol Chem 2006;281:5310–20.

Yoshida T, Fujita Y, Sayama H, Kidokoro S, Maruyama K, Mizoi J, et al. AREB1, AREB2 and ABF3 are master transcription factors that cooperatively regulate ABRE-dependent ABA signaling involved in drought stress tolerance and require ABA for full activation. Plant J 2010;61:672–80.

Zhang H, Mao X, Wang C, Jing R. Over-expression of a common wheat gene TaSnRK2.8 enhances tolerance to drought, salt and low temperature in *Arabidopsis*. PLoS ONE 2010;5:16041–50.

Zhang K, Gan SS. An abscisic acid-AtNAP transcription factor-SAG113 protein phosphatase 2C regulatory chain for controlling dehydration in senescing *Arabidopsis* leaves. Plant Physiol 2012;158:961–70.

Zhang K, Sridhar VV, Zhu J, Kapoor A, Zhu JK. Distinctive core histone post-translational modification patterns in *Arabidopsis thaliana*. PLoS ONE 2007., http://dx.doi.org/10.1371/journal.pone.0001210.

Zhang X, Wollenweber B, Jiang D, Liu F, Zhao J. Water deficit and heat shock effects on photosynthesis of a transgenic *Arabidopsis thaliana* constitutively expressing *ABP9*, a bZIP transcription factor. J Exp Bot 2008;59:839–40.

Zhou C, Miki B, Wu K. CHB2, a member of the SWI3 gene family, is a global regulator in *Arabidopsis*. Plant Mol Biol 2003;52:1125–30.

Zhou C, Zhang L, Duan J, Miki B, Wu K. Histone deacetylase 19 is involved in jasmonic acid and ethylene signaling of pathogen response in *Arabidopsis*. Plant Cell 2005;17:1196–200.

Zhou JL, Wang XF, He K, Charron JBF, Elling AA, Deng XW. Genome-wide profiling of histone H3 lysine acetylation and dimethylation in *Arabidopsis* reveals correlation between multiple histone marks and gene expression. Plant Mol Biol 2010;72:585–90.

Zhu J, Dong HC, Zhu JK. Interplay between cold-responsive gene regulation, metabolism and RNA processing during plant cold acclimation. Curr Opin Plant Biol 2007;10:290–300.

Zhu J, Jeong J, Zhu Y, Sokolchik I, Miyazaki S, Zhu JK, et al. Involvement of *Arabidopsis* HOS15 in histone deacetylation and cold tolerance. Proc Natl Acad Sci U S A 2008;105:4945–50.

American Journal of Plant Sciences, 2013, 4, 549-554
http://dx.doi.org/10.4236/ajps.2013.43071 Published Online March 2013 (http://www.scirp.org/journal/ajps)

Scientific Research

AACA Element Negatively Regulates Expression of Protein Phosphatase 2C (*PP2C*) like Promoter in *Arabidopsis thaliana*

Purva Bhalothia[1], Anshu Alok[2], Sandhya Mehrotra[1], Rajesh Mehrotra[1*]

[1]Department of Biological Sciences, Birla Institute of Technology & Sciences (BITS), Pilani, India; [2]National Agri-Food Biotechnology Institute (NABI), Department of Biotechnology, Government of India, Mohali, India.
Email: [*]rmehrotra@bits-pilani.ac.in, [*]rajmeh25@hotmail.com

Received January 5th, 2013; revised February 10th, 2013; accepted February 20th, 2013

ABSTRACT

Genome wide analysis of *Arabidopsis thaliana* reveals a unique genetic arrangement of ACGT motif in the protein phosphatase 2C (*PP2C*) like promoter (accession number AT5G59220). In the present study, full length, 900 base pair, 500 base pair, 400 base pair and NRM deletion variants of *PP2C* like promoter were constructed to investigate the activity of *PP2C* like promoter in the presence of inducers like abscisic, jasmonic acid and salicylic acid. The *PP2C* promoter has three ACGT elements in close vicinity. They are so positioned that they are separated by a spacer of 30 base pair and 5 base pair respectively from each other. The study has shown that $ACGT_{N30}ACGT$ genetic architecture is essential for the promoter to be induced in response to abscisic acid. The synergistic and antagonistic effects of *cis* elements were observed. AACA is a positive regulatory element in endosperm and is known to act as negative regulatory element in other tissues. In this study AACA, have been found to negatively regulate the expression of reporter gene EGFP in both induced and under uninduced conditions.

Keywords: Abscisic Acid; ACGT; AACA; Abiotic Stress

1. Introduction

Plants, being sessile in nature need to adopt different methods to combat stresses in comparison to animals, be it environmental or biological. Phytohormones like jasmonic acid (JA) and salicylic acid (SA) help the plant to face biotic stress. To cope up with the abiotic stress plants have two pathways, viz. abscisic acid (ABA) dependent and ABA independent [1]. The ABA dependent stress responsive genes have ABA responsive element (ABRE) which is present in its promoter region [2] and has ACGT as a core sequence [3]. ACGT is also the core sequence of the G box, involved in the red light signalling [4], UV response, anaerobiosis in plants and SA response elements [5].

The ABA mediated stress signaling pathway involves the receptor PYL/PYR/RCAR, a negative regulator *PP2C*, positive regulator SnRK2s, *cis* regulatory element known as ABRE and transcription factors [6]. *Cis* element is a consensus sequence present in the upstream region of the basal promoter, and it acts as the binding site for transcription factors. Gene expression is the result of the interaction between *cis* element and its complementary transcription factor, the latter being activated by a particular developmental stage or environmental cues.

A promoter consisting of TATA box is sufficient to transcribe a gene at its basal level. Diversity among promoters is an outcome of the presence of different *cis* regulatory elements, their variation in copy numbers, position in context to TATA box and the spacer length between two adjacent *cis* elements [7]. Copy number and length of the spacer is critical in determining the expression of the gene [8]. This is much evident with ACGT and GT *cis* elements, where both copy number and spacer length are responsible for high or low expression of gene. GT element acts as negative regulator on increasing its copy number with increase in spacer length. Reference number [7] has reported decrease in gene expression with an increasing copy number of adjacent GT motifs [9]. But unlike GT elements, an increase in copy number and spacer length to 5, 10, and 25 in case of ACGT element leads to an enhanced gene expression of the *β-glucuronidase* (*gus A*) reporter gene.

Reference number [10] showed that two ACGT elements separated by 25 base pair (bp) are induced by ABA. However, when separated by 5 nucleotides, they are induced by SA in transgenic tobacco. The genome

[*]Corresponding author.

Copyright © 2013 SciRes.

wide analysis of *Arabidopsis thaliana* revealed the frequency of occurrence of two ACGT motifs separated by a spacer of 25 bp to be 62 while for 5 bp, the occurrence was 72 [11]. One such promoter located at chromosome number 5 (accession number AT5G59220) is protein phosphatase 2C (*PP2C*) like promoter from *Arabidopsis thaliana* which has ABA and SA inducible ACGT motifs. *PP2C* like promoter has genetic architecture which includes three ACGT elements. Two adjacent ACGT elements are separated by 30 nucleotides, followed by another set of ACGT motifs separated by 5 nucleotides (**Figure 1**).

The ACGT core motif serves as a binding site for the bZIP class of transcription factors. The nucleotides flanking the ACGT motif determine the binding affinity and specificity of bZIP transcription factors, which in turn leads to distinct physiological functions as in stress responses, light regulation etc. [12,13]. It has been reported that the flanking sequence is also responsible for differential distribution of stress responsive regulatory elements in stress responsive promoters [14]. In this study, *PP2C* like promoter was cloned and several deletion constructs were made and bombarded on to the tobacco leaves to see their effect on the expression of EGFP reporter gene cloned downstream of it. Further, considering the fact that *PP2C* like promoter has ABA and SA response elements positioned in close proximity, the study could also give an insight in to the relation between abiotic and biotic stress mediated responses.

Earlier, constitutive promoters were widely used in development of transgenic plants, but they suffer from the drawback of being highly energy demanding as they are always active even in favorable conditions. The demand for high energy for promoter activity in relation to constitutive promoters leads to economic imbalance in terms of energy which may significantly affect overall expression of productive genes. The beneficial solution can be through the use of a stress inducible promoter

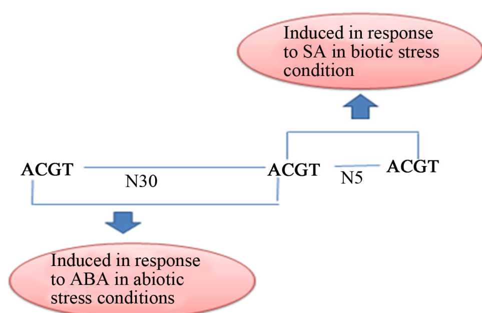


Figure 1. Genetic arrangement of NRM cassette. The bold ACGT element has been removed in NRM deletion construct of *PP2C* like promoter.

which is intended to stimulate under precise biotic and abiotic stresses [15].

PP2C like promoter has both biotic stress and abiotic stress response elements in close proximity and provides an added advantage over the constitutive promoters and it could be used as an inducible promoter.

In this study the expression of 400 bp construct have shown higher expression in comparison to the full length construct due to the decrease in the copy number of AACA element.

2. Material and Methods

2.1. Cloning of *PP2C* like Promoter from *Arabidopsis thaliana*

Growth conditions of *Arabidopsis thaliana* columbia ecotype were standardized and grown at 22˚C under 16 hr light and 8 hr dark conditions until silique formation. Total genomic DNA was isolated from the young leaves of *Arabidopsis thaliana* using CTAB method [16]. TAIR (The Arabidopsis Information Resource) data base was used to access the sequence of the protein phosphates 2C (*PP2C*) like promoter bearing the accession number AT5G59220. The primers used for amplification are mentioned in **Table 1**. The PCR conditions were standardized as initial denaturation at 94˚C for 2 minutes, final denaturation at 94˚C for 1 minutes, annealing temperature at 58˚C for 1.30 minutes, extension temperature at 72˚C for 1.30 minutes and final extension temperature at 72˚C for 4 minutes.

The resultant PCR amplicons were cloned into pEGFP (Clontech) vector between *PstI* and *BamHI* restriction sites (**Figure 2**). The *E. coli DH5α* competent cells were transformed with the recombinant plasmid. After transformation, plasmids were isolated and the presence of the inserted promoter was confirmed by sequencing.

Table 1. The primers used for amplification of the *PP2C* like promoter.

Name	Sequence of the primer
Forward Primer NA1	5'AACTGCAGAAGTATTCACGCACC AAGGT 3'
Reverse Primer NA3	5'CGGGATCCACAAACACACTCCAT CACTC 3'
900 bp (100 bp deleted from 5')	5'AACTGCAGTGTCCTTGAACACAC CAAAC 3'
500 bp (500 bp deleted from 5')	5'AACTGCAGAAATGGTAAGGTAAA TTTCCAC 3'
400 bp (600 bp deleted from 5')	5'AACTGCAGCACTTTAGGTCTGAGT AGTGT3'
NRM	5'AACTGCAGGGAAGTGGTAATAAA TTTACA 3'

Copyright © 2013 SciRes.

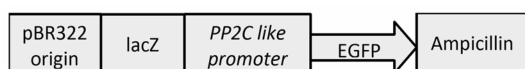


Figure 2. A pictorial representation of promoter reporter gene constructs used for plant transformation.

2.2. Transient Transformation and EGFP Assay

Transient transformation of *Nicotiana tabbacum* cv. Xanthi was done using Biolistic method [17]. DNA for bombardment was prepared by using Qiagen columns. The PP2C promoter reporter cassettes were coated onto gold microprojectiles and bombarded onto the tobacco leaves at 1100 psi using biolistic gene gun (BioRad PDS-1000/He). Before bombardment the leaves were incubated on MS [18] agar plates for 12 hr. After bombardment, leaves were incubated on Hoagland media containing inducers such as 100 µM ABA, 100 µM SA, 50 µM JA at 24˚C under16 hr light and 8 hr dark photoperiod. After 48 hrs of incubation, leaves were ground in liquid nitrogen and suspended in extraction buffer (1 M Tris pH 8, 0.5 M EDTA pH 8, 0.5 M EGTA pH 8, 0.5M NaF, 1M sodium orthovandate, 1 M beta mercaptoethanol, 10% v/v glycerol) and centrifuged at 13,000 rpm for 10 minutes at 4˚C. 200 µl of the supernatant was loaded to 96 well titer plates. The excitation and emission wavelength peaks for EGFP are 484 nm and 509 nm. EGFP concentration was determined by the absorbance at 484 nm and was calculated using the reported excitation coefficients 55,000/cm/M [19,20].

3. Results and Discussion

3.1. The Functional Classification of *Cis* Elements Present in Full Length *PP2C* like Promoter

To characterize the *PP2C* like promoter, a series of 5' deletion constructs were made like 900 bp, 500 bp, 400 bp and NRM. For analyzing the presence of *cis* elements in full length of *PP2C* like promoter, the PLACE database was used [21,22]. This analysis shows that different motifs responsive to inducers like ABA, JA, SA, light regulation etc. are present in the *PP2C* like promoter (**Table 2**). In our study inducers like ABA, JA and SA were used.

3.2. AACA Motif Negatively Regulates the Gene Expression of *PP2C* like Promoter

The full length *PP2C* like promoter and its deletion variants which include 900 bp, 500 bp, 400 bp and NRM constructs were bombarded on to the leaves of *Nicotiana tabacum* cv. Xanthi and were incubated in Hoagland medium supplemented with inducers ABA, JA and SA and without inducers. The activity of the promoter reporter construct was evaluated on the basis of the expression of EGFP reporter gene following the transient transformation of *Nicotiana tabacum* cv. Xanthi.

Non-recombinant pEGFP vector was used as control. The expression of EGFP was observed in the leaves bombarded with control construct which was low in comparison to other promoter reporter cassettes of *PP2C* like promoter, as expected. The expression of the reporter gene observed in case of the control was due to the lacZ promoter. However, in the presence of inducers like ABA, JA and SA, control does not show any induction (**Figure 3**).

Under uninduced conditions the full length as well as deletion variants of *PP2C* like promoter showed more activity than the control, *i.e.* pEGFP vector (**Figure 3**). It was observed that, the expression of EGFP is higher in 400 bp and NRM cassettes of *PP2C* like promoter whereas in full length, 900 bp and 500 bp constructs of *PP2C* like promoter the expression of EGFP was low, the latter being at par with each other. This is attributed to the presence of AACA elements in the full length, 900 and 500 bp constructs. AACA element is a tissue specific *cis* element and is known to express only in the endosperm [23]. Reference number [23] observed that deletion of AACA motif leads to decrease in the seed storage *Glu B*-1 promoter activity. The presence of AACA in

Table 2. *Cis* regulatory elements present in the *PP2C* like promoter.

Family	*Cis* element	Sequence	Occurrence (+ & − Strand)
ABA response element	ABRELATERD1	ACGTG	1 & −2
	ABRERATCAL	MACGYGB	1 & −1
	ACGTATERD1	ACGT	6 & −6
	DRE1COREZMRAB17	ACCGAGA	−1
Light regulated genes	ASF1MOTIFCAMV	TGACG	2 & −2
	GATABOX	GATA	5 & −3
	GT1CONSENSUS	GRWAAW	5 & −7
	GT1CORE	GGTTAA	1
	IBOXCORE	GATAA	1 & −1
Gbox binding site	CACGTGMOTIF	CACGTG	1 & −1
	CGACGOSAMY3[a]	CGACG	−1
	IRO2OS	CACGTGG	1
WRKY group	WBOXATNPR1	TTGAC	−2
	WBOXNTERF3	TGACY	−3
	WRKY71OS	TGAC	3 & −7
	ELRECOREPCRP1[b]	TTGACC	−2

M = A/C; K = G/T; Y = T/C; N = A/G/C/T; R = A/G; W = A/T; S = G/C; H = T/C/A; D = A/T/G. [a]Act as coupling element for G box element; [b]Elecitor response element.

Copyright © 2013 SciRes.

promoter of genes in any other tissue, acts as negative *cis* element for other elements [24]. The AACA element being present in full length, 900 bp and 500 bp could be responsible for reduced expression of reporter gene. This negative *cis* element AACA is present in *PP2C* like promoter in 9 copies. This number decreases to one in case of deletion cassettes like 400 bp and NRM construct and show high activity in terms of expression of the reporter gene (the position of the AACA motif is mentioned in **Table 3**).

3.3. Expression of *PP2C* like Promoter in the Presence of Elicitors

The induction was observed in each case of *PP2C* like promoter cassettes in comparison to the control incubated in Hoagland media supplemented with ABA, JA and SA, respectively.

The motifs which confer response towards ABA have ACGT as a core sequence. These motifs are like ACGTERD1 and ABRELATERD1 (as mentioned in **Table 2**). The successive constructs viz. full length and 900 bp of *PP2C* like promoter have six copies of ACGTATERD1 motif while 500 bp and 400 bp constructs have four copies of ACGTATERD1 motif. The ACGT motifs in close proximity lead to increase in the expression of the gene [7]. In the *PP2C* like promoter, other ACGT motifs are not located in close vicinity to each other. It could also be attributed to the presence of AACA, which acts as negative elements in these promoter reporter cassettes. Though the AACA motif does not have any direct relationship with the ACGT *cis* element, it decreases the overall expression of the reporter gene in constructs like full length, 900 bp and 500 bp deletion construct of *PP2C* like promoter.

The expression of NRM cassette in the presence of ABA when compared with the 500 bp and 400 bp constructs is low even though AACA copy number falls down to 1 from 9. The NRM cassette showed expression nearer to full length and 900 bp constructs. The genetic architecture of NRM construct (**Figure 1**) has been disturbed by removing one ACGT motif, and retains only two ACGT elements separated by 5 bp. The induction of NRM deletion construct was observed to be comparatively mild in the presence of ABA in comparison to 400 bp construct. Hence this shows that two ACGT elements separated by 30 bp are essentially responsible for induction by ABA.

In the induction studies in the presence of JA, the full length, 900 bp and NRM constructs of *PP2C* like promoter have not shown similar levels of EGFP expression. The highest level of expression is observed in the case of 500 bp cassette and very low expression observed in 400 bp cassette (shown in **Figure 3**). The deletion of 100 bp from 500 bp deletion construct led to deletion of *cis* regulatory elements and decrease in the copy number of the *cis* elements like WRKY elements which could have reduced the expression of 400 bp cassette.

When incubated in the presence of SA the full length, 500 bp, 400 bp and NRM cassette (**Figure 3**) showed more expression than 900 bp construct. Minimal machinery required for the induction by SA is present in full length, 900 bp and 500 bp. The 400 bp and NRM construct shows more expression, which have 1 copy of AACA element. The 400 bp and NRM has the basic structure *i.e.* two ACGT separated by 5 bp. This leads to increase in the expression of EGFP. In study conducted by reference number [25] similar observations were made, viz. the absence of regulatory elements in the −119/101 mutant did not lead to induction by methyl jasmonate (MJ) and SA. However, induction was observed in the case of the −112/101 mutant which had the regulatory elements, thus indicating the importance of the presence of regulatory element for MJ and SA dependent induction.

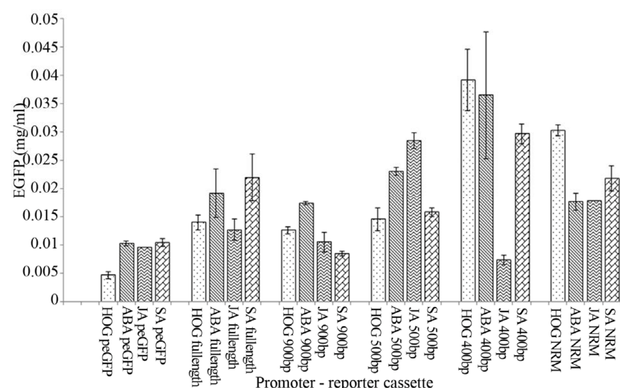


Figure 3. Expression of the *PP2C* like promoter (with different constructs) under un-induced and induced conditions (green: Hoagland media (H.M); blue: H.M + ABA; purple: H.M + JA; orange: H.M + SA). The vertical bar indicates standard deviation.

Table 3. The copy number and the position of the AACA motif in context to TATA box.

Constructs	No. of copies of AACA element	Position upstream to TATA box
Full length *PP2C* like promoter	9	−100, −322, −353, −432, −504, −533, −635, −675, −730
900 bp	9	−100, −322, −353, −432, −504, −533, −635, −675, −730
500 bp	2	−100, −322
400 bp	1	−100
NRM	1	−100

4. Conclusion

Up regulation and down regulation of expression is based on the synergistic action of *cis* elements present in the promoter region of the gene. Reference number [26] and [15] have reported that proper spatial arrangement of the *cis* elements is responsible for the transcription. The combinatorial activity of positive and negative *cis* elements is found to be responsible for the regulation of gene expression. The current study reveals that AACA element negatively regulates the expression of the *PP*2*C* like promoter, even in the presence of inducers viz. ABA, JA and SA. Reference number [10] had shown that two ACGT elements separated by 25 bp were induced in response to ABA and in present study; it has been shown that *PP*2*C* like promoter with an arrangement of two ACGT elements separated by 30 bp is also induced in response to ABA.

5. Acknowledgements

The authors are grateful to the Department of Science and Technology, New Delhi, India for Grant-in-Aid and financial support to carry out this work bearing the file no. SR/FT/LS-126/2008. The authors are thankful to Dr Rakesh Tuli and Dr Siddharth Tiwari for providing biolistic facilities. The authors are grateful to the administration of BITS, Pilani, Rajasthan and NABI, Mohali for providing logistic support. PB is also thankful to UGC/ BSR for providing financial support.

REFERENCES

[1] P. K. Busk and M. Pages, "Regulation of Abscisic Acid-Induced Transcription," *Plant Molecular Biology*, Vol. 37, No. 3, 1998, pp. 425-435. doi:10.1023/A:1006058700720

[2] Q. Shen and T. H. Ho, "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*, Vol. 7, No. 3, 1995, pp. 295-307.

[3] M. J. Guiltinan, W. R. Marcotte and R. S. Quatrano, "A Plant Leucine Zipper Protein that Recognizes an Abscisic Acid Response Element," *Science*, Vol. 250, No. 4978, 1990, pp. 267-271. doi:10.1126/science.2145628

[4] R. G. K. Donald and A. R. Cashmore, "Mutation of Either G Box or I Box Sequences Profoundly Affects Expression from the Arabidopsis rbcS-1A Promoter," *The EMBO Journal*, Vol. 9, No. 6, 1990, pp. 1717-1726.

[5] I. Jupin and N. H. Chua, "Activation of the CaMV as-1 *cis* Element by Salicylic Acid: Differential DNA-Binding of a Factor Related to TGA1a," *The EMBO Journal*, Vol. 15, No. 20, 1996, pp. 5679-5689.

[6] T. Umezawa, K. Nakashima, T. Miyakawa, T. Kuromori, M. Tanokura, K. Shinozaki and K. Yamaguchi-Shinozaki, "Molecular Basis of the Core Regulatory Network in ABA Responses: Sensing, Signaling and Transport," *Plant and Cell Physiology*, Vol. 51, No. 11, 2010, pp. 1821-1839. doi:10.1093/pcp/pcq156

[7] R. Mehrotra, K. Kiran, C. P. Chaturvedi, S. A. Ansari, N. Lodhi, S. Sawant and R. Tuli, "Effect of Copy Number and Spacing of the ACGT and GT *cis* Elements on Transient Expression of Minimal Promoter in Plants," *Journal of Genetics*, Vol. 84, No. 2, 2005, pp. 183-187. doi:10.1007/BF02715844

[8] T. Hobo, M. Asada, Y. Kowyama and T. Hattori, "ACGT Containing Abscisic Acid Response Element (ABRE) and Coupling Element 3 (CE3) Are Functionally Equivalent," *The Plant Journal*, Vol. 19, No. 6, 1999, pp. 679-689. doi:10.1046/j.1365-313x.1999.00565.x

[9] R. Mehrotra and J. Panwar, "Dimerization of GT Element Interferes Negatively with Gene Activation," *Journal of Genetics*, Vol. 88, No. 2, 2009, pp. 257-260. doi:10.1007/s12041-009-0037-7

[10] R. Mehrotra and S. Mehrotra, "Promoter Activation by ACGT in Response to Salicylic and Abscisic Acids Is Differentially Regulated by the Spacing between Two Copies of the Motif," *Journal of Plant Physiology*, Vol. 167, No. 14, 2010, pp. 1214-1218. doi:10.1007/s12041-009-0037-7

[11] R. Mehrotra, A. Yadav, P. Bhalothia, R. Karan and S. Mehrotra, "Evidence for Directed Evolution of Larger Size Motif in *Arabidopsis thaliana* Genome," *The Scientific World Journal*, Vol. 2012, 2012, Article ID: 983528. doi:10.1100/2012/983528

[12] T. Izawa, R. Foster and N. H. Chua, "Plant bZIP Protein DNA Binding Specificity," *Journal of Molecular Biology*, Vol. 20, No. 4, 1993, pp. 1131-1144. doi:10.1006/jmbi.1993.1230

[13] R. Foster, T. Izawa and N. H. Chua, "Plant Basic Leucine Zipper Proteins Gather at ACGT Elements," *The EMBO Journal*, Vol. 8, No. 2, 1994, pp. 192-200.

[14] M. Suzuki, M. G. Ketterling and D. R. McCarty, "Quantitative Statistical Analysis of *cis*-Regulatory Sequences in ABA/VP1- and CBF/DREB1-Regulated Genes of *Arabidopsis*," *Plant Physiology*, Vol. 139, No. 1, 2005, pp. 437-447. doi:10.1104/pp.104.058412

[15] R. Mehrotra, G. Gupta, R. Sethi, P. Bhalothia, N. Kumar and S. Mehrotra, "Designer Promoter: An Artwork of *cis* Engineering," *Plant Molecular Biology*, Vol. 75, No. 6, 2011, pp. 527-536. doi:10.1007/s11103-011-9755-3

[16] M. G. Murray and W. F. Thompson, "Thompson Rapid Isolation of High Molecular Weight Plant DNA," *Nucleic Acids Research*, Vol. 8, No. 19, 1980, pp. 4321-4326. doi:10.1093/nar/8.19.4321

[17] S. V. Sawant, P. K. Singh and R. Tuli, "Pretreatment of Micropojectiles to Improve the Delivery of DNA in Plant Transformation," *Biotechniques*, Vol. 29, No. 2, 2000, pp. 246-248.

[18] T. Murashige and F. Skoog, "A Revised Medium for Rapid Growth and Bioassays with Tobacco Tissue Cultures," *Plant Physiology*, Vol. 15, No. 43, 1962, pp. 473-497.

Copyright © 2013 SciRes.

[19] C. Reichel, J. Mathur, P. Eckes, K. Langenkemper, C. Koncz, J. Schell, B. Reiss and C. Maas, "Enhanced Green Fluorescence by the Expression of an Aequorea victoria Green Fluorescent Protein Mutant in Mono- and Dicotyledonous Plant Cells," *Proceedings of the National Academy of Sciences of the United States of America*, Vol. 93, No. 12, 1996, pp. 5888-5893. doi:10.1073/pnas.93.12.5888

[20] J. J. Wu, Y. W. Liu and M. X. Sun, "Improved and High throughput Quantitative Measurements of Weak GFP Expression in Transgenic Plant Materials," *Plant Cell Report*, Vol. 30, No. 7, 2011, pp. 1253-1260. doi:10.1046/j.1365-313x.2000.00797.x

[21] D. S. Prestridge, "SIGNAL SCAN: A Computer Program That Scans DNA Sequences for Eukaryotic Transcriptional Elements," *Computer Applications in the Biosciences*, Vol. 7, No. 2, 1991, pp. 203-206.

[22] K. Higo, Y. Ugawa, M. Iwamoto and T. Korenaga, "Plant *cis*-Acting Regulatory DNA Elements (PLACE) Database," *Nucleic Acids Research*, Vol. 27, No. 1, 1999, pp. 297-300. doi:10.1093/nar/27.1.297

[23] C. Wu, H. Washida, Y. Onodera, K. Harada and F. Takaiwa, "Quantitative Nature of the Prolamin-Box, ACGT and AACA Motifs in a Rice Glutelin Gene Promoter: Minimal *cis*-Element Requirements for Endosperm-Specific Gene Expression," *The Plant Journal*, Vol. 23, No. 3, 2000, pp. 415-421. doi:10.1046/j.1365-313x.2000.00797.x

[24] T. Yoshihara, H. Washida and F. Takaiwa, "A 45-bp Proximal Region Containing AACA and GCN4 Motif Is Sufficient to Confer Endosperm-Specific Expression of the Rice Storage Protein Glutelin Gene, *GluA*-3," *FEBS Letters*, Vol. 383, No. 3, 1996, pp. 2133-2138. doi:10.1016/0014-5793(96)00233-5

[25] S. R. Kim, Y. Kim and G. An, "Identification of Methyl Jasmonate and Salicylic Acid Response Elements from the Nopaline Synthase (*nos*) Promoter," *Plant Physiology*, Vol. 103, No. 1, 1993, pp. 97-103. doi:10.1104/pp.103.1.97

[26] S. Sawant, K. Kiran, R. Mehrotra, C. P. Chaturvedi, S. A. Ansari, P. Singh, N. Lodhi and R. Tuli, "A Variety of Synergistic and Antagonistic Interactions Mediated by *cis*-Acting DNA Motifs Regulate Gene Expression in Plant Cells and Modulate Stability of the Transcription Complex Formed on a Basal Promoter," *Journal of Experimental Botany*, Vol. 56, No. 419, pp. 2345-2353. doi:10.1093/jxb/eri227

Abbreviations

ABA: Abscisic Acid;
bp: Base Pair;
EGFP: Enhanced Green Fluorescent Protein;
JA: Jasmonic Acid;
*PP*2*C*: Protein Phosphatase 2C;
SA: Salicylic Acid.

Copyright © 2013 SciRes.

Mehrotra et al. BMC Genomics 2013, 14:203
http://www.biomedcentral.com/1471-2164/14/203

BMC Genomics

RESEARCH ARTICLE Open Access

Patterns and evolution of ACGT repeat *cis*-element landscape across four plant genomes

Rajesh Mehrotra*, Sachin Sethi†, Ipshita Zutshi†, Purva Bhalothia and Sandhya Mehrotra

Abstract

Background: Transcription factor binding is regulated by several interactions, primarily involving *cis*-element binding. These binding sites maintain specificity by means of their sequence, and other additional factors such as inter-motif distance and spacer specificity. The ACGT core sequence has been established as a functionally important *cis*-element which frequently regulates gene expression in synergy with other *cis*-elements. In this study, we used two monocotyledonous – *Oryza sativa* and *Sorghum bicolor*, and two dicotyledonous species – *Arabidopsis thaliana* and *Glycine max* to analyze the conservation of co-occurring ACGT core elements in plant promoters with respect to spacer distance between them. Using data generated from *Arabidopsis thaliana* and *Oryza sativa,* we also identified conserved regions across all spacers and possible conditions regulating gene promoters with multiple ACGT *cis*-elements.

Results: Our data indicated specific predominant spacer lengths between co-occurring ACGT elements, but these lengths were not universally conserved across all species under analysis. However, the frequency distribution indicated local regions of high correlation among monocots and dicots. Sequence specificity data clearly revealed a preference for G at the first and C at the terminal position of a spacer sequence, suggesting that the G-box motif is the most prevalent for the ACGT class of promoters. Using gene expression databases, we also observed trends suggesting that co-occurring ACGT elements are responsible for gene regulation in response to exogenous stress. Conservation in patterns of ACGT (N) ACGT among orthologous genes also indicated the possibility that emergence of functional significance across species was a result of parallel evolution of these *cis*-elements.

Conclusions: Although the importance of ACGT elements has been acknowledged for several plant species, ours is the first study that attempts to compare their occurrence across four species and analyze conservation among them. The apparent preference for particular spacer distances suggest that these motifs might be implicated in important physiological functions which are yet to be identified. Variations in correlation patterns among monocots and dicots might arise out of differences in transcriptional regulation in the two classes. In accordance with literature, we established the involvement of co-occurring ACGT elements in stress responses and showed how this regulation differs with variation in the ACGT (N) ACGT motif. We believe that our study will be an essential resource in determining optimum spacer length and spacer sequence between ACGT elements for promoter design in future.

Keywords: Gene expression, Promoter regulation, *Cis*-element, Stress, Spacer, *Arabidopsis*, Rice, Soybean, Sorghum, Inter-motif distance

* Correspondence: rmehrotra@pilanibits-pilani.ac.in
†Equal contributors
Biological Sciences Department, Birla Institute of Technology and Science, Pilani, RJ, India

BioMed Central

© 2013 Mehrotra et al.; licensee BioMed Central Ltd. This is an Open Access article distributed under the terms of the Creative Commons Attribution License (http://creativecommons.org/licenses/by/2.0), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Background

Eukaryotic gene expression and its regulation by means of transcription is one of the most significant areas of research currently. Regulation of gene transcription relies on interactions among transcription factors (TFs) which bind to specific DNA *cis* sites to form a conglomeration of proteins which guide the polymerase binding [1]. These *cis* sites include enhancers, core promoters, matrix or scaffold attachment regions, insulators and silencers. Of these, enhancer elements are an important class of *cis*-regulatory sequences that are usually present upstream of the transcription initiation site and contain multiple short binding site sequences for targeting several activators and repressors of transcription. These short binding site sequences are often referred to as sequence motifs and occur in recurring patterns across DNA [2]. It is generally accepted that sequence motifs co-evolve with their core promoter in both sequence specific and location specific ways to achieve a directed target function [3,4]. Comparative studies show conserved regions of promoters are expressed widely across genomes of various species, indicating directed evolution [5], correlating with the notion that functionally less important regions of DNA evolve (in terms of mutant substitutions) faster than more important ones [6]. For example, in a comparison of the 200 base pair early enhancer of *Hoxc8* in 29 species of mammals the complete nucleotide sequences of this region were 90% similar across all taxa, confirming that this enhancer sequence has been specifically conserved [7,8]. Apart from the sequence of these motifs, the positional and inter-motif distances within the enhancers also play a critical role in interactions between transcription factors [9], as the spacing and intermediary sequences control the size and strength of interactions of TF binding [10], subsequently affecting the gene expression [11]. In fact, unless helical phasing is conserved to provide allowance for protein binding, even a change in spacer length by one base pair can drastically alter gene expression levels [12]. Holistically, this implies that conserved sequences (conforming to the various spatial and positional constraints) occurring in higher numbers as compared to what is probabilistically expected may be of specific functional significance.

In plant genomes, one such sequence motif – the ACGT core sequence – is functionally important in a variety of promoters that respond to different stimuli like light [12], anaerobiosis [13], jasmonic acid [14] and hormones such as salicylic acid [15], abscisic acid [16-18] and auxin [19]. This core element is present at different relative positions in multiple copies upstream of the transcription start site [20], and any alterations in this core sequence reduce the overall promoter activity significantly, for it contributes synergistically to gene expression by stabilizing the transcription complex formed on the minimal promoter [21]. Co-occurring ACGT elements are over-represented in Arabidopsis and rice genomes, emphasizing their functional relevance when compared to single ACGT core elements [22]. As discussed, the inter motif distance between these co-occurring ACGT sequence is of particular importance as promoter activation by ACGT is differentially regulated by the spacing between two copies of the motif [23]. Additionally, the copy number of ACGT elements in a promoter and distance from the transcription start site also drastically alter gene expression [24]. While most reports on the ACGT core sequence are based on *Arabidopsis thaliana*, the ACGT family of promoters (ACEs) have also been identified in wheat [25] rice [26-28] and barley [29], suggesting that ACEs are conserved across plant species.

Given that the ACGT core sequence is dispersed across promoters of various plant species; and that they occur in multiple copies with a variable number of base pairs separating them, we have attempted to analyze patterns in occurrence of ACGT core element repeats in plant genomes. We performed an *in-silico* search for ACGT elements separated by spacers of varying lengths in all identified promoters for two monocots – Sorghum and Rice, and two dicots – Arabidopsis and Soybean. Our data indicated similarities in the frequency patterns across the four plant species, with correlations for particular spacer lengths between ACGT core elements. In order to analyze if a specific sequence motif is preferred as a spacer between multiple ACGT elements across all promoters, we developed consensus sequences from all spacers observed. Additionally, we studied the evolution of co-occurring ACGT elements by analyzing their prevalence among orthologous genes in Arabidopsis, Rice and Sorghum. Further, to understand the functional significance of these elements, we used microarray data to analyze which conditions might be responsible for regulation of genes consisting of multiple ACGT *cis*-elements.

Methods

Data extraction

Focusing our analysis on promoters, we extracted 1 kb sequences upstream of all identified chromosomal genes from the following genomes - *Arabidopsis thaliana* (The Arabidopsis Genome Initiative v. 10, 2011), *Oryza sativa* (Rice) (International Rice Genome Sequencing Project, Build 4.0, 2009), *Glycine max* (Soybean) (US DOE Joint Genome Institute (JGI-PGF), v. 1.0, 2010) and *Sorghum bicolor* (Sorghum) (Sorghum Consortium, v. 1.0, 2009) using the NCBI Reference Sequence database [30-33]. Using a code (Additional file 1), we extracted gene annotation information (Gene ID/ Arabidopsis TAIR ID and

ATG site) from Gene bank files and the corresponding 1 kb upstream region from the FASTA sequence. We searched for co-occurring ACGT elements of the form ACGT (N) ACGT, where $0 \leq N \leq 30$ in all extracted 1 kb regions for our analysis. As it has been previously seen that cooperatively binding transcription factors are usually spaced within 25 bp, we limited our analysis to a spacer distance of 30 bp. The sequence of each spacer (region between two ACGT core elements) was extracted and the total number of occurrences for each spacer length was determined for each species. In order to test the significance of these frequencies, we used four palindromic – TAGC, CGTA, GCTA, ATGC, and four non-palindromic – AGCT, TGCA, CTAG, GATC sequences as controls. Using the PLACE database, we ensured that each of these 4 bp sequences are not conserved *cis* element themselves [34]. By performing a similar analysis on each control sequence, we compared the frequency of the ACGT (N) ACGT motif with the corresponding frequencies of control sequences for the same N.

Spacer sequence analysis

Spacer sequences occurring in all promoters of *Arabidopsis thaliana* were analyzed to identify preferences for a particular nucleotide at each position within the spacer sequence. The percentage of A, G, C and T at each position for all spacer lengths (N = 0–30) was calculated to identify preferences at particular positions within the spacers. Since the genome wide GC content for *Arabidopsis thaliana* is known to be around 36% [35], we chose threshold occurrence percentage of 25% for C/G and 40% for A/T for a particular position in all spacers of the same length (N). Single letter IUPAC DNA codes were assigned to each position to generate consensus sequences for all spacer lengths.

Ortholog analysis

In order to understand the mechanism of evolution of the ACGT (N) ACGT *cis*-element in the aforementioned plant species, we analyzed its predominance for N = 0–30 in all identified in-paralogs/orthologs among Arabidopsis, Rice and Sorghum [36]. Gene names for Rice were converted from MSU annotation to RAP annotation prior to analysis [37]. Frequencies for co-occurring ACGT elements were analyzed from extracted genes.

Functional analysis

Using published microarray data from the EBI Gene Expression Atlas [38] for *Arabidopsis thaliana*, we analyzed whether genes containing multiple core ACGT *cis*-elements were unregulated/down regulated during developmental stages (embryo, seedling), by hormones (ABA, auxin, ethylene, gibberellin, jasmonic acid, salicylic acid), in different plant parts (cambium, flower, leaf, root, pollen, seed, sperm cell, stem, vegetative apex, apical root meristem) and by environmental conditions (at baseline growth temperature, disease, drought, low water potential, at optimum photosynthetic temperature, salt, 20% inhibition from optimum photosynthetic temperature, 30% inhibition from optimum photosynthetic temperature). In addition, we extended our functional analysis to Rice by extracting genes up regulated by salt and drought stress in 7 day old seedling from publically available microarray data [39]. By comparing genes regulated by a condition with genes containing multiple ACGT elements, we calculated the likelihood of occurrence by the following formula:

$$\text{Likelihood of occurrence} = X/Y \text{ for a particular spacer length } (N)$$

where $X = (A \cap B)/B$ and $Y = P(A)$;

A = event that a given gene is regulated (up/down) by a particular condition B = event that a given gene contains multiple ACGT elements separated by N base pairs. Further, we calculated the overall likelihood of occurrence for each condition (for N = 0–30). All conditions with likelihood of occurrence > 1.30 were subjected to further statistical analysis using the 8 control sequences described earlier.

Statistical analysis

Wherever possible, statistical analysis was performed to determine the significance of results. The frequency of ACGT (N) ACGT was assessed for significant peaks by box and whiskers plots, with 10% and 90% whiskers. The outliers were considered as potential peaks, especially if they were present across all species. The degree of correlation between the two monocot and dicot species was calculated by taking the frequencies for N consecutive spacer lengths at a time, beginning from 0 till 30 for each of the species, where N > =6. By assuming Gaussian distribution, the Pearson's correlation coefficient was calculated for each of the cases to determine significance. Consecutive spacer lengths of N with the highest degrees of correlation were interpreted as the most conserved spacer distances.

To identify conditions regulating ACGT (N) ACGT containing promoters over the 8 control elements, a Grubb's outlier's test was performed on the likelihood of occurrences for each condition, assuming the data set was normally distributed. If the ACGT (N) ACGT likelihood emerged significantly higher than the controls by this test for a certain condition, it was interpreted to be specifically regulated by that condition.

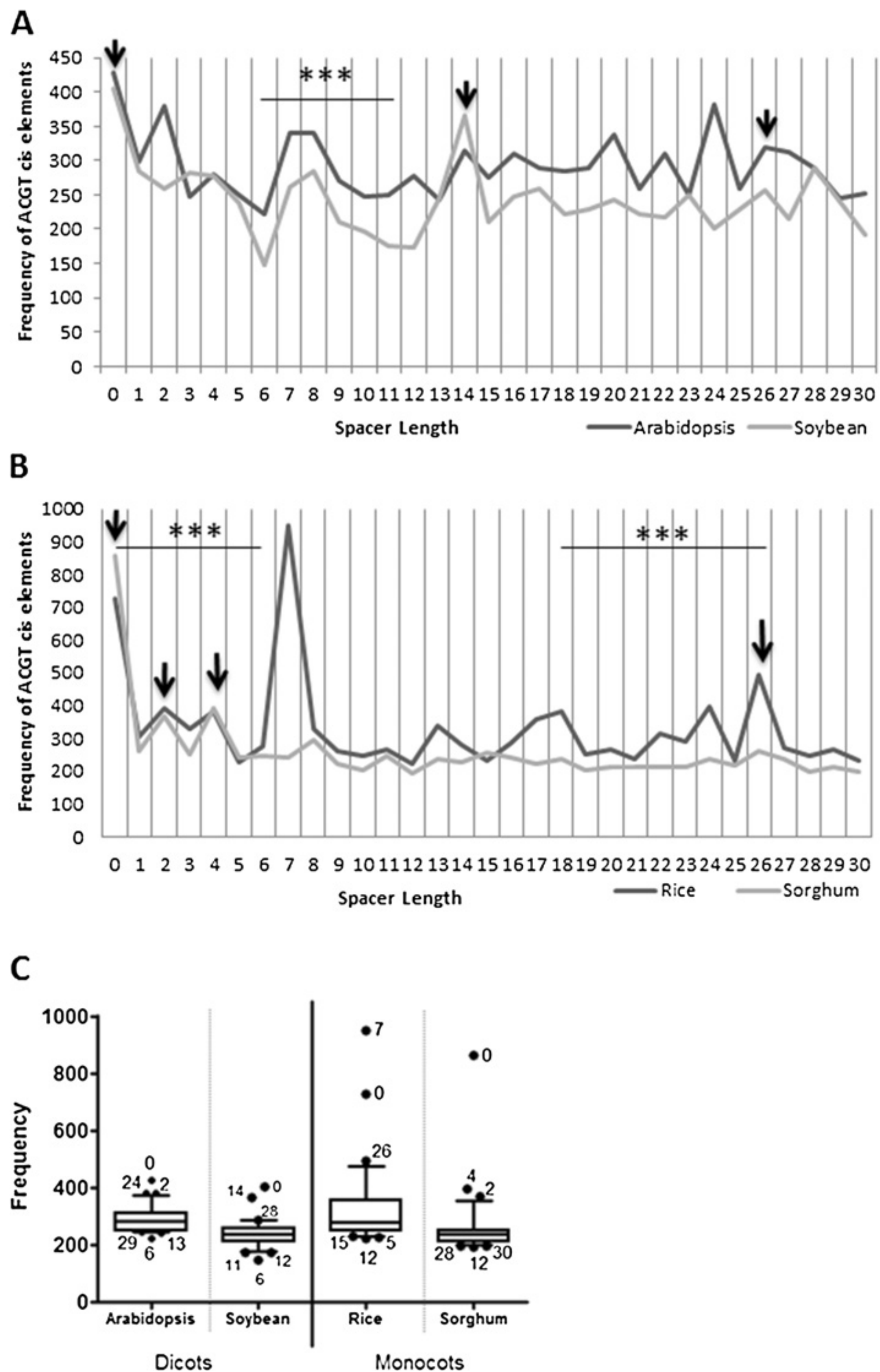


Figure 1 Variation in spacer length frequencies across species. (**A**) A comparison of ACGT elements separated by increasing spacer distance (<31) between two dicots – Arabidopsis and Soybean. Possible common peaks include 0, 14 and 26. Spacer lengths 6–11 show a high degree of correlation between the two genomes. (**B**) A similar comparison between two monocots – Rice and Sorghum, with possible peaks at 0, 2, 4 and 26. Spacer lengths 0–6 and 18–26 show significant correlation. (**C**) A 10–90% box and whiskers plot suggests 0 as the only common peak across all species. Interestingly, spacer distance 6 emerges as a common dip in dicots, whereas 12 exists as a common dip in monocots. Arrow – potential common peaks. *** $p < 0.001$.

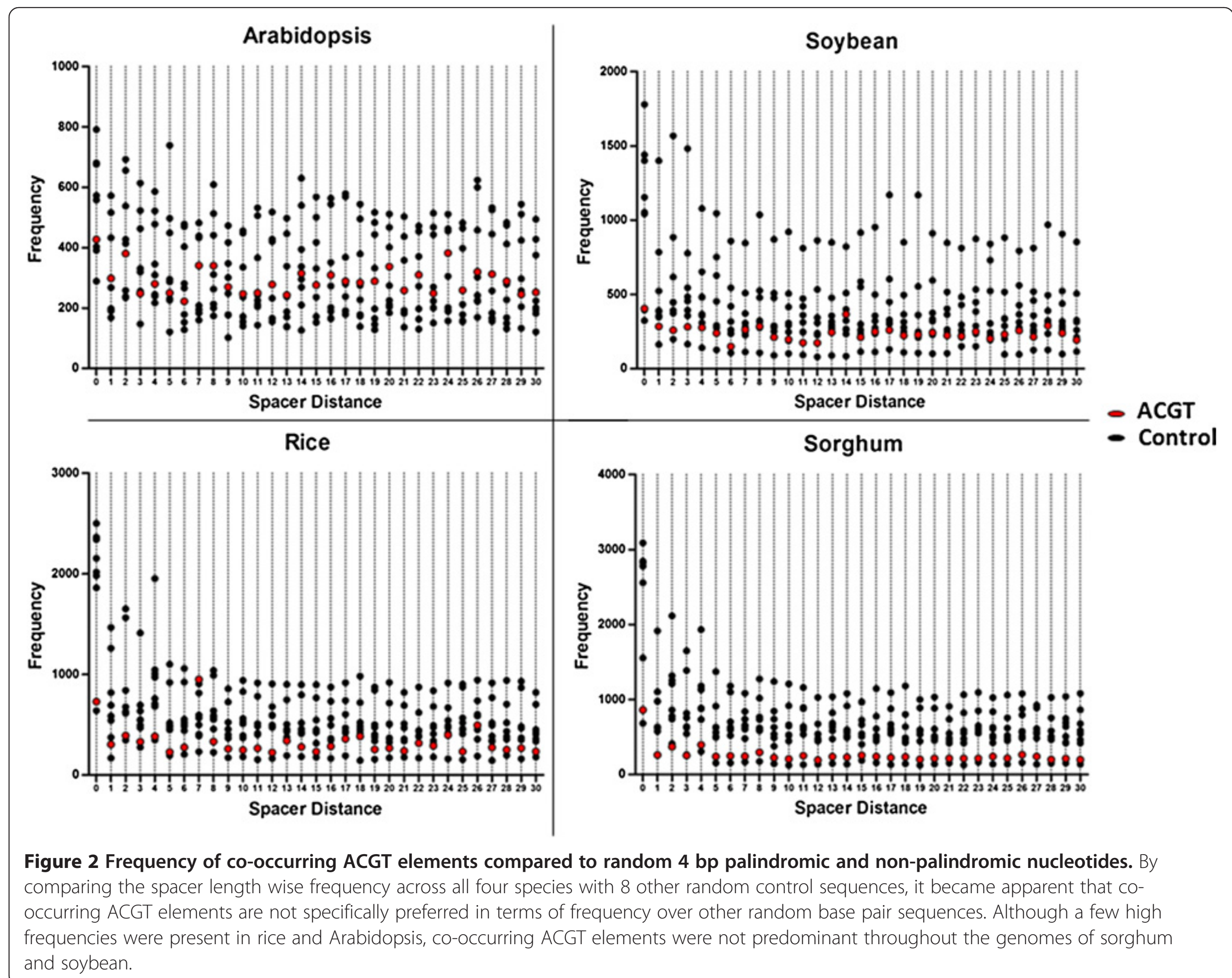


Figure 2 Frequency of co-occurring ACGT elements compared to random 4 bp palindromic and non-palindromic nucleotides. By comparing the spacer length wise frequency across all four species with 8 other random control sequences, it became apparent that co-occurring ACGT elements are not specifically preferred in terms of frequency over other random base pair sequences. Although a few high frequencies were present in rice and Arabidopsis, co-occurring ACGT elements were not predominant throughout the genomes of sorghum and soybean.

Results

Spacer frequency comparison shows common peaks and dips for different spacer lengths across species

We extracted 1 kb regions upstream of the ATG site for 33,323 genes in Arabidopsis, 49,841 genes in soybean, 30,294 genes in rice and 32,886 genes in sorghum. In total, the number of *cis*-elements of the form ACGT (N) ACGT ($N \leq 30$) per 1000 promoters was found to be 271 for Arabidopsis, 341 for Rice, 246 for Sorghum and 151 for Soybean. Our data indicated an overall frequency variation for $N = 0–30$ for the four species under analysis. Attempting to identify conservation between monocots and dicots, we identified potential peaks at spacer lengths of 0, 14 and 26 for dicots (Figure 1A) and lengths of 0, 2, 4 and 26 (Figure 1B) across monocots. Following this, the box and whisker's plot (10%–90%) interestingly indicated that while spacer length 0, i.e., two ACGT's in tandem, appeared to be significantly high across all species, there was no common peak across all four genomes (Figure 1C). Frequencies for $N = (0, 2, 24)$ for Arabidopsis, (0, 14, 28) for Soybean, (0, 7, 26) for Rice and (0, 2, 4) for Sorghum emerged to lie outside the 90th percentile. A particularly surprising observation was that spacer length 6 and the region from 11–13 had the lowest frequencies across both dicots, whereas spacer length 12 was present as a common dip between monocots.

Local regions of conservation in monocots and dicots

Based on trends in variation of frequencies in the four species, we observed patterns of high correlation specific to monocots and dicots (Figure 1A, B). The two dicotyledonous plants involved in our study - Arabidopsis and Soybean, depicted high correlation in frequencies for spacer distance (N) = 6–11($r = 0.974$; $t = 8.599$; $p = 0.0010$; $N = 6$) (Figure 1A). Similarly, Rice and Sorghum, both monocotyledonous plants, showed regions of high correlation for spacer distance (N) = 0–6 ($r = 0.984$; $t = 12.42$; $p = .0001$; $N = 7$) and (N) = 18–26 ($r = 0.934$; $t = 6.921$; $p = 0.0002$; $N = 9$). Comparisons across monocots

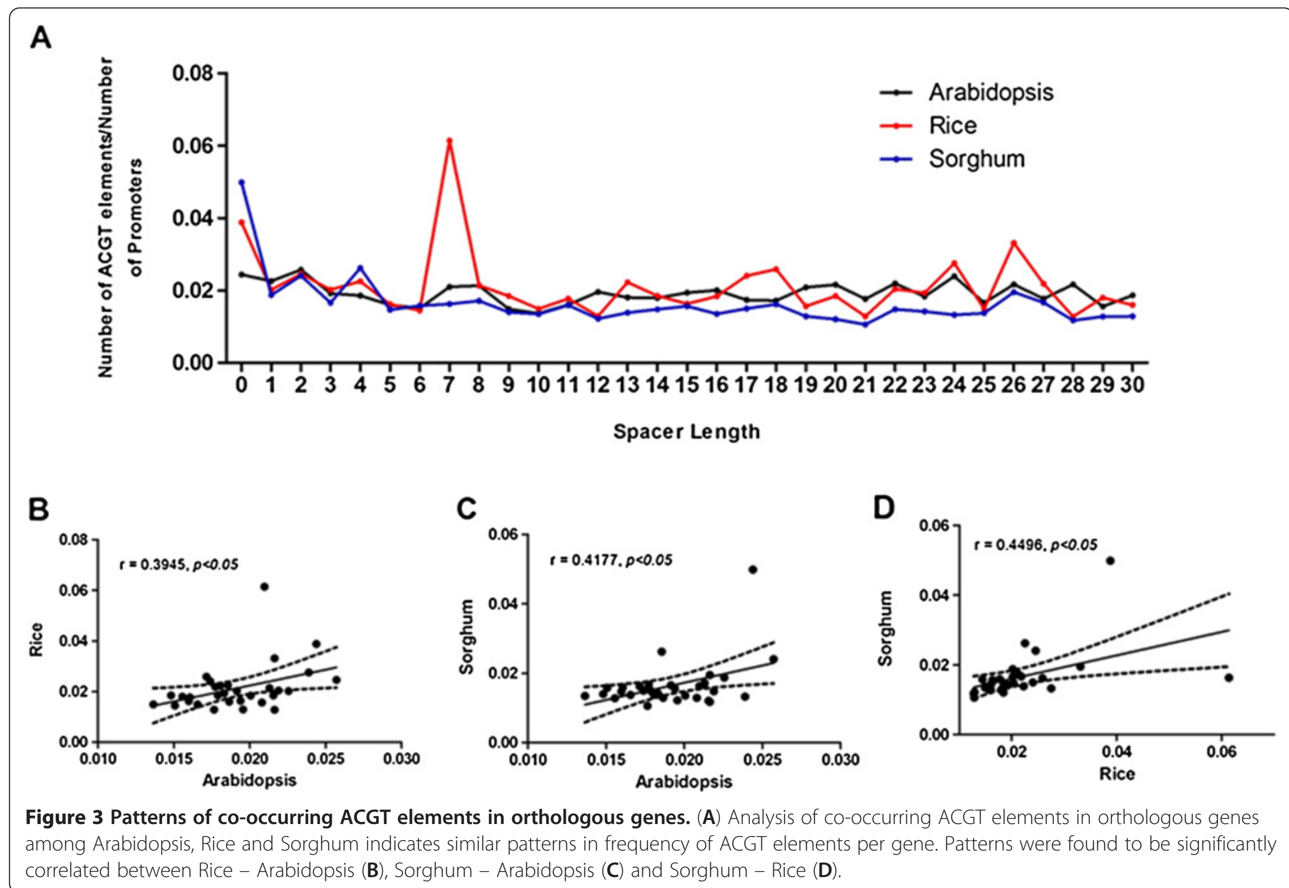


Figure 3 Patterns of co-occurring ACGT elements in orthologous genes. (**A**) Analysis of co-occurring ACGT elements in orthologous genes among Arabidopsis, Rice and Sorghum indicates similar patterns in frequency of ACGT elements per gene. Patterns were found to be significantly correlated between Rice – Arabidopsis (**B**), Sorghum – Arabidopsis (**C**) and Sorghum – Rice (**D**).

and dicots, for the most part, were found to be not as significantly correlated ($r < 0.8$).

Limited preference for ACGT elements compared to random elements

The frequency of occurrence of random 4 bp nucleotide sequences separated by varying spacer lengths in Arabidopsis, Soybean, Rice and Sorghum was determined, and compared within each species. A dot plot indicated that contrary to our hypothesis, ACGT elements are not drastically preferred over other elements across all species (Figure 2). In fact, ACGT elements usually lay around the mean of the frequencies of the control elements, with the exception of sorghum, for which the frequency of ACGT elements was lower. For Arabidopsis and Rice, spacer lengths 7 and 24 were present in higher frequencies, being surpassed only by ATGC, GATC and TGCA, which were predominantly high across all species (details in Additional file 2).

Patterns of ACGT *cis* elements are highly conserved among orthologous gene groups

We performed a frequency analysis for ACGT repeats in promoters of all orthologous gene groups reported between Arabidopsis (12013 genes), Rice (11453 genes) and Sorghum (10575 genes). Interestingly, orthologous genes in all three species exhibit similar trends in spacer distance frequency (Figure 3A), with significant correlations across all spacer lengths for Arabidopsis and Rice ($r = 0.3945$, $p < 0.05$, $n = 31$) (Figure 3B), Arabidopsis and Sorghum ($r = 0.4177$, $p < 0.05$, $n = 31$) (Figure 3C), and Rice and Sorghum ($r = 0.4495$, $p < 0.5$, $n = 31$) (Figure 3D).

