Plant Cis-regulatory elements: methods of identification and applications

Usman Ijaz¹, Tehreem Pervaiz¹, Temoor Ahmed¹, Riffat Seemab¹, Muhammad Shahid¹, Muhammad Noman¹, Majid Nadeem², Farrukh Azeem¹*
¹Department of Bioinformatics and Biotechnology, Government College University, Faisalabad, Pakistan
²Wheat Research Institute, Ayub Agriculture Research Institute, Faisalabad, Pakistan

Abstract
Biotic and abiotic factors badly affect the productivity and growth of plants and found the major yield-limiting factor in agriculture. In recent years, various cis-regulatory elements have been identified that can function as molecular switches to regulate the expression of stress-related genes. They comprise of insulators, silencers, promoters and enhancers. These elements mediate not only functional diversity, but also modulate plant physiology at all developmental stages. Thus, identifying and characterizing cis elements that are intricated in plant stress response is essential for the development of plants tolerance to stress. This review article presents characteristic properties of cis-acting regulatory elements. Two aspects of cis-elements are targeted; one is properties and examples of condition-specific cis-elements while the other is techniques used for their identification. This review will be helpful in elucidating recent advancements in cis-elements studies.

Keywords: Regulatory elements, Tissue specific, Promoter, Plants, Gene regulation

How to cite this:
DOI: https://doi.org/10.35495/ajab.2019.08.352

Introduction
The process of selective activation of a different subset of genes is key to cell differentiation, survival, and functional diversity. The fate of each cell is defined according to its function in living organisms. Gene expression is primarily governed by certain genomic sequences, which are specifically targeted by transcription factors. At each stage of genetic information, many elements control the level of gene expression through interaction with regulatory proteins or other DNA duplexes. On the basis of structural diversity, regulators are grouped into cis-sequences and trans-factors. Cis-regulatory sequences are array of non-coding DNA sequences. The distribution and position of regulatory elements within the genome depend on the gene type and activity (Venter and Botha, 2010). Specific regulatory proteins, known as trans factors form active complexes by interacting with cis sequences and other proteins. All eukaryotes have similar genome organization and universality in regulatory elements. However, the elements located in tissue/organ-specific promoters are considerably different from one another (Twyman et al., 2003; Venter and Botha, 2010). Transcription factors (TFs) mediate the
repression or activation of target gene within a specific tissue or organ type by interacting with regulatory elements and basal transcription machinery. Hence, identification of regulatory regions within genome is a significant step towards understanding the transcriptional logic of gene expression (Levine, 2010; Bilas et al., 2016). Different databases containing information about well-characterized cis-elements have been enlisted in Table 1.

Comparisons of gene regulation during development among morphologically divergent organisms and analyses of variations within members of the same species, support that regulatory DNA is major source of genetic diversity and forms the basis of evolutionary and morphological variations. The role of cis-elements in evolution can be formulated through the adaptive pattern of cis-regulatory systems (Arnone and Davidson, 1997). The interaction between these different regulatory elements, and their target promoters (i.e., nucleosomes, chromatin fibers, loops, rosettes, chromosomes and their location) regulate the expression of genes (Kellum and Schedl, 1991; Narlikar and Ovcharenko, 2009; Kolovos et al., 2012; Gup et al., 2018).

Cis-regulatory elements are linear fragments of non-coding DNA (6-10 bp) which have specific sequence sites that serve as the binding site for TFs in gene promoter regions (Bilas et al., 2016). Higher eukaryotes contain different types of cell, which display diverse cellular functions. Meanwhile, certain biotic and abiotic stimuli also influence cellular processes. Such functional differences are mediated through definite patterns of gene expression usually controlled by regulatory sequences (Weber et al., 2016).

Promoter sequences of a gene are present upstream of the coding region. The RNA Pol II enzyme attaches with specific binding sites in promoter sequence to enable initiation of transcription. A eukaryotic gene promoter is divided into two regions as core promoter (important for initiation of transcription) and distal promoter region (involved in modulation of gene expression). The core promoter region is usually defined by the absence or presence of specific cis-elements like initiator region (Inr), TATA box (present upstream of transcription start site or TSS), downstream promoter elements (DPE) and CAAT box. Like all other core promoters, DPE play a significant role in the initiation of transcription by RNA polymerase II. DPE is analogous to TATA-box and perform its function in TATA-less promoter. But sometimes DPE is also found along with TATA-box in single promoter. In promoters where DPE is absent, for example, plants and yeast promoters, the function of binding transcription Factor II D (TFIID) complex is carried out by its analog TATA-box (Kutach and Kadonaga, 2000; Pandey et al., 2019).

The distal promoter region is generally located thousands of base pairs upstream TSS or in intronic regions (for review see Bilas et al., 2016). Unlike core region, the distal promoter region is better known for gene expression regulation under specific environmental conditions or tissues. Therefore, different cis-sequences and their role in differential gene expression discussed in the following sections.

Types of Cis-elements in distal promoter regions

Gene expression in eukaryotes is a multiplex phenomenon, involving the TFs mediated activation and repression of genes in response to various environmental and developmental stimuli (Burke and Baniahmad, 2000). After the identification of eukaryotic enhancers (Banerji et al., 1981), different classes of cis-elements have been decided (Kellum and Schedl, 1991; Kolovos et al., 2012; Narlikar and Ovcharenko, 2009).

Enhancers

Enhancers are cis-regulatory DNA sequences that recruit RNA polymerase II-mediated transcription machinery (Barrett et al., 2012). Enhancers can be located and classify as intragenic (within an actively transcribed gene) or extragenic (outside actively transcribed gene) using tools like cis genome (Blinka et al., 2017). Transcription regulating sequences often comprise of several enhancer modules that vary in size from about 50bp to 1.5kbp (Blackwood and Kadonaga, 1998). They may be present far away from the transcription start site, which makes their detection more difficult (Visel et al., 2007). The enhancer sequences may also present downstream or upstream of gene and may function in an orientation-independent manner (Gupta, 2018). They show particular characteristics, like specific histone modifications, accessibility of chromatin, TF binding motifs, low DNA methylation, eRNA (enhancer transcribed long non-coding RNA) expression and physical interactions with target genes (Shlyueva et al., 2014).

Plant enhancers are present at distinct positions, often at a notable distance from downstream or upstream of
the promoter sequence. They enhance the gene expression by recruiting specific TFs (Biłas et al., 2016). It is reported that a particular protein binds with the enhancer sequence, which results in a loop formation. The interaction between promoters and enhancers depends upon the distance between them (Yang et al., 2011). Large-scale studies investigating some molecular features revealed that enhancers could be classified based on their activity state and features of chromatin. Based on activity state some additional types of enhancers are inactive enhancers, active enhancers and intermediate enhancers (Shlyueva et al., 2014). The basic phenomenon of enhancer regulated transcription is shown in (Fig 1).

**Figure-1: Schematic representation for mode of action of enhancer sequence (Li et al., 2015)**

**Silencer**
A silencer is a short nucleotide sequence, present upstream of a certain genes and exerts a negative (repressing or silencing) effect on the target gene expression (Aronow, 1992). These sequences have the ability to bind transcription factor proteins with promoter, which in turn perform specific functions. Several types of silencers exist which have the ability to control the expression of gene in various aspects, like intron splicing, activity of positive acting TFs, 3’ upstream untranslated signal recognition, and cytoplasmic retention of TFs (Clark and Docherty, 1993), and general transcription factor (GTF) assembly to eventually down-regulate gene expression. Silencers are known to be the binding sites for negative transcription factors (repressors). The function of repressor is to recruit the negative cofactors, so-called corepressors (Privalsky, 2004) and, in some conditions, differential cofactor recruitment change the activator into repressor. Two major types of silencers present in DNA sequence, i.e., negative regulatory elements (NREs) and silencer elements
Silencer elements are classical and position-independent elements that direct mechanism of active repression usually by interfering with GTF assembly (Ogbourne and Antalis, 1998). On the other hand, NREs are position-dependent and non-classical elements that initiate a passive repression mechanism usually by interfering with upstream elements (Ogbourne and Antalis, 1998).

**Insulator**
Insulators (also known as boundary elements) are ~0.5-3kb in length. An insulator acts as a dominant repressor and can work over long distances from the promoter region of target gene (Kellum and Schedl, 1991). When insulator sequence is located close to the promoter of gene, it serves to stabilize the interaction between enhancer and promoter elements. If it is present far away from the promoter sequence, they interfere with transcription activity of gene (Recillas-Targa et al., 2002) or interactions between cis-elements and inappropriate promoters (Kellum and Schedl, 1991). Two main properties of insulators include: (a) enhancer-blocking activity by blocking the interaction between enhancers and promoter of target gene and (b) heterochromatin-barrier activity that prevents the spread of repressive chromatin. Thus, they limit the partition of the genome into distinct realms of the expression and reduce the action of transcription regulatory elements to define domains (Recillas-Targa et al., 2002). A transformation booster sequence (TBS) from Petunia hybrida acts as an enhancer blocking insulator in Arabidopsis thaliana (Singer et al., 2011)

**Biological importance of cis-elements**
Cis-regulatory elements (CREs) are vital components of regulation of genetic network, which further control morphogenesis, anatomy and other aspects of developmental biology. In the following section, we have discussed the role of CREs in stress response, tissue-specific gene expression and morphological traits development.

**Role of cis-regulatory elements in plant Abiotic stress response**
Various abiotic stresses (drought, temperature, salinity and osmotic) have detrimental effects on plant growth and development (Zhu, 2002). The Intergovernmental Panel on Climate Change,
suggested that these stresses exert negative effects on plant growth in future (Hirayama and Shinozaki, 2010). For their survival, plants have to make changes at, cellular, biochemical, molecular and physiological levels. Tolerance to abiotic stress is an intricate phenomenon as these stresses may arise simultaneously and affect plant growth at different developmental stages (Park et al., 2003). Therefore, the basic mechanisms for adaptation and tolerance to abiotic stress have been the area of comprehensive research.

Several TFs and cis-elements not only function as molecular regulators for gene expression but also act as a terminal point for signal transduction pathways during abiotic stress responses (Gujjar et al., 2014). Abiotic stresses are significant for sessile organisms like plants because it enables the plants to deal with unfavorable environmental conditions (Hirayama and Shinozaki, 2010). Plants have to develop a variety of mechanisms to counteract the adverse environmental conditions. During abiotic stress, plants have to develop different mechanisms to recognize changes in growth, and prompt different signaling pathways, which in turn activates several genes at transcription and their products provide stress tolerance (Gao et al., 2007). Drought stress affects 10% of the total world’s agricultural land, which results in reducing 50% yield of important crops all over the world. During drought stress plants activates a variety of cellular and molecular mechanisms, which confer tolerance to plants against all type of abiotic stresses. The primary event in abiotic stress adopt by plants is the recognition of stress signals and activation of the signal transduction pathways, which ultimately lead to activation of several stress tolerance genes. 

In Arabidopsis thaliana, many cis-elements have been recognized, which are responding to abiotic stress. C-repeat binding factor (CBF), Dehydration responsive binding element (DREB1, DREB2) regulons function in ABA-independent gene expression (Lenka and Bansal, 2019), while ABA-responsive element (AREB) and ABA binding factor (ABF) is the major transcription factors activated in abiotic stress response. Moreover, various regulons, such as NAC (N-acetylcarnosine) and MYB (myeloblastosis)/MYC (myelocytomatosis) regulons, regulate gene expression in an ABA independent manner. Current studies described that NAC, AREB/ABF, DREB1/ CBF, DREB2 regulons have pivotal roles against abiotic stresses in rice (Bartels and Sunkar, 2005; Chinnusamy et al., 2004; Sunkar et al., 2007).

The unique class protein that plays a crucial role in abiotic stress is AP2/ERF protein. AP2/ERF protein is a large family that was further classified into three groups i.e. AP2, RAV, ERF. The DREB (dehydration responsive binding elements) also termed (C- repeat binding factor) belonging to ERF subfamily and show a significant role in cold stress. DREB/ERF transcription factors were mostly found in plants that are tolerant of cold stress i.e. Arabidopsis, Hordeum vulgare and Brassica napus (Leng and Zhao, 2019). Similarly, NAC transcription factors play a significant role in plant abiotic stress. Owing to the large number of NAC TFs, it is a great challenge for the researchers to uncover their role in abiotic stress response. Currently, transcriptome studies and whole-genome sequencing have enabled to identify the putative role of NAC TFs in abiotic stress such as 38 NAC genes that were found in soybean, which respond to drought stress. 33 NAC genes changed significantly under salt stress. In Oryza sativa, 40 NAC genes were found to express under salt and drought stress (Yuan et al., 2019).

Role of cis-regulatory elements in plant biotic stress response

In plants, different genes are expressed to combat biotic stress. This primary infection combats generally termed as the hypersensitive response (HR) is followed up by general protection mechanism that makes uninfected portion of plant tolerant to future attack, a process termed systemic acquired resistance (SAR) (Durrant and Dong, 2004). Natriuretic peptide receptor 1 (NPR1) protein is a stance that is highly conserved across many species and is involved in defense mechanisms (Cantu et al., 2013; Durrant and Dong, 2004). When AtNPR1 is overexpressed in carrot tissues, it showed resistance to several pathogens (Wally et al., 2009).

Plant cell defense is activated by the detection of pathogen invasion. When pathogens attack, salicylic acid (SA) accumulates in the cell resulting in changes in the cytosolic cellular redox (Garretón et al., 2002). These changes are followed by conversion of inactive NPR1 into active monomers in the absence of pathogen attack (Mou et al., 2003). Upon activation, NPR1 monomers reach nucleus where they interact with the TGA class of leucine-zipper TFs (Cao et al., 1994; Dong, 2001; van Loon et al., 2006). In turn, SA responsive-elements in the promoters of pathogen related (PR) genes bind to TGA factors, launching the onset of SAR (Fan and Dong, 2002) (Fig 2). WRKY
TFs family is also known for regulating the PR1 and NPR1 gene expression, by interacting with W box present in their promoters (Cao et al., 1998; Pandey and Somssich, 2009). In Arabidopsis it was found that 74 TFs were triggered in response to bacterial invasion and their expression was reduced in mutants that have defective signaling of SA, jasmonates (JA) or ethylene signaling indicated that these TFs have major role in plant defense (Chen and Widom, 2005). In tobacco, SINAC35 improved resistance to leaf curl virus in transgenic tobacco. In rice, OsEREBP1, which binds to the ethylene-responsive element (ERF) GAGCCGCC, has been reported to be expressed in Magnaporthe grisea (Yuan et al., 2019). In glycine max, GmERF5/GmERF113 was found to express by the attack of Phytophthora sojae. Similarly, in solanum tuberosum, SIERF gene block the replication of R. solanacearum (Abarca and Sánchez, 2019). Therefore, TF transcriptome and binding site mutation array will be an area to explore in the future.

Figure-2: Model summarizing one possible pathway of plant cell defense triggered by the recognition of pathogens.

It is known that upon pathogen attack, accumulation of SA causes a change in the cytosolic cellular redox, leading to the conversion of the inactive NPR1, present as cytosolic disulfide-bound oligomers in the absence of pathogen attack, into active monomers. NPR1 monomers are transported into the nucleus where they interact with the TGA class of basic leucine zipper transcription factors. This interaction in turn, stimulates the binding of TGA factors to SA responsive elements in the promoters of PR genes, launching the onset of SAR. It is also known that the WRKY family of transcription factors can regulate the expression of PR1 and NPR1 genes, interacting with W box elements present in their promoter regions. NPR1 is constitutively expressed and levels of its transcripts increased two-fold following SA treatment. Evidences also suggest that the climate change will also expand the host range of pathogens with increased chances of virulent strain development. Specifically studied the regulation of Arabidopsis transcription factors under different stresses (including defense-related stresses), and found that the expression of 74 transcription factor genes was responsive to bacterial pathogen infection, and was reduced or abolished in mutants that have defects in SA, JA, or ethylene signaling, suggesting that these transcription factors play an important role in plant defense (Kawano et al., 2010)

Role of cis-regulatory elements in plant tissue-specific gene expression
Patterns of tissue-specific gene expression, differences in gene expression at different developmental timing and quantitative levels due to the variations in cis-regulatory elements. Combinatorial effects of TFs binding with cis-elements is another way of gene expression regulation (Reményi et al., 2004; Singh, 1998; Wolberger, 1998). Combinatorial transcription factors constitute multi-protein complexes that possess their regulatory property from both intrinsic potential and the potential of their trans-acting partners (Singh, 1998). A single transcription factor can be a part of different complexes, single trans-acting factor can control numerous genes with different spatial and temporal expression patterns (Messenguy and Dubois, 2003; Yamaguchi and Hirano, 2006). The expression of phytoene synthase (Psy) was found in ripened fruits and breakers, as well as flowers but northern blot analysis confirms its absence in leaves or green fruits. Similarly, the phytoene desaturase (Pds) gene transcript was detected in leaves of barley and green fruit but expressed in flowers and breaker fruits. This shows that both are expressed in chromoplast containing tissues (Table 1). Nutrient transport related genes are involved in either influx or efflux of cellular nutrients (Stern and Orgogozo, 2008). A better understanding of the tissue-specific expression of genes and intelligent use of their tissue-specific cis-elements may ensure climate resilient nutritious crop production. In addition, synthetic promoters designed to initiate the transcription of eukaryotic genes.
### Table-1: Shows **Cis**-acting elements involved in Different Conditions

<table>
<thead>
<tr>
<th>Name</th>
<th>Consensus/Core Sequence</th>
<th>Transcription factor</th>
<th>Function</th>
<th>Reference</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Biotic</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>EIRE</td>
<td>TTCGACC</td>
<td>WRKY factors</td>
<td>Upregulate Chitinase genes</td>
<td>(Ramer et al., 1992; Fukuda, 1997; Passricha et al., 2016)</td>
</tr>
<tr>
<td>EL1-box3</td>
<td>AAACCAATT</td>
<td>ERF1</td>
<td>Response to fungal elicitor</td>
<td>(Passricha et al., 2016; Wang et al., 2016)</td>
</tr>
<tr>
<td>Box-WI</td>
<td>TTAGGACC</td>
<td>WRKY factors</td>
<td>Unknown nuclear factor</td>
<td>(Rushion et al., 1996; Passricha et al., 2016)</td>
</tr>
<tr>
<td>GC</td>
<td>CGCCCG</td>
<td>GC-binding protein</td>
<td>Enhancer-like element involved in anoxic specific inducibility</td>
<td>(Olive et al., 1991; Passricha et al., 2016)</td>
</tr>
<tr>
<td>GCC Box</td>
<td>AGCCGCC</td>
<td>Pt4</td>
<td>Function as an ethylene response element</td>
<td>(Ohme-Takagi and Shinshi, 1995; Büttner and Singh, 1997; Chakravartthy et al., 2003)</td>
</tr>
<tr>
<td>MRE-like</td>
<td>A(T)/C(A/T)/A(C)/C</td>
<td>MEP protein</td>
<td>Sites of fungal elicitor-inducible DNA–protein interactions</td>
<td>(Mishra et al., 2018)</td>
</tr>
<tr>
<td>SARE</td>
<td>TACGGACCTCC (core sequence)</td>
<td>Unknown Nuclear Factor (db)</td>
<td>A 76 bp fragment conferred a 20-fold induction by SA in transgenic tobacco plants</td>
<td>(Shah and Klessig, 1996)</td>
</tr>
<tr>
<td>PR1-motif</td>
<td>AGCTCAGAGATGTCGGGCCGCATATATTCTCCAGGA CTTTTT</td>
<td>BiP2</td>
<td>Cause induction Salicylic acid</td>
<td>(Carvalho et al., 2014; Lebel et al., 1998)</td>
</tr>
<tr>
<td>W box</td>
<td>(T)TGAC (CT)</td>
<td>WRKY</td>
<td>Fungal elicitor, oomycetes, fungi, bacteria</td>
<td>(Eulgem et al., 2000; Métraux et al., 2002)</td>
</tr>
<tr>
<td>JERE (jasmonic acid responsive element)</td>
<td>AGACCGCAGCAAGAGGA CCCAAGAT</td>
<td>ORCA factors (db)</td>
<td>Jasmonic acid, yeast caused induction derived elicitors, Phytophthora elicitor, oomycetes, fungi, bacteria</td>
<td>(Menke et al., 1999; Rushon et al., 2018)</td>
</tr>
<tr>
<td><strong>Tissue Specific</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>TRAB1</td>
<td>CAACGTGTGAC</td>
<td>bZIP</td>
<td>Binding site for ABA signal transaction gene TRAB1</td>
<td>(Hobo et al., 1999)</td>
</tr>
<tr>
<td>ABRE-like</td>
<td>ACGTGTGCTCCAT</td>
<td>OSBZ8</td>
<td>Binding site for Abscisic acid (ABA) signal transaction gene OSBZ8</td>
<td>(Nakagawa et al., 1996; Roy Choudhury et al., 2008)</td>
</tr>
<tr>
<td>DPBF COREDCD C3</td>
<td>ACACNG</td>
<td>DPBF-1 and 2 (Dc3 promoter-binding factor-1 and 2)</td>
<td>Binding core sequence found in the carrot embryo-specific Dc3 gene promoter, and induced by ABA</td>
<td>(Kim et al., 1997)</td>
</tr>
<tr>
<td>RYREPEATVFL EB4</td>
<td>CATGCATG</td>
<td>ABI3VP1</td>
<td>RY repeat motif, related to ABA-regulated gene expression during late embryo-genesis</td>
<td>(Hobo et al., 1999)</td>
</tr>
<tr>
<td>MYB1AT</td>
<td>WAACCA</td>
<td>MYB</td>
<td>MYB recognition site for dehydration-responsive gene and mediated by ABA</td>
<td>(Abe et al., 2003)</td>
</tr>
<tr>
<td>MYCATERD1</td>
<td>CATGTCG</td>
<td>MYC-domain transcription factors</td>
<td>MYC recognition sequence for early responsive to dehydration and mediated by ABA</td>
<td>(Abe et al., 2003)</td>
</tr>
<tr>
<td>MYCCONSENS USAT</td>
<td>CANNTG</td>
<td>MYC</td>
<td>MYC recognition site found in the promoters of the dehydration-responsive gene and mediated by ABA</td>
<td>(Abe et al., 2003)</td>
</tr>
<tr>
<td>MYBGAHV</td>
<td>TAACAAA</td>
<td>MYB101</td>
<td>GARC involved in gibberellin signal pathway and sugar suppression</td>
<td>(Gubler et al., 1995; Liang et al., 2013)</td>
</tr>
<tr>
<td>WRKY71OS</td>
<td>TGAC</td>
<td>WRKY71</td>
<td>A core of W-box, involved in gibberellin and ABA signaling pathways</td>
<td>(Eulgem et al., 1999; Zhang et al., 2006)</td>
</tr>
<tr>
<td>GATABOX</td>
<td>GATA</td>
<td>ASF-2</td>
<td>Conserved in light-regulated and tissue-specific expression genes</td>
<td>(Lam et al., 1989)</td>
</tr>
<tr>
<td>G-Box</td>
<td>CACGTG</td>
<td>bZIP proteins termed</td>
<td>ubiquitous, <strong>cis</strong>-acting DNA regulatory element</td>
<td>(Donald and Cashmore, 1990; Giuliano et al., 1988; Menkens et al., 1995)</td>
</tr>
<tr>
<td>GCN4</td>
<td>TGAGTCATG</td>
<td>bZIP</td>
<td>controlling seed-specific expression of the genes</td>
<td>(Wu et al., 1998)</td>
</tr>
<tr>
<td>TELO-Box</td>
<td>AAACCCCTAA</td>
<td>bZIP</td>
<td><strong>cis</strong>-acting DNA regulatory element</td>
<td>Korkurcu et al., 2014</td>
</tr>
</tbody>
</table>
Such promoters comprise of stretch of a combination of heterologous upstream regulatory elements and a core promoter region (Kassaw et al., 2018). The core promoter region (also termed as minimal region) contained a TATA-box required for recruiting RNA polymerase II to form a preinitiation complex. Synthetic promoters include repressors, inducers, and enhancers that bind to TFs and regulate gene expression under specific conditions. The utilization of synthetic promoters for targeted inducibility is of great interest for plant engineering strategies. Non-plants chemically inducible and transactivated (using transformation cassettes incorporating a core promoter and multimers of the upstream have gained much attention.

Role of cis-regulatory elements in plant morphological traits development

The elaborate network of genes controls the plant morphology and development (Prud’Homme et al., 2006). The functioning of genes that impart phenotypic diversity within or among species is resulted by the mutation of coding or regulatory sequences (Khan and Ali, 2013). Currently, in nature the intricacy of plant morphology has been increased. In these circumstances, the recognition of cis-regulatory modules (CRMs) is imperative for study of transcriptional regulation. Few studies were carried out on CRMs of plants, that they might involve in many mechanistic benefits for binding of TFs (Michael et al., 2008; Konishi and Yanagisawa, 2010;
Usman Ijaz et al.

Wang et al., 2010). Structurally and functionally the best-categorized families in plants are MADS (MCM1 agamous deficiency-serum response factor -box gene family that is suitable for CRMs studies (Smaczniak et al., 2012). Transcriptional regulators are encoded by the MADS-box, gene family, which are essential for the development of flowering plants (Theissen et al., 2000). MPF2 (Maturation Promoting Factor 2) like genes employ their role in leaf development, effects fertility, floral transition and calyx inflation in many Arabidopsis and Solanaceous species (He and Saedler, 2005; Sanjana et al., 2012). Epidermal specific expression of StFAR3 and StPOD72-like gene is directed by L1-Box element that helps in the development of cutin and suberin in Solanum tuberosum via long-chain fatty acid biosynthesis. The MBSII element present in StHAP3 and StCASP1-like gene regulates cellular differentiation (Vulavala et al., 2017).

Techniques for the identification of Cis-elements

Various experimental and bioinformatic based approaches have been developed and used for the analysis of cis-elements and promoters. Each technique has its limitations and methodological biases. In recent years, with the emergence of next-generation sequencing techniques (Hawkins et al., 2010), our ability to identify new enhancers has been increased dramatically. Following are the few frequently used techniques for identifying cis-elements:

ChIP-seq

This technique combines chromatin immune precipitation (ChIP) with extensive DNA sequencing to recognize the binding regions of DNA associated proteins (Li et al., 2015). It is used to determine cis-stromes (Lupien, 2008). ChIP-seq (chromatin immunoprecipitation) is an in vivo technique, in which DNA binding protein of interest co-precipitate with a genomic fragment are sequenced. (Shamloo-Dashtpagerdi et al., 2015). Subsequently, ChIP sequence data can be used to predict TF binding motifs. The protein or modifications targeted by ChIP, influence the type and number of identified enhancers (Weber et al., 2016; Puente and Peso, 2018). ChIP is treated with exonuclease which removes the DNA not bound to TF to identify the DNA bound to TF sites. This technique has been adapted in plants to identify the TFs sites (Zhu et al., 2015). However, the most appropriate approach combinations of TFs or histone marks for enhancer identification remained to be efficient (Haring et al., 2010).

SNP-based approach

Single nucleotide polymorphism (SNP) is variation at a single position in DNA sequence of more than 1% of population. SNP within a gene give rise to allele. SNPs are also found in non-coding DNA. It is based on certitude that there is high-level SNP density in regulatory regions of genes having a role in environmental response and signaling. While the regulatory region of housekeeping genes has low-level SNP density. It is further followed up that SNP density in known conserved cis-elements is low. Hexamers with more sequence conservation and evidence from positional preferences, annotation information and functional significance of co-related expressed genes are helpful in categorizing putative cis-elements (Gupta et al., 2019). It has been found that the greater the set of promoter sequences to be compared, larger will be the density and set of detected mutations. Thus, better the false results will be at which sequence of conserved elements can be detected (Korkuč et al., 2014).

Transcription factors (TFs) binding motif scan

Enhancers contain clusters of TF binding sites, so the enhancers can be detected by searching genomes for TF binding motifs (González et al., 2012; Spitz and Furlong, 2012). TF binding site is not always present in a sequence-specific manner and thus, it could be difficult to identify cis-elements by motif scanning. Transcription factors perform function regularly in multiplexes, identifying clusters of TF binding motifs decreases the digit of false positives (González et al., 2012). The benefits of transcription binding motif scan include identification of TF binding sites, and the limitations include high false-positives; requirement of previous information of TF binding motif (Reiss et al., 2015).

Enhancer trapping

In enhancer trapping, the basal promoter is fused with reporter gene and is randomly inserted into the genome and transformants with fascinating pattern of reporter expression are selected for further study (Aichinger et al., 2003). The Endogenous enhancer or activator of the gene presents adjacently to the reporter gene is selected for further studies (Chudalayandi, 2011). Possibly these are the genes that become stimulated during growth of the host plant.
Usman Ijaz et al.

(McGarry and Ayre, 2008). It is observed that tissue-specific patterns are mediated by endogenous cis-regulatory sequences and these are easy to locate via trapped enhancers (Yang et al., 2005).

DNase-seq

Another approach for studying the gene regulatory networks including transcription factor binding sites and genes regulatory sequence via their chromatin characteristics (Gupta, 2018). If the genome is sequenced, active regulatory elements from different cell-types can be identified by single experiment of DNase-seq. This technique is based on principle that high throughput method can identify DNase I hypersensitive sites (DHSs) across all genome by analyzing digested fragments with DNase-I and subsequently sequencing them through next-generation sequencing (Song and Crawford, 2010). Active enhancer sequences are mostly present in nuclear dbf2-related (NDRs) and are vulnerable to nuclease activity (Hesselberth, 2009). DHSs are recognized by partial digestion of DNA with DNase-I followed by sequencing of digested fragments that represents the accessible portion of the genome.

Bisulfite (BS)-seq

In DNA methylation, methyl group (-CH$_3$) is added to the fifth residue (5mC) of cytosine’s ring. This conversion is catalyzed by DNA methyl transferases (Law and Jacobsen, 2010; Pelizzola and Ecker, 2011). In animals and animals, DNA methylation is associated with the transcriptional inactivation (Law and Jacobsen, 2010). This mechanism is used by cells to control gene expression and when present at enhancers, it down regulates the expression of target genes (Stadler et al., 2011). Therefore, low DNA methylation levels is an indication of the enhancers (Stadler et al., 2011). The BS conversion can be used to measure the DNA methylation levels across the genome. This approach involves the conversion of unmethylated cytosines to thymines and subsequently by sequencing (BS-seq) (Krueger et al., 2012). BS-seq exhibits allow the accurate identification of partially methylated regions by using computational tools (Burger et al., 2013; Stevens et al., 2013).

Reporter Assay

A remarkable extend in throughput technique that is recently achieved include the parallel use of reporter assays and DNA sequencing technologies. These assays help in recognition of enhancers and their properties, when applied on native genomic sequences. A reporter assay is a tool for testing the enhancer’s sequence (Shlyueva et al., 2014). In situ gene fusion can be done by inserting the promoter-less reporter gene within a transcription unit of the genome. Individual genes and their regulatory elements can be identified on the basis of the differential expression patterns of tagged reported genes (Mathur, 1998; Yamamoto et al., 2003).

Particularly, this method can be used to cloned control fragment and solicitant enhancer upstream of promoter for controlling a reporter gene, along with the introduction into the specific tissue and activity of reporter gene can be measured (Belele et al., 2013; Herrera-Estrella et al., 1985; Sparkes et al., 2006).

DNA affinity purification

Transcription factor binding assay also known as DAP-seq (DNA affinity purification) which combines next-generation sequencing with in vitro expression of affinity-purified TFs to construct epicistrome and cistrome map for many species. DNA libraries were constructed from any source of native genomic DNA, such as preserving cells, gene of interest, tissue specific modifications i.e. DNA methylation that affect TF binding (Bartlett et al., 2017). Subsequently, DNA library is incubated with in-vitro expressed TF, and unbound DNA fragments are washed away. Bound genomic DNA is eluted form TF and sequence through next generation sequencing and resulting genome-wide binding sites were analyzed. This method is fast, inexpensive and easy as compared to ChIP-seq and does not require sample-specific reagents anti-bodies and gene specific primers (Jutras et al., 2012).

Pattern recognition method

This method combines a genetic programing and Finite State Automata to identify promoter sequence in a primary sequence data. This method can take long jump base pairs, which may enable it to help to identify gene specific cis-elements or genes in long genome sequences i.e. human genome (Huang and Ecker, 2018). This method can also use to automatically identify motifs of different length and allow to combine the motif matches using logical functions to find a cis-acting region and identification decision (Howard and Benson, 2003).

DNA binding micro array

DNA binding microarray (PBM) is a comprehensive,
high-throughput method that discover the DNA binding preference of a protein in an unbiased and inclusive manner. DNA binding site data from, PBMs combined with used to predict what function of a TF and it target gene, genes are regulated by a given TF, and how TF may fit into cell transcriptional regulatory network (Wong et al., 2017).

Conclusion

The progress in discovering cis-sequences have significantly extended our knowledge about gene expression and brought a lot of benefits in economical production planning. Cis-regulatory elements (CREs) are important components for the regulation of genetic network, which further control morphogenesis, anatomy and other aspects of developmental biology. Although, several types of cis-elements are reported, which are respond to abiotic and biotic stresses in plants i.e. In rice, OsEREBP1, which binds to the ethylene-responsive element (ERF) GAGCCGCC, has been reported to be expressed in Magnaporthe grisea. Current studies described that NAC, AREB/ABF, DREB1/ CBF and DREB2 regulons have pivotal roles against abiotic stresses in rice. Moreover, the combinations of transcription factors and cis-acting elements play a crucial role in determining the crosstalk between various stress signaling pathways. Some cis sequences are functional in both directions. Many aspects of gene expression regulation remained unexplained and required integrative research and detailed analysis. The cis-sequence diversity, numbers, and awareness that not all of them discovered trigger research. Identification of such sequences and elucidation of their role under stress condition is exceptionally significant for plant growth, and it may help to facilitate the construction of novel and valuable artificial promoters.

Disclaimer: None.
Conflict of Interest: None.
Source of Funding: None.

References

Abarca SMV and Sánchez EC, 2019. Diseño de un proceso para la obtención de una bebida nutritiva a base de Hortalizas como el tomate (Solanum lycopersicum) y el pepinillo (Cucumis sativus). Ciencia Digital. 3(3): 456-472.
41:e155-e155


Kutach AK and Kadonaga JT, 2000. The downstream promoter element DPE appears to be as widely used as the TATA box in Drosophila core


McGarry RC and Ayre BG, 2008. A DNA element between At4g28630 and At4g28640 confers companion-cell specific expression following the sink-to-source transition in mature minor vein phloem. Planta. 228:839-849.


Ohme-Takagi M and Shinshi H, 1995. Ethylene-inducible DNA binding proteins that interact with...


Singh KB, 1998. Transcriptional regulation in plants:
the importance of combinatorial control. Plant Physiol. 118:1111-1120.


Contribution of Authors
Ijaz U: Manuscript write up and literature search
Pervaiz T: Manuscript write up
Ahmed T: Manuscript write up
Seemab R: Manuscript write up
Shahid M: Data Analysis
Noman M: Final approval of article
Nadeem M: Final approval of article
Azeem F: Designed the study