Australian Journal of Crop Science AJCS

AJCS 17(12):950-958 (2023) ISSN:1835-2707
doi: 10.21475/ajcs.23.17.12.poj3789

Computational analysis of regulation and expression of plant DNA repair genes under various abiotic stresses reveals a stochastic correlation

Padinhatta Purayil Amritha[1], Rajeev Anjitha, Balu Sruthi[2], Meenakshi Anilkumar[3], Jasmine Madhukant Shah*[4]

[1]Department of Plant Science, Central University of Kerala, Kasaragod, Kerala, India
Rajeev Anjitha: Department of Plant Science, Central University of Kerala, Kasaragod, Kerala, India
[2]Department of Plant Science, Central University of Kerala, Kasaragod, Kerala, India
[3]Vellore Institute of Technology, Vellore, Tamil Nadu, India
[4]Department of Plant Science, Central University of Kerala, Kasaragod, Kerala, India

***Corresponding author:jasmine@cukerala.ac.in**

Abstract
Biotic/abiotic stresses are known to induce genome instability in all organisms. Plants being sessile, are more vulnerable to stresses. Thus, they have evolved mechanisms to protect their genome from the unpredictably changing environment. DNA repair genes play a critical role in preserving genome integrity. Plants have about 15 different DNA repair pathways comprising more than 100 genes. To gain a better understanding of how plant DNA repair genes perceive stress and regulate their expression, we performed a computational analysis of the putative binding sites of different abiotic stress-responsive transcription factors (ASRTFs) on their promoters, in the model plant *Arabidopsis thaliana*. Of the 12 different ASRTFs that we analysed, MYC2 had the highest number of 1876 putative binding sites. Diversity study in 135 DNA repair gene promoters revealed that these had about 6 to 12 types of putative binding sites for the ASRTFs. While *RAD51* promoter had the highest number of 207 putative binding sites, *RECQ4B* had only 11. Comparison of the *Arabidopsis* eFP browser-based temporal/spatial expression profiles of these genes under nine different abiotic stresses revealed that all of these genes had altered expression under various stresses. Some genes exhibited similar expression profiles. To identify the reasons for their similar expression pattern, we compared the promoter sequences and putative *cis*-regulatory elements such as binding sites for ASRTFs and methylation-prone nucleotide repeats. Our analysis indicated that the gene expression and gene regulation mediated by ASRTFs, and DNA methylation have a stochastic correlation.

Keywords: Abiotic stress-responsive transcription factors, DNA repair genes, Abiotic stress, Methylation, Gene expression, *Arabidopsis thaliana*.
Abbreviations: ASRTFs_Abiotic stress-responsive transcription factors; TFs_Transcription factors; ASRTFBSs_Abiotic stress-responsive transcription factor binding sites; TAIR_The *Arabidopsis* Information Resource; TSS_Transcription start site.

Introduction

Since plants are sessile, they are forced to face changing environmental conditions. Environmental stresses could be biotic or abiotic. The former type includes infections/attacks due to microbes and herbivores. The latter type includes changes in environmental conditions such as temperature, water availability, salinity, and exposure to harmful chemicals/radiations (Khan et al., 2018), which are often mutagenic (Bray and West, 2005). The DNA repair machinery of organisms protects the genome by recognizing and reversing the damage. Plants are known to have more than a hundred DNA repair genes, broadly classified into 15 pathways (Manova and Gruszka, 2015; Ciccia and Elledge, 2010). Because DNA replication and repair are highly conserved across higher eukaryotes (Liu and Huang, 2015) and plants can survive in highly mutagenic conditions (lethal to animals) (Yoshiyama et al., 2013), plants offer an excellent model system for studying the function and regulation of genes involved in DNA repair.

Transcription is the primary target for gene regulation. Gene expression during transcription is regulated by multiple *cis* and *trans* factors (Levy and Darnell, 2002; Latchman, 1990). The *trans* factors act on *cis*-elements located on promoters and regions in and around genes. *Trans* factors mostly include special transcription factors (TFs) and epigenetic regulators such as DNA/histone-modifying enzymes, including various methyltransferases, deacetylases, other chromatin remodeling enzymes, and small RNAs (Kim, 2019). TFs are DNA-binding proteins that bind to unique *cis*-elements, RNA polymerases, and/or other TFs, thereby activating or repressing transcription (Gonzalez, 2015). TFs act in response to various parameters such as growth, development, cell cycling, cell signaling, and stress (Gonzalez, 2015). Abiotic stress-responsive transcription factors (ASRTFs) include families such as MYB, MYC, NAC, DREB/CBF, ABA-responsive TFs, and WRKY (Nakashima et al., 2009; Khan et al., 2018; Hirayama and Shinozaki, 2010). Epigenetic factors also play an important role in gene

regulation during stressed conditions (Chinnusamy and Zhu, 2009). Epigenetic regulation has the potential to maintain long-term regulation of the gene expression. Cytosine methylation is one of the epigenetic mechanisms that provide enhanced tolerance to environmental stresses through symmetric (^{m}CpG and ^{m}CpHpG) and asymmetric (^{m}CpHpH) methylation of the DNA in plants (Sahu et al., 2013; Henderson and Jacobsen, 2007). Methylation status can vary (hypermethylated/hypomethylated), depending on the loci and/or stress. Epigenetic regulation is also conferred by chromatin remodeling (Kim et al., 2008). Small RNAs and microRNAs are important epigenetic regulators at the post-transcription level. (Filipowicz et al., 2005).

The availability of *Arabidopsis'* complete genome and whole transcriptome sequences (Kaul et al., 2000; Zeller et al., 2009) has improved our understanding of various pathways in plants. Though more than a hundred DNA repair genes are known to function in plants, reports on their regulation are scanty. In this work, we report an elaborate *in silico* survey of transcription regulation of almost all known DNA repair genes (about 135), belonging to 15 different pathways in the model plant *A. thaliana*. We generated a comparative profile of putative binding sites of 12 ASRTFs. *Arabidopsis* eFP browser-based temporal/spatial expression profile of DNA repair genes under nine different abiotic stresses such as cold, osmotic change, salt, drought, genotoxic, oxidative, UV-B, wounding, and heat was compared. Based on this pattern, we performed a promoter sequence analysis to understand the nature of binding sites of ASRTFs, and methylation-sensitive sequences. Our computational analysis, elaborate profiling, and comparison can aid in designing further experiments on functional confirmation of DNA repair expression/regulation, and stress tolerance in plants.

