Vietnam Journal of Biotechnology **22**(2): 289-304, 2024. DOI: 10.15625/vjbt-18581

COMPARATIVE ANALYSIS OF *CIS*-REGULATORY ELEMENTS ASSOCIATED WITH SALINITY AND DROUGHT TOLERANCE IN RICE (*ORYZA SATIVA* L.) USING *IN SILICO* ANALYSIS

Bui Thi Hai Hoa[1], Nguyen Huy Duong[2], Dinh Thi Thu Le[1], Trinh Thi Thu Hang[1], Vu Kim Thoa[1], Le Thu Thuy[2], Bui Van Ngoc[2,3,✉]

[1]*Institute of Food and Biotechnology - Ha Noi Open University, 101B Nguyen Hien, Hai Ba Trung District, Ha Noi, Vietnam*

[2]*Institute of Biotechnology, Vietnam Academy of Science and Technology, 18 Hoang Quoc Viet Road, Cau Giay District, Hanoi, Vietnam*

[3]*Graduate University of Science and Technology, Vietnam Academy of Science and Technology, 18 Hoang Quoc Viet Road, Cau Giay District, Hanoi, Vietnam*

✉To whom correspondence should be addressed. E-mail: bui@ibt.ac.vn

Received: 25.07.2023
Accepted: 18.06.2024

ABSTRACT

The critical roles of *cis*-regulatory elements (cREs) in the regulation of gene expression in response to environmental stress were reported in previous studies. Although transcription factor families to regulate gene expression in plants are well documented, there is a limited number of cREs related to salinity and drought tolerance in rice to be identified. Therefore, in this study, a comparative analysis and characterization of cREs associated with specific drought and salinity tolerance genes of rice, namely *OsNHX1*, *OsNHX5*, *OsHKT1;1*, *OsHKT2;1*, and *OsSOS1*, was performed using the PLACE and PlantPAN 3.0 databases, along with *in silico* methods. Several *cis*-elements within the core promoter region, including TATA-box, CAAT-box, G-box, DPE, and Y-Patch were identified. Additionally, eight other cis elements: ABRE, MYBRS, MYCRS, NAC-binding site, ACGTATERD1, GT1GMSCAM4, W-box, and DRE, were discovered and suggested to be potentially involved in drought and salinity tolerance in rice. Comparative analysis revealed that *OsNHX1* and *OsHKT1;1* exhibit a higher abundance of cREs compared to the other genes studied. The presence of an increased number of cREs suggests a more complex regulatory network, potentially enhancing the ability of these genes to cope with environmental stressors and fine-tune their responses to changing conditions. Furthermore, understanding the distribution and diversity of cREs across different genes can offer practical implications for genetic engineering and crop improvement strategies. Genes with desirable regulatory profiles, especially those associated with specific stress tolerances, may be prime candidates for genetic manipulation.

Keywords: *Cis*-regulatory elements (cREs), Ion transporters, Rice, Salinity tolerance, *In silico* analysis.

INTRODUCTION

Rice is an important food source, providing energy and nutrition to billions of people globally. However, its growth, yield, and quality are often negatively affected by abiotic stresses (Nguyen, Ferrero, 2006), especially drought and salinity. Drought and salt cause a variety of problems for rice plants (Liu et al., 2022), for example, homeostasis imbalance, ionic toxicity, disruption of photosynthesis, generation of reactive oxygen species (ROS), altered hormonal balance, reduced transport and uptake of nutrients.

Salt tolerance in plants is mediated by a variety of factors, pathways, and molecules. This mechanism requires precise coordination of numerous signaling molecules, such as modifying, adapter, and scaffold proteins. Small chemicals like calcium, glycine, betaine, proline, ROS, ABA, and ion pumps (such as antiporters, proton pumps, and ATP-dependent ion pumps) also play a role in the ability of plants to tolerate salt stress. Among them, several classes of transporters have been identified as key players in maintaining ion homeostasis and conferring salinity tolerance in plants (Yamaguchi et al., 2013; Miyoshi et al., 2010; Hauser, