

2012

Functional analysis of rice bidirectional promoters

Surendar R. Dhadi
Michigan Technological University

Follow this and additional works at: <https://digitalcommons.mtu.edu/etds>



Part of the [Biology Commons](#)

Copyright 2012 Surendar R. Dhadi

Recommended Citation

Dhadi, Surendar R., "Functional analysis of rice bidirectional promoters", Dissertation, Michigan Technological University, 2012.

<https://doi.org/10.37099/mtu.dc.etds/191>

Follow this and additional works at: <https://digitalcommons.mtu.edu/etds>



Part of the [Biology Commons](#)

FUNCTIONAL ANALYSIS OF RICE BIDIRECTIONAL PROMOTERS

By

Surendar R Dhadi

A DISSERTATION

Submitted in partial fulfillment of the requirements for the degree of

DOCTOR OF PHILOSOPHY

(Biological Sciences)

MICHIGAN TECHNOLOGICAL UNIVERSITY

2012

Copyright © Surendar R Dhadi 2012

This dissertation, “Functional Analysis of Rice Bidirectional Promoters”, is hereby approved in partial fulfillment of the requirements for the Degree of DOCTOR OF PHILOSOPHY IN BIOLOGICAL SCIENCES.

Department of Biological Sciences

Signatures:

Dissertation Advisor

Ramakrishna Wusirika

Department Chair

K. Michael Gibson

Date

TABLE OF CONTENTS

List of Figures.....	6
List of Tables.....	8
Preface.....	9
Acknowledgements.....	10
Abstract.....	11
Literature review	12
CHAPTER 1:.....	17
GENOME-WIDE COMPARATIVE ANALYSIS OF PUTATIVE BIDIRECTIONAL PROMOTERS FROM RICE, <i>ARABIDOPSIS</i> AND <i>POPULUS</i>	17
1.1 Abstract	18
1.2 Introduction	19
1.3 Methods	20
1.3.1 Promoter identification and selection.....	20
1.3.2 Promoter sequence analysis	21
1.4 Results and Discussion	22
1.4.1 Computational identification of putative bidirectional promoters from rice, <i>Arabidopsis</i> , and <i>Populus</i> and their high GC content.....	22
1.4.2 Overrepresented and abundant motifs in putative bidirectional promoters in rice, <i>Arabidopsis</i> and <i>Populus</i> genomes represented in the PLACE database	23
1.4.3 Overrepresented motifs in putative bidirectional promoters in three plant genomes represented in PLANTCARE database.....	28
1.4.4 Representative bidirectional promoters and divergent genes	29
1.4.5 Bidirectional promoters: a new toolkit to manipulate plant genomes.....	33
CHAPTER 2:.....	35

A NOVEL NON-WOUNDING TRANSIENT EXPRESSION ASSAY FOR CEREALS MEDIATED BY <i>AGROBACTERIUM TUMEFACIENS</i>	35
2.1 Abstract.....	36
2.2 Introduction	37
2.3 Materials and Methods	37
2.3.1 Constructs Used in the Present Study.....	37
2.3.2 Growth of Plants under Sterile Conditions	38
2.3.3 Growth of Young Cereal Plants and <i>Arabidopsis</i> in the Greenhouse	38
2.3.4 <i>Agrobacterium</i> Co-cultivation and Stress Treatments	38
2.3.5 Effect of <i>Agrobacterium</i> Co-cultivation and Plant Incubation Times in MS Salt Solution on Reporter Gene Expression.....	39
2.3.6 Downstream Process and Applications	39
2.3.7 Quantitative Expression Analysis of <i>gus</i> and <i>egfp</i> Reporter Genes	40
2.4 Results.....	40
2.4.1 Utility of AmTEA as a Universal Method for Cereals.....	40
2.4.2 Conserved Expression Patterns of Reporter Genes in Stable and Transiently Expressing <i>Arabidopsis</i> Plants	43
2.4.3 <i>Agrobacterium</i> Co-cultivation and Plant Incubation Times in MS Salt Solution Affect Reporter Gene Activity in Plants.....	43
2.4.4 AmTEA - A Useful and Reliable Tool for Deciphering Promoter and Gene Function in Plants.....	44
2.5 Discussion.....	47
2.8 Supplementary information	49
CHAPTER 3:	52
REGULATION OF BIDIRECTIONAL PROMOTER ACTIVITY IN RICE BY CIS- ELEMENTS IN 5' -UNTRANSLATED REGIONS.....	52
3.1 Abstract.....	53

3.2 Introduction	54
3.3 Methods	55
3.3.1 Construction of a vector with two reporter genes for testing BDPs	55
3.3.2 Gateway cloning of BDP-reporter gene constructs	55
3.3.3 Plant transient expression assay	56
3.3.4 Deletion analysis	56
3.3.5 Extraction of rice nuclear proteins	56
3.3.6 DNA footprinting assay	56
3.3.7 Motif analysis	57
3.3.8 Site-directed mutagenesis of the BDP constructs	57
3.3.9 Qualitative and quantitative expression analysis of reporter genes	58
3.4 Results	58
3.4.1 Structure of the divergent genes harboring the divergent promoters (BDP) chosen for functional evaluation	58
3.4.2 Deletion constructs identify bidirectional promoters	59
3.4.3 <i>Cis</i> -regulatory motifs: identification, deletion and expression analysis	62
3.5 Discussion	68
3.8 Supplementary information	72
CONCLUSION	76
FUTURE DIRECTIONS	77
REFERENCES	78
APPENDIX	90

LIST OF FIGURES

Fig. 1.1 A divergent gene pair with a bidirectional promoter.....	20
Fig. 1.2 Putative bidirectional promoters with cis-elements and divergent genes from rice.....	30
Fig. 1.3 Putative bidirectional promoters with cis-elements and divergent genes from <i>Arabidopsis</i> and <i>Populus</i>	31
Fig. 1.4 Putative bidirectional promoters with cis-elements and divergent genes with conserved gene order and orientation in other plant species.....	32
Fig. 2.1 Organization of the six promoter–reporter gene constructs used for AmTEA.....	41
Fig. 2.2 GUS expression driven by CaMV 35S promoter in pBI121 vector in young plants of rice, barley, maize, oats, rye, sorghum, and wheat using AmTEA.....	41
Fig. 2.3 Microscopic analysis of GUS expression driven by 35S promoter in (a) rice leaves ($\times 10$ magnification).....	42
Fig. 2.4 Reporter gene expression driven by the rice glutaredoxin promoter.....	42
Fig. 2.5 Conserved expression pattern of reporter genes driven by <i>Arabidopsis</i> promoter for the gene encoding DEAD-box RNA helicase family protein in transiently expressing plants and stable <i>Arabidopsis</i> lines subjected to drought stress.....	43
Fig. 2.6 <i>Agrobacterium</i> co-cultivation time influences GUS expression in rice plants.....	44
Fig. 2.7 Effect of the antibiotic carbenicillin on GUS expression in young rice plants.....	46
Fig. 2.8 Expression of an intron (catalase) containing gus gene driven by CaMV 35S promoter in pCAMBIA 2201 using AmTEA.....	46
Fig. 2.9 Full-length rice promoter R2-273 show high levels of GUS and eGFP expression compared to the 5’-truncated promoter.....	46
Fig. 3.1 eGFP-GUS and RFP expression pattern driven by the BDP 2-273 and its truncated versions in rice.....	60
Fig. 3.2 eGFP-GUS and RFP expression pattern driven by the	

BDP 4-431 and its truncated versions in rice.....	61
Fig. 3.3 eGFP-GUS and RFP expression pattern driven by the BDP 6-5 and its truncated versions in rice.....	62
Fig. 3.4 Cis-regulatory motifs in BDP 2-273 bound to nuclear proteins identified by DNA foot-printing.....	64
Fig. 3.5 eGFP-GUS and RFP expression pattern driven by the BDP 2-273 and its deletion construct Y in rice.....	65
Fig. 3.6 Cis-regulatory motifs in BDP 4-431 bound to nuclear proteins identified by DNA foot-printing.....	65
Fig. 3.7 eGFP-GUS and RFP expression pattern driven by the full-length and deletion versions of 4-431 BDP in rice.....	66
Fig. 3.8 Cis-regulatory motifs in BDP 6-5 bound to nuclear proteins identified by DNA foot-printing.....	66
Fig. 3.9 eGFP-GUS and RFP expression pattern driven by the full-length and deletion versions of 6-5 BDP in rice.....	68

LIST OF TABLES

Table 1.1 GC Content of bidirectional promoters.....	23
Table 1.2 Overrepresented motifs in rice bidirectional promoters from PLACE database.....	26
Table 1.3 Overrepresented motifs in <i>Arabidopsis</i> bidirectional promoters from PLACE database.....	27
Table 1.4 Overrepresented motifs in <i>Populus</i> bidirectional promoters from PLACE database.....	27
Table 1.5 Overrepresented motifs in bidirectional promoters from PLANTCARE database.....	28
Table 2.1 Taqman primers and probes used for qRT-PCR.....	40

PREFACE

The material used in chapter 1 of this dissertation was reproduced in its entirety with kind permission from Elsevier B.V. This article was previously published in the journal “GENE”, as a collaborative work with Nicholas Krom and Ramakrishna Wusirika. Entire data was produced in Dr. Ramakrishna Wusirika’s laboratory at MTU. My contribution included data retrieval and bioinformatic analysis. The initial draft was written by me with minor inclusions from Nicholas Krom and revised by Dr. Ramakrishna Wusirika to develop the final version.

The material used in chapter 2 of this dissertation was reproduced in its entirety with kind permission from Springer Science and Business Media. This article was previously published in the journal “Plant Molecular Biology Reporter”, as a collaborative work with Aparna Deshpande and Ramakrishna Wusirika. Entire method was developed in Dr. Ramakrishna Wusirika’s laboratory at MTU. My contribution included AmTEA method design and development, generation of promoter-reporter constructs, plant growing, real time PCR and GUS assays. pBGWSF7-RFP vector and 2-273 promoter-reporter construct used for this work was developed by Aparna Deshpande. The initial draft of the article was written by me and was revised by Dr. Ramakrishna Wusirika to develop the final version.

ACKNOWLEDGEMENTS

I would like to thank my advisor, Ramakrishna Wusirika through the depths of my heart for his patience and understanding in the difficult times of my graduate research. I greatly acknowledge his help and guidance throughout my graduate career without which this dissertation would have been an impossible task. Apart from this I would like to thank all the committee members, John Adler, Rupali Datta, Wan Jin Jahng and Hairong Wei for their insights and feedback on my work.

Many thanks to my various teaching supervisors: Ramakrishna Wusirika, John Adler, Mike Gretz and Ron Gratz. I am very much grateful for the tremendous learning experiences that I gained and cherished while teaching for them. I would also like to thank Guiliang Tang for helping me financially towards the end of my career.

I would also like to extend my gratitude towards the other helpful staff members of the department of biological sciences: Jeff Lewin, Patricia Asselin and Emily Betterly for their impeccable timing in sorting out all the difficult things for me and making my life easier at Tech.

I specially thank all my friends for lending me their ears as I express my concerns and worries as and when I required them. No names necessary but my heartfelt thanks for each one of you for being with me in the hardest times of my life.

I am deeply indebted to my parents for what I am today. And also to my siblings, my two brothers and sister for their small and large fights with me all the time, expressing how much they love and care for me.

ABSTRACT

Bidirectional promoters regulate adjacent genes organized in a divergent fashion (head to head orientation). Several Reports pertaining to bidirectional promoters on a genomic scale exists in mammals. This work provides the essential background on theoretical and experimental work to carry out a genomic scale analysis of bidirectional promoters in plants.

A computational study was performed to identify putative bidirectional promoters and the over-represented *cis*-regulatory motifs from three sequenced plant genomes: rice (*Oryza sativa*), *Arabidopsis thaliana*, and *Populus trichocarpa* using the Plant *Cis*-acting Regulatory DNA Elements (PLACE) and PLANT CARE databases. Over-represented motifs along with their possible function were described with the help of a few conserved representative putative bidirectional promoters from the three model plants. By doing so a foundation was laid for the experimental evaluation of bidirectional promoters in plants.

A novel *Agrobacterium tumefaciens* mediated transient expression assay (AmTEA) was developed for young plants of different cereal species and the model dicot *Arabidopsis thaliana*. AmTEA was evaluated using five promoters (six constructs) and two reporter genes, *gus* and *egfp*. Efficacy and stability of AmTEA was compared with stable transgenics using the *Arabidopsis* DEAD-box RNA helicase family gene promoter. AmTEA was primarily developed to overcome the many problems associated with the development of transgenics and expression studies in plants.

Finally a possible mechanism for the bidirectional activity of bidirectional promoters was highlighted. Deletion analysis using promoter-reporter gene constructs identified three rice promoters to be bidirectional. Regulatory elements located in the 5'-untranslated regions (UTR) of one of the genes of the divergent gene pair were found to be responsible for their bidirectional activity.

Literature review

Sequencing of plant, animal and human genomes has revealed a complex arrangement of genes, regulatory regions and various classes of repetitive DNA. These data provide valuable resources to identify DNA sequences required for the regulation of gene expression. Comparisons involving the human genome sequence with other related genomes resulted in the identification of novel promoters and regulatory elements. The most common architectural feature identified from the above information is that of a RNA polymerase II core promoter and its *cis* regulatory elements. Apart from this recently a unique class of promoters which act in a bidirectional manner has been characterized to some extent in the human genome. Understanding the sequence patterns and mechanisms used by bidirectional promoters in model plant genomes will not only provide insights into mechanisms unique or/ and common to humans but also enhance our ability to use them for better manipulation of genes. Before going to bidirectional promoters, it is necessary to understand as to how unidirectional promoters regulate the expression of genes in prokaryotes and eukaryotes. This review of the RNAP core promoter and its functional mechanism will help further in understanding the bidirectional promoter and its physical and functional architecture.

A promoter is a unique entity in the genome, one of the foremost pivotal regions defined in a transcriptional unit, where in it provides the space for the binding of the transcriptional apparatus. Capable of driving gene expression, its organization and regulation is complex and not yet fully understood. A promoter is the most dynamic unit of a genome. By regulating a promoter a cell can regulate and control all the inherent and coherent expressions symmetrically and simultaneously on a global scale. It can even extend this dynamism of a promoter to both temporal and spatial allocations. The most commonly studied promoter is that of RNA polymerase II core promoter transcribing 80% of the genes (Ptashne and Gann 1997).

Basic and comparative structure of RNA polymerase II core promoter

A transcriptional unit can be defined as a specific region in the genome where in synthesis of RNA takes place from the DNA. A typical transcription unit consists of three prominent regions, a promoter region, an RNA coding region and a terminator region (Ptashne and Gann 1997). Bacterial promoters typically harbor two important regions termed as the -10 and -35 elements upstream the RNA coding region. These two elements form the minimal RNA polymerase (RNAP) core promoter, also called as the minimal promoter or core promoter. These two regions are quite essential for recognition, binding, assembly and stabilization of the RNAP. The -10 element is also known as the Pribnow box. The -10 and the -35 regions are highly conserved in bacterial promoters. The

consensus sequence of -10 element is “TATAAT”, whereas the consensus sequence of -35 is “TTGACA” and these two elements are separated by a 17 base spacer (Lee and Young 2000). The tight conservation of the sequences in these two motifs dictates the strength of the promoter. Weak conservation of the motif sequences is the hall mark of weaker promoters. The -10 and the -35 elements each has a unique functional role to play in the transcription process. The -35 element stabilizes the RNAP complex by providing the necessary binding energy, so that the RNAP can remain bound to the promoter. Initiation of DNA melting and the transition from the closed to open complex and stabilization of the single strand DNA happens within the -10 element in coordination with the RNAP complex (Von-hippel 1998).

Apart from these two elements which are highly conserved in at least 60-80% of the bacterial promoters, other elements like the ‘UP’ elements were also found upstream the -35 element around positions -40 to -60. The promoters that harbor the UP element have an enhanced rate of transcription (Cramer 2004). Another element known as the discriminator positioned immediately downstream of the -10 element enhances and strengthens the stabilization of the RNAP complex to the core promoter. The UP and the discriminator elements are examples of few such elements among several other elements present in a promoter that can regulate the function of the promoter.

A eukaryotic minimal/core promoter spans approximately 40-60 bases either upstream or downstream from the Transcription Start Site (TSS). The core promoter essentially harbors two of the most important elements needed for the recognition, binding, and assembling and transcription initiation by the RNAP complex. The first and the foremost element is the TATA box and its position is approximately -30bp upstream of the Transcription Start Site (TSS). TATA box alone carries out both functional roles of the -10 and -35 elements in the eukaryotic promoter. TATA box has a consensus sequence of (TCACTATATATAG (Kiran et al. 2006; Joshi 1987) / TATATAT (T/A)A(T/A) (Zhu et al. 1995)). The TATA binding protein (TBP), a subunit of the TFIID recognizes and binds to the TATA box. TBP binding to the TATA box results in a total distortion of the TATA sequence, flattening it and bending the DNA by 80°, creates space to recruit several other TATA binding protein transcription factors known as the TBP associated factors (TAF) along with the RNAP and thus forms a pre-initiation complex. The “AT” base pairs of the TATA sequence are the most favored for this process as they involve two hydrogen bonds, less energy and allows for easy distortion and opening of the DNA minor groove by the TBP and its associated factors (Nikolov et al. 1995). Next in line downstream to the TATA box there is another element known as the Initiator element (Inr). The Inr element serves as the starting point for the initiation of transcription process. The TATA box and the initiator element are separated by a 30-40bp spacer. Apart from these two elements other regulatory elements like TFIIB

recognition element (BRE), downstream core elements (DCE) I, II and III and downstream promoter elements provide binding sites for the TAFs (Butler and Kadonaga 2002; Juven-Gershon et al. 2006). *In-vitro* the core promoter is enough to carry out the transcription process but *in-vivo*, due to the higher order complex structure of the chromatin additional transcriptional regulatory sequences are required. These regulatory sequences can be categorized into various orders like enhancers, insulators, silencers, chromatin remodelers, upstream activator elements, promoter proximal elements etc. All these regulatory sequences are bound and regulated by two general classes of regulatory proteins termed as the repressors and activators (Boeger et al. 2005). Mediator proteins/complex belongs to the class of regulatory proteins that are essential for achieving significant levels of transcription *in-vivo*. The mediator complex does not directly interact with the regulatory sequences on the DNA, but it coordinates the functional activity between RNAP and other regulatory proteins, especially the activator proteins. The proteins of the mediator complex are highly conserved in eukaryotic organisms like yeast and humans. It is found that several subsets of mediator complexes can work with different transcriptional machineries regulating the expression of the genes in temporal and spatial manner (Conaway et al. 2005; Malik et al. 2005).

Role of enhancer-activator complex in regulating gene expression

‘Action at a distance’ is the general phenomenon attributed to the functionality of enhancer-activator complex in regulating gene expression. An enhancer is a *cis* regulatory DNA element positioned either upstream or downstream of a gene either in close proximity or placed a few kilo bases or even tens and hundreds of kilo bases away from the respective promoter. Gene regulation by action at a distance phenomenon is made possible by DNA looping. DNA looping/bending happens when activator proteins bound to an enhancer element on the DNA interact with the mediator-RNAP complex bound to a promoter or recruit moderator-RNAP complex to a promoter, thereby enhancing or initiating transcription process (Vilar and Saiz 2005). DNA looping can be short or long range interaction depending upon the position of the regulatory elements with respect to the promoter. Short range DNA looping can be observed in the arabinose operon in bacteria. The AraC activator when not bound by arabinose obtains a different conformation and becomes a repressor of the arabinose promoter. In its repressor conformation, AraC protein binds to the operator *araO2* which is 194bp away from the initiator element *araI1* and makes a loop. In the looped configuration AraC does not bind to the *araI2* and thereby represses the arabinose operon (Ogden et al. 1980). Medium range DNA loop interactions can be found in the negative auto-regulation of lambda repressor during lysogenic establishment in the life cycle of the bacteriophage lambda. Lambda repressor binds to the operators left and right of the *CI* gene. The distance between the operators left and right is approximately 3.5kb. For proper repression of the

cro gene, it is essential that operators left and right interact and bind with each other. In order to accommodate for this medium range DNA interactions between operators right and left, DNA loops with the help of the lambda repressor. This establishes a proper shut down of the *cro* gene (Hochschild 2002). Long range DNA interactions can be observed in the fruit fly (*Drosophila*). The ‘CUT’ gene in fruit fly is activated by an enhancer placed about 100kb away from the gene. It is established that a protein by name ‘chip’ plays a major role in bringing the enhancer and the CUT gene together. How the chip protein mediates such a huge range DNA interactions is still under speculation. A simple model proposes that chip forms mini loops or loop on loop in the DNA by binding to the regulatory elements present between the enhancer and the CUT gene and there by brings the enhancer and the CUT gene promoter to interact with each other (Dorsett 1995; Dean 2006). Similar phenomenon was observed in the regulation of the bithorax genes in *Drosophila*, whose *cis*-regulatory regions are located far upstream of about 300kb (Cleard et al. 2006). Position–effect variegation is another example which utilizes long range DNA interactions in determining the eye color in *Drosophila* (Dernburg et al. 1996). ‘Action at a distance’ phenomenon was also seen in plants. Regulatory sequences for the transcription and paramutation of B’ are located about 100kb upstream of the transcription start site which convert a purple pigmented maize plant into a lightly colored one with colorless seed (Stam et al. 2002).

Bidirectional promoters: An anecdote

Genomes of organisms contain special inherent zones that are rich in GC content, and the genes in these regions are transcribed from closely spaced divergent promoters. The first report of such a divergent transcription was observed by Taylor et al. (1967) in the bacteriophage lambda genes *cI* and *cro*. These two genes were positioned in a head-head manner and the intergenic distance is about 103bp. This intergenic distance harbors the promoters P_R and P_{RM} . These promoters along with the genes *cI* and *cro* were the best understood divergent transcription and regulation pattern so far. The next report on divergent transcription was observed in the *E. coli* biotin locus by Guha et al. (1971). From 1967 to 1988 about 60 divergently transcribed regions have been observed in prokaryotes, viruses of eukaryotes and mitochondrial and chloroplast genomes. In the divergently transcribed regions, promoters were arranged in a back to back to fashion leaving some intervening DNA between them, in an overlapping fashion and in a face to face format (Beck and Warren 1988). Depending upon the type of gene products encoded by the divergently arranged genes, they can be classified into three broad categories. The structural polypeptide class (S-S), wherein both the gene transcripts code for structural elements (non-regulatory in function). The second one is the R-S class, wherein one of the gene transcripts code for a regulatory protein/RNA. The third class is a pure R-R type, where in both the gene transcripts code for regulatory proteins/RNA as in the case

of the bacteriophage *cI* and *cro* divergent genes (Beck and Warren 1988). The most important functional significance of a divergent transcription is the tight and coordinated regulation of the divergent genes. A single *cis* regulatory region positioned between the two genes can regulate the expression of both the genes in a temporal and spatial manner and in stoichiometric proportions. Other functions include topological regulation and stability of the divergent module during the process of evolution (Beck and Warren 1988).

Latest insights on bidirectional promoters

Human Genome was estimated to harbor about 10% genes in a divergent manner (transcriptional start sites [TSS] separated by <1 kb), which are regulated by bidirectional promoters (Trinklein et al. 2004). These promoters have unique characteristics such as higher GC content, under representation of TATA boxes, and mirror images (Trinklein et al. 2004; Engstrom et al. 2006). The transcripts of most of the divergent gene pairs were shown to be co-expressed (Trinklein et al. 2004). Deletion and mutagenesis analysis identified GA-binding protein (GABP), the ets-family transcription factor to regulate more than 80% bidirectional promoters in humans by binding to GABP motif (Lin et al. 2007). However, it is not required for the divergent genes regulated by a bidirectional promoter to be coexpressed. This was shown in case of an evolutionarily conserved bidirectional promoter of CDT2, which regulates DNA replication and INTS7 (integrator complex subunit 7), which interacts with RNA polymerase II in mammalian genomes (human, mouse and canine) (Nakagawa et al. 2008). Investigation of biological significance of bidirectional promoters resulted in the identification of several genes involved in breast and ovarian cancers with common transcription factor binding sites (Yang et al. 2007). A bidirectional promoter driving the expression of two human genes, SIRT3 and PSMD13, involved in aging harbor transcription factor sp1 binding sites, which regulates expression of the two flanking genes (Bellizzi et al. 2007). The arrangement of these genes is evolutionarily conserved in bird, rat, mouse, dog, chimpanzee, and human. Bidirectional promoters appear to regulate genes associated with diseases. Examples include divergent gene pairs whose expression is associated with brain cancer and Parkinson's disease (West et al. 2003; Chen et al. 2009). Although bidirectional promoters have been identified and characterized to some extent in the human genome (Adachi and Lieber, 2002; Trinklein et al. 2004; Collins et al. 2007; Lin et al. 2007; Yang et al., 2007), very little is known about this unique class of promoters in model plant genomes. A few bidirectional promoters have been identified in plants that regulate nuclear genes. The first report was a promoter of oleosin and methionine sulphoxide reductase genes in *Brassica* (Keddie et al. 1994; Sadanandom et al. 1996). In addition, CaTin1 and CaTin1-2 genes in *Capsicum annuum* were found to be regulated by a bidirectional promoter (Shin et al. 2003).

CHAPTER 1:

GENOME-WIDE COMPARATIVE ANALYSIS OF PUTATIVE BIDIRECTIONAL PROMOTERS FROM RICE, *ARABIDOPSIS* AND *POPULUS*

Surendar Reddy Dhadi, Nicholas Krom, and Wusirika Ramakrishna

This work was supported by the National Research Initiative of the USDA Cooperative State Research, Education and Extension Service, grant number 2007-35301-18036.

The material contained in this chapter was previously published in the journal
GENE

Volume 429, issues 1-2, 15 January 2009, Pages 65–73

doi.org/10.1016/j.gene.2008.09.034

Copyright © 2012 Elsevier B.V. All rights reserved

<http://www.sciencedirect.com/science/article/pii/S0378111908004861>

1.1 Abstract

A bidirectional promoter can regulate the expression of two flanking genes arranged in a divergent manner. Although reports pertaining to bidirectional promoters on a genomic scale exist in mammals, little progress has been made in plants. In the present study, we performed a computational analysis of this unique class of promoters to identify overrepresented *cis*-regulatory motifs from three sequenced plant genomes: rice (*Oryza sativa*), *Arabidopsis thaliana*, and *Populus trichocarpa* using the Plant *Cis*-acting Regulatory DNA Elements (PLACE) and PLANT CARE databases. We describe these overrepresented elements and their possible regulatory mechanisms. We also discuss similarities and differences with human bidirectional promoters. Furthermore, we describe in detail a few coexpressed and evolutionarily conserved divergent gene pairs and their bidirectional promoters. This study provides insights into bidirectional promoters in three plant species, thereby laying a foundation for their experimental analysis.

1.2 Introduction

Regulating the simultaneous expression of two or more genes is often necessary in plants for metabolic engineering. Repeated use of a single promoter would result in transcriptional gene silencing (DeWilde et al. 2000). Therefore, bidirectional promoters can prove to be an excellent choice in coordinating multi-gene expression. They can be very useful in gene stacking in transgenic plants and gene farming (Halpin 2005). However, the identification and functional characterization of naturally occurring bidirectional promoters has not been performed on a genome wide scale in plants. Divergent genes are organized head-to-head in opposite orientations (Fig. 1.1). These genes, when separated by a small distance, may share a single promoter that acts in a bidirectional manner. Bidirectional promoters provide a unique mechanism of regulation by controlling the expression of two genes. Several bidirectional promoters have been reported in mammals compared to only a few reported in plants. Recent studies have shown that about 10% of genes in human genome are divergent and have putative bidirectional promoters (Trinklein et al. 2004). DNA repair genes were found to be more than five-fold overrepresented in these divergent gene pairs (Adachi and Lieber 2002; Trinklein et al. 2004). Transcripts of most of these divergent gene pairs were shown to be co-expressed. Deletion analysis was carried out for ten of these promoters and all of them were found to be bidirectional (Trinklein et al. 2004). Furthermore, overrepresented motifs corresponding to transcription factor binding sites were identified in human bidirectional promoters (Lin et al. 2007). Genome wide expression analysis of mouse divergent genes with bidirectional promoters revealed that they are more likely to be coordinately expressed than random pairs of genes (Engström et al. 2006). Examples of other bidirectional promoters include a promoter for chicken transporters associated with antigen processing (TAP) genes, TAP1 and TAP2 in the MHC region (Walker et al. 2005). Furthermore, two divergent genes, ACACA encoding acetyl-CoA carboxylase- α involved in fatty acid biosynthesis and TADA2L encoding a component of chromatin-modifying complexes have a bidirectional promoter in both human and mouse (Travers et al. 2005). A few bidirectional promoters have also been identified in plants that regulate nuclear genes. The first report was a promoter of oleosin and methionine sulphoxide reductase genes in Brassica (Keddie et al. 1994; Sadanandom et al. 1996). In addition, a promoter regulating expression of CaTin1 and CaTin1-2 genes in Capsicum annuum which was inoculated with tobacco mosaic virus (TMV) was found to be bidirectional (Shin et al. 2003). Identification and functional analysis of naturally occurring bidirectional promoters on a genomic scale in plants will provide insights into their functional significance in the genome and their role in gene regulation.

Our previous study identified three elements: CGACG, E2FCONSENSUS (recognized by E2F transcription factors), and SURECORE (sulfur responsive element core sequence), to be overrepresented in both rice and *Arabidopsis* when divergent genes pairs (separated by <1kb) with strong correlation were compared with those with weak or no correlation (Krom and Ramakrishna 2008). In the same study, PRECONSCRHSP70A was identified as an overrepresented element in *Populus* and *Arabidopsis* promoters. Since coexpression cannot be used as the only criteria required for bidirectionality of a promoter, we used short intergenic distance (<250bp) between divergent genes as an additional criteria to identify regulatory elements in three plant genomes: rice, *Arabidopsis*, and *Populus* in the present study. Furthermore, overrepresentation of regulatory elements in coexpressed divergent genes was identified using a reference set of random promoters. Here, we discuss these overrepresented regulatory elements, which may contribute to their bidirectional activity in three diverse plant genomes.



Fig. 1.1 A divergent gene pair with a bidirectional promoter. Arrows represent divergent genes. The filled box represents a bidirectional promoter.

1.3 Methods

1.3.1 Promoter identification and selection

Sequence and annotation data for the *Oryza sativa* ssp. Japonica genome was downloaded from the TIGR Rice Genome Annotation Database (<http://www.tigr.org/tdb/e2k1/osa1>). Similar data for the *Arabidopsis thaliana* and *Populus trichocarpa* genomes was obtained from The *Arabidopsis* Information Resource (TAIR) (ftp://ftp.Arabidopsis.org/home/tair/Genes/TIGR5_genome_release) and Joint Genome Institute (JGI) (http://genome.jgi-psf.org/Poptr1_1/Poptr1_1.home.html) websites, respectively. A total of 28,287 and 24,019 genes were analyzed from rice and *Arabidopsis*, respectively, after filtering out genes annotated as hypothetical and transposons. 45,554 genes listed in version 1.1 of the JGI annotation of the *Populus trichocarpa* genome were analyzed. However, hypothetical genes and transposons were not removed from this data because predicted functions for genes were not available. Divergently arranged pairs of adjacent genes (head-to-head on opposite strands) were identified as described earlier with any pair containing genes annotated as transposon-related or hypothetical being excluded from further analysis (Krom and Ramakrishna

2008). The entire region between the genes' transcription start sites separated by <250bp was designated as a putative bidirectional promoter, and was analyzed for over- or under-represented regulatory motifs using a reference set of one thousand <250bp upstream sequences from randomly selected genes. Massively Parallel Signature Sequencing (MPSS) (Meyers et al. 2004) data was collected (<http://mpss.udel.edu/>) for all rice and *Arabidopsis* divergently arranged genes separated by <250bp. Only 17bp signatures of classes 1, 2, 5, and 7 that mapped to a single gene were used, and abundance values less than 5 were ignored as background interference. Correlated expression was identified by calculating the Pearson correlation coefficient from each gene's average abundance values across all libraries as described earlier (Krom and Ramakrishna 2008). Gene pairs with Pearson correlation greater than 0.5 were considered strongly correlated. Pearson correlation was also calculated using microarray expression data from the Yale Rice Project (<http://bioinformatics.med.yale.edu/rc/overview.jsp>), the Nottingham *Arabidopsis* Stock Centre microarray database (<http://affymetrix.Arabidopsis.info/>), and the NCBI Gene Expression Omnibus (<http://www.ncbi.nlm.nih.gov/geo/>) as described earlier (Krom and Ramakrishna 2008). Promoters from gene pairs displaying strong correlation based on either source of expression data were analyzed for over- or under-represented regulatory motifs as a distinct subclass. Conserved divergent gene pairs in each species were identified in the genomes of the other two species using a combination of BLASTP and TBLASTN searches (Krom and Ramakrishna 2008). If homologs were identified for both genes, and the original divergent arrangement was conserved, then the overall gene pair was considered to be conserved.

1.3.2 Promoter sequence analysis

The promoter sequences in all the sets were analyzed using two publicly available databases: Database of Plant *Cis*-acting Regulatory DNA Elements (PLACE) (<http://www.dna.affrc.go.jp/PLACE/>) and the Plant CARE database (<http://bioinformatics.psb.ugent.be/webtools/plantcare/html/>). The results from each database were then examined and the number of promoter sequences containing each regulatory motif or repeat sequence was calculated. Promoter sets from all identified divergent pairs and strongly correlated pairs were compared with the results from a set of 1,000 randomly selected promoters for statistical analysis. The binomial test (normal approximation) was used to test for significant ($P < 0.001$) deviation from the frequency observed in the random set.

1.4 Results and Discussion

1.4.1 Computational identification of putative bidirectional promoters from rice, *Arabidopsis*, and *Populus* and their high GC content

We identified 212, 462, and 141 divergent gene pairs separated by <250bp from rice, *Arabidopsis*, and *Populus* genomes, respectively (Krom and Ramakrishna 2008). The intergenic region between transcriptional start sites of these divergent genes was defined as a putative bidirectional promoter. We used stringent criteria to define a bidirectional promoter compared to the study of human genome where an intergenic distance of <1kb between divergent genes was considered to be a bidirectional promoter (Trinklein et al. 2004). Of these, 52, 75, and 76 divergent gene pairs flanking bidirectional promoters were coexpressed based on Pearson coefficient value of $R > 0.5$ using microarray or MPSS data from rice, *Arabidopsis*, and *Populus*, respectively (Krom and Ramakrishna 2008). Since only small numbers of conserved divergent gene pairs were identified, this data was not used to identify overrepresented motifs.

We selected 250bp as the upper limit compared to 1kb used for the identification of bidirectional promoters in the human genome because of the smaller genome sizes of rice (390 Mb), *Arabidopsis* (125Mb) and *Populus* (480Mb) compared to human (3000Mb). However, using this criteria, bidirectional promoters between divergent genes which are <250 bp apart would not be identified. Therefore, we performed an additional analysis using the criteria of <1 kb to identify this set of promoters, similar to that used for the identification of bidirectional promoters in the human genome. Using this approach, 1242, 2106, and 613 promoters were identified which regulate about 9%, 18%, and 3% of genes from rice, *Arabidopsis* and *Populus*, respectively. The percentages of rice and *Arabidopsis* genes with bidirectional promoters of <1kb are comparable to that of human (11%). The higher number of these promoters in the *Arabidopsis* genome may be due to its compact genome size. Of the three genomes, *Populus* is the most recently sequenced genome using shotgun approach whose annotation is far less refined compared to rice and *Arabidopsis*, which were sequenced using map-based approach. This is likely to be the reason for the low number of bidirectional promoters of <1kb identified in the *Populus* genome. GC content of these promoters was analyzed and compared to a set of random promoter sequences of 250bp from the transcriptional start site. Significantly higher GC content of 55%, 37%, and 48% was observed in rice, *Arabidopsis*, and *Populus* bidirectional promoters, compared to 47%, 33%, and 38% of rice, *Arabidopsis*, and *Populus* random promoter sets, respectively (Table 1.1). Increase in GC content was much higher in *Populus* and rice compared to *Arabidopsis*. Statistical significance increased in most cases when sets of these promoters which drive the expression of coexpressed genes were considered. Human bidirectional promoters showed a similar

trend with 66% GC content compared to random promoters with 53% GC content (Trinklein et al. 2004). These results suggest that plant bidirectional promoters share a higher GC content with human bidirectional promoters.

Table 1.1 GC Content of bidirectional promoters

	GC content	Z
Rice		
Promoters < 250bp	54.78%	3.35
Coexpressed divergent genes (Microarray R > 0.5)	55.56%	3.69
Coexpressed divergent genes (MPSS R > 0.5)	53.68%	2.88
1000 random 250bp promoters	47.00%	
Arabidopsis		
Promoters < 250bp	36.78%	1.78
Coexpressed divergent genes (Microarray R > 0.5)	38.00%	2.33
Coexpressed divergent genes (MPSS R > 0.5)	37.77%	2.23
1000 random 250bp promoters	32.90%	
Populus		
Promoters < 250bp	48.05%	4.42
Coexpressed divergent genes (Microarray R > 0.5)	50.95%	5.7
1000 random 250bp promoters	38.07%	

Z values were estimated for each subset by comparing its GC content with that of 1000 random promoters given at the end of each set. P value estimates: $|Z| > 3.0902$: $p < 0.001$, $|Z| > 2.3263$: $p < 0.01$, $|Z| > 1.6449$: $p < 0.05$.

1.4.2 Overrepresented and abundant motifs in putative bidirectional promoters in rice, *Arabidopsis* and *Populus* genomes represented in the PLACE database

Cis-acting regulatory DNA elements in bidirectional promoters were identified from rice, *Arabidopsis*, and *Populus* using the PLACE database (Higo et al. 1999; www.dna.affrc.go.jp/PLACE/index.html). Over or under-represented motifs in these promoters were compared to random sets of 1000 promoter sequences of 250 bases from the transcriptional start site for each of the three plant species. For computational analysis we divided the data into two sets comprised of 1) putative bidirectional promoters of <250bp flanked by divergent genes in the three plant genomes; 2) putative bidirectional promoters of <250bp flanked by divergent genes that display highly correlated expression based on Pearson coefficient value of $R > 0.5$ using the expression data from microarray or MPSS.

Three motifs were overrepresented in the first set of all rice bidirectional promoters compared to a random set of promoters based on Z values with a cut-off p-value of 0.001 (Table 1.2). In addition to a significant Z value, the presence of these motifs in 10% or

more bidirectional promoters was used as a second criterion. Here, we identified SORLIP2AT (GGGCC) to be significantly overrepresented in putative bidirectional promoters of rice. This motif is a subset of sequences over-represented in light-induced promoters (SORLIP), which were characterized in light induced promoters of phytochrome genes (*phyA*) of *Arabidopsis* and assigned a role in the regulation of a significant subset of the *phyA*-responsive transcripts (Hudson and Quail 2003). The second element, SITEIIATCYTC (TGGGCT) is present in promoters of rice and *Arabidopsis* proliferating cell nuclear antigen (PCNA) genes, which are expressed in meristematic tissues (Kosugi et al. 1995; Trémousaygue et al. 2003). These genes code for PCNA, which assists DNA polymerase δ and is involved in several other cellular functions (Jonsson and Hubscher 1997). Furthermore, this SITE motif is over-represented in the promoters of nuclear genes encoding components of the oxidative phosphorylation machinery of *Arabidopsis* and rice (Welchen and Gonzalez 2006). The third *cis*-regulatory element, UP1ATMSD (GGCCAWWW), was found to be associated with promoters of genes whose expression was up-regulated in axillary buds after stem decapitation (Tatematsu et al. 2005). This element shares a common core sequence “GGGCC” with the *cis* element, SITE II motif. Due to the presence of this common core sequence, these two elements become potential target sequences of the TEOSINTE BRANCHED 1-CYCLOIDEA-PCF (TCP) family of transcription factors (Kosugi and Ohashi 1997, 2002; Trémousaygue et al., 2003). This family is composed of two subfamilies, PCF (class I) and CYC/TB1 (class II). Although proteins belonging to these two subfamilies can bind to similar sequences, they have preferences for certain DNA sequences (Cubas et al. 1999; Kosugi and Ohashi 2002). Overall these three elements with a common core conserved sequence “GGGCC” were overrepresented in putative bidirectional promoters in the rice genome. Genome-wide quantitative analysis was performed using microarray and MPSS data to identify correlated expression levels across multiple tissues and treatments, as determined by Pearson correlation ($R > 0.5$) (Krom and Ramakrishna 2008). In order to make this data set more uniform in all three species and the different databases used, we combined the micro array and MPSS data into a single high correlation set. The three *cis*-regulatory elements, SORLIP2AT, SITEIIATCYTC, and UP1ATMSD, overrepresented in the first set of all bidirectional promoters separated by <250bp, were also overrepresented in this category (Table 2). Similar analysis in *Arabidopsis* identified six overrepresented motifs in all promoters set (Table 1.3). In addition to the three motifs identified in rice promoters in this category, CGACGOSAMY3, UP2ATMSD, WUSATAg elements were found in *Arabidopsis*. CGACG element was identified in the promoter of the α -amylase gene, Amy3D and is required for its expression during sugar starvation (Hwang et al. 1998). This element was also found to be part of an enhancer region in the promoter of *Arabidopsis* pathogen responsive gene, PDF1.2 (Brown et al. 2003). UP1 and UP2 act synergistically to up-

regulate expression of nuclear genes after main stem decapitation (Tatematsu et al. 2005). The *Arabidopsis* WUSCHEL homeo domain proteins have been shown to bind to WUSATAg motifs and regulate the formation and maintenance of shoot and root apical meristems (Mayer et al. 1998; Kamiya et al. 2003). In the *Arabidopsis* high correlation set, in addition to the three elements observed in rice, UP2ATMSD was also overrepresented. SORLIP2AT, SITEIIATCYTC, and UP1ATMSD were the only elements overrepresented in both *Arabidopsis* and rice. The fraction of *Populus* promoters with overrepresented motifs, especially in the all promoters set, was found to be much lower than rice and *Arabidopsis*. Seven motifs, CGACGOSAMY3, CGCGBOXAT, GCCCORE, LTRECOREATCOR15, PRECONSCRHSP70A, SITEIIATCYTC, and SORLIP2AT were found to be overrepresented in both sets in *Populus* (Table 1.4). CGACG element was the common overrepresented element in the first set of *Populus* and *Arabidopsis* but not in rice promoters. CGCG box is present in promoters of genes with a role in ethylene and abscisic acid signaling, and light signal perception (Yang and Poovaiah 2002). GCC box is present in the promoters of ethylene and defense responsive genes (Brown et al. 2003). Low temperature responsive element (LTRE) is required for the expression of cold induced genes (Baker et al. 1994; Medina et al. 1999). PRECONSCRHSP70A is a consensus motif present in a plastid response element (PRE), which is part of HSP70A gene promoter of *Chlamydomonas* and acts as an enhancer (Von Gromoff et al. 2006). Gene expression is up-regulated through this motif by a chlorophyll precursor, Mg-protoporphyrin and light. Two motifs, SITEIIATCYTC and SORLIP2AT are overrepresented in different sets of bidirectional promoters of all three plant species. These elements have in common the GGGCC conserved motif, which is a potential candidate for bidirectional activity of promoters. Human bidirectional promoters were found to have a similar motif GGGCGG that serves as a binding site for a zinc finger protein SP1, which is a transcription factor (Engström et al. 2006; Todd and Neidle 2008). Several motifs were found to be underrepresented in bidirectional promoters. A small number of motifs were found to be significantly underrepresented in all three plant genomes. The most prominent of these motifs are ARR1AT (NGATT), CACTFTPPCA1 (YACT), CAATBOX1 (CAAT), GATABOX (GATA), GT1CONSENSUS (GRWAAW), GTGANTG10 (GTGA), and POLLEN1LELAT52 (AGAAA). Furthermore, the number of TATA boxes was highly underrepresented. Most of these elements are bound by various transcription factors. These results suggest that the underrepresented regulatory motifs show a preference to unidirectional but not bidirectional promoters. Similar to our results, human bidirectional promoters were also found to have TATA boxes underrepresented in them. TATA box appears to be responsible for promoter unidirectionality in most cases whereas having few or no TATA boxes appears to be a hallmark of most of human bidirectional promoters (Dong et al. 2000; Kawai et al. 2003; Trinklein et al. 2004). Our analysis of three plant genomes shows a similar trend

suggesting a novel mechanism of regulation by bidirectional promoters compared to unidirectional promoters. Promoters of <1kb between divergent genes were analyzed using the PLACE database in order to identify overrepresented motifs ([Supplementary table S1.1](#)). Twenty two motifs were overrepresented in rice which include the three motifs in the <250bp dataset. Similar analysis in *Arabidopsis* identified ten motifs which includes all motifs in the <250bp dataset, except WUSATAg motif. However, analysis of *Populus* promoters identified only three overrepresented motifs, which include only two out of seven motifs identified in the <250bp dataset. In addition to SORLIP2AT, UP1 was found to be a common overrepresented element in all three genomes in promoters of <1kb. GCCCORE was found to be a common overrepresented element only in rice and *Populus* while five elements were common only in rice and *Arabidopsis*. Although most of the elements found to be overrepresented in <250bp promoters, were also identified in <1kb promoters in rice and *Arabidopsis*, several additional overrepresented elements were found in <1kb promoter dataset.

Table 1.2 Overrepresented motifs in rice bidirectional promoters from PLACE database

Motif	Sequence	Promoters		Z
		BD	Random	
All promoters (212)				
SITEIIATCYTC	TGGGCY	145	317	11.48
SORLIP2AT	GGGCC	155	346	11.79
UP1ATMSD	GGCCCAWWW	63	69	13.11
High correlation (52)				
SITEIIATCYTC	TGGGCY	37	317	6.11
SORLIP2AT	GGGCC	38	346	5.83
UP1ATMSD	GGCCCAWWW	15	69	6.24

All promoters represent promoters between divergent genes separated by <250 bp. High correlation represents promoters whose divergent genes are separated by <250 bp and show Pearson $R > 0.5$. The number in parenthesis indicates number of promoters in each subset. BD represents the number of putative bidirectional promoters with the specific cis element. Random represents the number of random promoters out of 1000 that have the specific cis-element. Overrepresented motifs present in >10% of promoters in each set are shown. Z values were estimated for each cis-element by comparing the number of putative bidirectional promoters with that of 1000 random promoters with the specific element. P value estimates: $|Z| > 3.0902$: $p < 0.001$, $|Z| > 2.3263$: $p < 0.01$, $|Z| > 1.6449$: $p < 0.05$. Cut-off p-value used was <0.001.

Table 1.3 Overrepresented motifs in *Arabidopsis* bidirectional promoters from PLACE database

Motif	Sequence	Promoters		Z
		BD	Random	
All promoters (462)				
CGACGOSAMY3	CGACG	97	156	3.2
SITEIIATCYTC	TGGGCY	198	204	11.98
SORLIP2AT	GGGCC	170	172	11.16
UP1ATMSD	GGCCCAWWW	132	96	13.84
UP2ATMSD	AAACCCTA	59	65	5.47
WUSATAg	TTAATGG	57	72	4.27
High correlation (75)				
SITEIIATCYTC	TGGGCY	37	204	6.22
SORLIP2AT	GGGCC	31	172	5.54
UP1ATMSD	GGCCCAWWW	23	96	6.19
UP2ATMSD	AAACCCTA	13	65	3.81

All the parameters are identical to those described in [Table 1.2](#).

Table 1.4 Overrepresented motifs in *Populus* bidirectional promoters from PLACE database

Motif	Sequence	Promoters		Z
		BD	Random	
All promoters (141)				
CGACGOSAMY3	CGACG	26	89	3.98
CGCGBOXAT	VCGGB	38	90	7.45
GCCCORE	GCCGCC	26	52	7.08
LTRECOREATCOR15	CCGAC	29	104	3.96
PRECONSCRHSP70A	SCGAYNRNNNNNNNNNNNNNNNNHHD	34	140	3.46
SITEIIATCYTC	TGGGCY	33	132	3.58
SORLIP2AT	GGGCC	41	141	5.11
High correlation (76)				
CGACGOSAMY3	CGACG	16	89	3.72
CGCGBOXAT	VCGGB	21	90	5.68
GCCCORE	GCCGCC	15	52	5.71
HEXAMERATH4	CCGTCG	8	21	5.12
LTRECOREATCOR15	CCGAC	20	104	4.55
PRECONSCRHSP70A	SCGAYNRNNNNNNNNNNNNNNNNHHD	23	140	4.09
SITEIIATCYTC	TGGGCY	22	132	4.06
SORLIP2AT	GGGCC	25	141	4.71

All the parameters are identical to those described in [Table 1.2](#).

1.4.3 Overrepresented motifs in putative bidirectional promoters in three plant genomes represented in PLANTCARE database

Overrepresented motifs in bidirectional promoters in rice, *Arabidopsis*, and *Populus* genomes were investigated using the PLANTCARE database with the same criteria as described for the PLACE database. None of the regulatory motifs were over or underrepresented in rice promoters. Only TGA-element (AACGAC), found in a set of auxin responsive genes in soybean (Nagao et al. 1993), was conserved in the all promoters set of *Arabidopsis* (Table 1.5). Three elements, including TGA-element, were overrepresented in the set of all *Populus* bidirectional promoters. Only two elements, GC-motif (GCCCCGG), required for Adh gene expression in *Arabidopsis* (Dolferus et al. 2002), and unnamed_2 (CCCCGG), were overrepresented in the high correlation set of *Populus* and shared with the first set. As seen in the case of PLACE database results, TATA boxes were highly underrepresented in bidirectional promoters of all three genomes based on PLANTCARE data. It is likely that the overrepresented elements play a prominent role in the bidirectional activity of these promoters. Similarly, underrepresented elements are not likely to be associated with these promoters suggesting a novel mechanism of regulating genes without using TATA boxes.

Table 1.5 Overrepresented motifs in bidirectional promoters from PLANTCARE database

Motif	Sequence	Promoters		Z
		BD	Random	
<i>Arabidopsis</i> motifs				
All promoters (462)				
TGA-element	AACGAC	93	108	6.46
<i>Populus</i> motifs				
All promoters (141)				
GC-motif	GCCCCGG	15	31	5.16
TGA-element	AACGAC	20	59	4.17
Unnamed_2	CCCCGG	14	28	5.13
High correlation (75)				
GC-motif	GCCCCGG	10	31	5.06
Unnamed_2	CCCCGG	11	28	6.17

All the parameters are identical to those described in Table 1.2.

1.4.4 Representative bidirectional promoters and divergent genes

Based on different criteria such as distance between divergent genes and their coexpression patterns, we identified several putative bidirectional promoters. Here we first describe a few representative putative bidirectional promoters and their flanking divergent genes from rice. Of the two genes regulated by one such promoter, the first gene, LOC_Os07g01540, annotated as calcineurin-like phospho-esterase family protein, is similar to the *Arabidopsis* gene (At4g31770) encoding a protein predicted to contain a carboxyl-terminal lariat debranching enzyme domain (Wang et al. 2004). The second gene, LOC_Os07g01550 encodes a protein similar to polygalacturonase-inhibiting protein (PGIP), a leucine rich repeat family protein, which is associated with plant cell walls and involved in plant defense (Di Matteo et al. 2003). Another divergent gene pair consists of LOC_Os03g53500 which codes for a protein with similarity to the C-terminal domain of a helicase with a possible role in ATP hydrolysis (Bird et al. 1998) and LOC_Os03g53510 which is annotated as a gene with similarity to *Arabidopsis* transducin/ WD40 repeat containing gene (At2g22040). Transducin is a hetero trimeric G protein whereas WD40 repeats are protein domains involved in a wide range of regulatory functions that include signal transduction and transcription. Similarly, a third pair consists of LOC_Os01g03650 encoding a protein containing sufB/sufD domains, which mediate sulfur mobilization, Fe-S cluster assembly and iron homeostasis (Xu et al. 2005), and LOC_Os01g03660, which encodes a protein similar to MYB like DNA-binding domain containing protein. MYB transcription factor family members in plants include several MYB genes involved in various developmental processes and defense responses (Yanhui et al. 2006). A fourth pair has LOC_Os01g08960 which encodes a protein similar to phosducin, which is a regulator of cytosolic G-protein. Phosducin sequesters G $\beta\gamma$ complex needed for the formation of a functional G protein trimer involved in several plant responses through signal transduction (Jones and Assmann 2004; Perfus-Barbeoch et al. 2004). The second gene, LOC_Os01g08970, codes for a protein similar to structure-specific recognition protein 1, which is a DNA bending protein (Hotze et al. 1995; Grasser et al. 2000). Four other divergent gene pairs which include genes that code for transcription factors, redox enzymes, and proteins whose functions are not known are shown in Fig. 1.2. All these promoters show multiple SORLIP2/SITEII elements. The other elements frequently present in these promoters are UP1 and SURECORE.

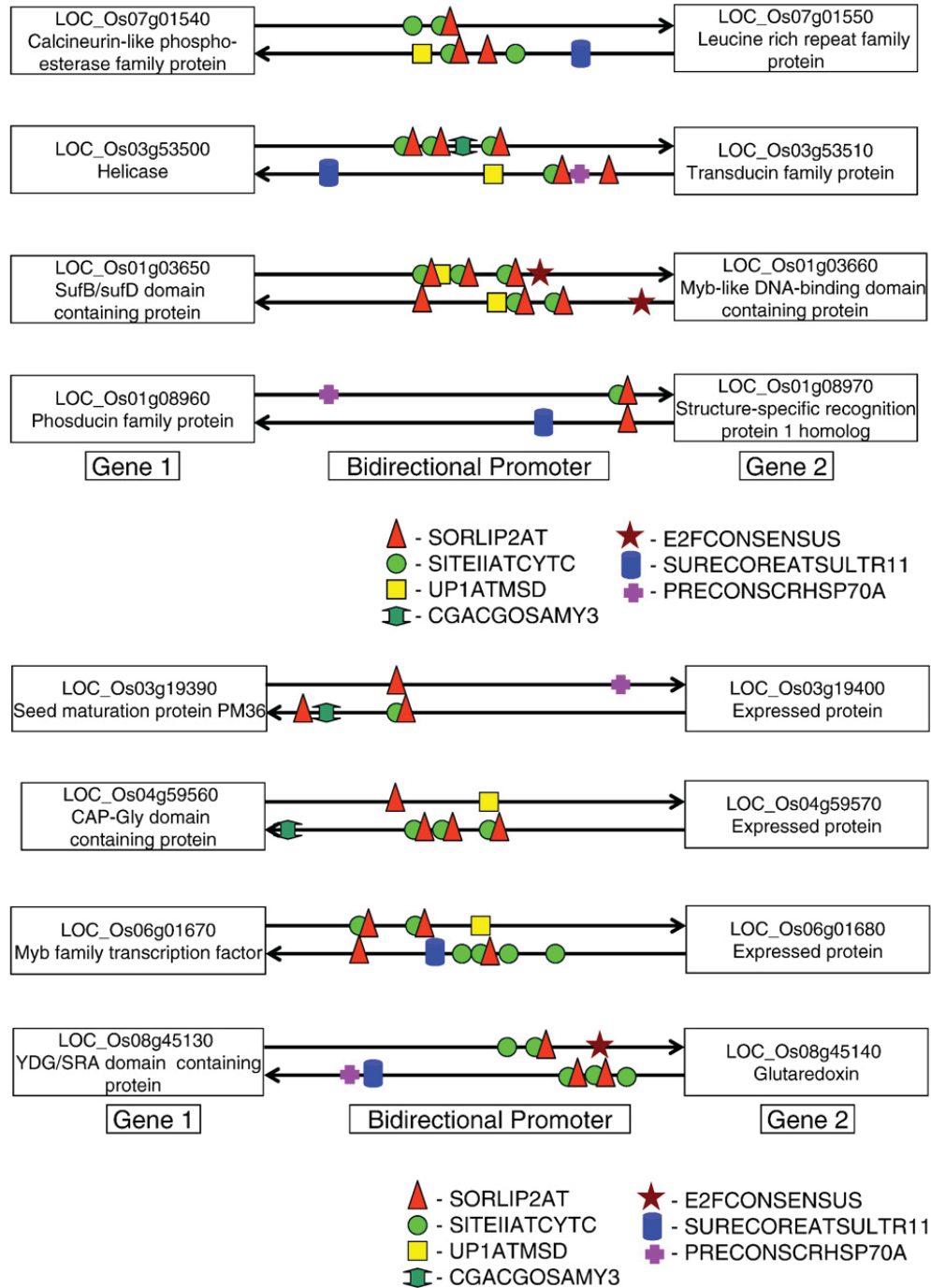


Fig. 1.2 Putative bidirectional promoters with cis-elements and divergent genes from rice. Locus identifiers (LOC) represent TIGR version 4 gene identification numbers. Distances between the genes in a pair were <250bp apart. These were estimated based on full-length cDNAs and ESTs, where available. Pearson R values were >0.5 based on microarray or MPSS data.

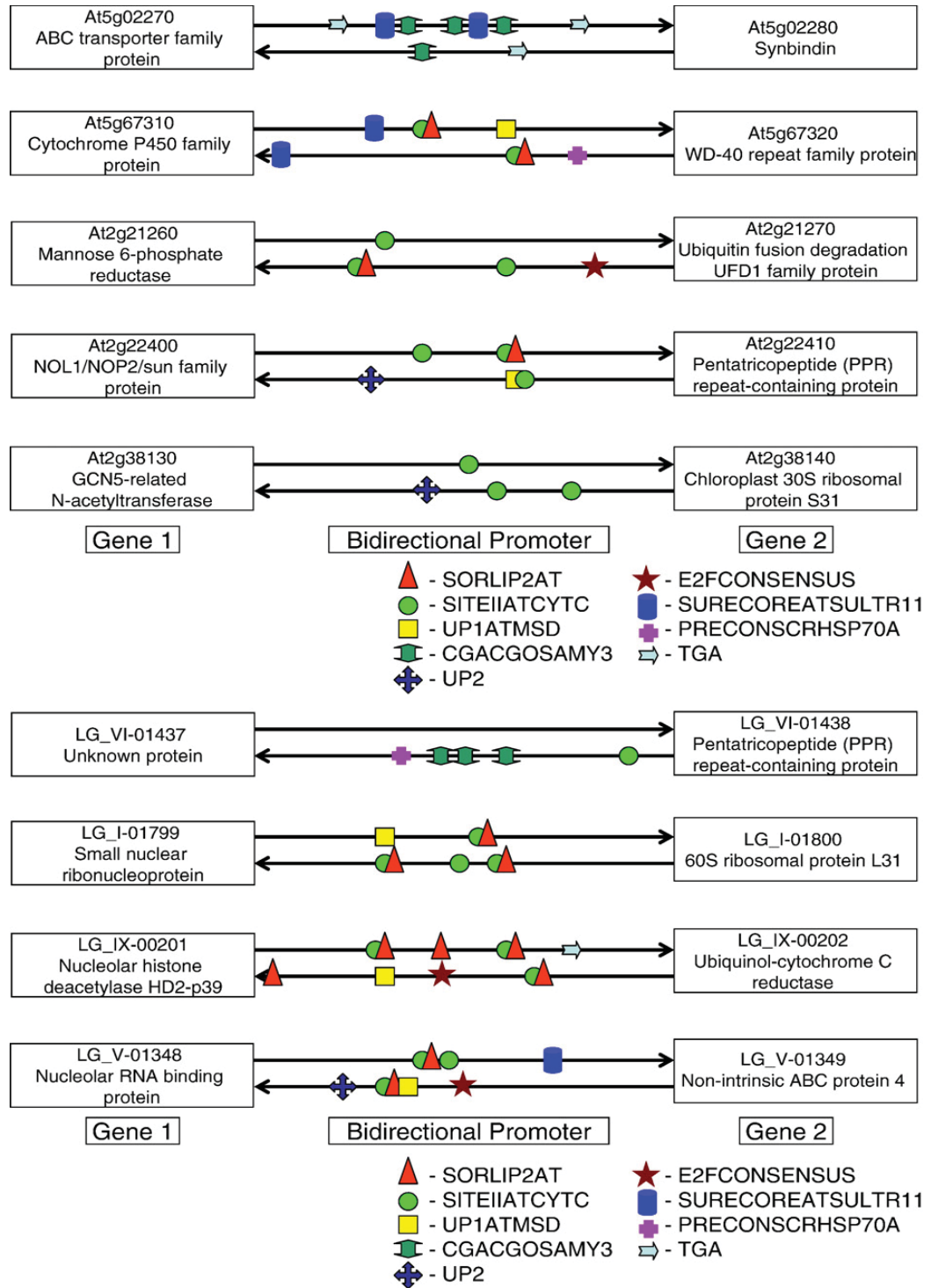


Fig. 1.3 Putative bidirectional promoters with cis-elements and divergent genes from *Arabidopsis* and *Populus*. Annotation of *Populus* genes is based on closest *Arabidopsis* homologs. Divergent genes are separated by <250 bp with Pearson R value>0.5.

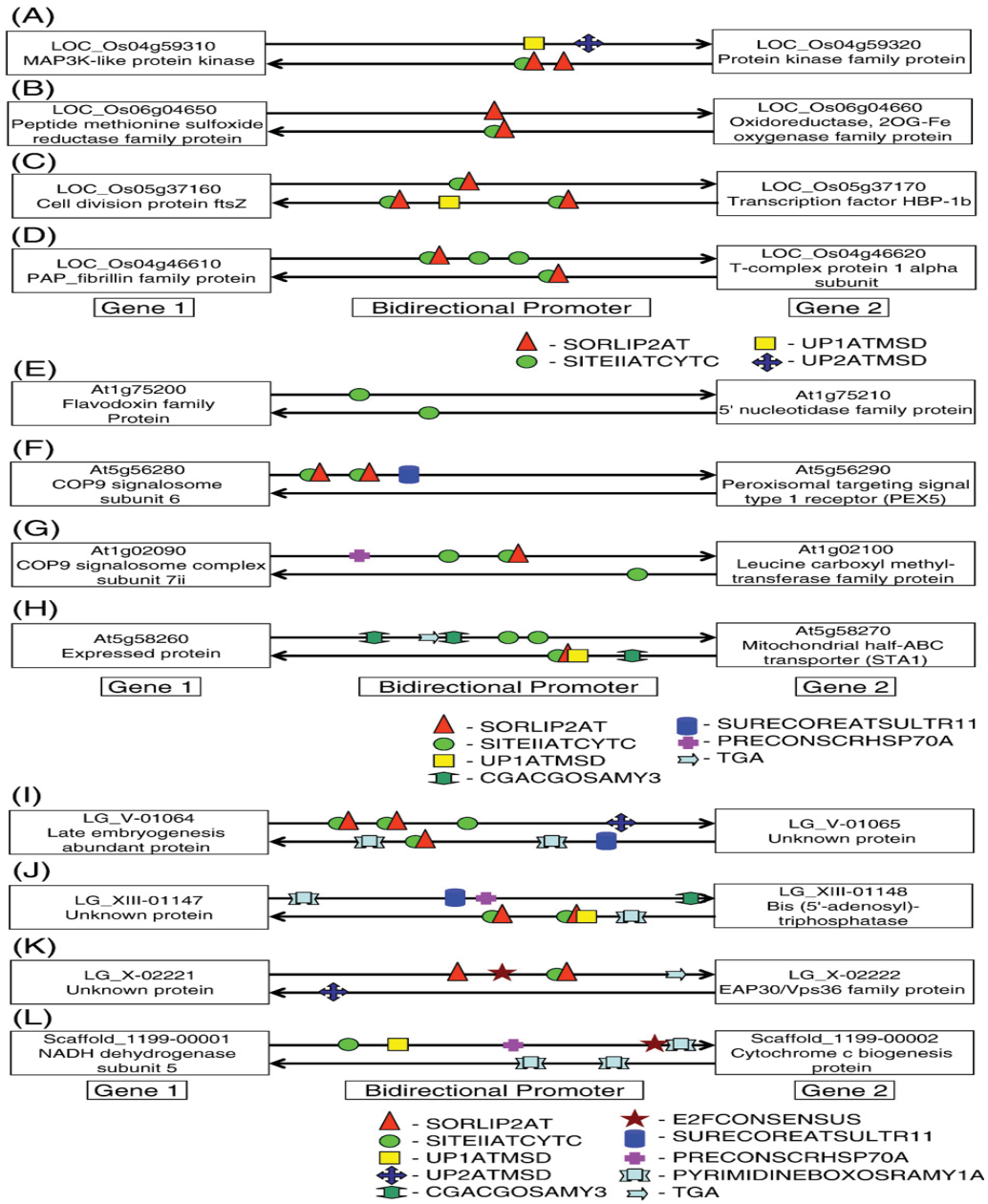


Fig. 1.4 Putative bidirectional promoters with cis-elements and divergent genes with conserved gene order and orientation in other plant species. (A, B) Rice gene pairs conserved in Arabidopsis. (C, D) Rice gene pairs conserved in Populus. (E, F) Arabidopsis gene pairs conserved in rice. (G, H) Arabidopsis gene pairs conserved in Populus. (I, J) Populus gene pairs conserved in Arabidopsis. (K, L) Populus gene pairs conserved in rice. Annotation of Populus gene pairs conserved in Arabidopsis and rice is based on their homologs in these two species.

A few representative *Arabidopsis* and *Populus* divergent genes and bidirectional promoters are shown in Fig. 1.3. *Arabidopsis* divergent genes with putative bidirectional promoters include those encoding ABC transporter, cytochrome P450 and WD-40 repeat family proteins, and oxidoreductases. *Populus* divergent genes shown in this figure include those that encode small nuclear ribonucleoprotein, histone deacetylase, a reductase, and an RNA binding protein. Most of these promoters have SORLIP2 or SITEII motifs as seen in case of rice. Divergent gene pairs displaying interspecies conservation are shown in Fig. 1.4. Rice divergent genes conserved in other species include protein kinases, reductases, and transcription factors (Figs. 1.4 A–D). *Arabidopsis* divergent genes conserved in other species include those encoding subunits of COP9 signalosome complex involved in protein degradation mediated by ubiquitin/proteasome (Karniol and Chamovitz 2000), oxidoreductases, and a mitochondrial half-ABC transporter involved in iron homeostasis (Figs. 1.4E–H). Finally, *Populus* divergent gene pairs conserved in other species include late embryogenesis abundant protein which protects plants from stress, bis (5'-adenosyl)-triphosphatase involved in purine metabolism, oxidoreductases, and EAP30/Vps36 family protein (Figs. 1.4I–L). Vps36 is involved in golgi to endosome trafficking while EAP30 is a subunit of the ELL complex which is a transcription factor (Winter and Hauser 2006). It would be interesting to evaluate functional relationships between the two genes of different divergent gene pairs. In addition to SORLIP2, SITEII and UP1 motifs, promoters in this set have few other elements. This includes PYRIMIDINEBOXOSRAMY1A present in *Populus* gene pairs conserved in other species. This motif (CCTTTT) was shown to be partially responsible for sugar repression of an alpha amylase gene (RAmy1A) in rice (Morita et al. 1998). The role of regulatory elements responsible for bidirectional activity of these promoters needs to be explored by targeted mutagenesis and making promoter-reporter gene deletion constructs.

1.4.5 Bidirectional promoters: a new toolkit to manipulate plant genomes

Plant genomes can regulate multiple genes involved in biological and biochemical pathways in an efficient manner using bidirectional promoters. This will result in the utilization of less energy in activating the expression of multiple genes. A bidirectional promoter can generate protein products from two adjacent related genes in stoichiometric quantities, which is biologically significant. Investigating the mode of action of these promoters will provide insights into novel gene regulatory mechanisms used by plants. Furthermore, these promoters can be used for tissue or cell specific expression based on the expression of flanking genes and will be useful in various biotechnological applications in plants.

Bidirectional promoters can be used for co-expressing multi-gene traits. They can regulate co-expression of genes functioning in the same or related biological pathways. In addition, a natural bidirectional promoter will aid in the expression of a gene close to its physiological conditions. Bidirectional promoters can be used to genetically engineer plants for crop improvement. Furthermore, bidirectional promoters will be useful in gene stacking where multiple genes are expressed in transgenic plants and in molecular farming in the production of vaccines, pharmaceuticals, and plastics.

Although bidirectional promoters have been better characterized in mammalian genomes, especially the human genome, very little is known about their biological significance. One study, which investigated this aspect, found that several genes which play a role in breast and ovarian cancers were regulated by bidirectional promoters with common transcription factor binding sites (Yang et al. 2007). This study suggested a role for these promoters for regulating cancer genes through epigenetic modifications. However, we know very little about the organization and regulatory mechanisms used by bidirectional promoters in plants. In this study, we identified motifs that are overrepresented in these bidirectional promoter sequences and may contribute to their bidirectional activity. This could be achieved through enhancer-like properties of regulatory motifs present in these promoters which may regulate expression of genes in both orientations. Our analysis has laid a foundation for the study of bidirectional promoters in plants in order to unravel their biological significance.

1.5 Acknowledgments

This work was supported by the National Research Initiative of the USDA Cooperative State Research, Education and Extension Service, grant number 2007-35301-18036.

CHAPTER 2:

A NOVEL NON-WOUNDING TRANSIENT EXPRESSION ASSAY FOR CEREALS MEDIATED BY *AGROBACTERIUM* *TUMEFACIENS*

Surendar Reddy Dhadi, Aparna Deshpande and Wusirika Ramakrishna

This work was supported by the National Research Initiative of the USDA Cooperative State Research, Education and Extension Service, grant number 2007-35301-18036.

The material contained in this chapter was previously published in the journal
PLANT MOLECULAR BIOLOGY REPORTER (2012)
Volume 30, Number 1, 36-45, DOI: 10.1007/s11105-011-0314-5
<http://www.springerlink.com/content/07m23860w34328vt/abstract/>
Copyright © 2011, Springer-Verlag

With kind permission from Springer Science and Business Media: Plant Molecular Biology Reporter, A novel non-wounding transient expression assay for cereals mediated by *Agrobacterium Tumefaciens*, Surendar Reddy Dhadi, Aparna Deshpande and Wusirika Ramakrishna, © Springer Science and Business Media 2012.

2.1 Abstract

A novel *Agrobacterium tumefaciens* mediated transient expression assay (AmTEA) was developed for young plants of different cereal species and the model dicot *Arabidopsis thaliana*. AmTEA was evaluated using five promoters (six constructs) and two reporter genes, gus and egfp. The constitutive 35S promoter and the promoter of the rice glutaredoxin gene showed gus and egfp expression in the cereals analyzed in the present study. A promoter for the DEAD-box RNA helicase family protein gene from *Arabidopsis* showed similar expression patterns of reporter genes in stable transgenic lines as well as in transient expression lines of *Arabidopsis*. *Agrobacterium tumefaciens* co-cultivation and plant incubation times were optimized using 35S and the rice expressed protein gene promoter (R2-273). The possibility of non-specific expression of the reporter genes was ruled out by using the antibiotic carbenicillin and the comparison of expression of the reporter genes driven by full-length and truncated R2-273 promoters. AmTEA considerably reduced time, space, labor, and cost requirements. Ease of use with stress treatments is another major advantage of this method. AmTEA can be automated and used for large-scale studies to decipher promoter and gene functions with the ultimate goal to enhance the performance of cereal crops against biotic and abiotic stresses.

2.2 Introduction

The majority of the world's population is dependent on cereal crops for their staple food. Improving the production of cereal crops is one of the most important missions of agricultural research. Plant genetic engineering offers a wide variety of tools to achieve this goal. These tools can be used to re-engineer genes and promoters to produce better cereal crops.

Agrobacterium tumefaciens has become a major tool in the hands of plant biotechnologists, for genetic engineering of plants. Binary vector systems based on Ti plasmid for the delivery of chimeric genes into various plant systems (Barton and Chilton 1983; Gelvin 2003) have revolutionized the field of plant genetic engineering. The concept of transient expression assay (TEA) was first developed using electroporation of plant protoplasts (Fraley et al. 1983; Fromm et al. 1985). This was followed by other methods such as biolistics (Li et al. 1993; Chlan et al. 1995) and *Agrobacterium*-mediated tissue culture methods (Barton and Chilton 1983) to transfer foreign DNA into plant cells. Typical transgenic stable line expression studies with tissue culture practices require 3–6 months compared to about 10–15 days to study the same expression pattern with transient expression assays. Production and analysis of stable transgenic lines for a large number of promoters and genes is time consuming and expensive. In addition, some plant species are recalcitrant to transformation. TEA is rapid, efficient, and successfully used in several plant systems (Kapila et al. 1997; Wroblewski et al. 2005; Li et al. 2009). *Agrobacterium* mediated TEA is becoming the prominent choice of TEA because it is highly efficient and easy to perform. The major drawback of the existing methods is the wounding of plants, which interferes with the functional evaluation of stress genes. Here, we describe a method that completely eliminates the need for wounding of plants. Our method is reproducible, inexpensive, and requires less time and labor. Furthermore, AmTEA is highly efficient, rapid, and allows for parallel screening of numerous genes and promoters to identify potential candidate genes or promoters for downstream applications. In this study, we developed and optimized a novel non-wounding AmTEA for cereal crop plants.

2.3 Materials and Methods

2.3.1 Constructs Used in the Present Study

Five promoters and six constructs were used in this study. Gus gene in pBI121 and pCAMBIA2201 vectors was driven by 35S Cauliflower Mosaic Virus promoter (CaMV 35S) with pCAMBIA2201 harboring a catalase intron in the gus gene. CaMV 35S promoter was used to study AmTEA in all the cereal plants because it is a constitutive

plant promoter. A promoter for a rice gene encoding putative expressed protein (LOC_Os02g16690), designated as R2-273, its 5'-truncated version, a putative glutaredoxin gene (LOC_Os08g45140) promoter, and *Arabidopsis* promoter for the gene encoding DEAD-box RNA helicase family protein (At3g58510) were the other four promoters used in this study. Rice promoters were used to establish the method for rice and other cereal plants. In order to confirm that this method generates results similar to those observed in stable transgenic lines, we used *Arabidopsis* as a model system to test *Arabidopsis* DEAD-box RNA helicase family protein encoding gene promoter in both AmTEA and stable transgenic lines. These promoters were ligated to *gus-egfp* fusion reporter gene system in the vector pBGWFS7 (Figure S2.1; Karimi et al. 2002). Rice and *Arabidopsis* promoters were amplified by PCR and cloned using Gateway cloning system (Invitrogen). *Agrobacterium* strain GV3101 was used for all co-cultivation procedures.

2.3.2 Growth of Plants under Sterile Conditions

Cereal seeds were dehusked and washed thrice with sterile water. Seeds were rinsed twice with 70% alcohol and incubated with 50% commercial bleach (5.25% hypochlorite) on a shaker at 150 RPM for 15–20 min. Seeds were rinsed further with sterile water to ensure complete removal of bleach from the seeds. The seeds were then dried on sterile filter paper for 10 min and sowed in magenta boxes containing basal MS/Chu's N6 salts (Phytotechnology Lab) solidified with phytigel (2 g/L; Sigma-Aldrich). Seeds were allowed to grow until two to three leaf stages (12–15 days) and these plants were subjected to *Agrobacterium* co-cultivation procedures. All the steps were carried out in a laminar air flow cabinet.

2.3.3 Growth of Young Cereal Plants and *Arabidopsis* in the Greenhouse

Arabidopsis and cereals were grown in soil. After 10 days, the young cereal plants were uprooted, cleaned with water, and soaked in sterile 0.5 MS solution. Before co-cultivation, seed remnants and decaying leaves were removed from the young plants and washed with sterile water to avoid subsequent contamination. *Arabidopsis* plants (20–25 days old) with flowers and immature fruits were uprooted from soil and cleaned with sterile water to remove soil particles.

2.3.4 *Agrobacterium* Co-cultivation and Stress Treatments

Young cereal plants were uprooted from the media and the roots were cleaned with sterile 0.5 MS salt solution to remove the solidified medium. Overnight grown *Agrobacterium* culture (~1 OD) in LB broth (Miller modification) was used for all co-cultivation procedures. Five to ten young cereal plants were co-cultivated with

Agrobacterium (10ml overnight grown culture and 30ml LB broth) harboring the construct of interest supplemented with 100 μ M Acetosyringone (20 μ l; Sigma-Aldrich) and Silwet L-77 (10 μ l; Lehle Seeds) at 28°C for 15h at 70rpm. After co-cultivation, the plants were rinsed thrice with sterile distilled water or 0.5 MS salt solution supplemented with 500mg/L carbenicillin (Phytotechnology Lab) to prevent bacterial contamination. These plants were placed in a sterile 15-mm petri dish and incubated for 8–12h with 15–20ml of MS salt solution supplemented with carbenicillin (500mg/L) and then subjected to salt or drought stress for 5h. For salt stress, 200mM NaCl solution was added in place of MS salt solution. For drought stress, the plants were placed on a sterile filter/Whatman paper and care was taken not to dry them completely. Four controls were used to ensure that the observed expression is a direct result of the promoter activity and not a false positive result from bacterial or fungal contamination or background expression from plants. The controls include a water control, *Agrobacterium* (GV3101) cells (no plasmid), *Agrobacterium* with pBGWFS7 vector (no promoter), and *E. coli* with pBGWFS7 vector (no promoter). For water controls, the plants were cleaned and incubated in 20–40ml of MS salt solution supplemented with Silwet L77 (10 μ l) and 100 μ M Acetosyringone (20 μ l). These plants were rinsed and subjected to the same stress conditions as those of the experimental plants. The plants subjected to the stress treatments were incubated for 3–6h with a photoperiod of 16:8 (light/dark) and an optimum temperature of 22–24°C (other than temperature stress). The procedures described above for *Agrobacterium* co-cultivation and stress treatments were also used for *Arabidopsis* plants. We also studied this method by wounding the plants with gentle surface scratching of the stem, leaves, and roots with a sterile needle. The rest of the procedure was the same as described above (non-wounding method).

2.3.5 Effect of *Agrobacterium* Co-cultivation and Plant Incubation Times in MS Salt Solution on Reporter Gene Expression

We analyzed the effect of *Agrobacterium* co-cultivation and plant incubation times in MS salt solution on the efficiency of expression of reporter genes. For this process, 12 days old rice plants grown under sterile conditions were incubated with *Agrobacterium* cultures for 2, 5, 8, 11, 15, and 20h. After co-cultivation, the plants were washed and incubated in sterile MS salt solution for 12, 24, 36, and 48h followed by evaluation of GUS expression.

2.3.6 Downstream Process and Applications

Plants were observed for GFP fluorescence under a stereo fluorescent dissecting microscope (Leica MZ10F) and subjected to histochemical GUS staining as described by

Jefferson (1987). Carbenicillin (500mg/L) was added to the GUS buffer to prevent non-specific GUS expression. The entire protocol is shown as a flowchart (Figure S2.2).

2.3.7 Quantitative Expression Analysis of gus and egfp Reporter Genes

Total RNA was isolated from three plants for each promoter using the RNeasy plant mini kit (Qiagen). Double DNase treatment (Qiagen) was carried out to eliminate contaminating DNA. First strand cDNA synthesis was performed using high-capacity cDNA reverse transcription kit (Applied Biosystems). qRT-PCR was performed using Taqman probes, primers, and Taqman gene expression master mix (Applied Biosystems) for gus, egfp, and EF1 α genes as per manufacturer's instructions. Two biological and three technical replicates were analyzed for each promoter. Expression of gus and egfp genes were normalized to the rice endogenous control EF1 α (Qi et al. 2010) and to the experimental control. Taqman primers and probes were designed using the Primer Express software (Applied Biosystems). Sequences for Taqman primers and probes used for this work are listed in Table 2.1.

Table 2.1 Taqman primers and probes used for qRT-PCR

Sequence name	5' Dye	Oligo sequence	3' Dye
GUSA_F		CAAGGTGCACGGGAATATTTTCG	
GUSA_R		GAACATTACATTGACGCAGGTGATC	
EF1ALPHA_F		CCCAAGAGGCCATCAGACAA	
EF1ALPHA_R		CCGATCTTGTACACGTCCTGAAG	
GFP_F		AGAACGGCATCAAGGTGAACTT	
GFP_R		CGCTGCCGTCCTCGAT	
GUSA_M	6FAM	TCGGGTCGAGTTTACG	MGBNFQ
EF1ALPHA_M	6FAM	CCCTGCGTCTTCCC	MGBNFQ
GFP_M	6FAM	TTGTGGCGGATCTTG	MGBNFQ

2.4 Results

2.4.1 Utility of AmTEA as a Universal Method for Cereals

A novel non-wounding AmTEA was developed for rapid assessment and functional characterization of promoters and genes since wounding and non-wounding AmTEA gave very similar results (Figure S2.3). This method can be easily adapted to different plant species. Twelve days old young plants were used for this method, which minimized

both time and space requirements. Cost and labor effectiveness are other major advantages of young plantlet based AmTEA. Two promoters, a constitutive 35S and a rice promoter for glutaredoxin gene, were used to show the utility of young-plantlet-based AmTEA for the functional analysis of promoters and genes (Fig. 2.1).

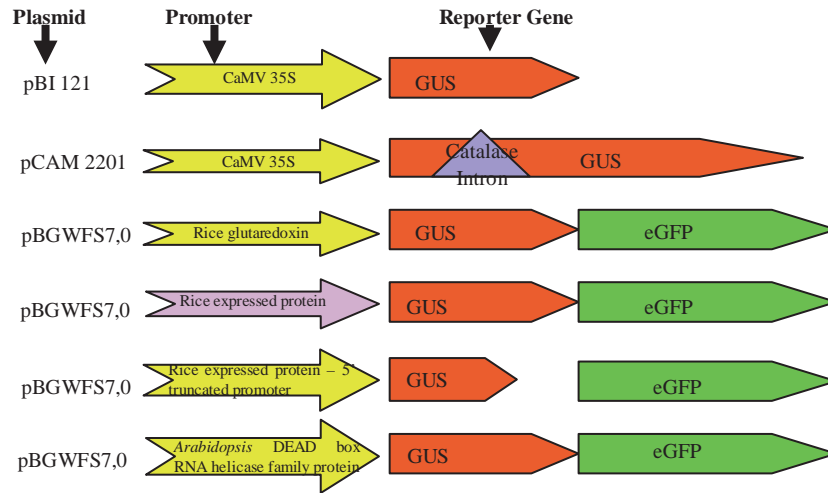


Fig. 2.1 Organization of the six promoter-reporter gene constructs used for AmTEA.

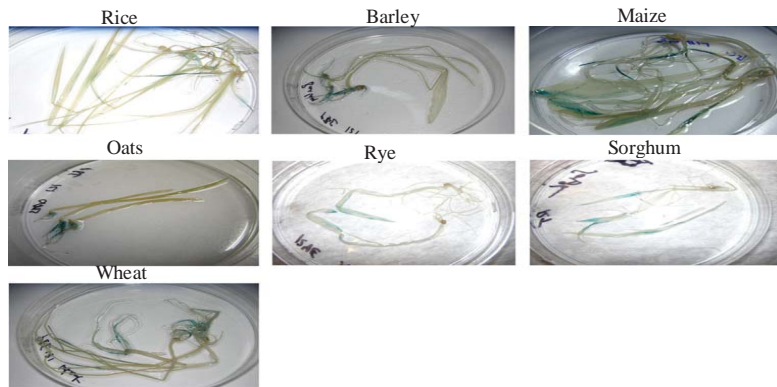


Fig. 2.2 GUS expression driven by CaMV 35S promoter in pBI121 vector in young plants of rice, barley, maize, oats, rye, sorghum, and wheat using AmTEA.

CaMV 35S promoter driven expression of gus reporter gene was consistent every time the assay was performed in cereals (rice, barley, maize, oats, rye, sorghum, and wheat) demonstrating its versatility as a constitutive promoter (Fig. 2.2). The above experiments establish the adaptability and applicability of AmTEA to different cereals. Microscopic analysis showed gus gene expression driven by 35S promoter in stomata, parenchymal, and mesenchymal cells of rice leaves (Fig. 2.3).

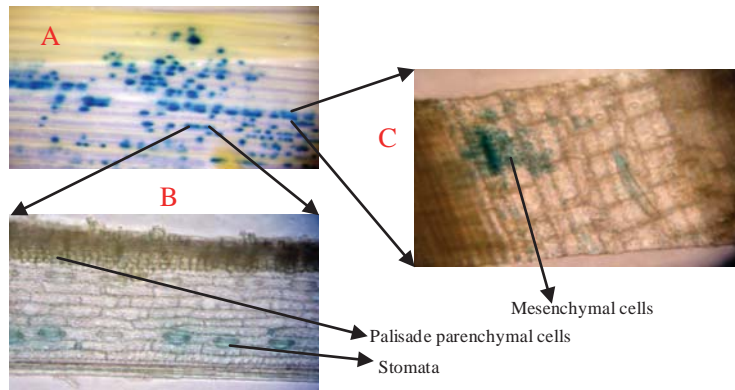


Fig. 2.3 Microscopic analysis of GUS expression driven by 35S promoter in (a) rice leaves ($\times 10$ magnification). The same sample was sectioned and viewed under $\times 100$ magnification. GUS expression was localized in (b) parenchymal cells and stomata and (c) mesenchymal cells.

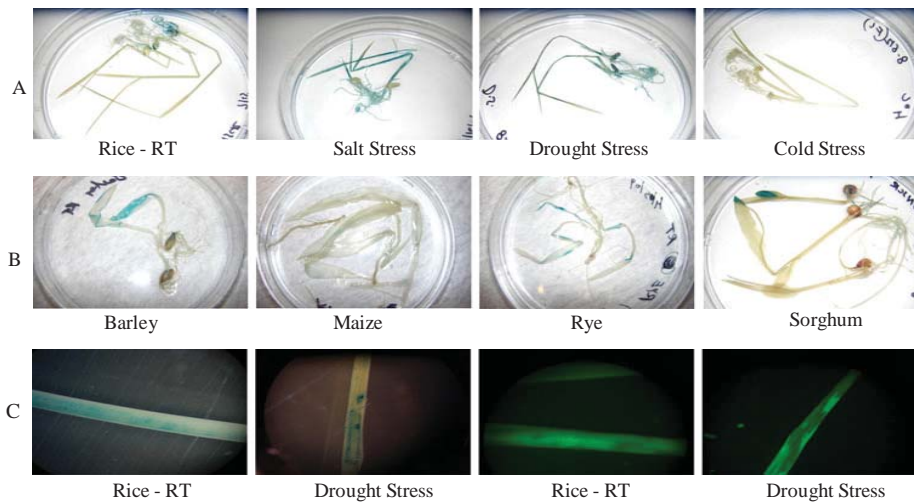


Fig. 2.4 Reporter gene expression driven by the rice glutaredoxin promoter. GUS expression in (a) rice plants at room temperature (RT), salt, drought, and cold stress, and (b) barley, maize, rye, and sorghum at RT. (c) GUS and eGFP expression in rice leaves at RT and drought stress observed under a fluorescent stereo microscope.

The putative glutaredoxin gene (LOC_Os08g45140) promoter mostly elicited constitutive expression in rice. Under salt and drought stress and to some extent at room temperature (25°C), this promoter drove high levels of reporter expression in rice plants (Fig. 2.4a). Cold stress (4°C) resulted in very little or no GUS expression. This promoter showed good expression levels at room temperature (RT) in other cereals: barley, maize, rye, and sorghum (Fig. 2.4b). The comparison of expression patterns between GUS and eGFP driven by the rice glutaredoxin gene promoter in rice leaves under RT and drought stress is shown in Fig. 2.4c. The observed expression pattern of gus gene driven by the rice glutaredoxin gene promoter was in coherence with the established functionality of

glutaredoxin (Eckardt 2007; Diao et al. 2011). Abiotic and biotic stresses induce the accumulation of reactive oxygen free radicals in plant cells which leads to increased levels of cellular glutaredoxin. High levels of glutaredoxin maintain the delicate redox homeostatic balance in plant cells (Eckardt 2007).

2.4.2 Conserved Expression Patterns of Reporter Genes in Stable and Transiently Expressing *Arabidopsis* Plants

The construct harboring *Arabidopsis* promoter for the gene encoding DEAD-box RNA helicase family protein (At3g58510) promoter was used to carry out both AmTEA and generation of stable transgenic lines in *Arabidopsis*. Reporter genes driven by this promoter showed the same expression pattern in leaves and flowers with respect to both AmTEA and stable transgenic lines under drought stress (Fig. 2.5). Expression of the reporter genes was not observed at room temperature. This demonstrates that the gene/promoter expression detected by AmTEA is in coherence with that of the stable transgenic lines. It also indicates the reduction in the time required for deciphering promoter/gene expression in plants using AmTEA.

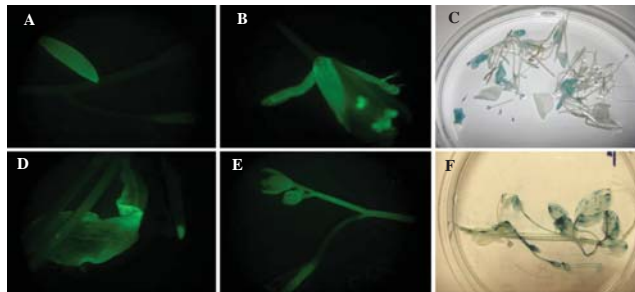


Fig. 2.5 Conserved expression pattern of reporter genes driven by *Arabidopsis* promoter for the gene encoding DEAD-box RNA helicase family protein in transiently expressing plants and stable *Arabidopsis* lines subjected to drought stress. Stable transgenic line showing expression of (a) eGFP in leaf, (b) eGFP in flower, and (c) GUS expression in the plant. The stigma and the anthers in the flower exhibit auto-fluorescence, which was also observed in the wild type. Transient expression of (d) eGFP in leaf, (e) eGFP in flower, and (f) GUS expression in the plant using AmTEA.

2.4.3 *Agrobacterium* Co-cultivation and Plant Incubation Times in MS Salt Solution Affect Reporter Gene Activity in Plants

CaMV 35S promoter and rice promoter R2-273 were used to study the effects of *Agrobacterium* co-cultivation and MS salt solution incubation times of plants on GUS expression in plants. Rice plants incubated for 2 and 5h showed several fold lower expression levels compared to those incubated for 11, 15, and 20h with *Agrobacterium* cultures (Fig. 2.6). Real-time RT-PCR analysis of *gus* expression levels confirmed the

histochemical GUS results. The second variable studied was the effect of different incubation times on reporter gene expression of *Agrobacterium*- treated young plants. Higher reporter gene expression was observed from 8 to 24h of incubation in MS salt solution and it decreased thereafter. Overall, 12 days old rice plants grown in sterile MS basal salt medium incubated for 11–20h in *Agrobacterium* cultures and 8–20h in MS salt solution were ideal for producing high expression of the transgenes.

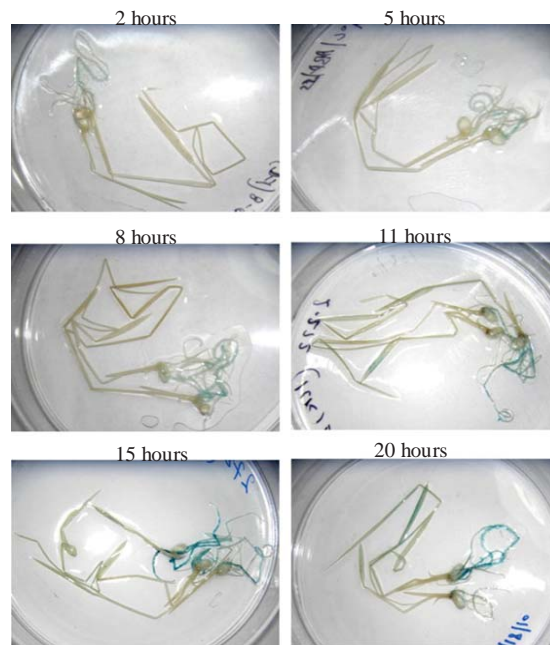


Fig. 2.6 *Agrobacterium* co-cultivation time influences GUS expression in rice plants. GUS assay performed with rice plants co-cultivated with *Agrobacterium* harboring the construct with promoter R2-273 for 2, 5, 8, 11, 15, and 20 h and incubated in MS salt solution for 12 h.

The effect of different concentrations of *Agrobacterium* cultures (overnight grown cultures, 1:1, 1:2, and 1:3 dilutions with LB broth) on *Agrobacterium* co-cultivated for 2, 5, 11, 15, and 20 h on GUS activity was evaluated. Similar results were observed in these experiments with different *Agrobacterium* concentrations.

2.4.4 AmTEA - A Useful and Reliable Tool for Deciphering Promoter and Gene Function in Plants

In order to establish the reliability of AmTEA and to prove that the observed expression is not a consequence of nonspecific bacterial expression, three independent experiments were performed. The first experiment assessed the activity of residual *Agrobacterium* cells and other bacterial contamination without the use of antibiotics. This

experiment resulted in non-specific GUS expression from residual *Agrobacterium* and endophytes (Fig. 2.7a). To eliminate this problem, three antibiotics, vancomycin, cefotaxime, and carbenicillin, were tested for their ability to inhibit *Agrobacterium* growth using antibiotic disk test method. *Agrobacterium* showed highest sensitivity to carbenicillin which is in accordance with the studies by da Silva and Fukai (2001). Treating young plants with GUS buffer supplemented with the antibiotic carbenicillin ensured elimination of non-specific GUS activity from residual *Agrobacterium* and endophytes (Fig. 2.7b). Based on these results, we used carbenicillin in all the experiments reported in this study. The second experiment investigated the expression of gus reporter gene with an intron (catalase), which cannot be processed by bacterial systems to demonstrate that AmTEA is effective in this case. Expression of gus gene driven by the 35S promoter observed in this experiment was comparable to the expression of gus gene with no intron (Fig. 2.8). For the third experiment, the 5'-truncated R2-273 promoter (339bp deletion out of 1,062bp full-length promoter) was studied for its ability to drive the expression of the reporter genes, and this resulted in drastic reduction of GUS and eGFP expression compared to the full-length R2-273 promoter (Fig. 2.9b and c). The above results were validated by qRT-PCR which identified about fivefold higher expression of gus and egfp driven by the full-length R2-273 promoter compared to its 5'-truncated version (Fig. 2.9d). This experiment verifies that the observed reporter gene expression driven by full-length R2-273 promoter with respect to the truncated R2-273 promoter is the result of a functional promoter and not due to nonspecific GUS activity. Summing up the essence of these independent experiments, AmTEA can be asserted as a promising method for deciphering gene and promoter expression patterns in cereals.



Fig. 2.7 Effect of the antibiotic carbenicillin on GUS expression in young rice plants. (A) Non-specific expression of GUS was observed when carbenicillin was not added to GUS buffer. (B) Complete suppression of non-specific GUS expression was observed on addition of carbenicillin

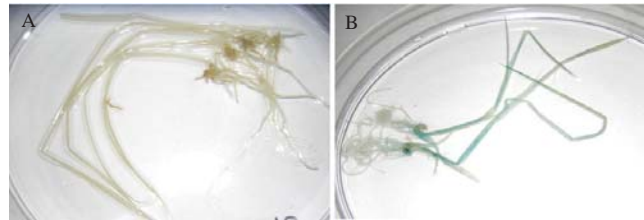


Fig. 2.8 Expression of an intron (catalase) containing *gus* gene driven by CaMV 35S promoter in pCAMBIA 2201 using AmTEA. Histochemical GUS staining (a) water control and (b) pCAMBIA 2201

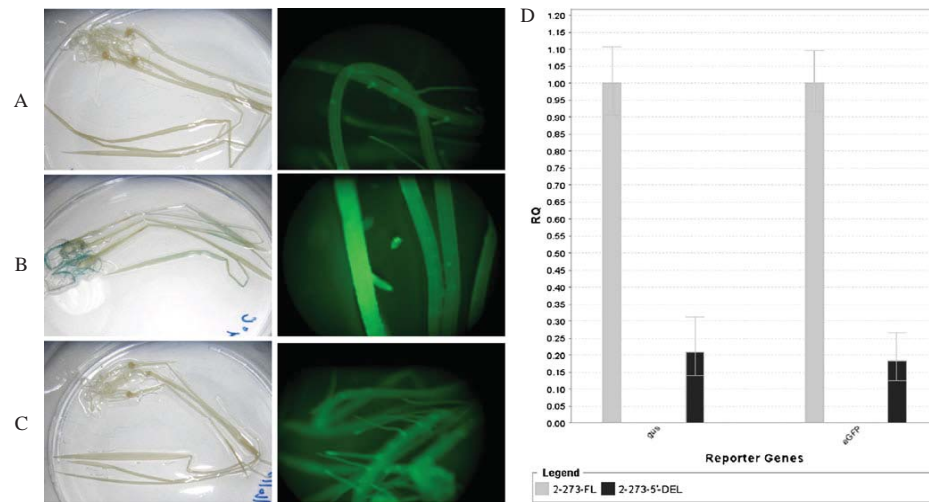


Fig. 2.9 Full-length rice promoter R2-273 show high levels of GUS and eGFP expression compared to the 5'-truncated promoter. Histochemical staining of GUS and fluorescence microscopic analysis of rice plants subjected to AmTEA showing (a) water controls, (b) full length promoter, and (c) 5'-truncated promoter. (d) Quantitative real-time RT-PCR of *gus* and *egfp* driven by the full-length promoter compared to its 5'-truncated version.

2.5 Discussion

AmTEA has been successfully employed to study transient gene expression in tobacco seedlings using vacuum infiltration (Rossi et al. 1993), plasmolysed rice embryos (Uze et al. 1997), and maize callus cultures (Amoah et al. 2001). We demonstrated the use of our novel nonwounding young-plantlet-based AmTEA for the analysis of promoters in different cereal plants. This method employed 12–15 days old intact young plants consisting of leaves, stem, and roots, which is similar to performing the experiment in a mature plant. AmTEA does not require vacuum infiltration, sonication, or wounding. Ease of use with respect to performing stress treatments is an added advantage of our method. It works equally well for flowering *Arabidopsis* plants (1-month-old plants with flowers and immature fruits). AmTEA offers the flexibility to study the expression patterns in reproductive tissues and developing fruits, which is not possible with callus or protoplasts. Furthermore, there is no need for advanced instrumentation or expensive and environmentally hazardous chemicals, demonstrating the cost-effectiveness and ecofriendliness of the method. Conservation of time, space, and resources are the major advantages of this method. AmTEA can be automated, based on the fact that it uses minimum labor and contact with young plants.

Agrobacterium co-cultivation with the young plants results in *Agrobacterium* carryover contamination, which interferes with downstream applications like GUS and qRT-PCR assays. The contamination problem can be eliminated by the use of the antibiotic carbenicillin and maintaining sterile conditions throughout the experiment and cleaning the plants with carbenicillin-supplemented sterile water. Previous transient expression assays did not mention the use of bactericidal or bacteriostatic agents to prevent nonspecific expression of reporter genes from residual *Agrobacterium* or endophytes (Kapila et al. 1997; Wroblewski et al. 2005; Li et al. 2009). Assessing the water controls for background GUS activity from plants due to endophytes is another key factor. Multiple repetitions of the experiments and proper interpretation of the results with respect to water controls is the key to success with AmTEA.

Based on our findings, the young-plantlet-based *Agrobacterium* mediated transient expression assay is a useful tool for functional analysis of promoters and genes. AmTEA can be used to study the role of cis-regulatory elements in abiotic (Bi et al. 2011; Wang et al. 2010) and biotic stresses (Meng et al. 2010), gene regulatory elements conserved in multiple plant species (Ann and Meagher 2010), characterize enhancers and insulators (Singer et al. 2010; Saha et al. 2011), and tissue-specific promoters (Kato et al. 2010). Furthermore, gene inter-relationships or “gene webs” can be studied under different stress treatments with the aid of microarrays. AmTEA can be easily adapted to evaluate

promoters and genes in dicots other than *Arabidopsis*, including plants with fruits such as citrus (Liu et al. 2011) and tomato. The small size and tender nature of the young plants offer the possibility of automation of this method for high throughput screening of promoters and genes.

2.6 Acknowledgment

We thank USDA-GRIN for providing the seeds of rice, barley, maize, Sorghum, rye, and oats. This work was supported by the National Research Initiative of the USDA Cooperative State Research, Education and Extension Service, grant number 2007-35301-18036. We thank the Michigan Technological University's Biotech Research Centre (BRC) for their continuous support and funding. We personally thank Lorie Bernhardt (Dale Bumpers National Rice Research Center) for providing bulk quantities of rice seeds. We thank Dr. Chandrashekar P. Joshi (Michigan Technological University) for the plasmid pBI121 and Dr. Victor Busov for *Agrobacterium* (GV3101). We also thank the Institute of Plant Systems Biology, Ghent University, Belgium, for pBGWFS7 and CAMBIA, Australia, for pCAMBIA2201.

2.8 Supplementary information

Figure S2.1 Organization of the vector pBGWFS7 (this figure can be found at the web site: <http://gateway.psb.ugent.be/vector/show/pBGWFS7/search/index>).

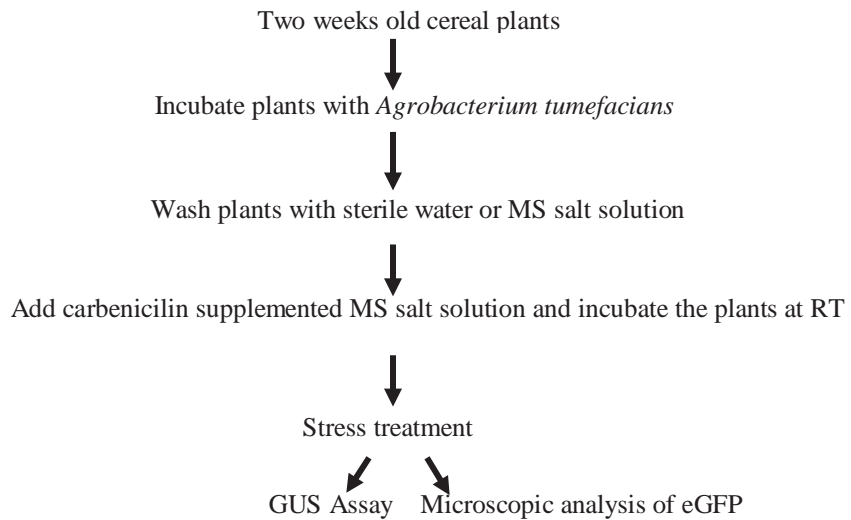


Figure S2.2 Flow chart enumerating important steps in AmTEA.

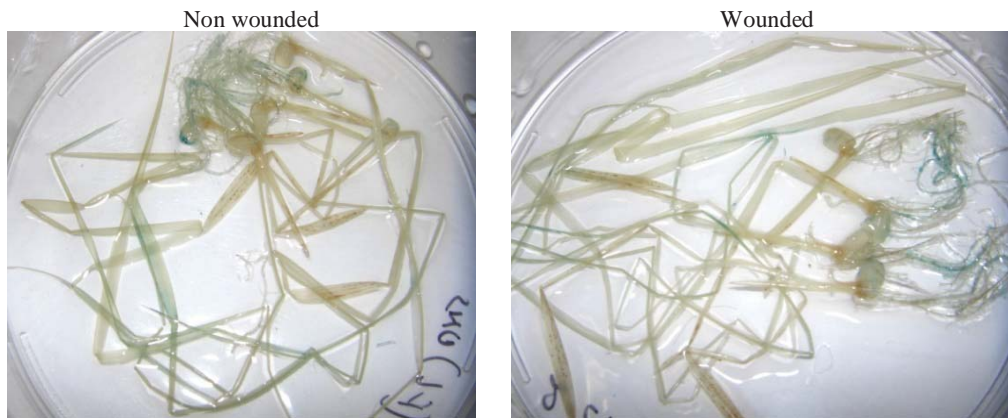


Figure S2.3 Evaluation of wounding and non-wounding AmTEA with R2-273 promoter using rice plants.

Table S2.1 Overrepresented motifs in rice, *Arabidopsis*, and *Populus* bidirectional promoters of <1 kb from PLACE

All promoters (2106) <i>Arabidopsis</i>				
Motif	Sequence	BD	Random	Z
CGACGOSAMY3	CGACG	1007	412	6.17
CGCGBOXAT	VCGCGB	470	196	3.14
CTRMCAMV35S	TCTCTCTCT	499	130	14.59
E2FCONSENSUS	WTTSSCSS	677	273	4.99
HEXAMERATH4	CCGTCG	331	105	7.81
SITEIIATCYTC	TGGGCY	1125	419	10.71
SORLIP2AT	GGGCC	960	341	11.12
TELOBOXATEEF1AA1	AAACCCTAA	443	92	18.79
UP1ATMSD	GGCCCAWWW	620	153	18.03
UP2ATMSD	AAACCCTA	655	149	20.88

Table S2.1 Overrepresented motifs in rice, *Arabidopsis*, and *Populus* bidirectional promoters of <1 kb from PLACE. (continued from above)

All promoters (1235) rice				
Motif	Sequence	BD	Random	Z
ABREATCONSENSUS	YACGTGGC	146	88	3.75
ABREOSRAB21	ACGTSSSC	230	142	4.45
BOXCPSAS1	CTCCCAC	184	109	4.51
BOXIIPCCHS	ACGTGGC	254	148	5.71
BS1EGCCR	AGCGGG	313	185	6.19
CBFHV	RYCGAC	758	551	4.43
CGACGOSAMY3	CGACG	939	601	11.43
CGCGBOXAT	VCGCGB	883	594	8.66
CTRMCAMV35S	TCTCTCTCT	232	150	3.73
DRECRTCOREAT	RCCGAC	567	389	5.05
E2FCONSENSUS	WTTSSCSS	619	424	5.49
GCCCORE	GCCGCC	863	401	21.35
HEXAMERATH4	CCGTCG	528	297	10.04
HEXAT	TGACGTGG	129	60	6.58
LTRECOREATCOR15	CCGAC	782	542	6.43
PRECONSCRHSP70A	SCGAYNRNNNNNNNNNNNNNNNNHHD	855	637	4.04
SITEIIATCYTC	TGGGCY	871	578	9.06
SORLIP2AT	GGGCC	914	629	8.08
UP1ATMSD	GGCCCAWWW	302	122	13.16
UP2ATMSD	AAACCCTA	249	87	14.29
UPRMOTIFIAT	CCACGTCA	129	60	6.58
UPRMOTIFIAT	CCNNNNNNNNNNNNCCACG	224	124	6.12

All promoters (613) *Populus*

Motif	Sequence	BD	Random	Z
GCCCORE	GCCGCC	77	82	3.94
SORLIP2AT	GGGCC	237	320	3.54
UP1ATMSD	GGCCCAWWW	58	47	5.57

All promoters represents promoters between divergent genes separated by <1 kb. Number in parenthesis indicates number of promoters analyzed. BD represents the number of putative bidirectional promoters (<1 kb) with the specific cis-element. Random represents the number of random promoters out of 1000 that have the specific cis-element. Overrepresented motifs present in >10% of promoters in each set are shown. Z values were estimated for each cis-element by comparing the number of putative bidirectional promoters with that of 1000 random promoters with the specific element. Cut-off p-value used for Z was < 0.001.

**CHAPTER 3:
REGULATION OF BIDIRECTIONAL PROMOTER ACTIVITY IN
RICE BY CIS-ELEMENTS IN 5'-UNTRANSLATED REGIONS**

Surendar Reddy Dhadi, Aparna Deshpande and Wusirika Ramakrishna

This work was supported by the National Research Initiative of the USDA Cooperative State Research, Education and Extension Service, grant number 2007-35301-18036.

3.1 Abstract

Bidirectional promoters regulate adjacent genes organized in a divergent fashion (head to head orientation). Deletion analysis using promoter-reporter gene constructs identified three rice promoters to be bidirectional. Regulatory elements located in the 5'-untranslated regions (UTR) of one of the genes (divergent gene pair) were found to be responsible for their bidirectional activity. DNA footprinting analysis identified unique protein binding sites on three bidirectional promoters. The first two motifs were positioned between 33 to 46bp upstream and the third motif was located 78-91bp upstream the transcription start site (TSS) of one of the divergent genes. Deletion/alteration of any one of these motifs resulted in 40-55% loss of expression of the reporter genes on either side of the promoter. Changes in the motifs at both the positions resulted in decrease of bidirectional activity by 85%, which was quantified using real-time RT-PCR of the two reporter genes flanking the promoter. Based on our results, we propose a novel mechanism for the bidirectionality of bidirectional promoters in rice.

3.2 Introduction

Bioinformatic and functional analysis of bidirectional promoters (BDP) is an actively pursued area of research both in plant and animal genomes. Divergent promoters were reported for the first time in lower classes of organisms like bacteria and viruses (Guha et al. 1971; Taylor et al. 1967). Divergent promoters have been extensively studied in the human genome (Trinklein et al. 2004). A few of these promoters were reported in plants (Keddie et al. 1994; Sadanandom et al. 1996). Non-availability of plant genome sequences was the primary hindrance towards genome-wide study of divergent promoters. Recently, with the availability of genome sequence of several plant genomes and expression data, several researcher groups investigated the organization and evolutionary trends of BDPs on a whole genome scale. We performed bioinformatic analysis of putative BDPs and divergent gene pairs in three plant genomes: rice, *Arabidopsis* and *Populus* (Dhadi et al. 2009; Krom and Ramakrishna 2008). This study was expanded to include maize, sorghum and *Brachypodium* (Krom and Ramakrishna 2010).

Unlike bacterial and animal model systems which have a fast generation time and it is relatively easy to generate transgenics, plants do not have these advantages. The primary challenge for a plant biologist is 1) to obtain quick expression data after the insertion of transgene, 2) reproducibility of the data and 3) offspring generation to study the phenotypic and genotypic effects of the transgene. We developed an *Agrobacterium* mediated Transient Expression Assay (AmTEA) for cereal plants, which facilitates rapid assay of reporter genes driven by promoters (Dhadi et al. 2012). On the other hand, cloning vector systems to study the expression pattern of BDPs are not available. Most researchers used a single reporter gene to analyze the expression pattern of BDPs by cloning it twice into the vector system in forward and reverse orientations (Trinklein et al. 2004). Further, identifying the *cis*-regulatory elements and the proteins that interact with BDPs is another major challenge. The traditional route of using electrophoretic mobility shift assay (EMSA) for DNA-protein binding interactions is cumbersome and will not identify the specific DNA sequence to which a protein binds. Chromatin immunoprecipitation (ChIP) assay is a technique that can identify protein DNA binding sites (Winter et al. 2011). Protein binding sites on DNA can be identified with a non-radioactive method using a capillary genetic analyzer (Yindeeyoungyeon and Schell 2000).

Bidirectional promoters play a crucial role in regulating the expression of divergent genes to maintain a stoichiometric relationship between gene products (Albig et al. 1997; Ahn and Gruen 1999) and co-expression of the genes functioning in common metabolic

pathways (Momota et al. 1998). BDPs regulate genes involved in diverse cellular processes including cell cycle (Guarguaglini et al. 1997). They govern a myriad of abiotic (environmental) and biotic (plant pathogen) responses (Taylor et al. 1967; Shin et al. 2003).

Our previous work identified 212 BDPs in the rice genome, out of which 52 were highly correlated and the intergenic distance between the two divergent genes was <250bp. In order to identify all potential BDPs, a criteria of <1kb intergenic distance between the divergent genes was employed and a total of 1242 BDPs were identified (Dhadi et al. 2009). A few bidirectional promoters were selected from both <250bp and <1kb classes of BDPs and were evaluated for bidirectional characteristics. Fragment deletion studies were performed to establish the bidirectional character of the BDPs. DNA foot-printing using a genetic analyzer identified *cis* regulatory motifs which were deleted and changes in expression patterns of the reporter genes revealed new insights into the functional mechanisms of divergent gene regulation by BDPs.

3.3 Methods

3.3.1 Construction of a vector with two reporter genes for testing BDPs

pBGWFS7 vector harbors the *egfp-gus* fusion reporter gene cassette in the forward orientation downstream of the attachment R2 site. In order to generate a vector suitable for the BDP expression analysis, the red fluorescent reporter gene, *rfp* from the pB7RWG2 vector was ligated upstream of the attachment R1 site on the pBGWFS7 vector in the reverse orientation to form a novel promoter expression vector, pBGWFS7-RFP consisting of two reporter genes that flank the inserted promoter on either sides.

3.3.2 Gateway cloning of BDP-reporter gene constructs

Primers harboring Gateway attachment sites were designed for the bidirectional promoter sequences and care was taken not to include the endogenous genes AUG translation start sites on both sides of the BDP. The primers used for this work are listed in supplementary table 1. Selected BDPs were PCR amplified and directionally cloned employing Gateway Technology (Invitrogen) into a modified Gateway cloning vector designated as pBGWFS7-RFP (Supplemental Fig. 3.1), generating an expression vector harboring a BDP between the two reporter genes.

3.3.3 Plant transient expression assay

AmTEA (Dhadi et al. 2012) was used for expression analysis of BDPs in rice plants. Rice plants grown for 10-15 days under sterile conditions were co-cultivated for 17-20 hrs with *Agrobacterium* culture harboring the insert of interest at 28° C and 70 RPM on a shaker incubator. An overnight grown *Agrobacterium* culture (~ 1 OD) (5-10mL) mixed with 30mL of fresh LB broth supplemented with 100 µM acetosyringone (20µL; Sigma-Aldrich) and Silwet L-77 (10µL; Lehle Seeds) was used for the above co-cultivation process. After the co-cultivation, rice plants were washed thoroughly with sterile 0.5 MS salt solution and incubated for 6-8hrs with 15-20mL of sterile 0.5 MS salt solution supplemented with carbenicillin to eliminate residual *Agrobacterium* contamination.

3.3.4 Deletion analysis

Based on the occurrence of over-represented *cis*-regulatory elements (Dhadi et al. 2009), each putative BDP was sequentially deleted and consequently 3 deletion promoter-reporter constructs were generated for each BDP (Supplemental Fig. 3.2). This was accomplished by designing primers at the appropriate junction points. List of primers used for this study are shown in supplemental table 3.1. Each fragment was PCR amplified and cloned into the destination vector (pBGWFS7-RFP) using Gateway Technology (Invitrogen).

3.3.5 Extraction of rice nuclear proteins

Rice nuclear proteins were extracted according to Escobar et al. (2001). 10-15 day old rice plants were homogenized using homogenization buffer and filtered through two layers of Miracloth. The filtrate was centrifuged and the pellet was suspended in a nuclear isolation buffer consisting of percoll and re-pelleted by centrifugation. The enriched nuclei pellet was disrupted and the nuclear proteins were precipitated by adding ammonium sulphate. Nuclear proteins were harvested by centrifugation and resuspended in 50µL reconstitution buffer for use with DNA interactions. The reconstitution buffer has 1:1 mixture of 1x PBS and 30% glycerol with 2µL 1M HEPES buffer added to the PBS/glycerol mixture.

3.3.6 DNA footprinting assay

Primers were designed corresponding to the sequence sites on the vector close to the insert with only one of the 5' end labeled either with 6-carboxyfluorescein (6FAM) or Hexachloro-fluorescein (HEX) fluorescent label. The primers used for DNA footprinting

assay are listed in Supplementary Table 3.2. DNA-protein interactions were studied on an ABI 310 genetic Analyzer as per the method devised by Yindeeyoungyeon and Schell (2000). Labeled DNA fragments were obtained by PCR using the labeled and unlabeled primer combination and purified using the Promega Wizard® SV Gel and PCR Clean-Up kit. Each BDP was analyzed with various nuclear protein concentrations ranging from 30ng-200ng to determine the optimal concentration of the protein required to produce a significant difference in peak intensities with respect to the control. 40ng of labeled DNA was treated with a series of rice nuclear protein concentrations viz 30, 60, 90, 120,160 and 200 ng in each of the 10µL reaction. DNA-protein reactions were equilibrated at 28⁰C for 1hr. Each separate reaction was treated with a diluted and optimized DNase I solution for 3min. DNase I was inactivated by heating the sample at 90⁰ C for 2min and chilled on ice for 10min. 3µL of the DNA-protein reaction was diluted with 20ul of Hi-Di™ Formamide and 0.2ul of the ROX size standard (Applied Biosystems). Each reaction was loaded on the ABI 310 Genetic Analyzer and was separated with the GS_POP4_D_1ml run module using the default parameters. Matrix standards and a size standard (GeneScan®-500-ROX (500-ROX) molecular weight standard) were also used. Controls include samples with only labeled DNA with no protein and no DNase I treatment, labeled DNA with DNase I treatment, and labeled DNA with 200ng of BSA and treated with DNase I. Addition of BSA to the samples was optional and did not affect the overall efficacy of the method. The raw data was analyzed with Gene Mapper software using default parameters. The protein binding bases were identified by comparing the data with the sequencing data.

3.3.7 Motif analysis

The three promoter sequences were analyzed using publicly available databases: Database of Plant Cis-acting Regulatory DNA Elements (PLACE) (<http://www.dna.affrc.go.jp/PLACE/>), Plant Promoter Analysis Navigator (Plant PAN) (<http://plantpan.mbc.nctu.edu.tw/index.php>) and the TRANSFAC database (<http://www.gene-regulation.com/pub/databases.html>).

3.3.8 Site-directed mutagenesis of the BDP constructs

QuikChange II XL Site-Directed Mutagenesis (Agilent Technologies) and the Change-IT multiple Mutation site directed mutagenesis kit (Affymetrix/USB) were employed to generate constructs with the required mutations that were identified by the DNA-protein footprinting reactions. Primers (Supplementary Table 3.3) were designed using the QuikChange® Primer Design Program (Agilent Technologies). According to the manufacturer's guidelines, mutant plasmids in the respective positions were generated

by using mutagenic primers employing a thermal cycler. Parental methylated and hemimethylated strands were digested with *DpnI* restriction enzyme. Mutated plasmid molecules were transformed into competent cells for nick repair and propagation. Mutations at respective positions were confirmed with sequencing.

3.3.9 Qualitative and quantitative expression analysis of reporter genes

AmTEA treated plants were observed for eGFP and mRFP fluorescent reporter proteins under a stereo fluorescent dissecting microscope (Leica MZ10F). Histochemical GUS staining was performed as described by Jefferson (1987). Quantitative expression analysis of the reporter genes was performed by qPCR using Taqman chemistry and EF1 α gene was used as an endogenous control. Taqman primers and probes used for this work are listed in the Supplementary Table 3.4 (Dhadi et al. 2012).

3.4 Results

3.4.1 Structure of the divergent genes harboring the divergent promoters (BDP) chosen for functional evaluation

Seven putative BDPs (Table 1) from rice were used for this study based on their length, expression data, presence of full-length cDNAs and conservation/ non-conservation of the flanking divergent genes described in Krom and Ramakrishna (2008) and Dhadi et al. (2009).

Promoter 2-273 is localized between the divergent genes bZIP transcription factor containing domain gene (LOC_Os02g16680) and an unknown expressed protein gene (LOC_Os02g16690). The bZIP transcription factor is a basic leucine-zipper class protein is conserved in grasses, maize, *Brachypodium* and sorghum. The second gene is absent in other grasses and is unique to rice. The full length cDNA corresponding to these two genes are AK070674 and AK120856.

Promoter 8-612 is localized between the divergent genes, SRA-YDG domain containing protein gene (LOC_Os08g45130) and a glutaredoxin coding gene (LOC_Os08g45140). The full-length cDNAs corresponding to these two genes are AK067640 and AK062431. These divergent gene proteins were conserved in sorghum, barley and *Brachypodium*.

Promoter 3-343 directs the expression of the divergent genes, a seed maturation protein PM36 gene (LOC_Os03g19390) and an unknown expressed protein gene (LOC_Os03g19400). The full-length cDNAs corresponding to these two genes are

AK073626 and AK243201. Proteins of both the genes were conserved in sorghum, barley, maize and *Brachypodium*.

Promoter 4-431 drives the divergent genes, a putative pentatricopeptide gene (prenyl transferase domain containing protein, LOC_Os04g41140) and a hypothetical expressed protein gene (LOC_Os04g41150). The full-length cDNAs corresponding to these two genes are AK069116 and AK073488. Protein of the first gene is conserved in *Brachypodium*, sorghum and barley, whereas the second gene protein was conserved in *Brachypodium*, maize, and barley.

Promoter 6-5 regulates the expression of the divergent genes, an unknown expressed protein gene (LOC_Os06g01460) and a terpenoid synthase domain containing protein gene (LOC_Os06g01470). The full-length cDNAs corresponding to these two genes are AK068490 and AK109910. Proteins of both the genes were conserved in sorghum, maize, and *Brachypodium*.

Promoter 6-8 regulates the expression of the divergent genes, a Myb family transcription factor protein gene (LOC_Os06g01670) and an unknown expressed protein gene (LOC_Os06 g01680,). The full-length cDNAs corresponding to these two genes are AK119640 and AK072638. This divergent gene pair and the proteins coded by them are conserved in maize, sorghum, *Brachypodium* and barley.

Promoter 1-914 governs the expression of the divergent genes, an unknown expressed protein gene (LOC_Os01g59080) and a thylakoid lumenal 20kDa protein gene (LOC_Os01g59090). The full-length cDNAs corresponding to these two genes are AK066239 and AK065795. Proteins coded by the above divergent gene pair is conserved in maize, sorghum, and *Brachypodium*.

Gene order and orientation of divergent genes driven by 6-5, 4-431 and 1-914 were conserved in sorghum, *Brachypodium* and maize, those driven by 3-343 in sorghum and *Brachypodium* and 8-612 in sorghum and maize respectively. Primarily this paper will focus on three BDPs: 2-273, 4-431 and 6-5.

3.4.2 Deletion constructs identify bidirectional promoters

The putative BDPs described above from rice were used to test for bidirectional activity employing AmTEA. 2-273 (1062bp) is the only large promoter that was selected for experimental validation. The remaining promoters were <500bp. Six of these promoters expressed the reporter genes in both directions while 1-914 promoter was

unidirectional expressing the eGFP-GUS fusion reporter gene. Expression of reporter genes in forward and reverse orientations by the inserted promoter does not necessarily constitute a bidirectional character. In order to identify promoter regions involved in bidirectional activity, a series of deletion constructs were generated for each of the promoter deleting the proximal, middle, and distal regions (Supplemental Fig. 2). Expression analysis of reporter genes of deletion constructs was used to identify the bidirectional activity of these promoters. Of the six promoters, one of the deletion constructs of 2-273, 4-431 and 6-5 lost the ability to express reporter genes in both directions.

The 5' end deletion construct of 2-273 promoter resulted in the total loss of qualitative expression of the reporter genes, while quantitative RT-PCR analysis showed about 50% loss of BDP activity (Fig. 3.1). This suggests that the 5' end of the promoter is essential for BDP activity and may harbor crucial *cis*-regulatory elements. On the contrary, 3' end deletion resulted in 40-50% higher activity compared to the full length promoter, which suggests the presence of promoter suppressor elements/insulators in this region. Figure 3.1 shows the expression of reporter proteins and co-localization of eGFP and RFP in rice plants with full-length and deletion constructs of 2-273 BDP.

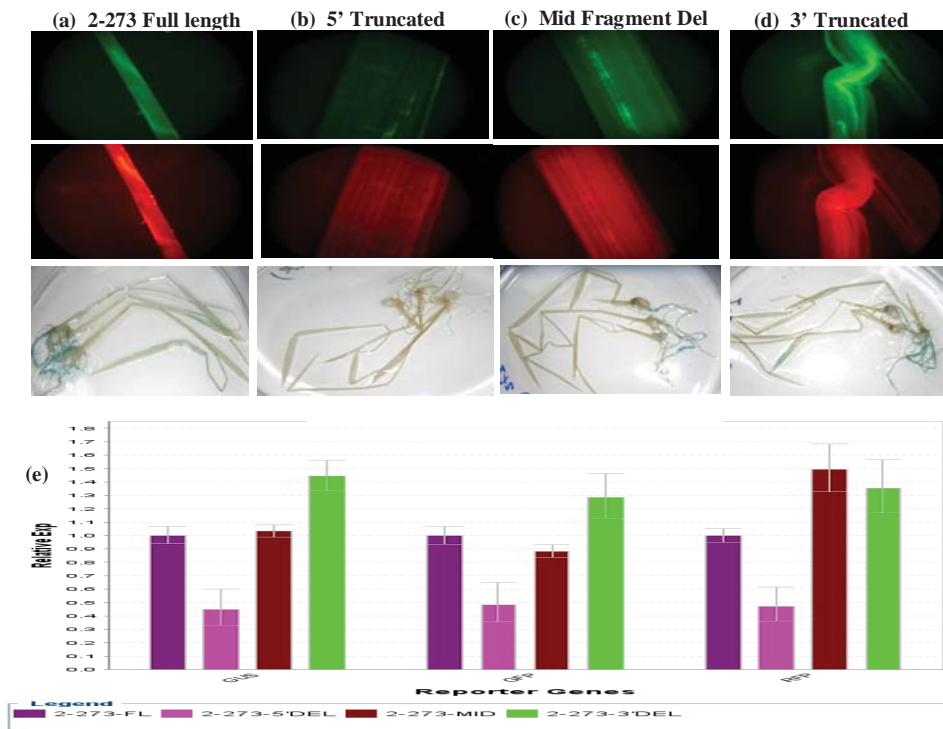


Fig. 3.1 eGFP-GUS and RFP expression pattern driven by the BDP 2-273 and its truncated versions in rice. Reporter gene expression of (a) full length, (b) 5' truncated, (c) mid fragment and (d) 3' truncated constructs. (e) Quantitative real time RT-PCR analysis of *gus*, *egfp* and *rfp* reporter genes driven by the full length 2-273 BDP compared to its truncated versions. (RQ: Relative Quantification)

Promoter 4-431 was verified as a BDP based on the 3' deletion construct which showed a significant loss in promoter activity in both directions (Fig. 3.2). Quantitative analysis revealed a loss of about 40% promoter activity based on the expression of the reporter genes (Fig. 3.2). 5' deletion construct enhanced the BDP activity by 70-90% in the *gus-egfp* reporter direction and by 50% in the *rfp* direction compared to the full-length promoter.

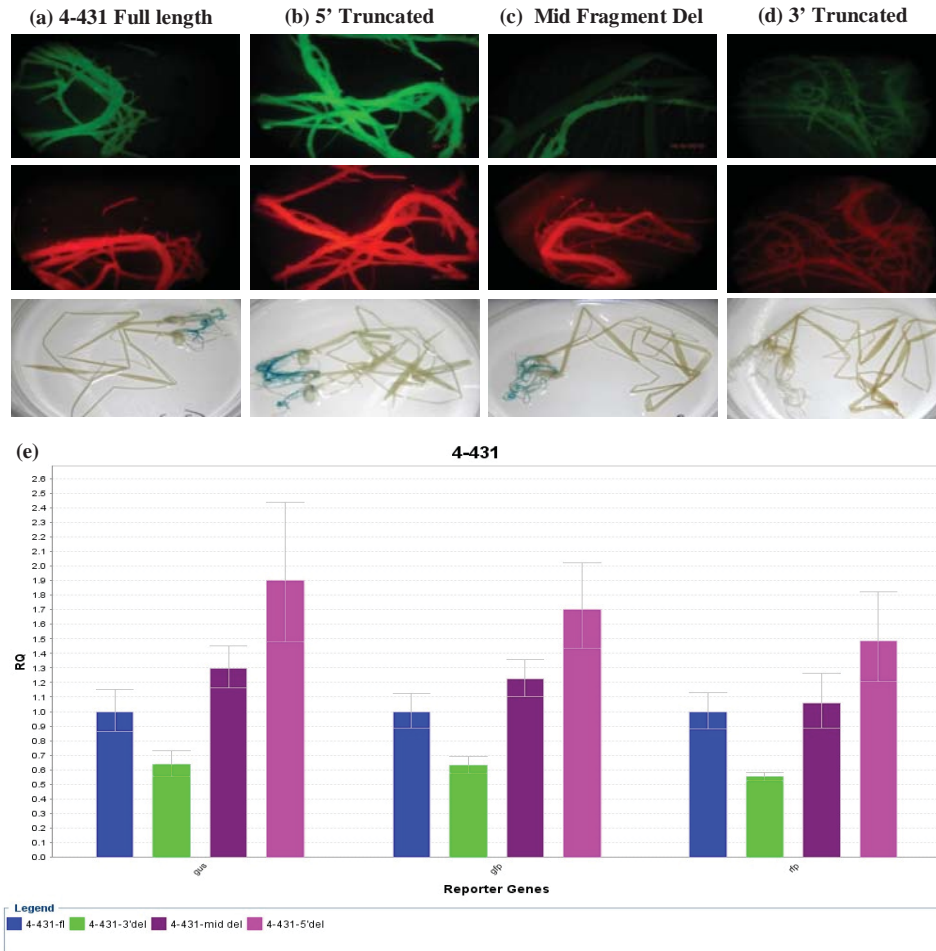


Fig. 3.2 eGFP-GUS and RFP expression pattern driven by the BDP 4-431 and its truncated versions in rice. (a-e) are same as described in figure 3.1.

6-5 promoter was established as a BDP due to significant decrease in the promoter functionality in its 3' deletion construct. Quantitative expression studies using q-PCR indicated a loss of about 85% and 60% loss in expression of *gus-egfp* and *rfp*, respectively (Fig. 3.3). 5' deletion construct showed similar q-PCR expression patterns of

the reporter genes as the 3' deletion construct. However, qualitative analysis of the 5' deletion construct showed higher expression of RFP and eGFP compared to 3' deletion construct. Therefore, we selected the 3' deletion construct for further characterization of BDP functionality.

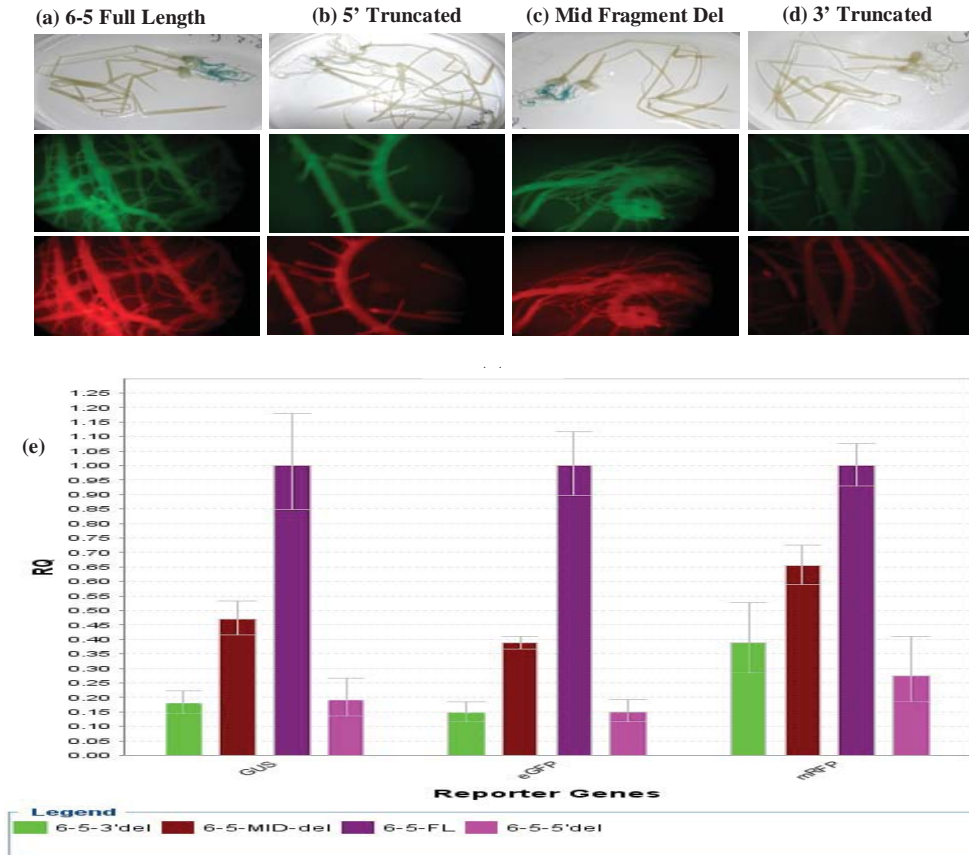


Fig. 3.3 eGFP-GUS and RFP expression pattern driven by the BDP 6-5 and its truncated versions in rice. (a-e) are same as described in figure 3.1.

3.4.3 *Cis*-regulatory motifs: identification, deletion and expression analysis

Cis-regulatory motifs bound by transcription factors and other proteins were identified by DNA footprinting analysis using genetic analyzer, which is fast, reliable and avoids the use of radioactivity. The three BDPs, 2-273, 4-431 and 6-5 gave maximal differences in peak intensities when compared to the control when incubated with 80, 60 and 90ng nuclear proteins, respectively. Full length promoters exhibited a large number of protein binding sites. Therefore, deleted fragments that were essential for regulating bidirectional expression patterns were analyzed. Rice nuclear protein binding DNA bases

in BDPs were deleted by designing deletion primers harboring the deletion of these bases. Bases to the right and left of the identified protein binding site were also included to ensure the deletion of supporting bases. This was done in accordance with the Gene Mapper software data, which identified the protein binding sites in fractions. For instance, for base number 83.6, we deleted the bases 82, 83, 84 and 85. Site-directed mutagenesis was used to generate 4, 2 and 1 *cis*-regulatory deletion constructs for 6-5, 4-431 and 2-273 BDPs, respectively. Our goal was to delete all the 5 marked motifs in case of 6-5 and 4-431 BDP and four motifs for 2-273 BDP. In spite of using two different methods to generate mutations, only partial deletions for some of the expected motifs was accomplished owing to the GC-rich nature of these BDPs (Fig. 3.4c, 3.6c, and 3.8c). The effect of these mutations on bidirectional promoter activity was tested using the above deletion constructs with AmTEA and the expression patterns of the reporter genes were analyzed both qualitatively as well as quantitatively.

Expression analysis of the *Cis* motif deletion constructs gave insights into the BDP functional mechanism. Analyses of the expression patterns combined with DNA footprint assays of 6-5, 4-431 and 2-273 BDPs identified that the initial two 4 base pair motifs were essential in coordinating the BDP functionality. These two 4bp motifs were found approximately between 33-48bp upstream of the 5' UTR of one of the divergent genes coordinated by the BDPs. These data identified pivotal regulatory motifs positioned between bases 33-40, 41-46, and 78-91 that regulate BDP functional activity. The location but not the motif itself was conserved in all the three BDPs. Detailed analyses for the three BDPs are given below:

2-273

For promoter 2-273, we were able to generate only one *cis* deletion construct (2-273-del-Y). In this construct bases in the third motif located from 88-91 bases were deleted (Fig. 3.4). Qualitative expression analysis shows highly reduced expression of the reporter genes in 2-273-del-Y construct compared to the full length 2-273 BDP (Fig. 3.5). Quantitative RT-PCR data shows a loss of 45-55% of promoter activity in both the reporter genes (Fig. 3.5C) in comparison to the full length BDP.

4-431

Two *cis* deletion constructs were generated for 4-431 BDP. In 4-431-del-G construct, base 'A' was inserted immediately after the first motif (located between 33-36 bases) and does not show any base changes with respect to the motifs located between 39-42 or 72-74 bases. Apart from this both the *cis* deletion constructs harbor common and unique base deletions in the three motifs located from 72-74, 78-81 and 83-87 bases (Fig. 3.6). Qualitative analysis of the *cis* deletion constructs of the 4-431BDP revealed very high loss of expression in both the reporter genes with respect to the 4-431-del-G construct

(Fig. 3.7c). Reduced expression of the reporter genes can be observed in case of the 4-431-del-E construct when compared to the full length BDP (Fig. 3.7b). Quantitative qRT-PCR data indicates a near total loss of expression (80-85%) in the 4-431-del-G compared to 25-45% loss of reporter gene expression with respect to the 4-431-del-E construct compared to the full length promoter (Fig. 3.7d).

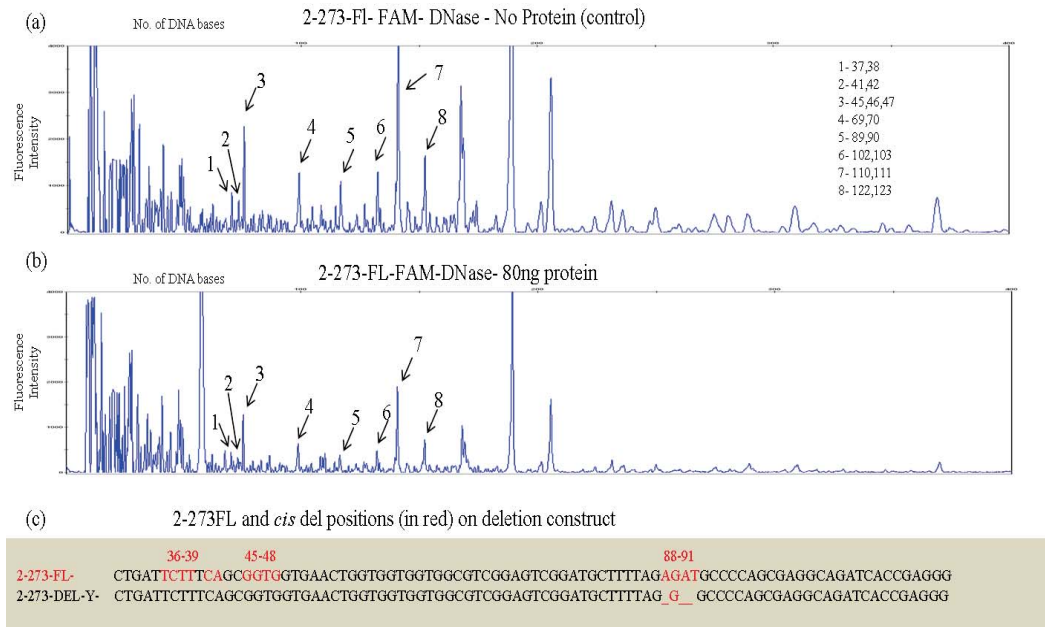


Fig. 3.4 *Cis*-regulatory motifs in BDP 2-273 bound to nuclear proteins identified by DNA foot-printing. The DNA was 5' end labeled with 6-FAM. Arrows indicate the differences in peak intensities in (a) control DNA when no protein was bound and (b) the 2-273 DNA-protein footprint at same positions where the nuclear protein was bound. Numerals indicate the base numbers corresponding to the peaks. Protein bound bases were marked only in the fragment corresponding to 5' truncated construct which showed considerable loss of BDP functional activity. (c) Protein binding sites in the BDP are marked in red with deleted bases shown in the deletion construct Y. Protein binding motifs are marked exclusively in the fragments that exhibited considerable loss of BDP functional activity.

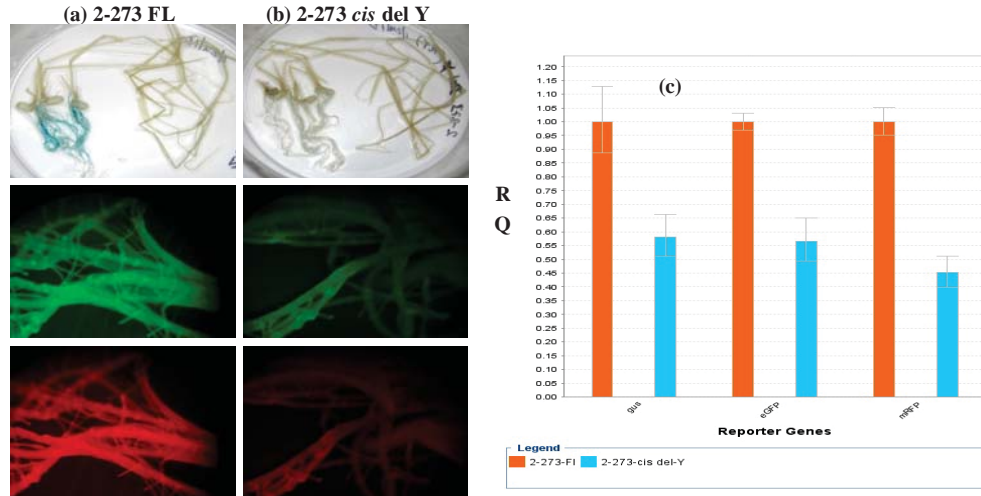


Fig. 3.5 eGFP-GUS and RFP expression pattern driven by the BDP 2-273 and its deletion construct Y in rice. Higher expression of reporter genes was observed in (a) full length BDP and lower expression was observed in (b) deletion construct Y. (c) Quantitative real time RT-PCR analysis of *gus*, *egfp* and *rfp* reporter genes driven by the full length 2-273 BDP compared to the deletion construct Y. (RQ: Relative Quantification)

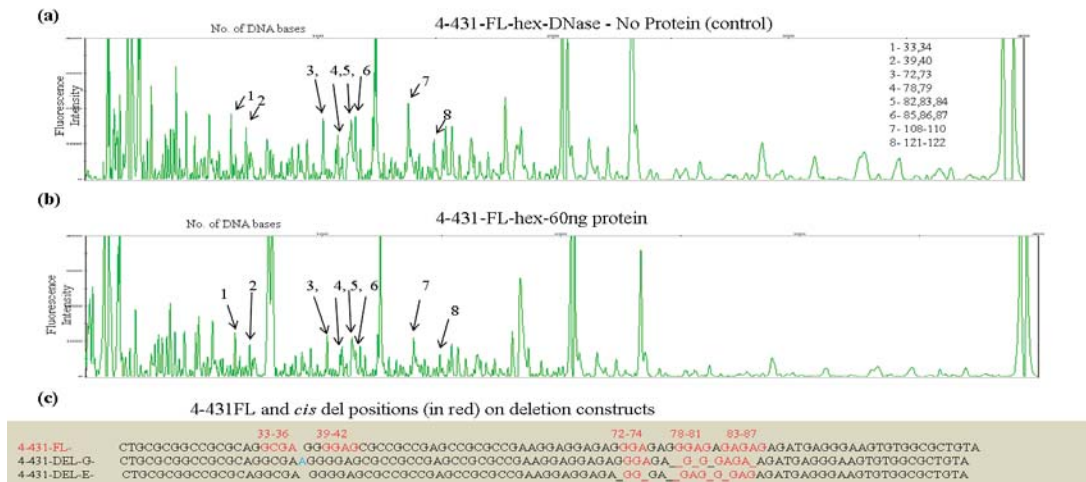


Fig. 3.6 *Cis*-regulatory motifs in BDP 4-431 bound to nuclear proteins identified by DNA foot-printing. The DNA was 3' end labeled with HEX. Arrows indicate the differences in peak intensities in (a) 4-431 full-length BDP DNA when no protein was bound and (b) 3' fragment of 4-431 DNA-protein footprint at same positions where the nuclear protein was bound. Numerals indicate the base numbers corresponding to the peaks. Protein bound bases were marked only in the fragment corresponding to 3' truncated construct which showed considerable loss of BDP functional activity. (c) Protein binding sites in the BDP are marked in red with deleted bases shown in the deletion constructs G and E.

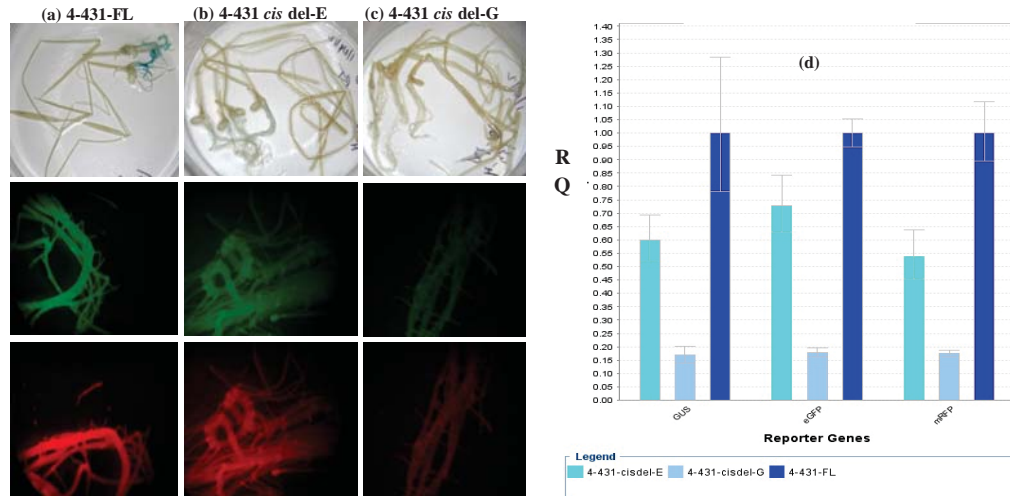


Fig. 3.7 eGFP-GUS and RFP expression pattern driven by the full-length and deletion versions of 4-431 BDP in rice. Expression of reporter genes driven by (a) full length promoter and deletion constructs (b) E and (c) G. (d) Quantitative real time RT-PCR analysis of *gus*, *egfp* and *rfp* reporter genes driven by the full length 2-273 BDP and the deletion constructs E and G. (RQ: Relative Quantification)

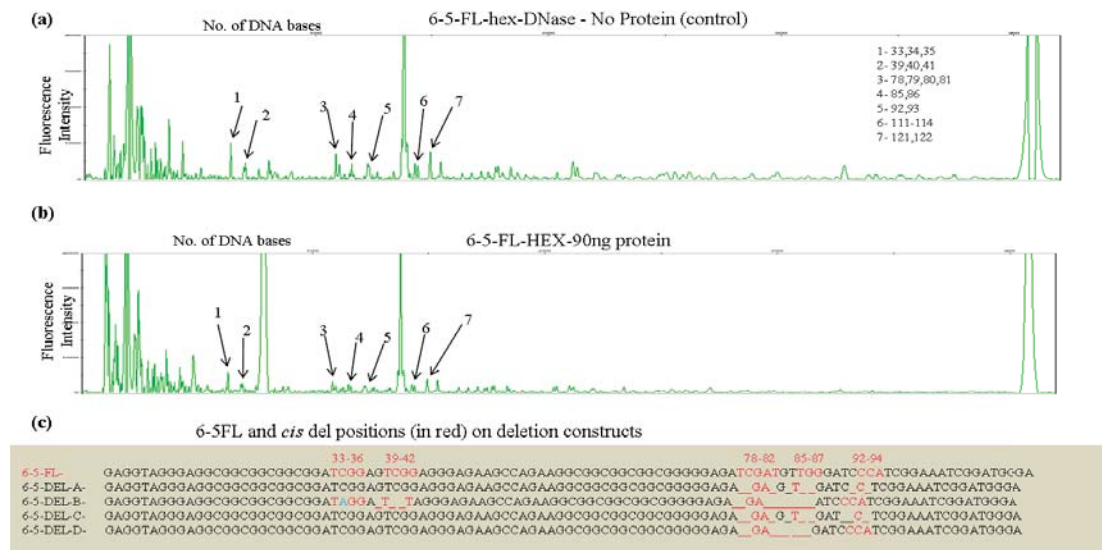


Fig. 3.8 *Cis*-regulatory motifs in BDP 6-5 bound to nuclear proteins identified by DNA foot-printing. The DNA was 3' end labeled with HEX. Arrows indicate the differences in peak intensities in (a) 6-5 full-length BDP DNA when no protein was bound and (b) 3' fragment of 6-5 DNA-protein footprint at same positions where the nuclear protein was bound. Numerals indicate the base numbers corresponding to the peaks. Protein bound bases were marked only in the fragment corresponding to 3' truncated construct which showed considerable loss of BDP functional activity. (c) Protein binding sites in the BDP are marked in red with deleted bases shown in the deletion constructs A through D.

6-5

Five *cis* deletion constructs were generated for the 6-5 BDP. Out of the five, only one construct (6-5-del-B) incorporated a base substitution from 'C' to 'A' in the first motif positioned from 33-36 bases and two base deletions in the second motif positioned between 39 and 42 bases. Apart from this, 6-5-del-B construct had an entire motif deleted between bases 85 and 87 (Fig. 3.8c). Rest of the constructs had base deletions in the motifs from 78-82, 85-87 and 92-94 bases (Fig. 3.8). Qualitative analysis of the *cis* deletion constructs of the 6-5 BDP revealed a striking loss of reporter gene expression in the 6-5-del-B construct whereas the expression of the reporter genes in the other *cis* deletion constructs was considerably reduced compared to the full length BDP (Fig. 3.9). On the other hand, analysis of the quantitative expression data (qRT-PCR) indicated a loss of 65-75% of promoter activity in the 6-5 Del B construct in both directions (Fig. 3.9e). The promoter activity was reduced by 40-50% in the rest of the *cis* deletion constructs compared to the full length promoter.

The 'CCA' motif in 6-5 BDP partially matches with the "GCCAC" core sequence of SORLIP1, which is a sequence over-represented in light-induced promoters (SORLIP) present mostly in the promoters of the phytochrome A (phyA) photoreceptor pathway (Hudson and Quail 2003). The 'GGTG' motif identified in 2-273BDP by DNA footprinting matches with the core binding consensus 'TG(G/N)NG' of Alfin1 motif. Alfin2 is a general and essential root growth regulator in plants which binds to the DNA in a sequence specific manner enhancing the expression of the MsPRP2 salt inducible gene in the roots of the plants. Over-expression of Alfin1 significantly increased plant growth and salt tolerance (Winicov and Bastola 1999). It is also shown that Alfin1 is highly conserved in rice, *Arabidopsis* and alfalfa.

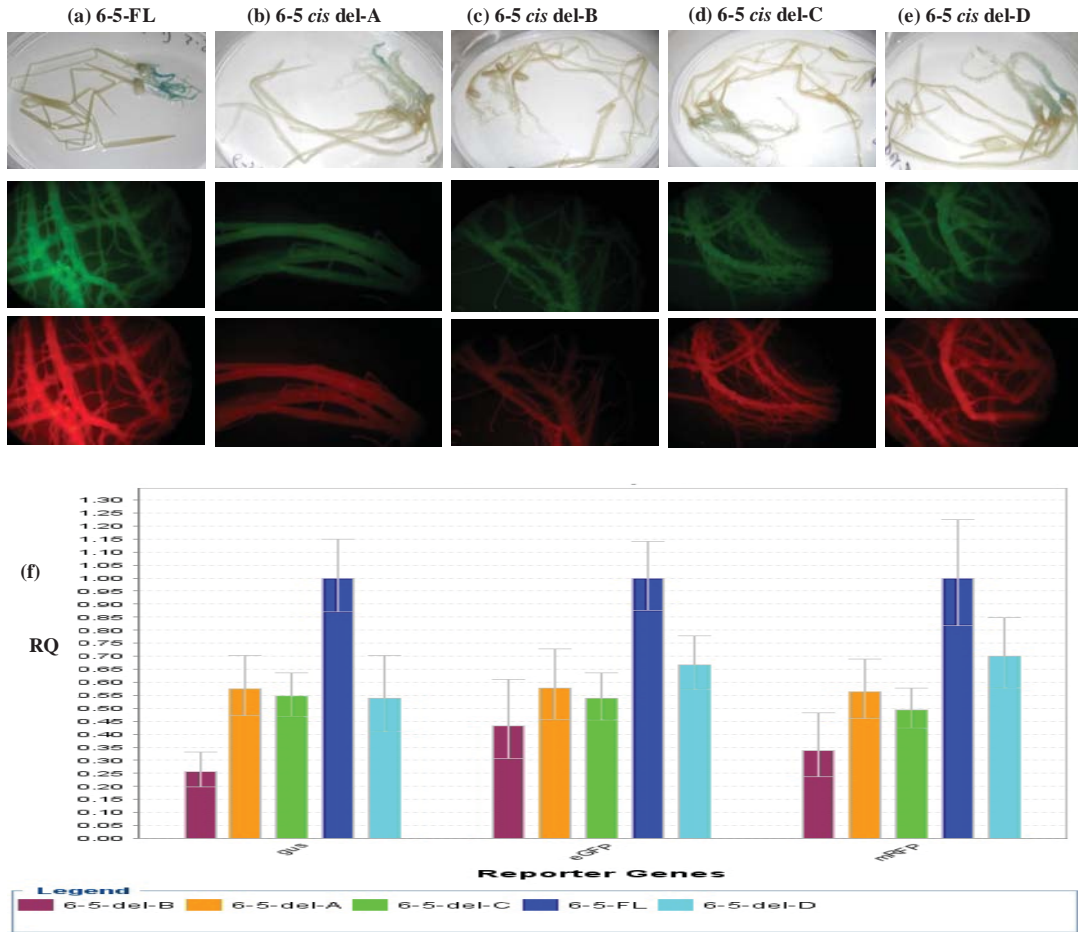


Fig. 3.9 eGFP-GUS and RFP expression pattern driven by the full-length and deletion versions of 6-5 BDP in rice. Expression of reporter genes driven by (a) full length promoter and deletion constructs (b) A, (c) B, (d) C, and (e) D. (f) Quantitative real time RT-PCR analysis of *gus*, *egfp* and *rfp* reporter genes driven by the full length 6-5 BDP and the deletion constructs A through D. (RQ: Relative Quantification)

3.5 Discussion

Evaluating bidirectional activity of promoters using a single reporter gene vector and flipping the directionality of the promoter has two major issues: 1) flipping the promoter can cause undesirable results primarily in the coexpression pattern of the divergent genes and is time consuming; 2) expression of the transgene is dependent on the position of the insertion in the plant genome (Matzke and Matzke 1995; 1998; Matzke and Matzke 1995, Matzke et al. 2000; Kumar and Fladung 2001). Large numbers of transgenic lines have to be screened for exact/approximate location of transgene insertion before the expression pattern of BDPs can be studied. Here, we engineered pBGWFS7-RFP, a novel promoter

expression plasmid with two reporter genes for BDP functional analysis, which addresses the above issues. The presence of the reporter genes like *egfp* on the 3' end and *rfp* on the 5' end of an inserted BDP eliminates the issue of flipping the promoter for evaluating BDP activity. Once the whole cassette has been transformed into plants, the insertional or positional effect will be experienced by the whole cassette as a single entity and hence it is no longer a hindrance in studying BDP expression. During transient expression assays using *Agrobacterium*, transgenes are inserted in a large number of plant cells (Dhadi et al. 2009), so positional effects are not a problem in this case. Moreover, it is easier to localize the expression pattern of the reporter genes in plant tissues. For instance, 2-273 BDP drives expression of the reporter genes in all tissues and hence constitutive in nature unlike 4-431 and 6-5 BDP where the expression is limited only to roots. Hence transient expression assay can be an effective tool in deciphering whether a promoter or gene is constitutive in nature or not. This work proves the usefulness of the pBGWFS7-RFP vector and AmTEA as indispensable tools in evaluating, characterizing and for studying the functional mechanism of BDPs in plants.

DNA footprinting using a genetic analyzer (Yindeeoungyeon and Schell 2000) has been regularly employed for evaluating DNA-protein interactions (Cagle et al. 2011; Moumita and Vincent 2012). Ours is the first report of employing this technology to identify the protein binding sites on BDPs to decode functional *cis* regulatory elements. Comparison of the quantitative as well as the qualitative expression data for all three promoters showed that the two 4 base pair motifs located between 33 and 46 bases contribute to at least 50% of the BDP activity. Comparing the DNA protein footprints along with the expression data of all three BDPs revealed another important *Cis* element located between 78 and 91 bases. Deletion or partial disruption of this motif resulted in approximately 50% loss in BDP activity in case of the 2-273 del-Y construct (Fig. 3.4 and 3.5) and constructs of other BDPs with deletions at similar positions. Another crucial fact to note is the spacer between the first two motifs and the third one is approximately 30-40bp which is similar to a normal TATA based promoter (Kiran et al. 2006). The radical loss of BDP activity in case of 6-5-del-B construct or 4-431-del-G construct can be attributed to the deletion/alteration of the motifs in the three specified positions disrupting the binding of the proteins responsible for regulating BDP activity. Furthermore, the deletions in the three motifs in the above constructs might have reduced the spacer distance between the first two and the third motif resulting in the recognition failure of the *Cis* regulatory motifs by the transcription enhancing proteins. Overall considering the expression data from the above BDPs it can be deduced that the three motifs located within the first 100bp upstream of the 5'UTR of one of the divergent genes act as "turn on" or "regulatory switches" for functional activity of BDPs.

Similar to a TATA box in unidirectional promoters, even BDPs harbor two 4 bp motifs from the transcription start sites in the 5' UTRs of the one of the divergent genes flanking the BDP. These two 4bp motifs are distributed among 10-15bp separated by a small spacer of 2-5bp, and the total length of the two motifs is 8bp which is equivalent to the TATA (TCACTATATATAG (Kiran et al. 2006; Joshi 1987) / TATA (T/A)A(T/A) (Zhu et al. 1995)) motif in unidirectional promoters or the GABPA (CCGGAARYR) motif identified in the human BDPs (Lin et al. 2007). The TATA box is rich in A/T bases whereas BDPs are mostly rich in G/C bases. The GC richness of these motifs can be accounted for the total high GC content of the BDPs both in humans (Trinklein et al. 2004) and plants (Dhadi et al. 2009). The TATA box along with the downstream *Cis* elements are required for the binding of the transcription factors and *trans*-acting proteins for modulation and formation of a stable initiation complex (Aso et al. 1994; Grayson et al. 1995). Considering the expression data obtained by the *Cis* motif deletion analysis of the three BDPs it is clear that there is a 40-50% loss of BDP activity in both the reporter genes, even if one of the three motifs was either deleted/mutated. Altering the motifs positioned between 33 and 46, and 78 and 91 bases, resulted in 70-80% loss of BDP activity which suggests it is not affecting the stability of the BDPs based on the reduction of expression equally of both the reporter genes but not just one. These motifs are likely to be essential in modulating the expression of the flanking divergent genes rather than responsible for the formation of a stable initiation complex. This reflects that the RNA polymerase assembly and stable initiation complex formation occurred somewhere else. It is possible that these motifs were just recruiting the RNA polymerase along with the initiation complex, thereby regulating the expression of the neighboring genes. Considering the role of downstream transcription elements from the TATA box in unidirectional promoters, we employed the same configuration of considering approximately 100-130bp of sequence on either sides of the TSS to understand the function of BDPs. Scrutinizing the sequences farther than 100bp beyond the TSS, especially towards the functional 5'UTR regions should open up more cues towards comprehensive understanding of BDPs functionality. Based on the present data it can be summed up that these motifs are functioning like enhancers in looping DNA for transcriptional activity. Enhancers generally have the inherent ability to drive, enhance the communication and mobility between promoter-enhancer in gene clusters, and stochastic communication between various promoters. Enhancers trigger transcriptional activation by recruiting transcriptional machinery to the site of transcription (Gondor and Ohlsson 2009). Based on our work, we propose that bidirectional promoters in rice regulate the expression of genes in both forward and reverse directions either by harboring binding regions for RNA polymerase or enhancer and repressor elements or both.

In summary, this work has underlined and explored efficient means of developing and using tools and methods to understand the functionality of BDPs. This paper provides first handed information on the identification of *Cis* protein targets on the DNA using a capillary genetic analyzer and corroborating the functionality of the motifs using real time gene expression studies. We identified that 5'UTR of one of the divergent genes plays an essential role in modulating the BDP activity. Extended research on the 5'UTRs should further enhance our understanding of the transcriptional regulation by BDPs in general.

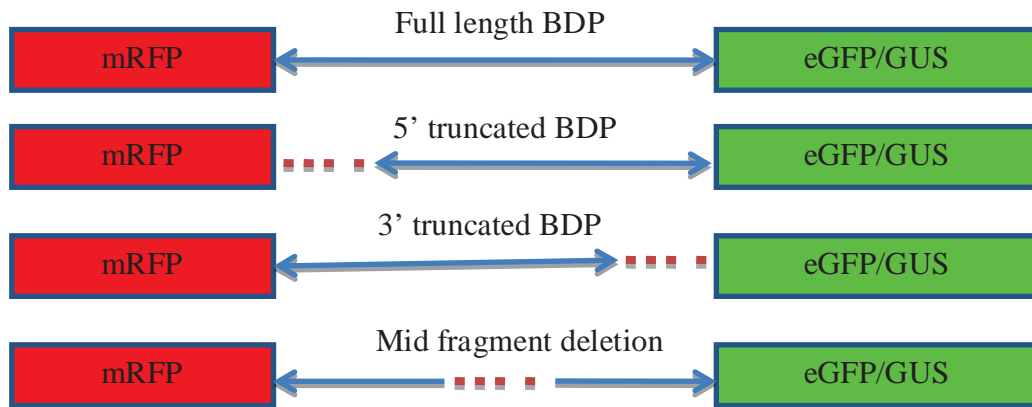
3.6 Acknowledgments

This work was supported by the National Research Initiative of the USDA Cooperative State Research, Education and Extension Service, grant number 2007-35301-18036. We thank the Michigan Technological University's Biotech Research Centre (BRC) for their continuous support and funding. We personally thank Lorie Bernhardt (Dale Bumpers National Rice Research Center) for providing bulk quantities of rice seeds. We thank USDA-GRIN for providing the seeds of rice. We thank Dr. Victor Busov for *Agrobacterium* (GV3101) and the Institute of Plant Systems Biology, Ghent University, Belgium, for pBGWFS7 and pB7RWG2 vector.

3.8 Supplementary information



Supplemental Figure 3.1 Expression cassette of pBGWFS7-RFP showing the insertion of a BDP between the two reporter genes. Arrows indicate direction of transcription in the reporter genes



Supplemental Figure 3.2 Schematic representation of large fragment deletions (truncations) in BDP functional analysis. Dotted lines indicate (sections) fragments of BDP deleted in the respective constructs. Each dotted fragment represents 100-150bp in BDPs <500bp in length and 300-350bp in length in BDPs >1kb in length.

Supplementary Table 3.1 Primers used for the generation of BDPs and their fragment deletion constructs

2-273										
2-273 F	5'-GGG GAC AAG TTT GTA CAA AAA AGC AGG CT	TCCTCGTGTGGCCTCTGCT-3'								
2-273-R	5'- GGG GAC CAC TTT GTA CAA GAA AGC TGG GT	GGAATGGTGGCTCTGATACCAGATTGT-3'								
2-273-F2-ATT	5'-GGG GAC AAG TTT GTA CAA AAA AGC AGG CT	GGTGGTCTCTTCCCTACCTCCCC-3'								
2-273-F2-0		5'-TCCCCACCTCTCCTCCTACC-3'								
2-273-R1-0		5'-TACAGCGTACCTTCCCTAACAAGTACC-3'								
2-273-R1-ATT	5'- GGG GAC CAC TTT GTA CAA GAA AGC TGG GT	AGGAGATTGCCAAGGTTCCAAGC-3'								
R-914										
914 F	5'-GGG GAC AAG TTT GTA CAA AAA AGC AGG CT	GGTGTGCGGAGCTGTGCTCTCT-3'								
914 R	5'- GGG GAC CAC TTT GTA CAA GAA AGC TGG GT	GCGGGCTCCGTGTCGTGTCT-3'								
914-R1-ATT	5'- GGG GAC CAC TTT GTA CAA GAA AGC TGG GT	CAACGACCCGTAGACGTACGTAC-3'								
914-F2-ATT	5'-GGG GAC AAG TTT GTA CAA AAA AGC AGG CT	CGGTCTCGTTTCCCTGCGAGAGAG-3'								
6-5-										
6-5-F	5'-GGG GAC AAG TTT GTA CAA AAA AGC AGG CT	ACGGAGAAGGAGAGTGGTGGCTT -3'								
6-5-R	5'- GGG GAC CAC TTT GTA CAA GAA AGC TGG GT	AGATGGAGATGGAGGTAGGGAGG -3'								
6-5-F2-ATT	5'-GGG GAC AAG TTT GTA CAA AAA AGC AGG CT	CAGGCAAGAGCAAGAGGATAACC-3'								
6-5-F2-0		5'-TCCCATCCGATTCCGATGGGAT-3'								
6-5-R1-0		5'-CCTATTGGA CCCGCGCGTTACC-3'								
6-5-R1-ATT	5'- GGG GAC CAC TTT GTA CAA GAA AGC TGG GT	TCAGACAGACTGGGCCGTTTCGTT-3'								
4-431										
4-431-F	5'-GGG GAC AAG TTT GTA CAA AAA AGC AGG	CTCGCGTAGTGTGCTACCGCCGCC -3'								
4-431-R	5'- GGG GAC CAC TTT GTA CAA GAA AGC TGG GT	CCGGCGCCTCCGCCCTACCGCTG -3'								
4-431-F2-ATT	5'-GGG GAC AAG TTT GTA CAA AAA AGC AGG CT	TGAGGGGAGAGGGGATTAACCCTAGGGAG-3'								
4-431-F2-0		5'-ACCCGAAGCTACAGCGCCACACTTCC-3'								
4-431-R1-0		5'-AACCTCCATGGACGGAGCTCCAG-3'								
4-431-R1-ATT	5'- GGG GAC CAC TTT GTA CAA GAA AGC TGG GT	GGTGGGGCTCACTGGCTCAG-3'								
3-343										
3-343-F	5'-GGG GAC AAG TTT GTA CAA AAA AGC AGG CT	CACTCTTGCACTTCGCTACAAC -3'								
3-343-R	5'- GGG GAC CAC TTT GTA CAA GAA AGC TGG GT	TCGCTTCTGGATGTCCAGCATCTTC -3'								

3-343-F2-ATT	5'-GGG GAC AAG TTT GTA CAA AAA AGC AGG CT TCACTTAGGCCCTCGAGCTTTTAG-3'
3-343-F2-0	5'-CGGAGTTGACGCTGCGACCATTCT-3'
3-343-R1-0	5'-ACTAGTCGAGTTATGGGGCTTCAA-3'
3-343-R1-ATT	5'- GGG GAC CAC TTT GTA CAA GAA AGC TGG GT CGGTTGCCGCGTGTACTCTAGA-3'
6-8-	
6-8-F	5'-GGG GAC AAG TTT GTA CAA AAA AGC AGG CTTCAGCCTCTGCC TAGGGTTGGGTT -3'
6-8-R	5'- GGG GAC CAC TTT GTA CAA GAA AGC TGG GTTCTCCGACTCCGATGAGCGAAA -3'
6-8-F2-ATT	5'-GGG GAC AAG TTT GTA CAA AAA AGC AGG CT TCCACGAGTTCCCCCGCTGAC-3'
6-8-F2-0	5'-TCCTTCGTTGCATCGGCGTGGG-3'
6-8-R1-0	5'-AGTGTAGCGCCCAAGAAGAGGAC-3'
6-8-R1-ATT	5'- GGG GAC CAC TTT GTA CAA GAA AGC TGG GT TGGAAGAAGCAGACAGCTGAGAG-3'
8-612	
8-612-F	5'-GGG GAC AAG TTT GTA CAA AAA AGC AGG CT GTCGGGCTTGGGAGTGAGGAG -3'
8-612-R	5'- GGG GAC CAC TTT GTA CAA GAA AGC TGG GT ATCGATAGGTATCTCGTCTCCGGCT -3'
8-612-F1-ATT	5'-GGG GAC AAG TTT GTA CAA AAA AGC AGG CT GACTAGTAGATGTCTCTCGCAGTC-3'
8-612-F1-0	5'-ATCAAAGAGGGGATTGCGTTGCG-3'
8-612-R1-0	5'-CACCTCACCTCACTATCCCTGAT-3'
8-612-R1-ATT	5'- GGG GAC CAC TTT GTA CAA GAA AGC TGG GT GGGCTGGGCCAATGGGCCGT-3'
Primer segments marked in black are the attachment sites required for gateway cloning recombination process	

Supplementary Table 3.2 primers used for the production of end labelled BDPs for DNA footprinting assays

cre-II-fam-F	6-Fam- 5'-GGG GAC AAG TTTGTA CAA AAA AG-3'
cre-II-R	5'-GGG GAC CAC TTT GTA CAA GAA-3'
cre-II-F	5'-GGG GAC AAG TTTGTA CAA AAA AG-3'
cre-II-hex-R	Hex- 5'-GGG GAC CAC TTT GTA CAA GAA-3'

Supplementary Table 3.3 Primers used for the generation of cis deletion constructs of the BDPs

2-273	CIS
DEL	
114-117-F	5'-GAGTCGGATGCTTTTAGGCCCCAGCGAGG-3'
114-117-R	5'-CCTCGCTGGGGCCTAAAAGCATCCGACTC-3'
143-149-F	5'-GGCAGATCACCGAGGAGGAGAAGAAGGAGA-3'
143-149-R	5'-TCTCCTTCTTCTCCTCCTCGGTGATCTGCC-3'
4-431	CIS
DEL	
112-116-F	5'-GAGGAGAGGGAGAGGGAGAAGATGAGGGA-3'
112-116-R	5'-TCCCTCATCTTCTCCCTCTCCCTCTCCTC-3'
62-71-F	5'-GGCCGCGCAGCGCCGCGAG-3'
62-71-R	5'-CTCGGCGGCGCTGCGCGGCC-3'
6-5-CIS DEL	
107-116-F	5'-GCGGCGGGGGAGAGATCCCATCGGAA-3'
107-116-R	5'-TTCCGATGGGATCTCTCCCCGCCGC-3'
62-71-F	5'-GGG AGG CGG CGG CGG CGG AA GGG AGA AGC CAG AAG GCG -3'
62-71-R	5'-CGC CTT CTG GCT TCT CCC TT CCG CCG CCG CCT CCC- 3'
107-116- MUT-2-F	5'-GGC GGC GGG GGA GAG ATC GGT GTT GAT CAA GTC GGA AAT CGG ATG GGA TCA-3'
107-116- MUT-2-R	5'-TGA TCC CAT CCG ATT TCC GAC TTG ATC AAC ACC GAT CTC TCC CCC GCC GCC-3'

Supplementary Table 3.4 Taqman primers and probes used for qRT-PCR

Sequence name	5' Dye	Oligo sequence	3' Dye
GUSA_F		CAAGGTGCACGGGAATATTTTCG	
GUSA_R		GAACATTACATTGACGCAGGTGATC	
EF1ALPHA_F		CCCAAGAGGCCATCAGACAA	
EF1ALPHA_R		CCGATCTTGATACACGTCCTGAAG	
GFP_F		AGAACGGCATCAAGGTGAACTT	
GFP_R		CGCTGCCGTCTCGAT	
GUSA_M	6FAM	TCGGGTCGAGTTTACG	MGBNFQ
EF1ALPHA_M	6FAM	CCCTGCGTCTTCCC	MGBNFQ
GFP_M	6FAM	TTGTGGCGGATCTTG	MGBNFQ

CONCLUSION

For the first time we provided a global scale presence of BDPs in three sequenced plant genomes: rice (*Oryza sativa*), *Arabidopsis thaliana*, and *Populus trichocarpa*. The numbers of BDPs were found to be proportional to the genome size of the respective plants and they were also in agreement with the proportion of BDPs in animal genomes. Similar to the BDPs in animals, the plant BDPs were exclusively concentrated in the 'CpG' islands and an underrepresentation of the "TATA" motif was observed, suggesting that their mode of function and regulation is different from the TATA based unidirectional promoters.

Analyzing gene expression in plants is a cumbersome and time consuming process. To mitigate this problem a 'novel *Agrobacterium tumefaciens* mediated transient expression assay' (AmTEA) was developed for young plants. AmTEA works efficiently for both dicots and monocots. Moreover the penetration and effectiveness of AmTEA was equivalent to that of the stable transgenic lines as shown with respect to the *Arabidopsis* DEAD-box RNA helicase family gene promoter.

Seven putative BDPs and their fragment deletion constructs were analyzed using AmTEA. Out of the seven, 2-273, 4-431 and 6-5 were established as true bidirectional promoters. DNA-protein foot-printing analysis using a genetic analyzer revealed three important regions in the first 100bp of the 5'UTR's of one of the divergent genes. Deletions or mutations in the three regions decreased the BDP functionality by 70-80%, whereas deletions /mutations in one of the regions decreased the BDP functionality by 40-50%. These differences can be attributed to the enhancer-like properties of the three regulatory motifs present in these promoters which may regulate expression of genes in both orientations. We conclude that the 5'UTR of one of the divergent genes regulate the functional activity of the BDPs.

FUTURE DIRECTIONS

Extensive characterization of bidirectional promoters have been done in animal genomes. This dissertation work summarizes the theoretical and experimental approaches towards exploration of BDPs in plants. Even though a part of the mechanism towards the functional activity of BDPs in plants has been revealed in this work, still a conclusive mechanism for BDPs has to be laid down. Moreover the actual biological significance of BDPs still remains an object unknown both in plants and animals.

This work has analyzed and outlined a few regulatory *cis* elements in the BDPs. Analyzing further elements either downstream or upstream of a BDP can reveal outstanding information into the regulatory functions offered by a BDP. DNA foot printing can be used to identify regulatory regions in the core BDPs and also in regions encompassing the two divergent genes. These regulatory regions can be mutagenized or deleted to study their regulatory effect on reporter gene expression using AmTEA. This process can be performed easily and helps in deciphering a conclusive regulatory mechanism for BDPs. Combining this strategy with generation of stable transgenic plants can help in understanding the biological significance of the BDPs in the genetic and physiological constitution of the plants and animals as a common entity.

Apart from identifying the regulatory *cis* motifs, uncoupling the proteins from the motifs during a DNA foot printing assay and analyzing the proteins by MS and MALDI-TOF can identify novel transcription factors that regulate a BDP. Knock out or silencing experiments can give a glimpse into the regulatory functions offered by these proteins.

Understanding the biological and functional significance of BDPs can open an efficient means of transforming plants and animals. It can help in metabolic engineering, in coordinating multi-gene expression and in tight stoichiometric controlling of the various gene products. This knowledge can be applied for molecular farming for producing better vaccines and pharmaceuticals. This technology can be used for producing better plants to withstand the changing global climate and result in high outputs in yield that can provide ample food for all the organisms on the planet.

REFERENCES

- Adachi N, Lieber MR. 2002. Bidirectional gene organization: a common architectural feature of the human genome. *Cell*. 109:807-809.
- Ahn J, Gruen JR. 1999. The genomic organization of the histone clusters on human 6p21.3. *Mammalian Genome*. 10:768-770.
- Albig W, Kioschis P, Poustka A, Meergans K, Doenecke D. 1997. Human histone gene organization: Nonregular arrangement within a large cluster. *Genomics*. 40:314-322.
- Amoah BK, Wu H, Sparks C, Jones HD. 2001. Factors influencing *Agrobacterium* mediated transient expression of uidA in wheat inflorescence tissue. *Journal of Experimental Botany*. 52:1135-1142.
- An Y-QC, Meagher RB. 2010. Strong expression and conserved regulation of ACT2 in *Arabidopsis thaliana* and *Physcomitrella patens*. *Plant Molecular Biology Reporter*. 28:481-490.
- Aso T, Conaway JW, Conaway RC. 1994. Role of core promoter structure in assembly of the RNA polymerase II preinitiation complex. *Journal of Biological Chemistry*. 269:26575-26583.
- Baker SS, Wilhelm KS, Thomashow MF. 1994. The 5'-region of *Arabidopsis thaliana* cor15a has cis-acting elements that confer cold-, drought- and ABA-regulated gene expression. *Plant Molecular Biology*. 24:701-713.
- Barton KA, Chilton MD. 1983. *Agrobacterium* Ti plasmids as vectors for plant genetic engineering. *Methods Enzymology*. 101:527-539.
- Beck CF, Warren RA. 1988. Divergent promoters, a common form of gene organization. *Microbiological Reviews*. 52(3):318-326.
- Bellizzi D, Dato S, Cavalcante P, Covello G, Di Cianni F, Passarino G, Rose G, De Benedictis G. 2007. Characterization of a bidirectional promoter shared between two human genes related to aging: SIRT3 and PSMD13. *Genomics*. 89:143-150.

- Bi C, Chen F, Jackson L, Gill BS, Li W. 2011. Expression of lignin biosynthetic genes in wheat during development and upon infection by fungal pathogens. *Plant Molecular Biology Reporter*. 29:149-161.
- Bird LE, Subramanya HS, Wigley DB. 1998. Helicases: a unifying structural theme? *Current Opinion Structural Biology*. 8:14-18.
- Boeger H, et al. 2005. Structural basis of eukaryotic gene transcription. *FEBS Letters*. 579:899-903.
- Brown RL, Kazan K, McGrath KC, Maclean DJ, Manners JM. 2003. A role for the GCC-box in jasmonate-mediated activation of the PDF1.2 gene of *Arabidopsis*. *Plant Physiology*. 132:1020-1032.
- Butler JEF, Kadonaga JT. 2002. The RNA polymerase II core promoter: a key component in the regulation of gene expression. *Genes and Development*. 16:2583-259.
- Cagle CA, Shearer JES, Summers AO. 2011. Regulation of the integrase and cassette promoters of the class 1 integron by nucleoid-associated proteins. *Microbiology-SGM*. 157:2841-2853.
- Chen PY, Chang WSW, Lai YK, Wu CW. 2009. c-Myc regulates the coordinated transcription of brain disease-related PDCD10–SERPINI1 bidirectional gene pair. *Molecular Cell Neuroscience*. 42:23-32.
- Chlan CA, Lin JM, Cary JW, Cleveland TE. 1995. A procedure for biolistic transformation and regeneration of transgenic cotton from meristematic tissue. *Plant Molecular Biology Reporter*. 13:31-37
- Cleard F, Moshkin Y, Karch F, Maeda RK. 2006. Probing long-distance regulatory interactions in the *Drosophila melanogaster* bithorax complex using Dam identification. *Nature Genetics*. 38:931-935.
- Collins PJ, Kobayashi Y, Nguyen L, Trinklein ND, Myers RM. 2007. The ets-related transcription factor GABP directs bidirectional transcription. *PLoS Genetics*. 3:e208.
- Conoway RC, Sato S, Tomomori-Sato C, Yao T, Conoway JW. 2005. The mammalian Mediator complex and its role in transcriptional regulation. *TRENDS in Biochemical Sciences*. 30:250-255.

- Cubas P, Lauter N, Doebley J, Coen E. 1999. The TCP domain: a motif found in proteins regulating plant growth and development. *Plant Journal*. 18:215-222.
- da Silva JAT, Fukai S. 2001. The impact of carbenicillin, cefotaxime and vancomycin on chrysanthemum and tobacco TCL morphogenesis and *Agrobacterium* growth. *Journal of Applied Horticulture*. 3:3-12.
- De Wilde CH, Van Houdt S, De Buck G, Angenon G, De Jaeger G, Depicker A. 2000. Plants as bioreactors for protein production: avoiding the problem of transgene silencing. *Plant Molecular Biology*. 43:347-359.
- Dean A. 2006. On a chromosome far, far away: LCRs and gene expression. *TRENDS in Genetics*. 22(1):38-45.
- Dernburg AF, et al. 1996 Perturbation of Nuclear Architecture by Long-Distance Chromosome Interactions. *Cell*. 85:745-759.
- Dhadi SR, Deshpande A, Ramakrishna W. 2012. A novel non-wounding transient expression assay for cereals mediated by *Agrobacterium tumefaciens*. *Plant Molecular Biology Reporter*. 30:36-45.
- Dhadi SR, Krom N, Ramakrishna W. 2009. Genome-wide comparative analysis of putative bidirectional promoters from rice, *Arabidopsis* and *Populus*. *Gene*. 429:65-73.
- Di Matteo D, et al. 2003. The crystal structure of polygalacturonase-inhibiting protein (PGIP), a leucine-rich repeat protein involved in plant defense. *Proceedings of National Academy of Sciences, USA*. 100:10124-10128.
- Diao G, Wang Y, Wang C, Yang C. 2011. Cloning and functional characterization of a novel glutathione S-transferase gene from *Limonium bicolor*. *Plant Molecular Biology Reporter*. 29:77-87.
- Dolferus R, Jacobs M, Peacock WJ, Lieber MR. 2002. Bidirectional gene organization: a common architectural feature of promoter elements in stress responses of the *Arabidopsis* Adh Gene. *Plant Physiology*. 105:1075-1087.

- Dong S, Lester L, Johnson LF. 2000. Transcriptional control elements and complex initiation pattern of the TATA-less bidirectional human thymidylate synthase promoter. *Journal of Cellular Biochemistry*. 77, 50–64.
- Dorsett D. 1999. Distant liaisons: long-range enhancer–promoter interactions in *Drosophila*. *Current Opinion Genetics and Development*. 9(5):505-514.
- Eckardt NA. 2007. Oxidation pathways and plant development: crosstalk between thioredoxin and glutaredoxin pathways. *Plant Cell*. 19:1719-1721.
- Engström PG, et al. 2006. Complex loci in human and mouse genomes. *PLoS Genetics*. 2:e47.
- Escobar C, Aristizabal F, Navas A, Del Campo FF, Fenoll C. 2001. Isolation of active DNA-binding nuclear proteins from tomato galls induced by root-knot nematodes. *Plant Molecular Biology Reporter*. 19:375a-375h.
- Fraley RT, Rogers SG, Horsch RB, et.al. 1983. Expression of bacterial genes in plant cells. *Proceedings of National Academy of Sciences. USA*. 80:4803-4807.
- Fromm M, Taylor LP, Walbot V. 1985. Expression of genes transferred into monocot and dicot plant cells by electroporation. *Proceedings of National Academy of Sciences. USA*. 82:5824–5828.
- Gelvin SB. 2003. *Agrobacterium*-mediated plant transformation: the biology behind the “gene-jockeying” tool. *Microbiology and Molecular Biology Reviews*. 67:16-37.
- Gondor A, Ohlsson R. 2009. Chromosome crosstalk in three dimensions. *Nature*. 461:212-217.
- Grasser KD, Krohn NM, Lichota J, Stemmer C. 2000. Chromatin-associated HMG1, HMG1/Y and SSRP1 proteins of higher plants. *Plant Physiology* 110:427-435.
- Grayson J, Williams RS, Yu YT, Bassel-Duby R. 1995. Synergistic interactions between heterologous upstream activation elements and specific TATA sequences in a muscle-specific promoter. *Molecular Cell Biology*. 15:1870-1878.
- Guarguaglini G, Battistoni A, Pittoggi C, Di Matteo G, Di Fiore B, Lavia P. 1997. Expression of the murine RanBP1 and Htf9-c genes is regulated from a shared

- bidirectional promoter during cell cycle progression. *Biochemistry Journal*. 325(Pt 1):277-286.
- Guha A, Saturen Y, Szybalski W. 1971. Divergent orientation of transcription from the biotin locus of *Escherichia coli*. *Journal of Molecular Biology*. 56:53-62.
- Halpin C. 2005. Gene stacking in transgenic plants — the challenge for 21st century plant biotechnology. *Plant Biotechnology*. 3:141-155.
- Higo K, Ugawa Y, Iwamoto M, Korenaga T. 1999. Plant *cis*-acting regulatory DNA elements (PLACE) database: 1999. *Nucleic Acids Research*. 27:297-300.
- Hochschild A. 2002. The λ Switch: cICloses the Gap in Autoregulation. *Current Biology*. 12:R87-89.
- Hotze M, Lurz G, Schriider J. 1995. A cDNA encoding a plant homologue to animal HMG box proteins involved in structure-specific recognition of DNA (SSRP family). *Gene*. 161:295-296.
- Hudson ME, Quail PH. 2003. Identification of promoter motifs involved in the network of phytochrome A-regulated gene expression by combined analysis of genomic sequence and microarray data. *Plant Physiology*. 133:1605-1616.
- Hwang YS, Karrer EE, Thomas BR, Chen L, Rodriguez RL. 1998. Three *cis*-elements required for rice α -amylase Amy3D expression during sugar starvation. *Plant Molecular Biology*. 36:331-341.
- Jefferson RA. 1987. Assaying chimeric genes in plants: the GUS gene fusion system. *Plant Molecular Biology Reporter*. 5:387-405.
- Jones AM, Assmann SM. 2004. Plants: the latest model system for G-protein research. *EMBO Reports*. 5:572-578.
- Jonsson ZO, Hubscher U. 1997. Proliferating cell nuclear antigen: more than a clamp for DNA polymerases. *BioEssays*. 19:967-975.
- Joshi C.P. 1987. An inspection of the domain between putative TATA box and translation start site in 79 plant genes. *Nucleic Acids Research*. 15:6643-6653.

- Juven-Gershon T, Cheng S, Kadonaga JT. 2006. Rational design of a super core promoter that enhances gene expression. *Nature Methods*. 3:917-922.
- Kamiya N, Nagasaki H, Morikami A, Sato Y, Matsuoka M. 2003. Isolation and characterization of a rice WUSCHEL-type homeobox gene that is specifically expressed in the central cells of a quiescent centre in the root apical meristem. *Plant Journal*. 35:429-441.
- Kapila J, De Rycke R, Van Montagu M, Angenon G. 1997. An *Agrobacterium*-mediated transient gene expression system for intact leaves. *Plant Sciences*. 122:101-108.
- Karimi M, Inze D, Depicker A. 2002. Gateway vectors for *Agrobacterium* mediated plant transformation. *Trends Plant Sciences*. 7:193-195.
- Karniol B, Chamovitz DA. 2000. The COP9 signalosome: from light signaling to general developmental regulation and back. *Current Opinion in Plant Biology*. 3:387-393.
- Kato H, Xie G, Sato Y, Imai R. 2010. Isolation of anther-specific gene promoters suitable for transgene expression in rice. *Plant Molecular Biology Reporter*. 28:381-387.
- Kawai Y, et al. 2003. Structure and promoter activity of the human glia maturation factor- γ gene: A TATA-less, GC-rich and bidirectional promoter. *Biochim. Biophys. Acta*. 1625:246-252.
- Keddie JS, Tsiantis M, Piffanelli P, Hatzopoulos P, Murphy DJ. 1994. A seed-specific *Brassica napus* oleosin promoter interacts with a G box-specific protein and may be bidirectional. *Plant Molecular Biology*. 24:327-340.
- Kiran K, Ansari AA, Srivastava R, Lodhi N, Chaturvedi CP, Sawant SV, Tuli R. 2006. The TATA-Box Sequence in the basal promoter contributes to determining light-dependent gene expression in plants. *Plant Physiology*. 142:364-376.
- Kosugi S, Ohashi Y. 1997. PCF1 and PCF2 specifically bind to *cis* elements in the rice proliferating cell nuclear antigen gene. *Plant Cell*. 9:1607-1619.
- Kosugi S, Ohashi Y. 2002. DNA binding and dimerization specificity and potential targets for the TCP protein family. *Plant Journal*. 30:337-348.

- Kosugi S, Suzuka I, Ohashi Y. 1995. Two of three promoter elements identified in a rice gene for proliferating cell nuclear antigen are essential for meristematic tissue specific expression. *Plant Journal*. 7:877-886.
- Krom N, Ramakrishna W. 2008. Comparative analysis of divergent and convergent gene pairs and their expression patterns in rice, *Arabidopsis*, and *Populus*. *Plant Physiology*. 147:1763-1773.
- Kumar S, Fladung M. 2001. Controlling transgene integration in plants. *Trends in Plant Science*. 6(4):155-159.
- Lee TI, Young RA. 2000. Transcription of eukaryotic protein coding genes. *Annual Review of Genetics*. 34:77-137.
- Li J-F, Park E, von Arnim AG, Nebenführ A. 2009. The FAST technique: a simplified *Agrobacterium*-based transformation method for transient gene expression analysis in seedlings of *Arabidopsis* and other plant species. *Plant Methods*. 5:6.
- Li L, Qu R, Kochko AD, Fauquet C, Beachy RN. 1993. An improved rice transformation system using the biolistic method. *Plant Cell Reports*. 12:250–255.
- Lin JM, Collins PJ, Trinklein ND, Fu Y, Xi H, Myers RM, Weng Z. 2007. Transcription factor binding and modified histones in human bidirectional promoters. *Genome Research*. 17:818-827.
- Liu Y-Z, Dong T, Lei Y, Deng X-X, Gu Q-Q (2011) Isolation of a polygalacturonase gene from *Citrus sinensis* fruit and its expression relative to fruit mastication trait, fruit development, and calcium or boron treatments. *Plant Molecular Biology Reporter*. 29:51–59.
- Malik S, Roeder RG. 2005. Dynamic regulation of pol II transcription by the mammalian Mediator complex. *TRENDS in Biochemical Sciences*. 30:256-263.
- Matzke AJM, Matzke MA. 1998. Position effects and epigenetic silencing of plant transgenes. *Current Opinion in Plant Biology*. 1:142-148.
- Matzke MA, Matzke AJM. 1995. How and why do plants inactivate homologous (trans)genes? *Plant Physiology*. 107:679-685.

- Matzke MA, Mette MF, Matzke AJM. 2000. Transgene silencing by the host genome defense: implications for the evolution of epigenetic control mechanisms in plants and vertebrates. *Plant Molecular Biology*. 43:401-415.
- Mayer KFX, Schoof H, Haecker A, Lenhard M, Jurgens G, Luax T. 1998. Role of WUSCHEL in regulating stem cell fate in the *Arabidopsis* shoot meristem. *Cell*. 95:805-815.
- Medina J, Bargues M, Terol J, Pérez-Alonso M, Salinas J. 1999. The *Arabidopsis* CBF gene family is composed of three genes encoding AP2 domain-containing proteins whose expression is regulated by low temperature but not by abscisic acid or dehydration. *Plant Physiology*. 119:463-470.
- Meng X, Li F, Liu C, Zhang C, Wu Z, Chen Y. 2010. Isolation and characterization of an ERF transcription factor gene from cotton (*Gossypium barbadense* L.). *Plant Molecular Biology Reporter*. 28:176-183.
- Meyers BC, et al. 2004. *Arabidopsis* MPSS. An online resource for quantitative expression analysis. *Plant Physiology*. 135:801-813.
- Momota R, Sugimoto M, Oohashi T, Kigasawa K, Yoshioka H, Ninomiya Y. 1998. Two genes, COL4A3 and COL4A4 coding for the human $\alpha_3(\text{IV})$ and $\alpha_4(\text{IV})$ collagen chains are arranged head-to-head on chromosome 2q36. *FEBS Letters*. 424:11-16.
- Morita A, Umemura T, Kuroyanagi M, Futsuhara Y, Perata P, Yamaguchi J. 1998. Functional dissection of a sugar-repressed alpha-amylase gene (Ramy1A) promoter in rice embryos. *FEBS Lett*. 423, 81–85.
- Moumita P, Vincent M. 2012. Use of an automated capillary DNA sequencer to investigate the interaction of cisplatin with telomeric DNA sequences. *Biomedical Chromatography*. 26:350-354.
- Nagao R-T, Goekjian VH, Hong JC, Key JL. 1993. Identification of protein-binding DNA sequences in an auxin-regulated gene of soybean. *Plant Molecular Biology*. 21:1147–1162.
- Nakagawa H, Tategu M, Yamauchi R, Sasaki K, Sekimachi S, Yoshida K. 2008. Transcriptional regulation of an evolutionary conserved intergenic region of CDT2-INTS7. *PLoS ONE*. 3:e1484.

Nikolov BD, et al. 1995. Crystal structure of a TFIIB–TBP–TATA-element ternary complex. *Nature*. 377:119-128.

Ogden S, Haggerty D, Stoner CM, Kolodrubetz D, Schleif R. 1980. The *Escherichia coli* L-arabinose operon: Binding sites of the regulatory proteins and a mechanism of positive and negative regulation. *Proceedings of National Academy of Sciences. USA*. 77:3346-3350.

Perfus-Barbeoch L, Jones AM, Assmann AM. 2004. Plant heterotrimeric G protein function: insights from *Arabidopsis* and rice mutants. *Current Opinions Plant Biology*. 7:719–731.

Ptashne M, Gann A. 1997. Transcriptional activation by recruitment. *Nature*. 386:569-577.

Qi J, Yu S, Zhang F, Shen X, Zhao X, Yu Y, Zhang D. 2010. Reference gene selection for real-time quantitative polymerase chain reaction of mRNA transcript levels in chinese cabbage (*Brassica rapa* L. ssp. *pekinensis*). *Plant Molecular Biology Reporter*. 28:597-604.

Rossi L, Escudero J, Hohn B, Tinland B. 1993. Efficient assay for TDNA dependent transient gene expression. *Plant Molecular Biology Reporter*. 11:220–229.

Sadanandom A, Piffanelli P, Knott T, Robinson C, Sharpe A, Lydiate D, Murphy D, Fairbairn D. 1996. Identification of a peptide methionine sulphoxidoreductase gene in an oleosin promoter from *Brassica napus*. *Plant Journal*. 10:235-242.

Saha D, Kumar V, Bhat SR, Srinivasan R. 2011. 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 Molecular Biology Reporter*. 29:265-77.

Shin R, Kim MJ, Paek KH. 2003. The CaTin1 (*Capsicum annuum* TMV-induced clone 1) and CaTin1-2 genes are linked head-to-head and share a bidirectional promoter. *Plant Cell Physiology*. 44:549-554.

Singer SD, Hily J-M, Liu Z (2010) A 1-kb bacteriophage lambda fragment functions as an insulator to effectively block enhancer–promoter interactions in *Arabidopsis thaliana*. *Plant Molecular Biology Reporter*. 28:69-76.

Stam M, Belele C, Ramakrishna W, Dorweiler JE, Bennetzen JL, Chandler VL. 2002. The Regulatory regions required for *B'* paramutation and expression are located far upstream of the maize *b1* transcribed sequences. *Genetics*. 162:917-930.

Tatematsu K, Ward S, Leyser O, Kamiya Y, Nambara E. 2005. Identification of *cis* elements that regulate gene expression during initiation of axillary bud outgrowth in *Arabidopsis*. *Plant Physiology*. 138:757-766.

Taylor K, Hradecna Z, Szybalski W. 1967. Asymmetric distribution of the transcribing regions on the complementary strands of coliphage lambda DNA. *Proceedings of National Academy of Sciences. USA*. 57:1618-1625

Todd AK, Neidle S. 2008. The relationship of potential G-quadruplex sequences in *cis* upstream regions of the human genome to SP1-binding elements. *Nucleic Acids Research*. 36:2700-2704.

Travers MT, Cambot M, Kennedy HT, Lenoir GM, Barber MC, Joulin V. 2005. Asymmetric expression of transcripts derived from the shared promoter between the divergently oriented ACACA and TADA2L genes. *Genomics*. 85:71-84.

Tre'mousaygue D, Garnier L, Bardet C, Dabos P, Herve' C, Lescure B. 2003. Internal telomeric repeats and ‘‘TCP-domain’’ protein binding sites cooperate to regulate gene expression in *Arabidopsis thaliana* cycling cells. *Plant Journal*. 33:957-966.

Trinklein ND, Aldred SF, Hartman SJ, Schroeder DI, O'tillar RP, Myers RM. 2004. An abundance of bidirectional promoters in the human genome. *Genome Research*. 14:62-66.

Uze M, Wunn J, Puonti-Kaerlas J, Potrykus I, Sautter C. 1997. Plasmolysis of precultured immature embryos improves *Agrobacterium* mediated gene transfer to rice (*Oryza sativa* L.). *Plant Science*. 130:87-95.

Vilar JMG, Saiz L. 2005. DNA looping in gene regulation: from the assembly of macromolecular complexes to the control of transcriptional noise. *Current Opinions in Genetics and Development*. 15:136-144.

Von Gromoff ED, Schroda M, Oster U, Beck CF. 2006. Identification of a plastid response element that acts as an enhancer within the *Chlamydomonas* HSP70A promoter. *Nucleic Acids Research*. 34:4767-4779.

Von Hippel PH. 1998. An integrated model of the transcription complex in elongation, termination, and editing. *Science*. 281:660-665.

Walker BA, Hateren AV, Milne S, Beck S, Kaufman J. 2005. Chicken TAP genes differ from their human orthologues in locus organisation, size, sequence features and polymorphism. *Immunogenetics*. 57:232-247.

Wang X, Dong J, Liu Y, Gao H. 2010. A novel dehydrationresponsive element-binding protein from *Caragana korshinskii* is involved in the response to multiple abiotic stresses and enhances stress tolerance in transgenic tobacco. *Plant Molecular Biology Reporter*. 28:664-675.

Wang H, Hill K, Perry SE. 2004. An *Arabidopsis* RNA lariat debranching enzyme is essential for embryogenesis. *Journal of Biological Chemistry*. 279:1468-1473.

Welchen E, Gonzalez DH. 2006. Overrepresentation of elements recognized by TCPdomain transcription factors in the upstream regions of nuclear genes encoding components of the mitochondrial oxidative phosphorylation machinery. *Plant Physiology*. 141:540-545.

West AB, Lockhart PJ, O'Farell C, Farrer MJ. 2003. Identification of a novel gene linked to parkin via a bi-directional promoter. *Journal of Molecular Biology*. 326:11-19.

Winicov I, Bastola DR. 1999. Transgenic over expression of the transcription factor Alfin1 enhances expression of the endogenous MsPRP2 gene in Alfalfa and improves salinity tolerance of the Plants. *Plant Physiology*. 120:473-480.

Winter N, Nimzyk R, Bösche C, Meyer A, Bullerdiek J. 2011. Chromatin immunoprecipitation to analyze DNA binding sites of HMGA2. *PLoS ONE*. 6(4):e18837.

- Winter V, Hauser M-T. 2006. Exploring the ESCRTing machinery in eukaryotes. *Trends Plant Sciences*. 11:115–123.
- Wroblewski T, Tomczak A, Michelmore R. 2005. Optimization of *Agrobacterium*-mediated transient assays of gene expression in lettuce, tomato and *Arabidopsis*. *Plant Biotechnology*. 3:259–273.
- Xu XM, Adams S, Chua N-H, Moller SG. 2005. AtNAP1 represents an atypical SufB protein in *Arabidopsis* plastids. *Journal of Biological Chemistry*. 280:6648-6654.
- YanG MQ, Koehly LM, Elnitski LL. 2007. Comprehensive annotation of bidirectional promoters identifies co-regulation among breast and ovarian cancer genes. *PLoS Computational Biology*. 3:e72.
- Yang T, Poovaiah BW. 2002. A calmodulin-binding/CGCG box DNA-binding protein family involved in multiple signaling pathways in plants. *Journal of Biological Chemistry*. 277:45049–45058.
- Yanhui C, et al. 2006. The MYB transcription factor superfamily of *Arabidopsis*: expression analysis and phylogenetic comparison with the rice MYB family. *Plant Molecular Biology*. 60:107-124.
- Yindeeyoungyeon W, Schell MA. 2000. Footprinting with an automated capillary DNA sequencer. *BioTechniques*. 29:1034-1041
- Zhu Q, Dabi T, Lamb C. 1995. TATA BQX and initiator functions in the accurate transcription of a plant minimal promoter in vitro. *Plant Cell*. 7:1681-1689.

APPENDIX

Role of other authors

Whole genome level computational analysis of bidirectional promoters was performed by Nicholas Krom in chapter 1. The rice promoter, R2-273 and the Arabidopsis promoter for the gene encoding DEAD-box RNA helicase family promoter-reporter constructs were generated by Aparna Deshpande. R2-273 and R1-914 promoter-reporter constructs and the engineering of the pBGWFS7-RFP vector were performed by Aparna Deshpande.

Chapter I of this dissertation was published in the journal Gene in 2009. Copyright © 2012 Elsevier B.V. I, as the author, am allowed to “include the article in full or in part in a thesis or dissertation”. Below are the excerpts from the Elsevier copyright declaration for authors (<http://www.elsevier.com/wps/find/authorsview.authors/rights>)

What rights do I retain as a journal author*?

As a journal author, you retain rights for a large range of author uses of your article, including use by your employing institute or company. These rights are retained and permitted without the need to obtain specific permission from Elsevier.

- the right to make copies (print or electronic) of the journal article for your own personal use, including for your own classroom teaching use;
- the right to make copies and distribute copies of the journal article (including via e-mail) to research colleagues, for personal use by such colleagues for scholarly purposes*;
- the right to post a pre-print version of the journal article on Internet websites including electronic pre-print servers, and to retain indefinitely such version on such servers or sites for scholarly purposes* (with some exceptions such as The Lancet and Cell Press. See also our information on [electronic preprints](#) for a more detailed discussion on these points)*;
- the right to post a revised personal version of the text of the final journal article (to reflect changes made in the peer review process) on your personal or institutional website or server for scholarly purposes*, incorporating the complete citation and with a link to the Digital Object Identifier (DOI) of the article (but not in subject-oriented or centralized repositories or institutional repositories with mandates for systematic postings unless there is a specific agreement with the publisher. [Click here](#) for further information);

- the right to present the journal article at a meeting or conference and to distribute copies of such paper or article to the delegates attending the meeting;
- for your employer, if the journal article is a ‘work for hire’, made within the scope of the author’s employment, the right to use all or part of the information in (any version of) the journal article for other intra-company use (e.g. training);
- patent and trademark rights and rights to any process or procedure described in the journal article;
- the right to include the journal article, in full or in part, in a thesis or dissertation;
- the right to use the journal article or any part thereof in a printed compilation of your works, such as collected writings or lecture notes (subsequent to publication of the article in the journal); and

Chapter 2 of this dissertation was published in the Plant Molecular Biology reporter Journal in 2012. The author has obtained permission from the publisher to reprint this material as shown below:

Springer License

This is a License Agreement between Surendar Dhadi ("You") and Springer ("Springer"). The license consists of your order details, the terms and conditions provided by Springer, and the [payment terms](#) and [conditions](#).

[Get the printable license.](#)

License Number	2882890157767
License date	Apr 06, 2012
Licensed publisher	content Springer
Licensed publication	content Plant Molecular Biology Reporter
Licensed title	content A Novel Non-wounding Transient Expression Assay for Cereals Mediated by <i>Agrobacterium tumefaciens</i>
Licensed author	content Surendar Reddy Dhadi
Licensed date	content Jan 1, 2011
Volume number	30
Issue number	1
Type of Use	Thesis/Dissertation
Portion	Full text
Number of copies	1
Author of this Springer article	Yes and you are a contributor of the new work

Title of your thesis Functional Analysis of Rice Bidirectional promoters
/ dissertation

Expected completion date	May 2012
Estimated size(pages)	100
Total	0.00 USD

Terms and Conditions

Introduction

The publisher for this copyrighted material is Springer Science + Business Media. By clicking "accept" in connection with completing this licensing transaction, you agree that the following terms and conditions apply to this transaction (along with the Billing and Payment terms and conditions established by Copyright Clearance Center, Inc. ("CCC"), at the time that you opened your Rightslink account and that are available at any time at <http://myaccount.copyright.com>).

Limited

License

With reference to your request to reprint in your thesis material on which Springer Science and Business Media control the copyright, permission is granted, free of charge, for the use indicated in your enquiry.

Licenses are for one-time use only with a maximum distribution equal to the number that you identified in the licensing process.

This License includes use in an electronic form, provided its password protected or on the university's intranet or repository, including UMI (according to the definition at the Sherpa website: <http://www.sherpa.ac.uk/romeo/>). For any other electronic use, please contact Springer at (permissions.dordrecht@springer.com or permissions.heidelberg@springer.com).

The material can only be used for the purpose of defending your thesis, and with a maximum of 100 extra copies in paper.

Although Springer holds copyright to the material and is entitled to negotiate on rights, this license is only valid, provided permission is also obtained from the (co) author (address is given with the article/chapter) and provided it concerns original material which does not carry references to other sources (if material in question appears with credit to another source, authorization from that source is required as well).

Permission free of charge on this occasion does not prejudice any rights we might have to charge for reproduction of our copyrighted material in the future.

Altering/Modifying Material: Not Permitted
You may not alter or modify the material in any manner. Abbreviations, additions, deletions and/or any other alterations shall be made only with prior written authorization of the author(s) and/or Springer Science + Business Media. (Please contact Springer at (permissions.dordrecht@springer.com or permissions.heidelberg@springer.com))

Reservation of Rights
Springer Science + Business Media reserves all rights not specifically granted in the combination of (i) the license details provided by you and accepted in the course of this licensing transaction, (ii) these terms and conditions and (iii) CCC's Billing and Payment terms and conditions.

Copyright Notice: Disclaimer
You must include the following copyright and permission notice in connection with any reproduction of the licensed material: "Springer and the original publisher /journal title, volume, year of publication, page, chapter/article title, name(s) of author(s), figure number(s), original copyright notice) is given to the publication in which the material was originally published, by adding; with kind permission from Springer Science and Business Media"

Warranties: None

Example 1: Springer Science + Business Media makes no representations or warranties with respect to the licensed material.

Example 2: Springer Science + Business Media makes no representations or warranties with respect to the licensed material and adopts on its own behalf the limitations and disclaimers established by CCC on its behalf in its Billing and Payment terms and conditions for this licensing transaction.

Indemnity

You hereby indemnify and agree to hold harmless Springer Science + Business Media and CCC, and their respective officers, directors, employees and agents, from and against any and all claims arising out of your use of the licensed material other than as specifically authorized pursuant to this license.

No Transfer of License
This license is personal to you and may not be sublicensed, assigned, or transferred by you to any other person without Springer Science + Business Media's written permission.

No Amendment Except in Writing
This license may not be amended except in a writing signed by both parties (or, in the case of Springer Science + Business Media, by CCC on Springer Science + Business Media's behalf).

Objection to Contrary Terms
Springer Science + Business Media hereby objects to any terms contained in any purchase order, acknowledgment, check endorsement or other writing prepared by you, which terms are inconsistent with these terms and conditions or CCC's Billing and Payment terms and conditions. These terms and conditions, together with CCC's Billing and Payment terms and conditions (which are incorporated herein), comprise the entire agreement between you and Springer Science + Business Media (and CCC) concerning this licensing transaction. In the event of any conflict between your obligations established by these terms and conditions and those established by CCC's Billing and Payment terms and conditions, these terms and conditions shall control.

Jurisdiction

All disputes that may arise in connection with this present License, or the breach thereof, shall be settled exclusively by arbitration, to be held in The Netherlands, in accordance with Dutch law, and to be conducted under the Rules of the 'Netherlands Arbitrage Instituut' (Netherlands Institute of Arbitration).**OR:**

All disputes that may arise in connection with this present License, or the breach thereof, shall be settled exclusively by arbitration, to be held in the Federal Republic of Germany, in accordance with German law.