Results

The frequency and diversity of putative binding sites for abiotic stress-responsive transcription factors (ASRTFs)

All ASRTFs had sites in any of the 135 DNA repair genes (Table S3). MYC 2 had the highest number of 1876 putative binding sites, followed by others (Fig. 1), with ABI3 having the least number of only three. MYC 2 also had sites invariably in all 135 genes, followed by WRKYs 18, 33, and 40, with sites in 134 genes (Table S3). The DNA repair pathway-wise analysis revealed that the highest frequency of putative binding sites was for MYC 2 or WRKY 40 (Fig. S1). The highest number of 207 putative ASRTFBSs, for 8 different ASRTFs, was observed in *RAD51* (Fig. 2 G) of the HR pathway. The number of putative ASRTFBSs often did not correlate with their diversity (Fig. 2) because the highest diversity (for 12 ASRTFs) was exhibited by two other genes, *MFD* and *MBD4* (Fig. 2 B and C). *RECQ4B,* belonging to the HR pathway, had the least number and diversity of 11 and 6 putative sites, respectively (Fig. 2 G).

Distribution of putative abiotic stress-responsive transcription factor binding sites

TF binding sites can be located both upstream and downstream of the transcription start site (TSS) (Porto et al., 2014). Their numbers and position are important in regulating transcription. To determine if there was any preference for their location, their numbers were studied in 5 regions of the promoter (Table S4). The distribution profiles generated (Fig. 3) indicated that most ASRTFs were scattered in all the regions, with more numbers upstream of TSS.

Expression profiling of DNA repair genes under abiotic stresses

Since all DNA repair genes exhibited the presence of many putative ASRTFBSs, we compared their expression profiles under nine abiotic stresses (Table S5). Most genes showed altered expression under all stresses. Some gene pairs/groups, belonging to the same/different pathways, exhibited similar expression patterns. It could be possible that they have a similar regulatory mechanism. To obtain a deeper insight, such gene profiles were identified (Table 1) and chosen for further analysis.

Comparison of sequence homology in the promoter region

Since the majority of transcription-related regulatory mechanisms depend on promoters, it could be possible that the gene groups with comparable expression patterns (Table 1) share similar sequences in the promoter regions. To verify this hypothesis, we compared the promoters of gene groups exhibiting similar as well as dissimilar expression patterns and counted the number of nucleotide stretches having ≥5 bases homology (Table S6). The closely related gene pair *BRCA2A*/*BRCA2B* shared the maximum number of 48 homologous stretches, with the longest stretch of 109 bp. Rest of the similarly expressing gene pairs had 0-31 numbers of homologous stretches, with length up to 15 bp. Groups containing more than two genes did not have any sequence with ≥ 5 nucleotide homology. The average number of stretches with ≥ 5 nucleotide homology was 8.06.

To see whether such stretches were present in promoters with dissimilar expression patterns, we analysed the status of 31 gene pairs/groups. Similar to the former case, here the number of homologous stretches ranged from 1-35, and the longest stretch of 12 nucleotides was shared by *RECQ4B*/*KU70* (Table S6). The average number of stretches with ≥ 5 nucleotide homology was 9.55, which was higher than the former case. Thus, in contrast to our hypothesis, the expression pattern was not related to the overall sequence similarity of promoters.

Comparison of putative ASRTFBs in the regions of homology

Genes exhibiting similar as well as dissimilar expression patterns randomly shared a few common putative ASRTFBSs (Table S7). *POLQ*/*Mag1p*, the similarly expressing gene pair, shared the highest number of 27 putative ASRTFBSs. *XRCC1/ATMSH7*, the pair with dissimilar expression patterns shared a higher number of 36 putative ASRTFBSs.

Analysing the methylation-prone sequences in the promoters

DNA methylation is an important gene regulation in plants. Sequence repeats such as CG, CHG, and CHH (H represents A, T, or C) in the promoter regions are prone to methylation/demethylation in plants. Therefore, we wanted to see if there was any correlation in the distribution of such sequences in the promoters of genes exhibiting similar/dissimilar expression patterns. The promoters showed the presence of about 13 to 42 numbers of randomly distributed repeats, belonging to any of the 12 categories (CAT, CTT, CCT, CTC, CAA, CTA, CCA, CAGCTT, CAGCTC, CAG, CTTCCT, CAGCAA) (Table S8). The overall frequency of these sites was 25.9 and 28.41 in genes

showing similar and dissimilar expression patterns, respectively. CTC repeats were present more frequently (4-17 times) in most of the genes analysed (Table S8). Similar distribution of CAT, CTT, CCT, CTC, and CAA repeats was observed only in one gene pair *RAD51C* and *DINB1* (Table S9), which also had similar expression patterns.

Co-expression analysis

Co-expression analysis was done for all 135 repair genes under abiotic stress. Expression correlation varied in the root (Fig. 4a) and shoot (Fig. 4b); seven and five correlation networks were obtained, respectively. The largest network was from roots and comprised eight genes, *BRCA1, TSO2, RAD51, PARP1, DINB1, RAD17, ADPRT,* and *RPA1*. The co-expression data obtained here did not correlate with our gene expression pattern analysis.

Discussion

The sessile nature of plants demands greater adaptability and crucial maintenance of genome integrity under stressful environments (Tuteja et al., 2009). Many previous studies have reported the stress-induced alteration of gene expression in plants (Bhargava and Sawant, 2013; Haak et al., 2017). The availability of high-throughput data on gene expression, and various sequence analysis/characterization-related databases has opened up new vistas for a deeper understanding of gene regulation (Sanchita and Sharma, 2015; Ko and Brandizzi, 2020). Hence, we performed a detailed *in silico* study of almost the entire DNA repair system of *A. thaliana*, comprising 135 genes of 15 pathways. A comparative analysis of putative ASRTFBSs on their promoters, and gene expression patterns under nine different abiotic stresses was performed. We then tried to understand the reason for the similarity in expression patterns by analysing their correlation with putative regulatory regions such as ASRTFBSs and methylation-prone sequences.

Our detailed *in silico* study on profiling the numbers and distribution of putative ASRTFBSs indicated their invariably wide range of presence on the promoters of all DNA repair genes. There are previous reports of other gene promoters, such as those of the heat-shock protein-encoding genes in *Arabidopsis* (Li et al., 2010), genes for photoperiod and vernalization in *Arabidopsis*, *Brachypodium*, wheat, and barley (Peng et al., 2016), and RNA-dependent RNA polymerases in *Arabidopsis* and other crop plants (Prakash and Chakraborty, 2019), whose profiles for putative binding sites of various TFs have been studied. In our study, MYC2 and ABI3 had the highest and least number of putative binding sites, respectively. MYC2 is a negative regulator of salt stress tolerance and is involved in the jasmonic acid-associated defence signalling pathway (Kazan and Manners, 2013; Verma et al., 2019). ABI3 is a seed germination TF, involved in the dehydration stress response, including stress recovery (Bedi et al., 2016). The sites for MYC2 were invariably present in all 135 genes analysed, followed by those for three WRKY TFs, 18, 33, and 40, all with sites in 134 genes, except *RECQ4B*. However, *RECQ4B* had an exceptionally low number of putative ASRTFBSs in comparison to all other genes. *RECQ4B* codes for a helicase and previous reports indicated that it is not required to repair DNA damage due to various genotoxic agents (Hartung et al., 2007). Probably, that could be the reason for the fewer numbers of putative ASRTFBSs on its promoter. The popularity of MYC 2 and WRKY 18, 33, and 40 in DNA repair genes indicates that these TFs could be playing a crucial role in their transcription regulation under abiotic stresses. There are many previous reports confirming the involvement of various WRKY TFs in regulating gene expression under abiotic stresses (Phukan et al., 2016; Shahzad et al., 2021). DNA repair pathway-wise analysis revealed that both the highest and least numbers of putative ASRTFBSs were located in genes belonging to the HR pathway (in *RAD51* and *RECQ4B*, respectively). The number of sites did not correlate with the diversity, as *MBD4* and *MFD* genes from BER and NER-related categories exhibited the highest diversity. A previous report conveyed that most of the DNA repair gene's promoters in *Arabidopsis* plants are rich in the distribution of putative binding sites for biotic stress-related TFs (Joseph et al., 2021). Other *in silico* reports include those regarding the presence of biotic/abiotic stress-related TF binding sites in the promoters of other genes such as germin-like protein genes (Das et al., 2019) and CAMTA gene family (Gain et al., 2022) of rice, *Pathogenesis-related* genes of *Arabidopsis* (Joseph et al., 2021), and various drought/salt-responsive genes in *Solanum tuberosum* (Ain-Ali et al., 2021; Sanchita and Sharma, 2015). In most of these reports, the expression of these genes also changed under various abiotic/biotic stresses. Hence, the presence of a large number of putative ASRTFBs on all DNA repair genes indicates that their expression could be altered by abiotic stresses.

Previously many studies reported on stress-induced expression profiles of multiple genes based on transcriptome as well as quantitative PCR (qPCR) analysis in plants (Samarina et al., 2020; Sharma et al., 2009). Thus the availability of expression profiles of various plant genes allows functional studies under developmental as well as stressed conditions. Accordingly, we complied the expression patterns of all 135 genes, under nine different abiotic stresses, obtained from the *Arabidopsis* eFP Browser database. Since most of the previous studies in plants focused on the effect of individual or a couple of stresses, the reports on the comparison of expression under multiple abiotic stresses were scanty (Chen et al., 2019; Arslan et al., 2021). A previous report on transcriptome dataset-based expression comparison of DNA repair genes indicated altered expression patterns under biotic stresses (Joseph et al., 2021). These genes also varied in the distribution of putative binding sites for biotic stress-responsive transcription factors. Our compilation showed that the expression of most genes was invariably altered due to all nine abiotic stresses, and some gene pairs/groups, belonging to the same or different repair pathways, exhibited similar patterns under the same stresses. Expression analysis of *histone deacetylase* gene family members in cotton plant under four different abiotic stresses such as metal, salt, cold, and drought resulted in similar expression patterns under same stress (Imran et al., 2020). In *Arabidopsis*, the expression of 13 genes from the *aquaporin* family under abiotic stresses such as drought, cold, high salinity, and ABA treatment was similar in certain cases depending upon the plant part and the duration of stress (Jang et al., 2004). Studies on the effect of multiple stresses can represent the genetic correlation between multiple traits (Thoen et al., 2017). Also, the same ASRTFs can bind on the same *cis*-elements under multiple stresses (Lata and Prasad, 2011).

Table 1. Genes having similar expression patterns under various stresses, based on the eFP browser data

Sl. No.	Category	Gene name	Accession Number**	S/ R*	Abiotic stresses								
					Cold	Osmotic	Salt	Drought	Genotoxic	Oxidative	UV-B	Wounding	Heat
1	NER	*XPB1*	AT5G41730	R	↕	↕	↕	↕	↗	↕	↕	↕	↕
	NER	*RAD23D*	AT5G38470	R	↕	↕	↕	↕	↗	↕	↕	↕	↕
	NER	*XPG*	AT3G28030	R	↕	↕	↕	↕	↗	↕	↕	↕	↕
2	NER	*GTF2H1*	AT3G61420	R	↘	↕	↕	↕	↕	↕	↕	↕	↕
	BER	*FPG*	AT1G52500	S	↘	↕	↕	↕	↕	↕	↕	↕	↕
	HR	*RAD51C*	AT2G45280	S	↘	↕	↕	↕	↕	↕	↕	↕	↕
	ODP	*DINB1*	AT1G49980	S	↘	↕	↕	↕	↕	↕	↕	↕	↕
	HR	*RAD52-1A*	AT1G71310	R	↘	↕	↕	↕	↕	↕	↕	↕	↕
3	HR	*RAD5A*	AT5G22750	R	↕	↕	↕	↕	↕	↕	↕	↕	↕
	HR	*ATM*	AT3G48190	R	↕	↕	↕	↕	↕	↕	↕	↕	↕
	ODP	*POLB*	AT1G10520	S	↕	↕	↕	↕	↕	↕	↕	↕	↕
	MMR	*MSH4*	AT4G17380	R	↕	↕	↕	↕	↕	↕	↕	↕	↕
	NER	*XPG*	AT3G28030	S	↕	↕	↕	↕	↕	↕	↕	↕	↕
	OIGWASDRF	*SNM1B*	AT1G27410	S	↕	↕	↕	↕	↕	↕	↕	↕	↕
4	HR	*XRCC3*	AT5G57450	S	↕	↘	↕	↕	↕	↕	↕	↕	↕
	NER	*AtNAP1;1*	AT4G26110	S	↕	↘	↕	↕	↕	↕	↕	↕	↕
5	EAPN	*EX01*	AT1G29630	S	↕	↕	↘	↕	↕	↕	↕	↕	↕
	PR	*PHR1*	AT4G28610	S	↕	↕	↘	↕	↕	↕	↕	↕	↕
	BER	*p3MAG1*	AT1G75090	R	↕	↕	↘	↕	↕	↕	↕	↕	↕
6	ODP	*POLH*	AT5G44740	R	↕	↕	↕	↕	↑	↕	↕	↕	↕
	OBF	*APE1*	AT2G41460	S	↕	↕	↕	↕	↑	↕	↕	↕	↕
	NER-RELATED	*CSB*	AT2G18760	S	↕	↕	↕	↕	↑	↕	↕	↕	↕
	NER	*RAD 16*	AT1G02670	S	↕	↕	↕	↕	↑	↕	↕	↕	↕
7	ODP	*POLQ*	AT4G32700	S	↕	↕	↕	↕	↓	↕	↕	↕	↕
	BER	*Mag1p*	AT1G75230	S	↕	↕	↕	↕	↓	↕	↕	↕	↕
8	HR	*BRCA2*	AT5G01630	S	↕	↘	↕	↕	↕	↕	↕	↗	↕
		BRCA2A	AT4G00020	S	↕	↘	↕	↕	↕	↕	↕	↗	↕
9	HR	*BRCA2*	AT5G01630	R	↕	↘	↕	↕	↓	↕	↓	↓	↕
	HR	*BRCA2A*	AT4G00020	R	↕	↘	↕	↕	↓	↕	↓	↓	↕
10	HR	*SMC6B*	AT5G61460	S	↕	↕	↕	↕	↕	↑	↕	↕	↕
	NHEJ	*PRKDC*	AT1G50030	R	↕	↕	↕	↕	↕	↑	↕	↕	↕
11	OTHER BER FACTORS	*APEI*	AT2G41460	R	↓	↕	↕	↕	↕	↕	↘	↕	↕
	NHEJ	*OSB1*	AT1G47720	R	↓	↕	↕	↕	↕	↕	↘	↕	↕
12	NER	*PCNA*	AT2G29570	S	↕	↕	↘	↘	↕	↘	↕	↕	↕
		LIG1	AT1G08130	S	↕	↕	↘	↘	↕	↘	↕	↕	↕

*S and R indicate shoot and root, respectively; ** Accession numbers from TAIR database; NER is nucleotide excision repair; BER is base excision repair; HR is homologous recombination; ODP is OTHER DNA POLYMERASES (catalytic subunits); MMR is mismatch repair; OIGWASDRF is other identified genes with a suspected DNA repair function; EAPN is editing and processing nucleases; PR is photo reactivation; OBF is BER factors; NHEJ is non-homologous end joining. ↕ indicate varied expression either increase or decrease irrespective of the time of exposure; ↗ indicate a decrease in expression in the initial time of exposure then gradual increase; ↘ indicate an increase in expression in the initial time of exposure then gradual decrease; ↑ indicate a gradual increase; ↓ indicate a gradual decrease.

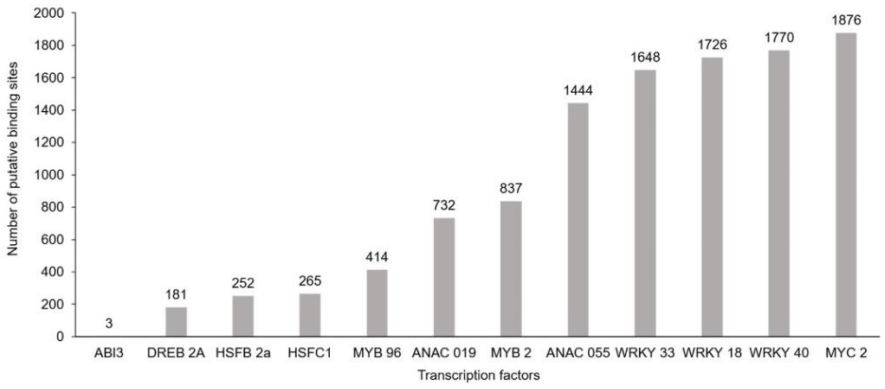

Fig 1. Total number of abiotic stress-responsive transcription factors binding sites in 135 DNA repair genes

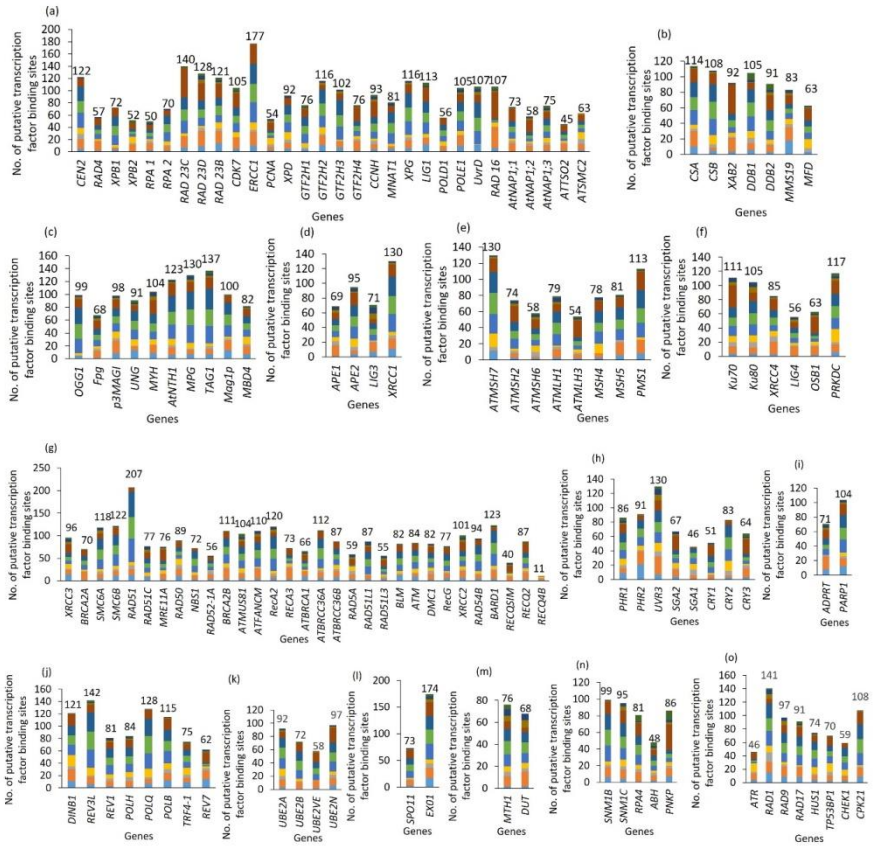

Fig 2. Diversity of abiotic stress-responsive transcription factors binding sites in the promoter regions of 135 DNA repair genes belonging to various categories **a** Nucleotide excision repair (NER) **b** NER-related **c** Base excision repair (BER) **d** Other BER factors **e** Miss-match repair **f** Non-homologous end joining **g** Homologous recombination **h** Photo reactivation **i** Poly (ADP ribose) polymerase enzymes **j** Other DNA polymerases (catalytic subunits) **k** RAD6 pathway **l** Editing and processing nucleases **m** Sanitization of nucleotide pools **n** Other identified genes with a suspected DNA repair function **o** Other conserved DNA damage response genes. ANAC0 19 ■, ANAC0 55 ■, MYB 96 ■, MYB 2 ■, WRKY 18 ■, WRKY 33 ■, WRKY 40 ■, MYC 2 ■, ABI3 ■, HSFB 2A ■, HSFC 1 ■, DREB 2A ■

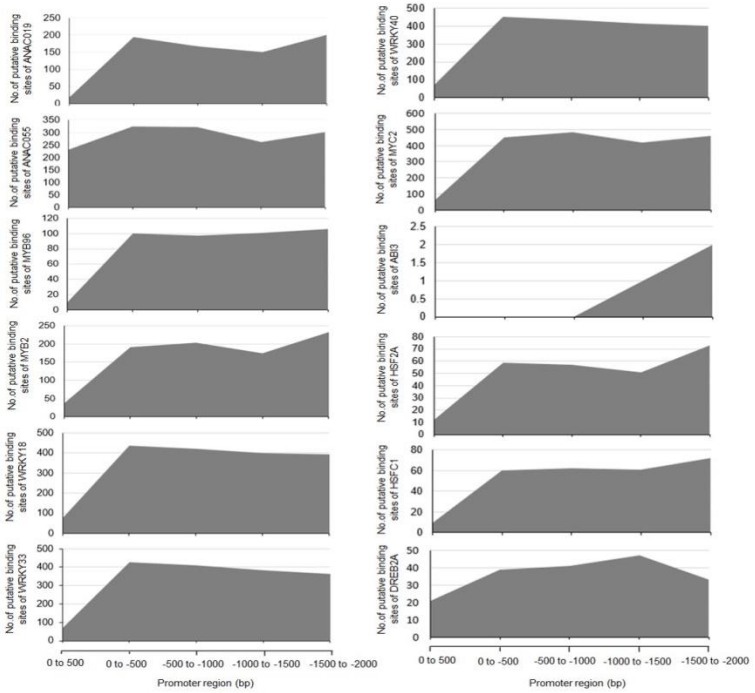

Fig 3. Distribution area graph of putative binding sites of abiotic stress-responsive transcription factors (ASRTFs). The X-axis indicates five regions (in bp) of the promoter sequence, one downstream of the transcription start site (0 to 500), and four upstream (0 to -500, -500 to -1000, -1000 to -1500, and -1500 to -2000). The Y-axis indicates number of putative binding sites for the respective ASRTF.

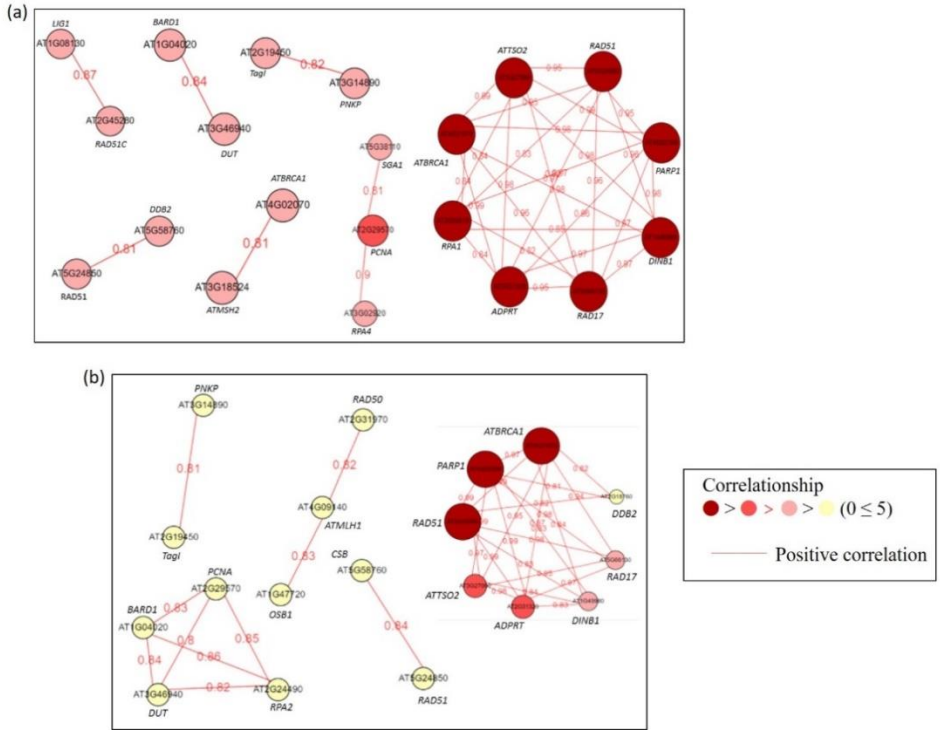

Fig 4. Correlation network under abiotic stress. **a** In root **b** In shoot.

For example, DREB activates various stress-inducible genes under cold, drought, salt as well as heat stresses (Dubouzet et al., 2003; Kidokoro et al., 2015). In *Arabidopsis*, a comparison of gene expression responses under nine abiotic stresses indicated similar expression peaks at different time intervals, that differ with the plant part to which the stress is exposed (Swindell, 2006). We postulated that gene pairs/groups with similar expression patterns might have similar regulatory mechanisms, and we started excavating the sequence information in this direction. A previous study in *Caenorhabditis elegans* revealed that similar expression patterns in the excretory cells were mediated by distinct regulatory elements (Zhao et al., 2005), which supports our hypothesis.

To see whether the promoter regions of similarly expressing genes shared more sequence similarity, we performed an elaborate sequence comparison of the promoter regions of genes exhibiting similar as well as dissimilar expression patterns. Though there were numerous homologous stretches ≥5 bp across the promoter lengths, the overall frequencies of these stretches did not vary between the promoters belonging to both expression categories. This could be probably because of the routine *cis*-regulatory requirements of most DNA repair gene promoters. Since our survey on putative ASRTFBSs indicated that their numbers and distribution vary between the DNA repair genes, we went on to analyse the frequencies of putative ASRTFBSs in the homologous stretches of promoters of genes with similar as well as dissimilar expression patterns. The overall frequencies of most ASRTF binding sites did not vary between the two gene categories. The distribution of TFBS in evolutionarily-related genes can differ due to the neofunctionalization of gene duplicates (Das et al., 2019). We did not find reports on the reason for the correlation between the presence of regulatory elements and their expression similarity.

Apart from TFs, transcription is also maintained by *cis*-elements involved in epigenetic regulation. DNA methylation/demethylation acts as a switch to gene expression. DNA methyltransferases act on cytosine residues, which are then converted to 5-methyl cytosine. 5-methyl cytosine is a sign of epigenetic gene silencing, which mostly hinder the physical interaction of TFs on the promoter region (Pikaard and Scheid, 2014). Among eukaryotes, plant DNA is highly methylated and most of the epigenetic regulation-related reports were based on studies done on plants (Eriksson et al., 2020). The cytosine methylation mainly occurs in repetitive regions like transposable elements, centromeric regions, and also in the gene body either in the promoter or coding regions. There are many reports of abiotic stresses resulting in the altered epigenetic status of many gene promoters in plants (Steward et al., 2002; Pan et al., 2011; Xu et al., 2015). Towards this end, like our analysis with ASRTFBSs, we compared the frequencies of 12 types of cytosine methylation-prone sequences in the promoters of genes exhibiting similar as well as dissimilar expression patterns. It was interesting to note that most gene promoters did have any of the 12 epigenetic regulatory repeats and, CTC was the most common of all. However, we did not observe any obvious difference in their overall frequencies in the promoters of both gene categories. This indicates that expression pattern maintenance probably does not involve the distribution of the above epigenetic regulatory regions. In fact, environmental stresses can alter the pattern of methylation in plant DNA (Peng and Zhang, 2009).

Plant gene regulation involves a complex network of many factors, and some of them have a combinatorial action on gene regulation (Samad et al., 2017; Bemer et al., 2017; Besseau et al., 2012; Roy, 2016). Hence, there could be many more unknown factors responsible for the pattern maintenance of transcription in plants. Our results are based on *in silico* analysis and hence need to be confirmed *in vitro*. Nevertheless, *in silico* predictions in plants show an appreciable correlation when verified *in vitro* (Franco-Zorrilla et al., 2014).

DNA repair genes being highly conserved across the eukaryotes (Lopez-Camarillo et al., 2009), it could be possible that similar to our results in plants, promoters of these genes in animals could also carry an appreciable diversity of ASRTFs. DNA repair being a crucial mechanism for maintaining the genetic integrity of an organism, elaborate profiling can be used for further studies in deeper *in vitro* aspects. There are several reports where overexpression, silencing, or genome editing of various TFs enhanced tolerance to various abiotic stresses in plants (Shahzad et al., 2021; Wang et al., 2016). Detailed *in silico* studies on TF binding sites on the promoters are important for the prediction of regulatory features of gene expression. Such analysis paves the way for *in vitro* functional characterization, and also crop development with enhanced tolerance to abiotic stress.

Materials and methods

Selection of DNA repair genes and abiotic stress-responsive transcription factors (ASRTFs)

A total of 135 DNA repair genes of *A. thaliana* belonging to 15 different categories were selected (Table S1). Details of various repair genes in *A. thaliana* were retrieved from The *Arabidopsis* Information Resource (TAIR) database (https://www.arabidopsis.org/). Twelve different ASRTFs belonging to seven different families were selected for this study (Table S2). Their details were retrieved from the Stress Responsive TranscrIption Factor Database, and previous literature (Lindemose et al., 2013; Nakashima et al., 2012; Chen et al., 2012; Wunderlich et al., 2014).

Identification of the putative promoter sequences and transcription factor binding sites

The putative promoter sequences ranging from 2000 bp upstream to 500 bp downstream of TSS for all DNA repair genes were obtained from ExPATH 2.0 (http://expath.itps.ncku.edu.tw) database. The putative ASRTFBSs were identified using ExPATH 2.0 database, the online tool AthaMap (Steffens et al., 2004; www.athamap.de/), and manually as well.

Calculating the frequency of putative transcription factor binding sites

As the total number of sites in each DNA repair category depended on the number of genes, their frequency was calculated by the formula,

Frequency of ASRTFBSs=Total number of putative ASRTFBSs ÷ Total number of genes analysed

Expression profiling of DNA repair genes

Expression patterns of DNA repair genes under nine different abiotic stresses, cold, osmotic, salt, drought, genotoxic, oxidative, UV-B, wounding, and heat, were obtained from the online tool *Arabidopsis* eFP Browser 2.0 BAR (http://bar.utoronto.ca/efp2/Arabidopsis/Arabidopsis_eFPBrowser2.html).

Finding sequence similarity in the promoter regions

Sequence homology between the promoter regions was analysed using the multiple sequence alignment tool Clustal Omega < EMBL-EBI (https://www.ebi.ac.uk/Tools/msa/clustalo/). The length of the sequences showing 100% homology was calculated manually. Average number of similar stretches was calculated by the formula,

Average number of stretches with ≥5bp homology=
Total number of stretches with≥5bp homology÷Total number of genes

Identification of putative transcription factor binding sites in similar stretches of promoter sequences

The putative sequences of ASRTF binding sites in the homologous stretches were identified using ExPATH 2.0 database. Their locations were identified manually. Their frequencies were calculated by the formula,

Frequency of putative ASRTFBSs in the similar stretches=
Total number of putative ASRTFBSs in the similar stretches÷
Total number of gene pairs analysed

Screening of methylation-prone sequence repeats

Methylation-prone sequences were identified from literature (Bartels et al., 2018), and screened on the promoter region using the tool SnapGene viewer. The average frequency was calculated by the formula,

Average frequency of methylation-prone sequence repeats=
Total number of sequence repeats÷
Total number of selected methylaion-prone sequences

Co-expression analysis

The correlation networks of co-expressing DNA repair genes were analysed using the online tool ExPath 2.0, in reference to Pearson Correlation Coefficient. The correlated genes with an input cut-off value 0.8 were considered for the construction of correlation network. For co-expression analysis, a group of input genes were matched with the microarray datasets under abiotic stressed condition and analysed separately in root and shoot.

Acknowledgments

Authors acknowledge Central University of Kerala for granting the facility for the study, and the Kerala State Council for Science, Technology, and Environment for Amritha PP's fellowship.

References

Ain-Ali QU, Mushtaq N, Amir R, Gul A, Tahir M, Munir F (2021) Genome-wide promoter analysis, homology modeling and protein interaction network of *Dehydration Responsive Element Binding* (*DREB*) gene family in *Solanum tuberosum*. Plos One. 16: e0261215.

Arslan B, İncili ÇY, Ulu F, Horuz E, Bayarslan AU, Öçal M, Kalyoncuoğlu E, Baloglu MC, Altunoglu YC (2021) Comparative genomic analysis of *expansin* superfamily gene members in zucchini and cucumber and their expression profiles under different abiotic stresses. Physiol Mol Biol Pla. 27:2739–2756.

Bartels A, Han Q, Nair P, Stacey L, Gaynier H, Mosley M, Huang QQ, Pearson JK, Hsieh T-F, An Y-QC, Xiao W (2018) Dynamic DNA methylation in plant growth and development. Int J Mol Sci. 19:2144.

Bedi S, Sengupta S, Ray A, Chaudhuri RN (2016) ABI3 mediates dehydration stress recovery response in *Arabidopsis thaliana* by regulating expression of downstream genes. Plant Sci. 250:125–140.

Bemer M, van Dijk AD, Immink RG, Angenent GC (2017) Cross-family transcription factor interactions: an additional layer of gene regulation. Trends Plant Sci. 22:66–80.

Besseau S, Li J, Palva ET (2012) WRKY54 and WRKY70 co-operate as negative regulators of leaf senescence in *Arabidopsis thaliana*. J Exp Bot. 63:2667–2679.

Bhargava S, Sawant K (2013) Drought stress adaptation: metabolic adjustment and regulation of gene expression. Plant breeding. 132:21–32.

Bray CM, West CE (2005) DNA repair mechanisms in plants: crucial sensors and effectors for the maintenance of genome integrity. New Phytol. 168:511–528.

Chen L, Song Y, Li S, Zhang L, Zou C, Yu D (2012) The role of WRKY transcription factors in plant abiotic stresses. Biochim Biophys Acta Gene Regul Mech. 19:120–8.

Chen Y, Zhang B, Li C, Kong C, Yang Y, Gong M (2019) A comprehensive expression analysis of the *expansin* gene family in potato (*Solanum tuberosum*) discloses stress-responsive *expansin-like B* genes for drought and heat tolerances. Plos One. 14:e0219837.

Chinnusamy V, Zhu JK (2009) Epigenetic regulation of stress responses in plants. Curr Opin Plant Biol. 12:133–139.

Ciccia A, Elledge SJ (2010) The DNA damage response: making it safe to play with knives. Mol Cell. 40:179–204.

Das A, Pramanik K, Sharma R, Gantait S, Banerjee J (2019) *In-silico* study of biotic and abiotic stress-related transcription factor binding sites in the promoter regions of rice *germin-like protein* genes. Plos One. 14:e0211887.

Dubouzet JG, Sakuma Y, Ito Y, Kasuga M, Dubouzet EG, Miura S, Yamaguchi-Shinozaki K (2003) *OsDREB* genes in rice, *Oryza sativa* L., encode transcription activators that function in drought-, high-salt-and cold-responsive gene expression. Plant J. 33:751–763.

Eriksson MC, Szukala A, Tian B, Paun O (2020) Current research frontiers in plant epigenetics: an introduction to a Virtual Issue. New Phytol. 226:285–288.

Filipowicz W, Jaskiewicz L, Kolb FA, Pillai RS (2005) Post-transcriptional gene silencing by siRNAs and miRNAs. Curr Opin Struc Biol. 15:331–341.

Franco-Zorrilla JM, López-Vidriero I, Carrasco JL, Godoy M, Vera P, Solano R (2014) DNA-binding specificities of plant transcription factors and their potential to define target genes. P Natl Acad Sci USA. 111:2367–2372.

Gain H, Nandi D, Kumari D, Das A, Dasgupta SB, Banerjee J (2022) Genome-wide identification of *CAMTA* gene family members in rice (*Oryza sativa* L.) and *in silico* study on their versatility in respect to gene expression and promoter structure. Funct Integr Genomics. 22:193–214.

Gonzalez, DH (ed) (2015) Plant transcription factors: evolutionary, structural and functional aspects. Academic Press.

Haak DC, Fukao T, Grene R, Hua Z, Ivanov R, Perrella G and Li S (2017) Multilevel regulation of abiotic stress responses in plants. Front Plant Sci. 8:1564.

Hartung F, Suer S, Puchta H (2007) Two closely related *RecQ* helicases have antagonistic roles in homologous recombination and DNA repair in *Arabidopsis thaliana* P Natl Acad Sci USA. 104:18836–18841.

Henderson I, Jacobsen S (2007) Epigenetic inheritance in plants. Nature. 447: 418–424.

Hirayama T, Shinozaki K (2010) Research on plant abiotic stress responses in the post-genome era: past, present and future. Plant J. 61:104–152.

Imran M, Shafiq S, Naeem MK, Widemann E, Munir MZ, Jensen KB, Wang RR (2020) *Histone deacetylase* (*HDAC*) gene family in allotetraploid cotton and its diploid progenitors: *In silico* identification, molecular characterization, and gene expression analysis under multiple abiotic stresses, DNA damage, and phytohormone treatments. Int J Mol Sci. 21:321.

Jang JY, Kim DG, Kim YO, Kim JS, Kang H (2004) An expression analysis of a gene family encoding plasma membrane aquaporins in response to abiotic stresses in *Arabidopsis thaliana*. Plant Mol Biol. 54:713–725.

Joseph JT, Ajith TG, Shah JM (2021) *In silico* analysis and comparison of pathogenesis-related transcription factor binding sites in *Arabidopsis* DNA repair gene promoters. P Natl A Sci India B. 91:929–945.

Kaul S, Koo HL, Jenkins J, Rizzo M, Rooney T, Tallon LJ, Feldblyum T, Nierman W, Benito MI, Lin X, Town CD (2000) Analysis of the genome sequence of the flowering plant *Arabidopsis thaliana*. Nature. 408: 796–815.

Kazan K, Manners JM (2013) MYC2: The Master in Action. Mol Plant. 6:686–703.

Khan SA, Li MZ, Wang SM, Yin HJ (2018) Revisiting the role of plant transcription factors in the battle against abiotic stress. Int J Mol Sci. 19:1634.

Kidokoro S, Watanabe K, Ohori T, Moriwaki T, Maruyama K, Mizoi J, Yamaguchi-Shinozaki K (2015) Soybean DREB 1/CBF-type transcription factors function in heat and drought as well as cold stress-responsive gene expression. Plant J. 81:505–518.

Kim JH (2019) Chromatin remodeling and epigenetic regulation in plant DNA damage repair. Int J Mol Sci. 20:4093.

Kim JM, To TK, Ishida J, Morosawa T, Kawashima M, Matsui A, Toyoda T, Kimura H, Shinozaki K, Seki M (2008) Alterations of lysine modifications on the histone H3 N-tail under drought stress conditions in *Arabidopsis thaliana*. Plant Cell Physiol. 49:1580–1588.

Ko DK, Brandizzi F (2020) Network-based approaches for understanding gene regulation and function in plants. Plant J. 104:302–317.

Lata C, Prasad M (2011) Role of DREBs in regulation of abiotic stress responses in plants. J Exp Bot. 62:4731–4748.

Latchman DS (1990) Eukaryotic transcription factors. Biochem J. 270:281–289.

Levy DE, Darnell JE (2002) Stats: transcriptional control and biological impact. Nat Rev Mol Cell Bio. 3:651–662.

Li M, Berendzen KW, Schöffl F (2010) Promoter specificity and interactions between early and late *Arabidopsis* heat shock factors. Plant Mol Biol. 73:559–567.

Lindemose S, O'Shea C, Jensen MK, Skriver K (2013) Structure, Function and Networks of Transcription Factors Involved in Abiotic Stress Responses. Int J Mol Sci. 14:5842–5878.

Liu L, Huang M (2015) Essential role of the iron-sulfur cluster binding domain of the primase regulatory subunit Pri2 in DNA replication initiation. Protein Cell. 6:194–210.

López-Camarillo C, Lopez-Casamichana M, Weber C, Guillen N, Orozco E, Marchat LA (2009) DNA repair mechanisms in eukaryotes: special focus in *Entamoeba histolytica* and related protozoan parasites. Infect Genet Evol. 9:1051–1056.

Manova V, Gruszka D (2015) DNA damage and repair in plants–from models to crops. Front Plant Sci. 6:885.

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

Nakashima K, Takasaki H, Mizoi J, Shinozaki K, Yamaguchi-Shinozaki K (2012) NAC transcription factors in plant

abiotic stress responses. BBA-Gene Regul Mech. 1819:97–103.

Pan Y, Wang W, Zhao X, Zhu L, Fu B, Li Z (2011) DNA methylation alterations of rice in response to cold stress. Plant Omics J. 4:364–369.

Peng FY, Hu Z, Yang RC (2016) Bioinformatic prediction of transcription factor binding sites at promoter regions of genes for photoperiod and vernalization responses in model and temperate cereal plants. BMC Genomics. 17:1–16.

Peng H, Zhang J (2009) Plant genomic DNA methylation in response to stresses: Potential applications and challenges in plant breeding. Prog Nat Sci. 19:1037–1045.

Phukan UJ, Jeena GS, Shukla RK (2016) WRKY transcription factors: molecular regulation and stress responses in plants. Front Plant Sci. 7:760.

Pikaard CS, Mittelsten Scheid O (2014) Epigenetic regulation in plants. CSH Perspect Biol. 6:a019315.

Porto MS, Pinheiro MPN, Batista VGL, dos Santos RC, de Albuquerque Melo Filho P, de Lima LM (2014) Plant promoters: an approach of structure and function. Mol Biotechnol. 56:38–49.

Prakash V, Chakraborty S (2019) Identification of transcription factor binding sites on promoter of RNA dependent RNA polymerases (RDRs) and interacting partners of RDR proteins through *in silico* analysis. Physiol Mol Biol Pla. 25:1055–1071.

Roy S (2016) Function of MYB domain transcription factors in abiotic stress and epigenetic control of stress response in plant genome. Plant Signal Behav. 11: e1117723.

Sahu PP, Pandey G, Sharma N, Puranik S, Muthamilarasan M, Prasad M (2013) Epigenetic mechanisms of plant stress responses and adaptation. Plant Cell Rep. 32: 1151–1159.

Samad AFA, Sajad M, Nazaruddin N, Fauzi IA, Murad AMA, Zainal Z, Ismail I (2017) MicroRNA and transcription factor: key players in plant regulatory network. Front Plant Sci. 8:565.

Samarina LS, Bobrovskikh AV, Doroshkov AV, Malyukova LS, Matskiv AO, Rakhmangulov RS, Orlov YL (2020) Comparative expression analysis of stress-inducible candidate genes in response to cold and drought in tea plant [*Camellia sinensis* (L.) Kuntze]. Front Genet. 11:1613.

Sanchita, Sharma A (2015) *In silico* identification of regulatory motifs in co-expressed genes under osmotic stress representing their co-regulation. Plant Gene. 1:29–34.

Shahzad R, Jamil S, Ahmad S, Nisar A, Amina Z, Saleem S, Wang X (2021) Harnessing the potential of plant transcription factors in developing climate resilient crops to improve global food security: Current and future perspectives. Saudi J Biol Sci. 28: 2323–2341.

Sharma R, Mohan Singh RK, Malik G, Deveshwar P, Tyagi AK, Kapoor S, Kapoor M (2009) Rice cytosine DNA methyltransferases–gene expression profiling during reproductive development and abiotic stress. FEBS J. 276:6301–6311.

Steffens NO, Galuschka C, Schindler M, Bülow L, Hehl R (2004) AthaMap: an online resource for *in silico* transcription factor binding sites in the *Arabidopsis thaliana* genome. Nucleic Acids Res. 32:368–372.

Steward N, Ito M, Yamaguchi Y, Koizumi N, Sano H (2002) Periodic DNA methylation in maize nucleosomes and demethylation by environmental stress. J Biol Chem. 277:37741–37746.

1.Swindell WR (2006) The association among gene expression responses to nine abiotic stress treatments in *Arabidopsis thaliana*. Genetics. 174:1811–1824.

Thoen MP, Davila Olivas NH, Kloth KJ, Coolen S, Huang PP, Aarts MG, Bac-Molenaar JA, Bakker J, Bouwmeester HJ, Broekgaarden C, Bucher J, Busscher-Lange J, Cheng X, Fradin EF, Jongsma MA, Julkowska MM, Keurentjes JJ, Ligterink W, Pieterse CM, Ruyter-Spira C, Dicke M (2017) Genetic architecture of plant stress resistance: multi-trait genome-wide association mapping. New Phytol. 213:1346–1362.

Tuteja N, Ahmad P, Panda BB, Tuteja R (2009) Genotoxic stress in plants: shedding light on DNA damage, repair and DNA repair helicases. Mutat Res. 681:134–149.

Verma D, Jalmi SK, Bhagat PK, Verma N, Sinha AK (2019) A bHLH transcription factor, MYC2, imparts salt intolerance by regulating proline biosynthesis in *Arabidopsis*. FEBS J. 287:2560–2576.

Wang H, Wang H, Shao H, Tang X (2016) Recent advances in utilizing transcription factors to improve plant abiotic stress tolerance by transgenic technology. Front Plant Sci. 7:67.

Wunderlich M, Groß-Hardt R, Schöffl F (2014) Heat shock factor HSFB2a involved in gametophyte development of *Arabidopsis thaliana* and its expression is controlled by a heat-inducible long non-coding antisense RNA. Plant Mol Biol. 85:541–550.

Xu R, Wang Y, Zheng H, Lu W, Wu C, Huang J, Yan K, Yang G, Zheng C (2015) Salt-induced transcription factor MYB74 is regulated by the RNA-directed DNA methylation pathway in *Arabidopsis*. J Exp Bot. 66:5997–6008.

Yoshiyama KO, Sakaguchi K, Kimura S (2013) DNA damage response in plants: conserved and variable response compared to animals. Biology. 2:1338–1356.

Zeller G, Henz SR, Widmer CK, Sachsenberg T, Rätsch G, Weigel D, Laubinger S (2009) Stress-induced changes in the *Arabidopsis thaliana* transcriptome analyzed using whole-genome tiling arrays. Plant J. 58:1068–1082.

Zhao Z, Fang L, Chen N, Johnsen RC, Stein L, Baillie DL (2005) Distinct regulatory elements mediate similar expression patterns in the excretory cell of *Caenorhabditis elegans*. J Biol Chem. 280:38787–38794.