The first and last positions within a spacer sequence are conserved for all spacer lengths in Arabidopsis

Consensus spacer sequences for spacers of length 0–30 were determined to be majorly composed of random nucleotides, shown as N in the sequence (Figure 4). However, the consensus sequences indicate a marked preference for G at the first position in the spacer sequences (shown in red). Similarly, a majority of the sequences indicate the presence of a C (shown in green) at the terminal end of the sequence. As an exception from all the other spacer sequences, the consensus sequence of spacer length 24 (shown in blue) is composed of conserved nucleotides at most (17/24) of the positions. Such a trend is not noted in any other spacer length,

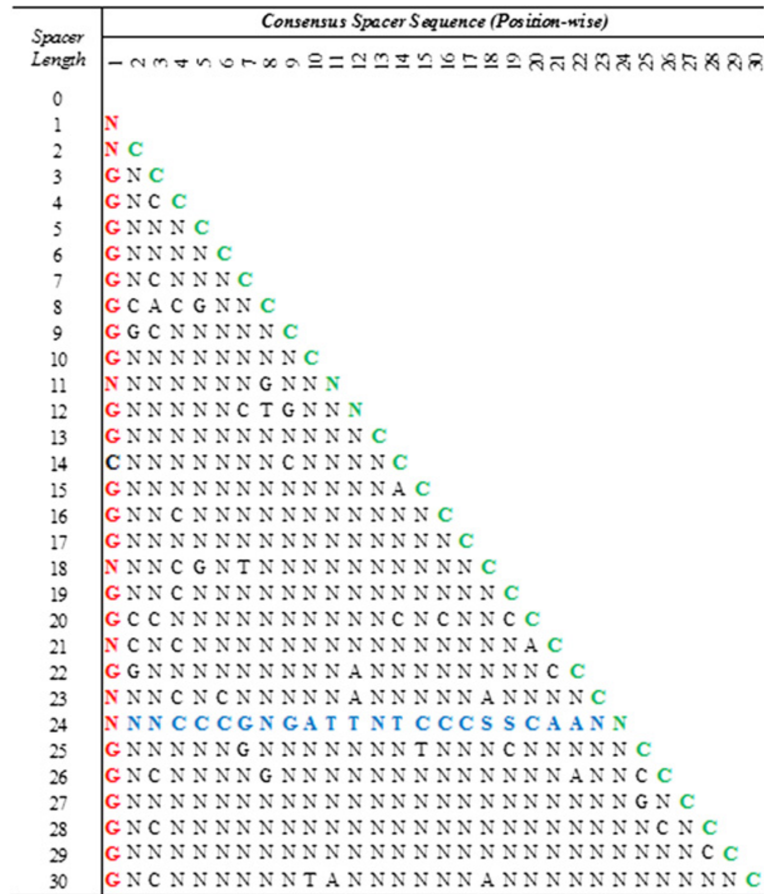


Figure 4 Consensus spacer sequences of varying length with ACGT elements flanking them. The first nucleotide for each spacer shows a clear preference for G (red), whereas the terminal nucleotide shows the conservation of C (green). Spacer length 24 is observed to be conserved to a greater extent, as compared to other spacer lengths (blue). *A* Adenine; *T* Thymine; *G* Guanine; *C*, Cytosine; *S* Guanine/Cytosine; *N* None of the bases met threshold requirements (A/T/G/C).

indicating a high degree of conservation for the sequence of this particular spacer length.

ACGT repeat elements across species are preferentially regulated by salt and drought stress conditions

We identified the genes regulated by specific conditions, which also contain an upstream ACGT (N) ACGT element. The overall likelihood of occurrence was calculated for each condition, with a likelihood of 1 being that of random chance. Therefore, conditions greater than 1.3 were selected for further analysis (Figure 5A); including environment conditions such as salt and drought, and hormones such as jasmonic acid. The likelihood for each of these conditions was compared with that of the same condition for the previously described 8 control sequences. Dot plots for each of the conditions and a Grubbs' test for outliers indicated a significant effect for ACGT containing promoters to be up-regulated by salt (Mean = 0.9791; Z = 2.21: $p < 0.5$; n = 9) and drought (Mean = 0.9539; Z = 2.21: $p < 0.5$; n = 9) (Figure 5B). An individual spacer distance -wise split up revealed considerable fluctuation in the likelihoods for each spacer length, and despite no clear pattern emerging over all functions (A 3), suggested potential spacer length specific gene regulation. Further, from microarray data for *Oryza sativa* we performed an identical analysis for our two candidate functions – salt and drought up-regulation. Both conditions showed likelihood of occurrences greater than 1.20, and the Grubbs' test for outliers emerged to be significant for both salt (Mean = 0.9611; Z = 2.21: $p < 0.5$; n = 9) and drought(Mean =0.8453; Z = 2.21; $p < 0.5$; n = 9) up-regulation (Figure 5C). Interestingly, similar fluctuating spacer length-wise patterns were observed for this dataset.

Discussion

The ACGT core motif forms an important class of *cis*-elements implicated in a variety of functions.

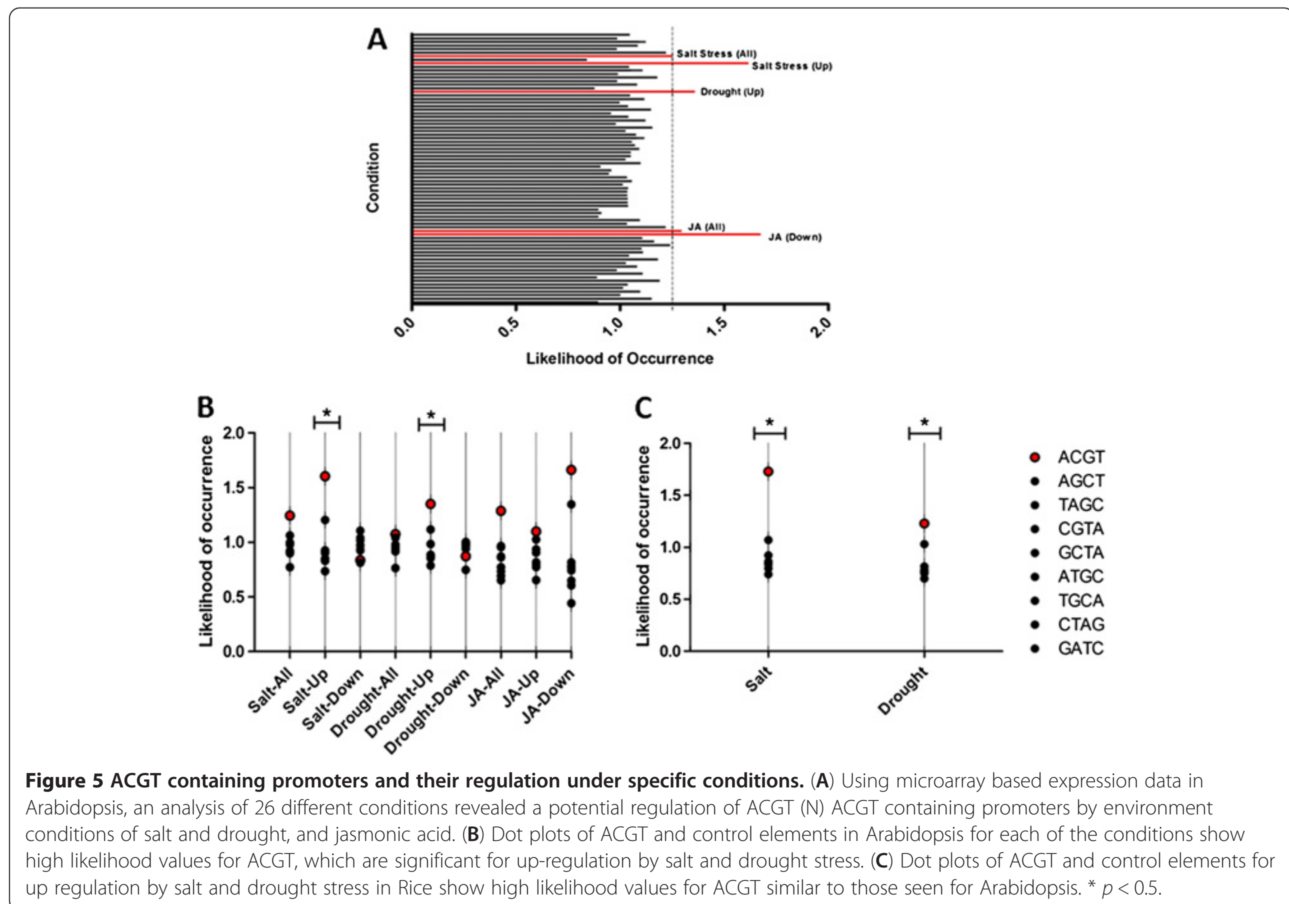


Figure 5 ACGT containing promoters and their regulation under specific conditions. (**A**) Using microarray based expression data in Arabidopsis, an analysis of 26 different conditions revealed a potential regulation of ACGT (N) ACGT containing promoters by environment conditions of salt and drought, and jasmonic acid. (**B**) Dot plots of ACGT and control elements in Arabidopsis for each of the conditions show high likelihood values for ACGT, which are significant for up-regulation by salt and drought stress. (**C**) Dot plots of ACGT and control elements for up regulation by salt and drought stress in Rice show high likelihood values for ACGT similar to those seen for Arabidopsis. * $p < 0.5$.

Multiple ACGT motifs have been shown to form enhancer elements which bind synergistically to transcription factors for gene regulation [40]. Our data indicates that certain spacer lengths are preferred over others in plant promoters. It is possible that these spacer lengths are present in abundance due to extra stability conferred by helical phasing at these particular lengths [41]. A major finding was the peak at spacer distance (N) = 7 for rice, which is noteworthy as it is almost twice of the next highest frequency observed in the distribution. To the best of our knowledge, there are no previous reports describing an interaction between *cis*-elements in rice with such spatial constraints. Therefore, a further investigation for this observation could be interesting. Unfortunately, the peaks observed in frequency distributions were not consistent throughout the four plant species, or even among individual classes. However, a consistency in the frequency dips observed among monocots and dicots suggests a class specific mode of gene regulation. A possible explanation for the consistent dips could be that certain spacer lengths might cause sterically incompatible binding of transcription factors, explaining why they could be less preferred as compared to an otherwise compatible binding caused by other spacer lengths [42]. The theory of a class specific mode of regulation is also supported by differences in regions of high correlation in monocots and dicots.

The correlations in the patterns of ACGT repeat frequencies led us to speculate the precise mechanism of conservation of these repeats. The similar trends and subsequent functional significance of ACGT elements suggest two possible mechanisms of progression - either parallel or convergent evolution. We expect that a significant correlation in the spacer patterns of orthologous/paralogous genes groups would confirm the evolution of ACGT elements being a result of parallel evolution, whereas no correlation would suggest convergent mechanisms. Our results strongly suggest proof for parallel evolution, as patterns in ACGT elements appear to have evolved from a common ancestral gene and subsequently persisted in the descendent genes across species.

The subsequent comparison of co-occurring ACGT elements with random 4 bp nucleotides indicated that ACGT elements are not as predominant in plant promoters as compared to other random elements. This observation indicates that although a higher frequency results from a

preferred occurrence, it might not be a result of conservation in the genome. Nevertheless, on analyzing the same random elements for functional preference using microarray data, ACGT (N) ACGT elements were found to be predominant in spite of lower frequencies than random elements in all promoters. This fact underscores the functional relevance of ACGT (N) ACGT *cis*-elements.

Based on our functional analysis, we deduced that co-occurring ACGT elements are involved in gene regulation in response to stress conditions in both Arabidopsis and Rice, suggesting species-wide functional significance. Genes which were up-regulated by salt and draught stress were much more likely to contain *cis*-elements of the form ACGT (N) ACGT in their promoters. This observation is supported by previous reports that multiple basic leucine zipper transcription factors, which recognize the ACGT core site, have been implicated in response under drought and high salinity conditions in Arabidopsis [16,43]. Similarly, bzip transcription factors have been shown to be involved in regulation under drought stress in Rice and Soybean [44,45]. Further, although regulation by jasmonic acid could not satisfy the criteria for the Grubs' outlier tests, the likelihood of occurrence of genes regulated by jasmonic acid was higher than the cutoff of 1.30. With jasmonic acid's conventional involvement in mediating stress responses in plants [46], this observation is extremely interesting in light of our findings.

Spacer sequence results showed a clear preference for G at the first and C at the terminal position for almost all spacer lengths. This validates previous reports which state that the sequence requirement of the ACGT-containing ABRE is ACGT-G G/T C [47]. The CACGTG motif, or the G-box, is recognized in rice (Kumar A, 2009), and our results clearly show a predominance of G at +2 and C at −2 positions. This result also corresponds to reports which state that the bZip class of TFs show enhanced binding to ACGT elements with the presence of the G box [48]. It can therefore be inferred, that a majority of TFs which bind to ACGT elements have stronger interactions if the flanking nucleotides are C and G. While the overall spacer sequence did not show any clear consensus sequence, spacer distance 24 showed a large amount of conservation. Spacer sequence between two ACGT motifs in a *cis*-element can be crucial for gene regulation [49]. From our frequency analysis, we determined that spacer length 24 also appears to be preferred in Arabidopsis. In light of these observations, it is possible that a spacer of 24 bases might be functionally relevant in Arabidopsis.

Determining the optimum spacer length and preferred spacer sequence could dramatically enhance promoter designing techniques. If a particular spacer length is confirmed to be implicated in regulation of a particular function, e.g. - stress response, a *cis*-element containing the ACGT repeated motif can be incorporated within promoters to give rise to sturdier and more resistant genetically modified crops [50]. Therefore, identifying the mechanism and implications of the conservation of specific spacer lengths and sequences is of prime importance for various genetic engineering techniques. Having identified spacer lengths between ACGT elements which can up-regulate gene expression in conditions of draught and salt stress, these results suggest improved methods for promoter design and creating hardy plant varieties.

Conclusions

This is the first study which has attempted to analyze patterns of ACGT repeat *cis*-elements in four plant genomes. We established that each species exhibits preferences for particular spacer lengths and demonstrated the existence of spacer lengths preferentially avoided in monocots and dicots. This suggests that a class specific mechanism of gene regulation might be present for ACGT (N) ACGT elements. We further identified parallel evolution to be the underlying mechanism for ACGT co-occurrences across species. Moreover, by indicating that genes up-regulated by salt and drought stress are more likely to contain ACGT repeat *cis*-elements in their promoters, our *in-silico* results suggest a significant role of these elements in these pathways.

Additional files

Additional file 1: Table S2. Frequency analysis of control cis-elements across four plant species. A comparison of the frequency of occurrence of each of the 4 palindromic and 4 non-palindromic control elements used for N=(0-30) in -1000 bp regions of Arabidopsis, Soybean, Sorghum and Rice.

Additional file 2: Table S3. Functional analysis of Arabidopsis and Rice genes containing ACGT *cis*-elements. The (A) total number of ACGT(N) ACGT (0<=N<=30) containing promoters in Arabidopsis that are also regulated by certain conditions, with a spacer-length wise frequency analysis, (B) likelihood of occurrence to contain ACGT(N)ACGT, calculated for each of these conditions, (C) identical analysis conducted for each of the eight control elements used previously (Additional file 1), (D) comparison of the likelihoods for ACGT and each of the controls, with conditions highlighted in green being those for which the likelihood of ACGT is significant in the Grubbs' test for outliers (E) total number of ACGT(N)ACGT (0<=N<=30) containing promoters in Rice that are up-regulated by salt and drought stress, with a spacer length wise analysis and a similar analysis for control elements and (F) comparison of the likelihoods for ACGT and for each of the controls in Rice with both conditions emerging significant in Grubbs' test for outliers.

Competing interest
Authors are not having any competing interest.

Authors' contributions
RM conceived the study and participated in its co-ordination. RM also gave critical inputs on designing of experiments. SS and IZ designed the experiments, conducted the analysis and drafted the manuscript. PB and SM

provided valuable inputs in shaping the manuscript to its final form. All authors read and approved the final manuscript.

Acknowledgements
We thank the Biological Sciences Department at Birla Institute of Technology and Science, Pilani for their cooperation. We also thank BITS Pilani, Pilani campus for their support in terms of infrastructure and facilities which were critical for completion of this project. We thank the editor and the reviewers for their valuable inputs.

Received: 30 July 2012 Accepted: 18 March 2013
Published: 25 March 2013

References

1. Ptashne M, Gann AA: **Activator and targets.** *Nature* 1990, **346**:329–333.
2. Stormo GD: **DNA binding sites: representation and discovery.** *Bioinformatics* 2000, **16**:16–23.
3. Butler JE, Kadonanhga JT: **Enhancer-promoter specificity mediated by DPE or TATA core promoter motifs.** *Genes Dev* 2001, **15**:2515–2519.
4. D'haeseleer P: **What are DNA sequence motifs?** *Nat Biotechnol* 2006, **24**(4):423–425.
5. Duret L, Bucher P: **Searching for regulatory elements in human noncoding sequences.** *Curr Opin Struct Biol* 1997, **7**:399–406.
6. Kimura M: *The neutral theory of molecular evolution.* Cambridge: Cambridge University Press; 1983.
7. Shashikant CS, Ruddle FH: **Combinations of closely situated *cis*-acting elements determine tissue-specific patterns and anterior extent of early Hoxc8 expression.** *Proc Natl Acad Sci U S A* 1996, **93**(22):12364–12369.
8. Carter AJ, Wagner GP: **Evolution of functionally conserved enhancers can be accelerated in large populations: a population-genetic model.** *Proc Biol Sci* 2002, **269**(1494):953–960.
9. Vardhanabhuti S, Wang J, Hannenhalli S: **Position and distance specificity are important determinants of *cis*-regulatory motifs in addition to evolutionary conservation.** *Nucleic Acids Res* 2007, **35**(10):3203–3213.
10. Ondek B, Gloss L, Herr W: **The SV40 enhancer contains two distinct levels of organization.** *Nature* 1988, **333**(6168):40–45.
11. Wu L, Berk A: **Constraints on spacing between transcription factor binding sites in a simple adenovirus promoter.** *Genes Dev* 1988, **2**(4):403–411.
12. Donald RG, Cashmore AR: **Mutation of either G box or I box sequences profoundly affects expression from the Arabidopsis rbcS-1A promoter.** *EMBO J* 1990, **9**(6):1717–1726.
13. McKendree WL Jr, Ferl RJ: **Functional elements of the Arabidopsis Adh promoter include the G-box.** *Plant Mol Biol* 1992, **19**(5):859–862.
14. Mason HS, DeWald DB, Mullet JE: **Identification of a methyl jasmonate-responsive domain in the soybean vspB promoter.** *Plant Cell* 1993, **5**(3):241–251.
15. Jupin I, Chua NH: **Activation of the CaMV as-1 *cis*-element by salicylic acid: differential DNA-binding of a factor related to TGA1a.** *EMBO J* 1996, **15**(20):5679–5689.
16. Guiltinan MJ, Marcotte WR Jr, Quatrano RS: **A plant leucine zipper protein that recognizes an abscisic acid response element.** *Science* 1990, **250**(4978):267–271.
17. Shen Q, Uknes SJ, Ho TH: **Hormone response complex in a novel abscisic acid and cycloheximide-inducible barley gene.** *J Biol Chem* 1993, **268**(31):23652–23660.
18. Busk PK, Pages M: **Regulation of abscisic acid-induced transcription.** *Plant Mol Biol* 1998, **37**:425–435.
19. Liu B, Ulmasov T, Shi X, Hagen G, Guilfoyle TJ: **Soybean GH3 promoter contains multiple auxin-inducible elements.** *Plant Cell* 1994, **6**:645–657.
20. Sawant SV, Singh PK, Madanala R, Tuli R: **Designing of an artificial expression cassette for high level expression of transgenes in plants.** *Theor Appl Genet* 2001, **102**:635–644.
21. Sawant SV, Kiran K, Mehrotra R, Chaturvedi CP, Ansari SA, Singh P, Lodhi N, *et al*: **A variety of synergistic and antagonistic interactions mediated by *cis* acting DNA motifs regulate gene expression in plant cells and modulate stability of the transcription complex formed on basal promoter.** *J Exp Bot* 2005, **56**:2345–2353.
22. Gómez-Porras JL, Riaño-Pachón DM, Dreyer I, Mayer JE, Mueller-Roeber B: **Genome-wide analysis of ABA-responsive elements ABRE and CE3 reveals divergent patterns in Arabidopsis and rice.** *BMC Genomics* 2007, **8**:260.
23. Mehrotra R, Mehrotra S: **Promoter activation by ACGT in response to salicylic and abscisic acids is differentially regulated by the spacing between two copies of the motif.** *J Plant Physiol* 2010, **167**:1214–1218.
24. Mehrotra R, Kiran K, Chaturvedi CP, Ansari SA, Lodhi N, Sawant S, Tuli R: **Effect of copy number and spacing of the ACGT and GT *cis* elements on transient expression of minimal promoter in plants.** *J Genet* 2005, **84**:183–187.
25. Marcotte WR Jr, Russel SH, Quatrano RS: **Abscisic acid-responsive sequences from the Em gene of wheat.** *Plant Cell* 1989, **1**:969–976.
26. Mundy J, Yamaguchi-Shinozaki K, Chua N-H: **Nuclear proteins bind conserved elements in the abscisic acid-responsive promoter of a rice rab gene.** *Proc Natl Acad Sci* 1990, **87**:1406–1410. USA.
27. Hattori T, Terada T, Hamasuna S: **Regulation of the Osem gene by abscisic acid and the transcriptional activator VP1: analysis of *cis*-acting prmoter elements required for regulation by abscisic acid and VP1.** *Plant J* 1995, **7**:913–925.
28. Ono A, Izawa T, Chua NH, Shimamoto K: **The rab76B promoter of rice contains two distinct abscisic acid-responsive elements.** *Plant Physiol* 1996, **11**(2):483–491.
29. Shen Q, Zhang P, Ho TH: **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* 1996, **8**:1107–1119.
30. *The Arabidopsis Information Resource (TAIR), v.10, RefSeq Assembly ID: GCF_000001735.3.* http://www.ncbi.nlm.nih.gov/assembly/GCF_000001735.3/.
31. *Oryza sativa Japonica Group, International Rice Genome Sequencing Project, Build 4.0,RefSeq Assembly ID: GCF_000005425.2.* http://www.ncbi.nlm.nih.gov/assembly/313038/.
32. *Sorghum Consortium, v. 1.0,RefSeq Assembly ID: GCF_000003195.2.* http://www.ncbi.nlm.nih.gov/assembly/GCF_000003195.2/.
33. *Glycine Max, US DOE Joint Genome Institute (JGI-PGF), RefSeq Assembly ID: GCF_000004515.1.* http://www.ncbi.nlm.nih.gov/assembly/310798/.
34. Higo K, Ugawa Y, Iwamoto M, Korenaga T: **Plant *cis*-acting regulatory DNA elements (PLACE) database.** *Nucleic Acids Res* 1999, **27**(1):297–300.
35. Mishra AK, Agarwal S, Jain CK, Rani V: **High GC content: critical parameter for predicting stress regulated miRNAs in Arabidopsis thaliana.** *Bioinformation* 2009, **4**(4):151–154.
36. *MSU Rice Genome Annotation Project.* http://rice.plantbiology.msu.edu/annotation_pseudo_apk.shtml.
37. Sakai H, Lee SS, Tanaka T, Numa H, Kim J, Kawahara Y, Wakimoto H, Yang CC, Iwamoto M, Abe T, Yamada Y, Muto A, Inokuchi H, Ikemura T, Matsumoto T, Sasaki T, Itoh T: **Rice annotation project database (RAP-DB): an integrative and interactive database for rice genomics.** *Plant Cell Physiol* 2013, **54**(2):e6.
38. *The Gene Expression Atlas.* http://www.ebi.ac.uk/gxa/.
39. Tyagi AK, Kapoor S, Khurana JP, Ray S: *Expression data for stress treatment in rice seedlings.* NCBI; 2007.
40. Tokuhisa JG, Singh K, Dennis ES, Peacock WJ: **A DNA-binding protein factor recognizes two binding domains within the octopine synthase enhancer element.** *Plant Cell* 1990, **2**(3):215–224.
41. Kim TK, Maniatis T: **The mechanism of transcriptional synergy of an in vitro assembled interferon- β enhanceosome.** *Mol Cell* 1997, **1**(1):119–129.
42. Spitz F, Furlong EE: **Transcription factors: from enhancer binding to developmental control.** *Nat Rev Genet* 2012, **13**(9):613–626.
43. Uno Y, Furihata T, Abe H, Yoshida R, Shinozaki K, Yamaguchi-Shinozaki K: **Arabidopsis basic leucine zipper transcription factors involved in an abscisic acid-dependent signal transduction pathway under drought and high-salinity conditions.** *Proc Natl Acad Sci U S A* 2000, **97**(21):11632–11637.
44. Ning T, Hua Z, Xianghua L, Jinghua X, Xiong L: **Constitutive activation of transcription factor OsbZIP46 improves drought tolerance in rice.** *Plant Physiol* 2012, **158**(4):1755–1768.
45. Pereira SS, Guimarães FC, Carvalho JF, Stolf-Moreira R, Oliveira MC, Rolla AA, Farias JR, Neumaier N, Nepomuceno AL: **Transcription factors expressed in soybean roots under drought stress.** *Genet Mol Res* 2011, **10**(4):3689–3701. doi:10.4238/2011. October.21.5.
46. Delker C, Stenzel I, Hause B, Miersch O, Feussner I, Wasternack C: **Jasmonate biosynthesis in Arabidopsis thaliana–enzymes, products, regulation.** *Plant Biol (Stuttg)* 2006, **8**(3):297–306.
47. Hattori T, Totsuka M: **Experimentally determined sequence requirement of ACGT-containing abscisic acid response element.** *Plant Cell Physiol* 2002, **43**(1):136–140.
48. Krawczyk S, Thurow C, Niggeweg R, Gatz C: **Analysis of the spacing between the two palindromes of activation sequences-1 with respect to**

binding to different TGA factors and transcriptional activation potential. *Nucleic Acids Res* 2002, **30**:775–781.
49. Singh K, Tokuhisa JG, Dennis ES, Peacock WJ: **Saturation mutagenesis of the octopine synthase enhancer: correlation of mutant phenotypes with binding of a nuclear protein factor.** *Proc Natl Acad Sci U S A* 1989, **86**(10):3733–3737.
50. Mehrotra R, Gupta G, Sethi R, Bhalothia P, Kumar N, Mehrotra S: **Designer promoter: an artwork of *cis* engineering.** *Plant Mol Biol* 2011, **75**(6):527–536.

doi:10.1186/1471-2164-14-203
Cite this article as: Mehrotra *et al.*: Patterns and evolution of ACGT repeat *cis*-element landscape across four plant genomes. *BMC Genomics* 2013 **14**:203.

Submit your next manuscript to BioMed Central and take full advantage of:

- **Convenient online submission**
- **Thorough peer review**
- **No space constraints or color figure charges**
- **Immediate publication on acceptance**
- **Inclusion in PubMed, CAS, Scopus and Google Scholar**
- **Research which is freely available for redistribution**

Submit your manuscript at
www.biomedcentral.com/submit

The Scientific World Journal
Volume 2012, Article ID 983528, 5 pages
doi:10.1100/2012/983528

The ScientificWorldJOURNAL

Research Article

Evidence for Directed Evolution of Larger Size Motif in *Arabidopsis thaliana* Genome

Rajesh Mehrotra,[1,2] Amit Yadav,[1] Purva Bhalothia,[1] Ratna Karan,[2] and Sandhya Mehrotra[1]

[1] *Department of Biological Sciences, Birla Institute of Technology and Science, Pilani, Rajasthan 333031, India*
[2] *Department of Plant Environmentand Soil Sciences, Louisiana State University, Baton Rouge, LA 70894, USA*

Correspondence should be addressed to Rajesh Mehrotra, rajmeh25@hotmail.com

Received 14 October 2011; Accepted 15 December 2011

Academic Editors: Z.-G. Han, Y. Y. Shugart, and W. A. Thompson

Copyright © 2012 Rajesh Mehrotra et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Transcription control of gene expression depends on a variety of interactions mediated by the core promoter region, sequence specific DNA-binding proteins, and their cognate promoter elements. The prominent group of *cis* acting elements in plants contains an ACGT core. The *cis* element with this core has been shown to be involved in abscisic acid, salicylic acid, and light response. In this study, genome-wide comparison of the frequency of occurrence of two ACGT elements without any spacers as well as those separated by spacers of different length was carried out. In the first step, the frequency of occurrence of the *cis* element sequences across the whole genome was determined by using BLAST tool. In another approach the spacer sequence was randomized before making the query. As expected, the sequence ACGTACGT had maximum occurrence in *Arabidopsis thaliana* genome. As we increased the spacer length, one nucleotide at a time, the probability of its occurrence in genome decreased. This trend continued until an unexpectedly sharp rise in frequency of (ACGT)N25(ACGT). The observation of higher probability of bigger size motif suggests its directed evolution in *Arabidopsis thaliana* genome.

1. Introduction

Gene expression in eukaryotic organisms has been a topic of great interest. Careful regulation and recruitment of transcription factors (TFs) to *cis* regulatory elements in promoter regions lead to generation of specificity and diversity [1] in genetic regulation. Promoters are arrays of *cis* regulatory elements present upstream of a gene arranged with other specific *cis* elements. At present 469 *cis* elements have been reported in the plant *cis* regulatory element (PLACE) database. The prominent group of *cis* acting elements in plants contains an ACGT core. Several *cis* elements with this core have been shown to be responding to abscisic acid [2–4], salicylic acid [5], and light signals [6]. It has been reported by Foster et al. [7] that bZIP class of transcription factors binds to this core motif. In an elegant study Krawczyk et al. [8] showed deletion of two base pairs between activator sequence-1 (as1) palindromes does not affect binding of activator sequence binding factor (ASF-1) and TGA factors (which binds to TGACG sequence), whereas insertion decreases factor binding *in vitro*. In their study the distance between palindromic centers was 12 base pairs. Mehrotra et al. [9, 10] have shown that this motif functions even when they are placed out of the native context. R. Mehrotra and S. Mehrotra [11] have shown that promoter activation by ACGT in response to salicylic and abscisic acids is differentially regulated by the spacing between these motifs. It contributes synergistically to gene expression by stabilising the transcription complex formed on minimal promoter [10]. The present study is an extension of aforementioned work. In this study, genome-wide comparison of the frequency of occurrence of two ACGT elements without any spacers and also separated by spacers of different lengths was done. Based on the data obtained we report that there is a directed evolution of bigger size of motif in the *Arabidopsis thaliana* genome.

2. Materials and Methods

The objective was to find out the frequency of the recurring sequences and then use these recurring sequences with

Table 1: Frequency of occurrence of the various promoter sequences in which spacer sequence length between two ACGT palindromes is gradually increased from 5 to 25 nucleotides.

	Cis element	Chromosome 1	Chromosome 2	Chromosome 3	Chromosome 4	Chromosome 5	Total
$(ACGT)_2$	ACGTACGT	469	312	367	327	410	1885
$(ACGT)_8$	ACGTACGTACGTACGTAC-GTACGTACGTACGT	70	31	12	28	59	200
$(ACGT)_{N5}(ACGT)$	ACGTGGCTAACGT	16	11	13	13	19	72
$(ACGT)_{N10}(ACGT)$	ACGTGGCTATGGCGACGT	8	5	10	4	12	39
$(ACGT)_{N25}(ACGT)$	ACGTGGCTATGGCGGAGC-AAGATTCACTCACGT	15	12	13	9	13	62
$(ACGT)_{RN5}(ACGT)$	ACGT--GCTAG--ACGT	7	5	5	2	4	23
$(ACGT)_{RN10}(ACGT)$	ACGT--TGGGGCCGAT--ACGT	2	2	4	3	3	14
$(ACGT)_{RN25}(ACGT)$	ACGTAGACACGTTGGGGG-AACTTACTGCCACGT	3	1	7	5	5	21
$(ACGT)_{RN25}(ACGT)$	ACGT-ATATGAGATCGGCGCT-TCACGGAGC-ACGT	4	14	6	4	4	32
$(ACGT)_{N5}(ACGT)$ randomized	GGAATCCTTGGCA	41	24	30	19	23	137
$(ACGT)_{N10}(ACGT)$ randomized	GCGGGCTATCGGTAGCAT	2	5	2	0	1	10
$(ACGT)_{N25}(ACGT)$ randomized	TAAGGCTTAGCCACGCTT-AGGGTGTGAGCACAC	6	6	3	0	3	18
$(TGCA)_{N25}(TGCA)$	TGCAGGCTATGGCGGAGC-AAGATTCACTCTGCA	13	12	9	12	9	55

N5, N10, N25 denote sequence length between two ACGT palindromes. RN5, RN10, RN25—signify only spacer sequence being randomized. $(ACGT)_{N_}(ACGT)$ randomized—signify complete sequence being randomized.

a random minimal promoter to predict transcription factors likely to interact with them.

The genomic sequence database of *Arabidopsis thaliana* at http://www.arabidopsis.org/ (The *Arabidopsis* Information Resource, TAIR) was analyzed using software BLASTn (available at NCBI website). All sequences were run in BLASTn against whole *Arabidopsis thaliana* genome to find their frequency of occurrence. Accession numbers of *Arabidopsis thaliana* chromosomes are as follows: chromosome 1: NC_003070.9, chromosome 2: NC_003071.7, chromosome 3: NC_003074.8, chromosome 4: NC_003075.7, and chromosome 5: NC_003076.8.

Randomization of the sequence was carried out using SHUFFLE program [12]. Different sequences obtained are listed in Table 1. In the next step we found the transcription factors binding to these *cis* elements separated by different length of nucleotides. A 139 bp long minimal promoter *Pmec* [13] was used in this study. The minimal promoter sequence as shown below was suffixed to the sequences shown in Table 1;

TCACTATATATAGGAAGTTCATTTCATTTGGAA-TGGACACGTGTTGTCATTTCTCAACAATTACCAACA-ACAACAAACAACAAACAACATTATACAATTACTATT-TACAATTACATCTAGATAAACAATGGCTTCCTCC.

These extended sequences were used in JASPAR core database [14] to scan for transcription factors and then these TFs were crosschecked with results obtained from CONSITE [15].

3. Results and Discussion

3.1. Promoters with Greater Length between ACGT Motifs Are More Frequent. It has been reported that ACGT *cis* elements function even when they are placed out of native sequence context [9, 10]. When the distance of separation between two ACGT elements are 5 base pairs, and 10 base pairs, they are induced in response to salicylic acid (SA) and abscisic acid (ABA), respectively. Interestingly, SA mimics biotic stress response and ABA mimics abiotic stress response in plants and thus is of great interest to plant biologists. Paixão and Azevedo [16] showed that multiplicity of *cis* element evolved through transitional forms showing redundant *cis* regulation. In this study, when the frequency of occurrence of two ACGT elements without any spacers and also separated by the spacer of different lengths was observed, we found that the total frequency of occurrence of two ACGT element in tandem is 1885 (Table 1), while the e value was same for all alignments obtained on a particular chromosome. When two ACGT elements were separated by spacer of 5, 10, and 25 nucleotides their frequency of occurrence was 72, 39, and 62, respectively. An unexpectedly high frequency of occurrence was observed when two ACGT elements were separated by 25 base pairs. According to the rule of probability the frequency of two ACGT elements separated by 25 base pairs should be less than when they are separated by 10 base pairs or lesser. Hobo et al. [17] have earlier reported that in ABA responsive promoters the distance between ACGT elements

Table 2: Frequency of occurrence of nitrogenous bases when spacer sequence length between two ACGT palindromes is gradually increased from 5 to 25 nucleotides.

		A	C	G	T	Seq. used	Gap	Count
$(ACGT)_{N5}(ACGT)$	ACGTGGCT_ACGT	72	42	33	34	72	5	690
$(ACGT)_{N6}(ACGT)$	ACGTGGCTA_ACGT	98	65	45	44	44	6	611
$(ACGT)_{N7}(ACGT)$	ACGTGGCTAT_ACGT	92	91	77	80	77	7	824
$(ACGT)_{N8}(ACGT)$	ACGTGGCTATG_ACGT	97	30	64	55	64	8	852
$(ACGT)_{N9}(ACGT)$	ACGTGGCTATGG_ACGT	39	32	22	32	32	9	602
$(ACGT)_{N10}(ACGT)$	ACGTGGCTATGGC_ACGT	34	36	39	66	39	10	600
$(ACGT)_{N11}(ACGT)$	ACGTGGCTATGGCG_ACGT	36	23	38	29	38	11	681
$(ACGT)_{N12}(ACGT)$	ACGTGGCTATGGCGG_ACGT	56	54	65	45	56	12	638
$(ACGT)_{N13}(ACGT)$	ACGTGGCTATGGCGGA_ACGT	78	50	77	59	77	13	652
$(ACGT)_{N14}(ACGT)$	ACGTGGCTATGGCGGAG_ACGT	86	53	96	52	53	14	841
$(ACGT)_{N15}(ACGT)$	ACGTGGCTATGGCGGAGC_ACGT	56	67	44	66	56	15	709
$(ACGT)_{N16}(ACGT)$	ACGTGGCTATGGCGGAGCA_ACGT	60	34	52	34	60	16	843
$(ACGT)_{N17}(ACGT)$	ACGTGGCTATGGCGGAGCAA_ACGT	39	41	42	39	42	17	830
$(ACGT)_{N18}(ACGT)$	ACGTGGCTATGGCGGAGCAAG_ACGT	49	47	58	48	49	18	719
$(ACGT)_{N19}(ACGT)$	ACGTGGCTATGGCGGAGCAAGA_ACGT	50	38	49	44	44	19	695
$(ACGT)_{N20}(ACGT)$	ACGTGGCTATGGCGGAGCAAGAT_ACGT	34	30	44	37	37	20	821
$(ACGT)_{N21}(ACGT)$	ACGTGGCTATGGCGGAGCAAGATT_ACGT	36	40	42	43	40	21	717
$(ACGT)_{N22}(ACGT)$	ACGTGGCTATGGCGGAGCAAGATTC_ACGT	53	42	42	46	53	22	726
$(ACGT)_{N23}(ACGT)$	ACGTGGCTATGGCGGAGCAAGATTCA_ACGT	91	55	60	61	55	23	771
$(ACGT)_{N24}(ACGT)$	ACGTGGCTATGGCGGAGCAAGATTCAC_ACGT	77	64	57	53	53	24	1171
$(ACGT)_{N25}(ACGT)$	ACGTGGCTATGGCGGAGCAAGATTCACT_ACGT	76	62	58	69	62	25	708

Table 3: Alterations in transcription factor binding sites when spacer sequence length between two ACGT palindromes is gradually increased from 5 to 25 nucleotides.

	Minimal promoter sequence (MPS)	(ACGT)	(ACGT) (MPS)	$(ACGT)_2$ (MPS)	$(ACGT)_{N5}$ (ACGT)(MPS)	$(ACGT)_{N10}$ (ACGT)(MPS)	$(ACGT)_{N25}$ (ACGT)(MPS)
Model name	Frequency						
ARR10	0	0	0	0	0	0	1
AGL3	2	0	2	2	2	2	2
ATHB-5	1	0	1	2	1	1	1
bZIP910	0	0	0	0	1	1	1
Dof3	1	0	1	1	1	1	2
EmBP-1	2	0	2	1	2	2	2
Gamyb	5	0	5	5	5	5	5
HAT5	2	0	2	2	2	2	2
HMG-1	6	0	6	6	6	6	6
HMG-I/Y	6	0	6	6	6	6	6
id1	5	0	5	5	5	5	5
myb.Ph3	1	0	1	1	2	1	1
PEND	1	0	1	1	1	1	1
squamosa	2	0	3	3	3	3	3
TGA1A	1	0	1	1	2	2	2
	35	**0**	**36**	**36**	**39**	**38**	**40**

is 30 base pairs. To address this discrepancy in the data obtained, we randomized the spacer sequence keeping the ACGT motif unchanged. The logic of this randomization was to identify how important is the distance between the binding sites for transcription factors. After randomization of the spacer there was a drop in the frequency of occurrence to 23, 14, and 21 from 72, 39, and 62 for $(ACGT)_{N5}(ACGT)$, $(ACGT)_{N10}(ACGT)$, and $(ACGT)_{N25}(ACGT)$, respectively. This means that along with the distance between binding motifs there has been a positive selection for the sequence of the spacer in transcriptional regulation. In the next step we completely randomized the sequence and we observed that there is a drop in frequency of occurrence of two ACGT elements when separated by 10 and 25 base pairs while there was an unexpected increase in the frequency when ACGT elements were separated by five base pairs. This happened because randomization generated a motif that has been positively selected in evolution.

3.2. A and G Are the Preferred Bases. We increased the spacer length one residue at a time and looked for the frequency of each resultant sequence in the database. As shown in Table 2, there has been preference for A and G in the spacer region between two ACGT sequences.

3.3. Increasing Spacing between Motifs Increases Transcription Factor Binding Sites. Potential transcription factor binding sites for all experimental sequences when predicted using JASPAR CORE software and subsequently crosschecked with CONSITE revealed the minimal promoter sequence to be possessing 35 potential TF binding sites (Table 3, MPS). Interestingly the sequence ACGT as such has no site for binding of transcription factors but when minimal promoter is suffixed to it, an extra site for squamosa is generated and the total transcription factor binding site increases from 35 to 36 in minimal promoter alone (Table 3, (ACGT)(MPS)). When two ACGT elements in tandem are placed over minimal promoter sequence no extra site for binding of transcription factor is generated (Table 3, $(ACGT)_2(MPS)$). However, when ACGT elements are separated by five base pairs (Table 3, $(ACGT)_{N5}(ACGT)(MPS)$), four additional transcriptional binding sites are generated while ATHB-5 binding site which existed in the earlier cases is lost. The new sites generated are for transcription factors bzip9-10, EmBP-1, myb.Ph3, and TGA1a. Placement of two ACGT elements separated by 10 base pairs, however, resulted in loss of one myb.Ph3 site and the total transcriptional binding site decreased to 38 (Table 3, $(ACGT)_{N10}(ACGT)(MPS)$). In case when ACGT elements are separated by 25 base pairs followed by minimal promoter an additional site for ARR10 and dof3 was generated (Table 3, $(ACGT)_{N25}(ACGT)(MPS)$).

Based on the data obtained in this study, we report here that there has been directed evolution of bigger size of the motif in the *Arabidopsis thaliana* genome.

4. Conclusions

The central question in promoter evolution is to know how does *cis* regulatory element multiplicity evolved. The promoter regions of many genes contains multiple binding sites for the same transcription factor. Multiplicity may have evolved through transitional forms showing redundant *cis* regulation. In this paper, we focused on multiplicity of ACGT *cis* element and the distances between them which occurs in natural promoters. We found that ACGT element separated by 25 base pairs is more frequent than those by 10 base pairs which is against the law of probability. It signifies that under some evolutionary forces this interval was favoured since this distance may cause changes in the level of gene expression or in its robustness against variation in transcription factor concentration. Selection for different levels of expression of certain genes in certain environment could, over time, generates a positive association between *cis* element multiplicity and expression level.

Acknowledgments

The authors are grateful to the Department of Science and Technology, New Delhi, India for Grant-in-Aid and financial support to carry out this work bearing the file no. SR/FT/LS-126/2008. The authors are grateful to the administration of Birla Institute of Technology and Sciences, Pilani, Rajasthan for providing logistic support. They are thankful to Professor C. Gatz for critically reading the paper.

References

[1] G. A. Wray, M. W. Hahn, E. Abouheif et al., “The evolution of transcriptional regulation in eukaryotes,” *Molecular Biology and Evolution*, vol. 20, no. 9, pp. 1377–1419, 2003.

[2] M. J. Guiltinan, W. R. Marcotte, and R. S. Quatrano, “A plant leucine zipper protein that recognizes an abscisic acid response element,” *Science*, vol. 250, no. 4978, pp. 267–271, 1990.

[3] Q. Shen and T. H. Ho, “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*, vol. 7, no. 3, pp. 295–307, 1995.

[4] P. K. Busk and M. Pagès, “Regulation of abscisic acid-induced transcription,” *Plant Molecular Biology*, vol. 37, no. 3, pp. 425–435, 1998.

[5] I. Jupin and N. H. Chua, “Activation of the CaMV as-1 cis-element by salicylic acid: differential DNA-binding of a factor related to TGA1a,” *EMBO Journal*, vol. 15, no. 20, pp. 5679–5689, 1996.

[6] R. G. K. Donald and A. R. Cashmore, “Mutation of either G box or I box sequences profoundly affects expression from the *Arabidopsis* rbcS-1A promoter,” *EMBO Journal*, vol. 9, no. 6, pp. 1717–1726, 1990.

[7] R. Foster, T. Izawa, and N. H. Chua, “Plant basic leucine zipper proteins gather at ACGT elements,” *FASEB Journal*, vol. 8, pp. 192–200, 1994.

[8] S. Krawczyk, C. Thurow, R. Niggeweg, and C. Gatz, “Analysis of the spacing between the two palindromes of activation sequence-1 with respect to binding to different TGA factors and transcriptional activation potential,” *Nucleic Acids Research*, vol. 30, no. 3, pp. 775–781, 2002.

[9] R. Mehrotra, K. Kiran, C. P. Chaturvedi et al., “Effect of copy number and spacing of the ACGT and GT cis elements on transient expression of minimal promoter in plants,” *Journal of Genetics*, vol. 84, no. 2, pp. 183–187, 2005.

[10] S. V. Sawant, K. Kiran, R. Mehrotra et al., "A variety of synergistic and antagonistic interactions mediated by cis-acting DNA motifs regulate gene expression in plant cells and modulate stability of the transcription complex formed on a basal promoter," *Journal of Experimental Botany*, vol. 56, no. 419, pp. 2345–2353, 2005.

[11] R. Mehrotra and S. Mehrotra, "Promoter activation by ACGT in response to salicylic and abscisic acids is differentially regulated by the spacing between two copies of the motif," *Journal of Plant Physiology*, vol. 167, no. 14, pp. 1214–1218, 2010.

[12] L. Doelz, *BioCompanion*, Biocomputing Essentials Series, Dr. Ing. U. Doelz, Basel, Switzerland, 1990.

[13] S. Sawant, P. K. Singh, R. Madanala, and R. Tuli, "Designing of an artificial expression cassette for the high-level expression of transgenes in plants," *Theoretical and Applied Genetics*, vol. 102, no. 4, pp. 635–644, 2001.

[14] J. C. Bryne, E. Valen, M. H. E. Tang et al., "JASPAR, the open access database of transcription factor-binding profiles: new content and tools in the 2008 update," *Nucleic Acids Research*, vol. 36, no. 1, pp. D102–D106, 2008.

[15] A. Sandelin, W. W. Wasserman, and B. Lenhard, "ConSite: web-based prediction of regulatory elements using cross-species comparison," *Nucleic Acids Research*, vol. 32, pp. W249–W252, 2004.

[16] T. Paixão and R. B. R. Azevedo, "Redundancy and the evolution of cis-regulatory element multiplicity," *PLoS Computational Biology*, vol. 6, no. 7, Article ID e1000848, 2010.

[17] T. Hobo, M. Asada, Y. Kowyama, and T. Hattori, "ACGT-containing abscisic acid response element (ABRE) and coupling element 3 (CE3) are functionally equivalent," *Plant Journal*, vol. 19, no. 6, pp. 679–689, 1999.

Plant Mol Biol (2011) 75:527–536
DOI 10.1007/s11103-011-9755-3

REVIEW

Designer promoter: an artwork of *cis* engineering

**Rajesh Mehrotra · Gauri Gupta ·
Riccha Sethi · Purva Bhalothia ·
Narayan Kumar · Sandhya Mehrotra**

Received: 22 January 2011 / Accepted: 2 February 2011 / Published online: 14 February 2011
© Springer Science+Business Media B.V. 2011

Abstract Advances in systematic computational biology and rapid elucidation of synergistic interplay between *cis* and *trans* factors governing transcriptional control have facilitated functional annotation of gene networks. The generation of data through deconstructive, reconstructive and database assisted promoter studies, and its integration to principles of synthetic engineering has started an era of designer promoters. Exploration of natural promoter architecture and the concept of *cis* engineering have not only enabled fine tuning of single or multiple transgene expression in response to perturbations in the chemical, physiological and environmental stimuli but also provided researchers with a unique answer to various problems in crop improvement in the form of bidirectional promoters.

Keywords Designer promoter · *cis* engineering · *cis* elements · Bidirectional promoters · Inducible promoters · Chemical and light switches · Pathogen inducible promoters · Gene silencing

Introduction

Revolution in sequencing technologies during the post genomic era spawned a huge amount of whole genome data that facilitated functional annotation of genes, proposing their evolution and expression in specific type of cells under certain environmental conditions. Expression of genes is regulated by a number of factors like promoter strength, *cis*- and *trans*-acting factors, cell growth stage, the expression level of RNA polymerase associated factors and other gene-level regulation. Efforts to characterize various promoters provided insights into control and modulation of gene expression, identifying *cis* elements as the major regulatory factors. The proximal and distal promoter regions include *cis*-regulatory elements that contain binding sites for *trans*-acting regulatory proteins known as transcription factors (TF) (reviewed in Novina and Roy 1996). In order to regulate the expression of gene under conditions like abiotic stress, biotic stress etc., we need to decipher the meaning of *cis* regulatory elements within the promoter region. Advances in promoter technology provide a framework for designing an expression cassette that could not only provide precise control of transgene activity but also modulate expression of a transgene in various contexts. A synthetic promoter is constructed using an array of *cis*-acting element from various sources as building blocks for the promoter engineering strategy. The concept may allow the researcher to reuse selectable markers after each gene modulation step, improve the expression characteristics, reduce unwanted background expressions and delimitate the number of genes that can be modulated simultaneously.

Naturally occurring promoters have long been available. They can be isolated and used to control the gene of interest put under their control. These constructs can then be inserted into appropriate host and used for analysing expression profile. But, however diverse the strengths of these promoters might be, they cannot provide a wide range of promoter strengths in a continuous manner. Constitutive promoter is a matter of choice in cases where steady state expression of gene is needed. A number of strong constitutive promoters have been derived from caulimoviruses, particularly cauliflower mosaic virus

R. Mehrotra (✉) · G. Gupta · R. Sethi · P. Bhalothia · N. Kumar · S. Mehrotra
Department of Biological Sciences, BITS, Pilani, Rajasthan 333-031, India
e-mail: rajmeh25@hotmail.com

(CaMV). Some of the examples of plant promoters are CaMV 35S and 19S (Odell et al. 1985), the figwort mosaic virus (FMV) for full length transcript (Gowda et al. 1989), the nopaline synthase promoter (An et al. 1990), the octopine synthase promoter (Ellis et al. 1987) etc. CaMV 35S promoter is used extensively for the expression of plant genes as it is well characterised and is active in diverse variety of plants (Odell et al. 1985). The CaMV 35S promoter contains three domains: (1) Minimal core promoter containing TATA box from −46 to +8 bp with respect to transcription initiator site; (2) Domain A (also designated as sub-domain A1) expanding from −90 to −46 bp region that performs accessory role in increasing the transcriptional activity of upstream enhancers; (3) Domain B region (−343 to −90 bp) is further divided into five sub-domains (designated as B1–B5) that are recognised based on their interactions with various transcription factors. This region of upstream *cis*-regulatory sequence can function as transcription enhancer and also perform accessory role in increasing the transcriptional activity of enhancers (Benfey and Chua 1990; Fang et al. 1989). According to Benfey and Chua (1990) domains A and B regulate the tissue-specific gene expression patterns and their function were studied through combinatorial gain-of-function studies. However, domain A fused to minimal promoter alone doesn't express gene. But domain A combined with domain B allows expression of transgene in plant. This indicates that complete domain B contains important sequences at the interface of sub-domains and the additive effect of sub-domains of domain B along-with domain A and minimal core promoter allows regulated expression of transgene. Recently, three studies conducted by Bhullar et al. (2003, 2007, 2010) have revealed that domain swapping or rearrangement of *cis*-elements in CaMV 35S promoter region creates synthetic CaMV 35S promoters with minimum sequence homology whose transgene activity is equivalent to that of the wild type CaMV 35S promoter. Hence, CaMV 35S promoter is a significant promoter model for studying combinatorial *cis*-engineering in synthetic plant promoters.

More recently many tissue specific promoters have been characterised that specifically direct gene expression in certain tissues only. These promoters are defined by specific *cis*-elements that allow expression of a gene. Zavallo et al. (2010) have isolated and functionally characterised two novel seed specific promoters namely HaFAD2-1 and HaAP10, from sunflower. These promoters contained seed specific motifs like AACA motif, ACGT element, E-Boxes, SEF binding sites and GCN4 motif. They further proved that promoter HaFAD2-1 has GUS activity in transformed seeds equivalent to constitutive CaMV 35S promoter in *Arabidopsis* transgenic plant. Hence this promoter has expression level same as CaMV 35S promoter.

Rediscovering the natural models

There are two approaches to study promoter structure and function in conjugation with different *cis*-elements and transgenes:

(1) Deconstructive approach includes 5′ and 3′ deletion, analysis of activity of gene with respect to addition as well as removal of module or regulatory elements and specific mutations in nucleotide sequence of *cis*-elements. For example, certain auxin induced *cis*-regulatory elements like AUXRE (Auxin-Responsive Element) and ABRE (Abscisic Acid-Responsive Element) were characterised using above-mentioned as well as gain-of-function strategies (Liu et al. 1994). Kamisugi and Cuming (2005) used site directed mutagenesis to characterize a promoter in *Physcomitrella patens* and studied the evolution of the Abscisic acid-response in land plants.

(2) Reconstructive approach allows addition of specific *cis*-regulatory elements individually or in combination, to the minimal strong constitutive promoter, thereby creating a synthetic promoter (Venter 2007; Venter and Botha 2010). It also involves generation of synthetic bidirectional promoter expressing more than one transgene under the regulation of specific and identified *cis*-elements. For example, Rushton et al. (2002), designed pathogen-inducible plant promoters by placing multiple *cis*-regulatory elements (that can mediate gene expression under pathogen attack) upstream to strong wild type CaMV 35S promoter. The study also revealed that pathogen induced gene transcription can also activate local wound-induced gene expression in plant. Cazzonelli and Velten (2008) also designed synthetic promoter carrying short direct repeated enhancer elements from Geminivirus, Nanoviruses, Badnaviruses and Caulimoviruses. These enhancer elements or short regulatory cassettes individually or in combination were placed upstream to minimal CaMV 35S promoter carrying reporter gene. This strategy resulted in higher expression of gene when direct repeat cassettes were placed in combination to the minimal promoter. Reconstructive approach also includes study of *cis*-elements in conjugation with minimal promoter. Mehrotra et al. (2005), Mehrotra and Panwar (2009) and Mehrotra and Mehrotra (2010) have studied two *cis*-motifs i.e. ACGT and GT which regulate pathogen defence and their respective functions when they

are placed 50 nucleotides or 100 nucleotides upstream to TATA box. It was interpreted that single ACGT motif enhanced promoter expression by 2.39-fold when placed 100 nucleotides upstream to TATA box while two ACGT motifs separated by 5 nucleotides enhanced promoter expression by sixfold when placed 50 nucleotides upstream to minimal promoter.

A recent study by Mehrotra and Mehrotra (2010) indicated that two copies of ACGT when separated by 5 nucleotides allowed promoter activation by salicylic acid and when separated by 25 nucleotides, promoter was induced by abscisic acid but not salicylic acid. The differential induction is expected to involve the recruitment of different bZIP transcription factors (Lam and Lam 1995) and this change in spacing between two copies of a given motif can alter the signal pathway to which a promoter responds. Furthermore, single GT elements enhanced expression by nearly twofold when placed 50 nucleotides or 100 nucleotides upstream to promoter, but when two GT elements separated by 5, 10 or 25 nucleotides were placed, their expression plummeted significantly. Hence, it was proved that second copy of GT negatively regulated the expression of promoter when placed in close proximity to minimal promoter (Mehrotra and Panwar 2009).

(3) Database Assisted Promoter Analysis: Merely the identification and conjugation of *cis*-elements to one type of promoter is not sufficient to predict transcriptional product of a transgene. That is because a *cis*-element active under the control of one promoter may not transcribe desired protein with another promoter (Tiwari et al. 2003). Hence, it is of utmost importance to annotate all the sequences in a genome and represent introns, exons, *trans*-acting factors, *cis*-regulatory elements in promoters with the help of regulatory networks and logic functions. Ettwiller et al. (2003) had predicted *cis*-elements in *Sacchromyces cerevisiae* genome using functional networks. They discovered *cis*-elements using information from protein–protein interactions and metabolic networks. They presented them as patterns using pattern discovery tool such as Teiresias and represented them as motifs. Furthermore, the motifs from metabolic networks were compared to motifs from protein–protein interaction information and gave an 'overlap score'. Based upon this score and standard deviation of patterns from start codon, they were able to find 42 degenerate motifs from 647 original patterns in *S. cerevisiae*. They also verified that *cis*-regulatory sequences were composed of Adenine–Thymine rich patterns that offer transcription factor sites in the genome.

Certain regulatory sequence analysis tools are also freely available and allow retrieval of sequences in genomes, pattern discovery using oligo-analysis and dyad-analysis tools, pattern matching with putative transcription factor sites in sequenced set of genes and their upstream regions, Transcription Factor Binding Sites using Strings or Position-Specific Scoring Matches (PSSM) etc. (Helden 2003). Some of the scientists have linked transcription factor binding sites to defined set of basic functions. They have related the *cis*-regulatory module target sites individually or in combinations as logic operators like AND-OR-NOR. As per Istrail and Davidson (2004), each *cis*-regulatory sequence is occupied by a particular transcription factor and this site is given a particular 'occupancy score' that specifies equilibrium constants for the interaction of transcription factors and target sites. The scientist have represented interaction of transcription factors and target sites into four categories based upon the operation that they carry out, for example: (1) D, transcriptional activation operators, (2) F, basal transcription apparatus control operators, (3) G, combinatorial logic operators and (4) E, external control operators. For example, G operator can be represented as AND operator under the condition when a gene is expressed as a combined effect of two or more transcription factors. If any one of the transcription factor is not active, the gene would not be transcribed. All the operators for the *cis*-regulatory modules are further compiled and functionally interpreted.

Till date there are three main databases that identify transcription factor binding sites in upstream *cis*-regulatory elements in plant promoters. First is PLACE (Plant *cis*-Acting Regulatory DNA Elements) that compiles more than 469 *cis*-acting regulatory DNA motifs as reported on January 8, 2007. The database compiles all the motifs in plants that have been reported (Higo et al. 1999). Another database is PlantCARE (Plant *cis*-Acting Regulatory Elements) that compiles information about transcription factor sites, motif sequence, function, species, cell type, genes and represent *cis*-elements by positional matrices, consensus sequences and individual sites on particular sequences. The database also allows identification of new *cis*-elements from in silico data from transcriptome (Lescot et al. 2002). Third database TRANSFAC (Transcriptional Regulation from Patterns to Profiles) compiles eukaryotic transcription factors, their target genes and regulatory binding sites. It is the only database that provides structural and functional information of transcription factors (Hehl and Wingender 2001; Matys et al. 2003).

Fauteux and Strömvik (2009) generated a seeding DNA motif discovery algorithm and analyzed 54 seed storage protein (SSP) gene promoters from three plant families, namely *Brassicaceae* (mustards), *Fabaceae* (legumes) and

Poaceae (grasses) with respect to certain representative species in *Arabidopsis thaliana*, soybean and rice, respectively. They could identify three conserved motifs, two RY-like and one ACGT-like. In comparison to other approaches that use a position weight matrix motif model, the seeder DNA motif effectively utilizes a string-based approach that identifies motifs that are statistically significant compared to a background set of sequences.

Studies characterizing different natural promoters (Kim et al. 2010; Saha et al. 2010; Xu et al. 2010b) are continuously increasing and provide the researchers with more and more *cis*-motifs which control gene expression. Saha et al. (2010) lately identified and characterized the *cis*-regulatory motifs responsible for tissue-specific expression in the −673 and +90 bases upstream of the LOJ (Lateral organ Junction) gene recognized as LOJ promoter. They found enhancer-like element in the distal region which stimulates a minimal promoter. Such examples where novel promoter elements are identified promises potential applications in genetic engineering.

Promoter studies to *cis* engineering…a designer shift

Synthetic promoters were initially created by three strategies: (1) By combining defined *cis*-regulatory element with strong constitutive promoter (Ito et al. 1998; Rushton et al. 2002; Gurr and Rushton 2005) or by duplicating the upstream enhancer domains in conjunction with strong promoter (Maiti et al. 1997); (2) By combining *cis*-regulatory elements from different promoters (Sawant et al. 2001); (3) By fusing two strong constitutive characterised promoters to develop hybrids that allow both the promoters to be active in either direction or by developing bidirectional promoter (Comai et al. 1990; Chaturvedi et al. 2006).

Sawant et al. (1999) carried out a computational analysis of conserved nucleotide sequences observed in promoter of highly expressing genes in plants and designed a 'minimal expression casette', *Pmec* and 'Transcription activation module', TAM (Sawant et al. 2001). TAM contains many *cis*-acting motifs responsive to salicylic acid, auxins, salt, abscisic acid, ethylene, gibberellic acid, jasmonate, a variety of abiotic as well as biotic stresses, development and nutrition related factors. Various *cis*-acting DNA motif could function as an activator by itself as well as a synergizing activator in the presence of other neighbouring homologous as well as heterologous motifs. This synergestic effect is well illustrated in the work of Sawant et al. (2005) where the effect of eight *cis*-acting motifs on transcription from the basal promoter (*Pmec*) was studied by placing these upstream of the TATA-box at the −38 position as in plant genes. Multimers of the eight different sequence elements were inserted, taking one at a time such that each of these caused 2–8-fold activation of the basal transcription. The complete module brought enhancement of 110-fold in transcription levels. This showed that all such factors have to be arranged in an orderly manner to give a fully functional enhanceosome-like complex. Chaturvedi et al. (2006) constructed a synthetic bidirectional expression module by placing *Pmec* on 5′ and 3′ side of TAM. They analyzed the expression in both directions by transient and stable transformation in tobacco. As a result, it was found that TAM worked as an activating enhancer functionally compatible to minimal promoter. Schlabach et al. (2010) used synthetic design to develop strong promoter where 10 mer DNA sequences in tandem were placed on microarray as 100 mers which yieled them an oligo nucleotide library of 52,429 unique sequence and were screened for the ability to activate GFP transcription from a CMV promoter.

An important approach to tune gene expression was developed by Jensen and Hammer (1998), Mijakovic et al. (2005) and Hammer et al. (2006). They attempted to control gene expression through construction of synthetic promoter libraries by introducing changes in the sequences flanking the −35 and −10 consensus sequences of bacterial promoters. Alper et al. (2005) devised a methodology for quantitative evaluation of gene expression by constructing library of synthetic promoters of varying strength. These promoters were constructed via directed evolution through mutagenesis of a constitutive promoter. The further analysis was made with various metrics and a study through integration of these constructs into genome to find out the dependency of phenotype on gene expression, was conducted. In this case, a derivative of the constitutive bacteriophage PL-lambda promoter was mutated through error-prone PCR, cloned into a reporter plasmid upstream of a low-stability GFP gene, and screened in *E. coli.* Based on GFP fluorescence a functional library of 22 mutants was obtained. The instabilities and inherent mutation rates associated with the over expression of endogenous genes by earlier used plasmid based systems were avoided with the use of this system. This and other promoter libraries have a broad host range, perhaps due to involvement of general polymerase machinery in the cell and a heterologous constitutive promoter.

Designer promoter…the result of *cis* engineering

Recent advances in synthetic promoter technology are enabling the production of novel promoters optimised to suit the requirements of a particular transgene and exercise tighter control. Inducible promoters are activated by one or more stimuli such as hormones (for example gibberellin, abscisic acid, jasmonic acid, salicylic acid, ethylene,

auxin), environmental conditions (light, temperature), abiotic stress (water stress, salt stress, wounding) and biotic stress (microbes, insects, nematodes). Although inducible, these may sometimes direct expression in the absence of the stimulus.

Pathogen inducible promoters

Rushton et al. (2002) found that defence signaling could be well conserved across species boundaries at the promoter element level. An array of *cis*-acting elements (boxes W1, W2, GCC, JERE, S, Gst1, and D) recognised by specific transcription factors (WRKYs, ERFs, bZIPs, Mybs, Dofs and bHLHs) can mediate local gene expression in plants after pathogen attack. Hence, defined synthetic promoters were constructed. These contain tetramers of only a single type of element and monitored expression during interactions with a number of pathogens, including compatible, incompatible, and non host interactions. As it is known that spacing between individual *cis*-acting elements and/or between these elements is difficult to predict (Wray 1998), and needs to be determined experimentally, there were major differences in the inducibility of the various promoters for pathogens tested, the speed of induction and the basal expression levels. Although all of the promoter elements originated from other plants, they worked in *Arabidopsis thaliana* retaining their function in this heterologous background. Thereby, these elements can be used to make promoters for use in different plants and against different pathogens.

Heise et al. (2002) studied pathogen inducible Arabidopsis CMPG1 gene promoter and found that it's also responsive to wounding. They found another pathogen-responsive element, F involved in the process. Hence; they constructed a synthetic promoter containing only F and were able to separate pathogen inducibility from wound inducibility. Rushton et al. (2002) also developed improved second-generation promoters by varying several parameters like the number of copies of an individual element in a promoter and hence, varied the strength and inducibility of a promoter. Importantly, this also can have the effect of reducing/eliminating some background expression because pathogen inducibility appears stronger than basal or wound-induced expression. What remains ahead is development of 'lifestyle-specific' promoters that are selectively inducible by biotrophic, hemibiotrophic or necrotrophic pathogens.

Römer et al. (2009) studied transcription- activator like (TAL) effector proteins that manipulate the hosts' transcriptome to promote disease. Certain resistance (*R*) genes may be specifically activated by the respective TAL effectors by interaction with the corresponding *UPT* (*UP*regulated by *T*AL effectors). They showed that *UPT* boxes from different plant promoter can be assembled in vitro into one complex promoter in which each *UPT* box retains its TAL effector specificity and results in a single plant resistance gene promoter engineered to recognize multiple TAL effectors from disparate pathogens.

Recently, Kovalchuk et al. (2010) cloned the genomic sequences of seven γ-thionin (defensin) genes that are mainly expressed in developing grain from wheat and rice. While they studied the spatial and temporal activities of four defensin promoters in wheat and rice stable transformants, a strong basal activity was detected in the tissues which are most vulnerable to fungal and bacterial infections. All wheat and rice promoters were strongly induced by wounding in transgenic rice plants. Their work suggest that these PRPI (Pathogen Responsive and Pathogen Inducible) promoters will be useful for specific targeting and accumulation of pathogen resistance proteins in young, vulnerable tissues of developing and germinating grain.

'Chemical switches' that regulate transcription

A flexible system has evolved in the form of chemical-inducible systems for regulating gene expression. These chemical switches remains quiescent in the presence or absence of inducers, and therefore, does not inhibit physiological activities. Chemicals that can regulate transgene expression include the antibiotic tetracycline (tc), the steroids dexamethasone (dex), estradiol, copper, ethanol, the inducer of pathogen-related proteins benzothiadiazol, herbicide safeners, and the insecticide methoxyfenozide. When used in combination with tissue-specific promoters, the gene expression can be restricted to a given tissue at a specific time. To constitutively or conditionally express an inactive chimeric transcription activator, promoter activity system based on transcriptional activation, containing a heterologous DNA-binding domain (DBD), an activation domain (AD), a nuclear localization signal (NLS) or a regulatory domain of an animal steroid nuclear receptor, can be used. Steroids such as the glucocorticoid dexamethasone exhibit high specificity for the transcriptional activator, the glucocorticoid receptor (GR). Using domains from different proteins, Unger et al. (2002) established a steroid-inducible promoter for transgenic maize. In this system, the DNA-binding domain and the ligand-binding domain were taken from the estrogen receptor, whereas the activation domain was taken from the maize transcription factor C1. Male sterility could thus, be chemically controlled by conditionally expressing the male fertility gene *MS45* in a male-sterile background.

Another strategy to regulate gene expression through inducers includes promoter inactivation system that utilizes the repression principle based on sterical interference of a repressor protein with proteins important for transcription.

For example, in tetracycline inactivation system, the construction of fusion proteins between transcriptional transactivation domains and bacterial repressor proteins such as the Lac repressor or TetR (Tetracycline repressor) are undertaken. TetR binds to the tetracycline operator, *Tet,* in the absence of tetracycline. But when associated with tetracycline, dimeric TetR is converted to the monomeric form and released from its operator.

In contrast to above strategy, Gatz et al. (1992) developed a de-repression system in plants where a mutant TetR showing a "reverse phenotype" was employed. This reverse repressor binds DNA only in the presence of tetracycline. The target promoter, a modified 35S promoter, consisted of one or two copies of the *tet* operator placed upstream and downstream from the TATA-box, respectively. In the absence of tetracycline, over-expressed TetR binds to the *Tet* operator, and thereby prevents target gene expression but upon tetracycline binding; TetR is released from the operator, relieving the repression.

Ordiz et al. (2002) designed an ethanol inducible system based on the regulatory elements of the *Aspergillus nidulans alcA* promoter and adapted for use in plants. The target promoter contains the TATA box as well as upstream sequences of the *alcA* promoter fused to position 223 of the CaMV 35S promoter. A DNA binding protein of C6 zinc binuclear cluster family, AlcR binds to its target sequences within this promoter in the presence of ethanol or other inducers such as ethyl methyl ketone. When stably transformed into tobacco, these constructs mediate ethanol-dependent expression of transgenes. While using such systems, a risk of false expression exists because oxygen limitation in case of submerged environment of suspension culture may induce *alcA* in the absence of exogenously applied inducer. This has been overcome by development of an *alc*-GR system. Here, the rat glucocorticoid receptor (GR) domain has been fused to the AlcR transcription factor; conferring steroid-inducible control over *alc*-mediated gene expression (Roberts et al. 2005). Moore et al. (2006) discussed about various technologies for chemically inducible gene expression in plants.

Synthetic light switches

Gilmartin et al. (1990) reported about the molecular light switches for plant genes. However, Puente et al. (1996) studied the minimal promoter elements sufficient to mediate responses to light and developmental signals, and systematically analyzed the ability of four well characterized Light responsive Elements (LREs), individually or in selected combinations, to confer light responsiveness to non-light-regulated basal promoters. Their study revealed that individual motifs cannot confer any light responsiveness. However, if three distinct pair wise combinations of these motifs were made considering appropriate cell type, and developmental context, light responsiveness was observed. This was independent of the basal promoters used. Although the role of individual LREs can be quite different depending on the promoter context in which they are located but when in pairs these LREs tend to behave as units and signal integration points mediating both light and developmental control of gene expression.

Logemann and Hahlbrock (2002) while studying the crosstalk among stress responses in plants, found that ACE/ACE (ACGTcontaining element) functions as a Light-Regulatory as well as Elicitor-Responsive Unit. Their studies on *Petroselinum crispum*, have shown that one of the involved genes, encoding acyl-CoA oxidase (AOC), responds positively to UV light and negatively to a pathogen-derived elicitor through a promoter unit consisting of two almost identical ACEs. When introduced into an unrelated promoter or subjected to mutation, have shown the same type of response pattern. Such studies have generated considerable interest in elucidating the hierarchy of networks of transcription factors, and in identifying the key regulatory elements in different light-responsive developmental processes. A range of LREs have been studied in different promoters, many of which positively or negatively mediate gene expression in response to light. No single element is found in all light-regulated promoters, suggesting that a complex light-regulation network exists (Jiao et al. 2007).

Similarly, Evrard et al. (2009) proposed that FORCA, the regulatory *cis*- element found in a large number of Arabidopsis promoters, can serve as integration point of light and defense-related signals and thereby bring about changes at the transcriptome level in response to surrounding environmental conditions. FORCA is a frequently occuring hexameric promoter motif in case of many Arabidopsis genes that occur in clusters and are coexpressed in response to fungal or oomycete pathogens as well as defined light treatments. This element interacts with nuclear Arabidopsis proteins. It has been reported that such interactions are suppressed by defense-related stimuli and enhanced by prolonged exposure to constant light. Hence, FORCA like light-responsive promoter provides the building blocks of the strategy that can be used to develop synthetic promoters capable of coping up with two or more environmental conditions simultaneously.

Other types

Tissue-specific promoters confine transgene expression to a single plant part, tissue or cell-type. Ni et al. (1995) studied the strength and tissue specificity of chimeric promoters derived from the octopine and mannopine synthase (*mas*) genes. It was found that a chimeric promoter, *Mac*,

incorporating the *mas* region from +65 to −301 and 35S enhancer region from −90 to −941 expressed GUS activity at a level that was several times of double 35S promoter in transgenic tobacco plants. The changes in strength of expression were investigated for different combinations. For example if *ocs* activator was added to the *mas* promoter and activator or *mas* activator in combination with *ocs* promoter and activator. These combinations resulted in observance of GUS activity in a larger number of cell types, including xylem vessels and leaf non-specialized epidermal cells. The strongest promoter that they constructed consisted of a trimer of the *ocs* activator added to the mas activator and promoter ((*ocs) 3mas*).

An improved promoter for gene expression in cereal cells was developed by Last et al. (1991). It is based on truncated maize *Adh1* promoter, with multiple copies of the Anaerobic Responsive Element from the maize *Adh1* gene and *ocs*-elements from the octopine synthase gene of *Agrobacterium tumefaciens.* Promoter activity when measured in five different monocot species [wheat, maize, rice, einkorn *(Triticum monococcum),* and *Lolium multiflorum]* and one dicot *(Nicotiana plumbaginifolia)*, the most highly expressing construct (pEmuGN) gave 10- to 50-folds higher expression than the CaMV 35S promoter in all the monocot species. The pEmu promoter should be valuable where a high level of gene expression is required in monocots.

With the work done by Pietrzak et al. (1989), it was found that transcriptional activity of a strong constitutive plant promoter, CaMV 35S promoter, could be modulated by addition of multiple synthetic oligonucleotides, carrying the heat shock promoter consensus sequence (CTXGAAX XTTCXAG, Pelham 1982). Under heat shock conditions such a promoter shows a threefold increase in activity. Similarly when an enhancerless 35S promoter was engineered, it gave increased activity under both normal and heat conditions, without a significant induction by the heat shock.

Coutu et al. (2007) developed 14 binary vectors pORE, suitable for *Agrobacterium*-mediated transformation of dicotyledonous plants and adaptable for biolistic transformation of monocotyledonous plants by using a combination of promoters (PHPL *Arabidopsis thaliana* hydroperoxide lyase promoter, PENTCUP2 *Triticum aestivum* lipid transfer protein promoter and PTAPADH *Triticum aestivum* lipid transfer protein promoter fused to an alcohol dehydrogenase intron). In 2008 Chung and Parish reported combinatorial interactions of multiple *cis* elements regulates the induction of the *Arabidopsis* XERO2 dehydrin gene by ABA and cold, Liu et al. (1994) constructed tuber-specific and cold-inducible chimeric promoters (CIPP) in potato. Deletion analysis of the CIPP and *A. thaliana*, cold inducible (cor15a) promoter showed the presence of a *cis*-element for tuber-specific and sucrose-responsive activity named TSSR and LTRE (Low Temperature Responsive element), respectively. So, they constructed two chimeric promoters using TSSR containing sequence and LTRE and measured the activity of a GUS reporter gene with and without cold-inducible conditions in transgenic potatoes. Results indicated that this chimeric promoter possessed a substantial cold-inducibility. So far, there are few promoters that can be used to control the expression of transgenes in potato in a tuber-specific manner combined with cold inducibility and this chimeric promoter may provide valuable tool for minimizing the accumulation of reducing sugars in cold-stored tubers. Recently, Tittarelli et al. (2009) have digitally identified various cold regulated promoters from peach using EST dataset.

Xu et al. (2010a) isolated and studied sequence for upstream region of endosperm-specific LPAAT (Lysophosphatidyl acytransferase) gene from coconut (*Cocos nucifera* L.) and searched for the promoter related elements. These elements were functionally analyzed in transgenic rice plants and were found to respond to fatty acid metabolism in endosperm cells. The special function and subcellular location of this LPAAT promoter make it a unique promoter candidate in crop endosperm modification in future. Sun et al. (2010) identified a new 130 bp *cis*-acting element in the *TsVP1* promoter involved in the salt stress response from *Thellungiella halophila*. The promoter activity in case of leaves, roots, stems and flowers was as strong as CaMV 35S promoter but no promoter activity was detected in the seeds. This is beneficial for its application in crop engineering. Novel motifs responding to salt stress may exist in this 130 bp region.

Discussion

Critical aspects in *cis* engineering…how long can the magic of synthetic promoters last?

The synthetic promoter approach might be easy to implement because a collection of building block elements could be built up from a variety of plant species but it increases the possibility of Homology Based Gene Silencing (HBGS) due to sequence homology in transgene or promoter or *cis*-elements. HBGS is the inactivation of transgene either at transcription level or at post-transcription level caused due to presence of duplicated sequence in transgene and promoter, repeated sequence in transgene and native gene or presence of multiple copies of transgene in genome. It can lead to RNA-directed DNA methylation and inactivation of promoters (Wassenegger et al. 1994; Vaucheret et al. 2001) as well as co-suppression of another closely linked gene

(Brusslan et al. 1993). So the challenge exists to develop synthetic promoters having minimum sequence homology or in other words to develop minimal core promoter with specific *cis*-regulatory element for specific expression. Bhullar et al. (2003) developed novel synthetic 35S promoter by two approaches: firstly, by swapping of domain A of CaMV 35S promoter in three ways: one by introduction of direct repeats of *as*-1 element (TGACG) in the synthetic DNA context with different flanking and intervening nucleotides, second by keeping the flanking and spacer nucleotides conserved as 7- and 5-bp, respectively, and third by replacement of sub-domain A1 with *octopine synthase* element (*ocs*) from *ocs* promoter of *Arabidopsis* sp.; secondly, by replacing native *cis*-elements with *cis*-elements having same organization but no intervening homologous region to existing promoters. Results proved that novel promoters comprising new *cis*-elements expressed GUS gene levels comparable to wild-type 35S promoter but synthetic promoter with domain swapping revealed lower GUS gene activity.

More recently Bhullar et al. (2010) designed an approach to circumvent HBGS. They conjugated the silencing loci like 271 loci from tobacco to CaMV 35S promoter (carrying GUS reporter gene) that have reduced similarity to target promoter sequence. As a result, nearly 67% of transgenic lines escaped *trans*-inactivation of CaMV 35S promoter and there was delay in induction of silencing in synthetic promoter construct. Another approach to reduce HBGS is to develop gene cassettes having very less sequence homology to the endogenous gene or to one another if two or more transgenes are introduced. However, it is believed that artificial gene cassettes in the genome deplete basal transcription factors (TF) levels in genome. But in contrast, Sawant et al. (2005) revealed that many *cis*-elements together may provide additional TF-binding sites and contribute to the stability of Pre Initiation Complex (PIC) at TATA-box. Thus all the interactions mediated by upstream *cis*-regulatory modules allow higher expression of gene in plants. However, some other driving forces that can cause inactivation of transgene in plants, such as paramutation, position-effect variegation and chromatin mediated inactivation (Leeuwen et al. 2001; Meyer and Saedler 1996) may still change the picture.

Synthesizing bidirectional promoter: as new toolkits to manipulate plant genomes

Expression of more than one transgene in plants covers an important aspect of plant genetic engineering and genetic improvement of crop plants. However, regulation of multiple genes by single promoter can induce homology-based gene silencing of the transgene or native gene. Hence, gene stacking in plants can be achieved by the development of bidirectional promoter that allows simultaneous expression of two transgene. A bidirectional promoter can be constructed by the fusion of two unidirectional minimal promoters in opposite direction while they are regulated by common set of *cis*-motifs. Several scientists have developed bidirectional promoter using this strategy facilitating the expression of multiple transgene (Chaturvedi et al. 2006; Frey et al. 2001; Li et al. 2004). Bidirectional promoters not only solve the problem of gene silencing in plants but also minimize the amount of foreign DNA in transgenic plants.

Previously, artificial construction of bidirectional promoter using constitutive CaMV 35S promoter had lead to gene silencing of transgene in few cases (Xie et al. 2001). Now many reports have cited that bidirectional promoters exist naturally in plants. Hence, the naturally occurring bidirectional promoters in plants are now used to prevent homology based gene silencing and allow stacking of various genes. Most recently, Mitra et al. (2009) have proved that naturally occurring bidirectional promoter in *Arabidopsis* could be used to express two genes simultaneously. 2 Kb intergenic region of *Arabidopsis* chlorophyll *a*/*b*-binding protein *cab*1, *cab*2 genes was reported to be regulated by bidirectional promoter. This group checked the expression of GUS and GFP proteins fused to *Arabidopsis cab*1 and *cab*2 genes and found that both the genes were expressed in tissue specific and light induced manner, respectively.

Now that whole genome sequence of many plants have been deciphered, it is possible to use bioinformatic tools to predict putative bidirectional promoters in plants. Dhadi et al. (2009) have used PLACE and PlantCARE databases to identify *cis*-regulatory elements from three different sequenced plant genomes i.e. rice (*Oryza sativa*), *Arabidopsis thaliana*, and *Populus trichocarpa*. Based upon their knowledge on over-represented *cis*-elements in bidirectional promoters in respective genomes, they defined the functions of various *cis*-elements. Furthermore, based upon intergenic distances between promoters in bidirectional promoter and their co-expression patterns they predicted bidirectional promoters and also compared their expression with bidirectional promoters in humans.

Acknowledgments RM is thankful to Department of science and Technolgy,Government of India for financial support under fast track scheme for young scientists.We are also thankful to the facilities provided by BITS, Pilani.

References

Alper H, Fischer C, Nevoigt E, Stephanopoulos G (2005) Tuning genetic control through promoter engineering. Proc Natl Acad Sci USA 102:12678–12683

An G, Costa MA, Ha S-B (1990) Nopaline synthase promoter 1S wound inducible and auxin inducible. Plant Cell 2:25–33

Benfey PN, Chua N-H (1990) The cauliflower mosaic virus 35S promoter: combinatorial regulation of transcription in plants. Science 250:959–966

Bhullar S, Chakravarthy S, Advani S, Datta S, Pental D, Burma PK (2003) Strategies for development of functionally equivalent promoters with minimum sequence homology for transgene expression in plants: cis-elements in a novel DNA context versus domain swapping. Plant Physiol 132:988–998

Bhullar S, Datta S, Advani S, Chakravarthy S, Gautam T, Pental D, Burma PK (2007) Functional analysis of the cauliflower mosaic virus 35S promoter: re-evaluation of the role of subdomains B5, B4, and B2 in promoter activity. Plant Biotechnol J 5:696–708

Bhullar S, Datta S, Burma PK (2010) Delayed trans-inactivation of synthetic domain A 35S promoters by "Tobacco 271 Locus" due to reduced sequence homology. Plant Mol Biol Rep 29:1–11

Brusslan JA, Karlin-Neumann GA, Huang L, Tobin EM (1993) An *Arabidopsis* mutant with a reduced level of cab140 RNA is a result of cosuppression. Plant Cell 5:667–677

Cazzonelli CI, Velten EJ (2008) In vivo characterization of plant promoter element interaction using synthetic promoters. Trans Res 17:437–457

Chaturvedi CP, Sawant SV, Kiran K, Mehrotra R, Lodhi N, Ansari SA, Tuli R (2006) Analysis of polarity in the expression from a multifactorial bidirectional promoter designed for high-level expression of transgenes in plants. J Biotechnol 123:1–12

Chung S, Parish RW (2008) Combinatorial interactions of multiple *cis*-elements regulating the induction of the *Arabidopsis* XERO2 dehydrin gene by abscisic acid and cold. Plant J 54:15–29

Comai L, Moran P, Maslyar D (1990) Novel and useful properties of a chimeric plant promoter combining CaMV 35S and MAS elements. Plant Mol Biol 15:373–381

Coutu C, Brandle J, Brown D, Brown K, Miki B, Simmonds J, Hegedus DD (2007) pORE: a modular binary vector series suited for both monocot and dicot plant transformation. Trans Res 16:771–781

Dhadi SR, Krom N, Ramakrishna W (2009) Genome-wide comparative analysis of putative bidirectional promoters from rice, *Arabidopsis* and *Populus*. Gene 429:65–73

Ellis JG, Llewellyn DJ, Walker JC, Dennis ES, Peacock WJ (1987) The *ocs* element: a 16 base pair palindrome essential for activity of the octopine synthase enhancer. EMBO J 6:3203–3208

Ettwiller LM, Rung J, Birney E (2003) Discovering novel *cis*-regulatory motifs using functional networks. Genome Res 13:883–895

Evrard A, Ndatimana T, Eulgem T (2009) FORCA, a promoter element that responds to crosstalk between defense and light signaling. BMC Plant Biol 9:1–13

Fang RX, Nagy F, Subramaniam S, Chua NH (1989) Multiple *cis* regulatory elements for maximal expression of the cauliflower mosaic virus 35S promoter in transgenic plants. Plant Cell 1:141–150

Fauteux F, Strömvik MV (2009) Seed storage protein gene promoters contain conserved DNA motifs in *Brassicaceae*, *Fabaceae* and *Poaceae*. BMC Plant Biol 9:126–136

Frey PM, Scharer-Hernandez NG, Futterer J, Potrykus I, Puonti-Kaerlas J (2001) Simultaneous analysis of the bidirectional African cassava mosaic virus promoter activity using two different luciferase genes. Virus Genes 22:231–242

Gatz C, Frohberg C, Wendenberg R (1992) Stringent repression and homogeneous depression by tetracyclin of a modified CaMV 35S promoter in intact transgenic tobacco plants. Plant J 2:397–404

Gilmartin PM, Sarokin L, Memelink J, Chua NH (1990) Molecular light switches for plant genes. Plant Cell 2:369–378

Gowda S, Wu FC, Herman HB, Shepherd RJ (1989) Gene VI of figwort mosaic virus (caulimovirus group) functions in posttranscriptional expression of genes on the full-length RNA transcript. Proc Natl Acad Sci USA 86:9203–9207

Gurr SJ, Rushton PJ (2005) Engineering plants with increased disease resistance: how are we going to express it? Trends Biotechnol 23:283–290

Hammer K, Mijakovic I, Jensen PR (2006) Synthetic promoter libraries—tuning of gene expression. Trends Biotechnol 24: 53–55

Hehl R, Wingender E (2001) Database-assisted promoter analysis. Trends Plant Sci 6:251–255

Heise A, Lippok B, Kirsch C, Hahlbrock K (2002) Two immediate-early pathogen-responsive members of the *AtCMPG* gene family in *Arabidopsis thaliana* and the W-box-containing elicitor-response element of *AtCMPG1*. Proc Natl Acad Sci USA 99:9049–9054

Helden JA (2003) Regulatory sequence analysis tools. Nucl Aci Res 31:3593–3596

Higo K, Ugawa W, Iwamoto M, Korenaga T (1999) Plant *cis*-acting regulatory DNA elements (PLACE) database: 1999. Nucl Aci Res 27:297–300

Istrail S, Davidson EH (2004) Logic functions of the genomic *cis*-regulatory code. Proc Natl Acad Sci USA 102:4954–4959

Ito M, Iwase M, Kodama H, Lavisse P, Komamine A, Nishihama R, Machida Y, Watanabea A (1998) A novel *cis*-acting element in promoters of plant B-type cyclin genes activates M phase–specific transcription. Plant Cell 10:331–341

Jensen PR, Hammer K (1998) The sequence of spacers between the consensus sequences modulates the strength of prokaryotic promoters. Appl Environ Microbiol 64:82–87

Jiao Y, Lau ON, Deng XW (2007) Light-regulated transcriptional networks in higher plants. Nat Rev Genet 8:217–230

Kamisugi Y, Cuming AC (2005) The evolution of the abscisic acid-response in land plants: comparative analysis of group 1 LEA gene expression in moss and cereals. Plant Mol Biol 59:723–737

Kim Y-H, Bae JM, Huh G-H (2010) Transcriptional regulation of the cinnamyl alcohol dehydrogenase gene from sweet potato in response to plant developmental stage and environmental stress. Plant Cell Rep 29:779–791

Kovalchuk N, Li M, Wittek F, Reid N, Singh R, Shirley N, Ismagul A, Eliby S, Johnson A, Milligan AS, Hrmova M, Langridge P, Lopato S (2010) Defensin promoters as potential tools for engineering disease resistance in cereal grains. Plant Biotechnol J 8:47–64

Lam E, Lam YKP (1995) Binding site requirements and differential representation of TGA factors in nuclear ASF-1 activity. Nucl Acids Res 23:3778–3785

Last DI, Bretteli RIS, Chamberlain DA, Chaudhary AM, Larkin PJ, Marsh EL, Peacock WJ, Dennis ES (1991) pEmu: an improved promoter for gene expression in cereal cells. Theo App Genet 81:581–588

Leeuwen WV, RuttInk T, Borst-Vrenssen AWM, Plas LHWVD, Krol ARVD (2001) Characterization of position-induced spatial and temporal regulation of transgene promoter activity in plants. J Exp Bot 52:949–959

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

Li ZT, Jayasankar S, Gray DJ (2004) Bi-directional duplex promoters with duplicated enhancers significantly increase transgene expression in grape and tobacco. Transgenic Res 13:143–154

Liu Z-B, Ulmasov T, Shi X, Hagen G, Guilfoyle TJ (1994) Soybean GH3 promoter contains multiple auxin-inducible elements. Plant Cell 6:645–657

Logemann E, Hahlbrock K (2002) Crosstalk among stress responses in plants: pathogen defense overrides UV protection through an

inversely regulated ACE_ACE type of light-responsive gene promoter unit. Proc Natl Acad Sci USA 99:2428–2432
Maiti IB, Gowda S, Kiernan J, Ghosh SK (1997) Promoter/leader deletion analysis and plant expression vectors with the figwort mosaic virus (FMV) full length transcript (FLt) promoter containing single or double enhancer domains. Transgenic Res 6:143–156
Matys V, Fricke E, Geffers R, ßling EG, Haubrock M, Hehl R, Hornischer K, Karas D, Kel AE, Kel-Margoulis OV, Kloos D-U, Land S, Lewicki-Potapov B, Michael H, Munch R, Reuter I, Rotert S, Saxel H, Scheer M, Thiele S, Wingender E (2003) TRANSFAC: transcriptional regulation, from patterns to profiles. Nucl Acids Res 31:374–378
Mehrotra R, Mehrotra S (2010) Promoter activation by ACGT in response to salicylic and abscisic acids is differentially regulated by the spacing between two copies of the motif. J Plant Physiol 167:1214–1218
Mehrotra R, Panwar J (2009) Dimerization of GT element interferes negatively with gene activation. J Genet 88:257–260
Mehrotra R, Kiran K, Chaturvedi CP, Anjuman S, Lodhi N, Sawant S, Tuli R (2005) Effect of copy number and spacing of the ACGT and GT *cis* elements on transient expression of minimal promoter in plants. J Genet 84:183–187
Meyer P, Saedler H (1996) Homology-dependent gene silencing in plants. Ann Rev Plant Physiol Plant Mol Biol 47:23–48
Mijakovic I, Petranovic D, Jensen PR (2005) Tunable promoters in systems biology. Curr Opin Biotechnol 16:329–335
Mitra A, Han J, Zhang ZJ, Mitra A (2009) The intergenic region of *Arabidopsis thaliana cab*1 and *cab*2 divergent genes functions as a bidirectional promoter. Planta 229:1015–1022
Moore I, Samalova M, Kurup S (2006) Transactivated and chemically inducible gene expression in plants. The Plant J 45:651–683
Ni M, Cui D, Einstein J, Narasimhulu S, Vergara CE, Gelvin SB (1995) Strength and tissue specificity of chimeric promoter derived from octopine and manopine synthase gene. Plant J 7:661–676
Novina CD, Roy AL (1996) Core promoters and transcriptional control. Trends Genet 12:351–355
Odell JT, Nagy F, Chua N-H (1985) Identification of DNA sequences required for activity of the cauliflower mosaic virus 35S promoter. Nature 313:810–812
Ordiz MI, Barbas CF-III, Beachy RN (2002) Regulation of transgene expression in plants with polydactyl zinc finger transcription factors. Proc Natl Acad Sci USA 99:13290–13295
Pelham HRB (1982) A regulatory upstream promoter element in the Drosophila Hsp 70 heat-shock gene. Cell 30:517–528
Pietrzak M, Burri M, Herrero J-J, Mosbach K (1989) Transcriptional activity is inducible in the cauliflower mosaic virus 35S promoter engineered with the heat shock consensus sequence. FEBS Lett 249:311–315
Puente P, Wei N, Deng XW (1996) Combinatorial interplay of promoter elements constitutes the minimal determinants for light and developmental control of gene expression in Arabidopsis. EMBO J 15:3732–3743
Roberts GR, Garoosi GA, Koroleva O, Ito M, Laufs P, Leader DJ, Caddick MX, Doonan JH, Tomsett AB (2005) The alc-GR system. A modified alc gene switch designed for use in plant tissue culture. Plant Physiol 138:1259–1267
Römer P, Recht S, Lahaye T (2009) A single plant resistance gene promoter engineered to recognize multiple TAL effectors from disparate pathogens. Proc Natl Acad Sci USA 106:20526–20531
Rushton PJ, Reinstädler A, Lipka V, Lippok B, Somssich IE (2002) Synthetic plant promoters containing defined regulatory elements provide novel insights into pathogen- and wound-induced signaling. Plant Cell 14:749–762
Saha D, Kumar V, Bhat SR, Srinivasan R (2010) Characterization of upstream sequences of the LOJ gene leads to identification of a novel enhancer element conferring lateral organ junction-specific expression in *Arabidopsis thaliana*. Plant Mol Biol Rep 28. doi:10.1007/s11105-010-0229-6
Sawant SV, Singh PK, Gupta SK, Madanala R, Tuli R (1999) Conserved nucleotide sequence in highly expressed genes in plants. J Genet 78:1–8
Sawant SV, Singh PK, Madanala R, Tuli R (2001) Designing of an artificial expression cassette for the high-level expression of transgenes in plants. Theor Appl Genet 102:635–644
Sawant S, Kiran K, Mehrotra R, Chaturvedi CP, Ansari SA, Singh P, Lodhi N, Tuli R (2005) A variety of synergistic and antagonistic interactions mediated by *cis*-acting DNA motifs regulate gene expression in plant cells and modulate stability of the transcription complex formed on a basal promoter. J Exp Bot 56:2345–2353
Schlabach MR, Hu JK, Li M, Elledge SJ (2010) Synthetic design of strong promoters. Proc Natl Acad Sci USA 107:2538–2543
Sun Q, Gao F, Zhao F, Li K, Zhang J (2010) Identification of a new 130 bp *cis*-acting element in the *TsVP1* promoter involved in the salt stress response from *Thellungiella halophila*. BMC Plant Biol 10:90–102
Tittarelli A, Santiago M, Morales A, Meisel LA, Silva H (2009) Isolation and functional characterization of cold-regulated promoters, by digitally identifying peach fruit cold-induced genes from a large EST dataset. BMC Plant Biol 9:121–136
Tiwari SB, Hagen G, Guilfoyle T (2003) The roles of auxin response factor domains in auxin-responsive transcription. Plant Cell 15:533–543
Unger E, Cigan AM, Trimnell M, Xu RJ, Kendall T, Roth B, Albertsen M (2002) A chimeric ecdysone receptor facilitates methoxyfenozide-dependent restoration of male fertility in ms45 maize. Transgenic Res 11:455–465
Vaucheret H, Béclin C, Fagard M (2001) Post-transcriptional gene silencing in plants. J Cell Sci 114:3083–3091
Venter M (2007) Synthetic promoters: genetic control through *cis* engineering. Trends Plant Sci 12:118–124
Venter M, Botha FC (2010) Synthetic promoter engineering. Plant Develop Biol Biotechnol Persp 2(4):393–414
Wassenegger M, Heimes S, Riedel L, Sänger HL (1994) RNA-directed de novo methylation of genomic sequences in plants. Cell 76:567–576
Wray GA (1998) Promoter logic. Science 279:1871–1872
Xie M, He Y, Gan S (2001) Bidirectionalization of polar promoters in plants. Nat Biotechnol 19:677–679
Xu L, Ye R, Zheng Y, Wang Z, Zhou P, Lin Y, Li D (2010a) Isolation of the endosperm-specific LPAAT gene promoter from coconut (*Cocos nucifera* L.) and its functional analysis in transgenic rice plants. Plant Cell Rep 29:1061–1068
Xu W, Yu Y, Ding J, Hua Z, Wang Y (2010b) Characterization of a novel stilbene synthase promoter involved in pathogen- and stress-inducible expression from Chinese wild *Vitis pseudoreticulata*. Planta 231:475–487
Zavallo D, Bilbao ML, Hopp HE, Heinz R (2010) Isolation and functional characterization of two novel seed-specific promoters from sunflower (*Helianthus annuus L.*). Plant Cell Rep 29: 239–248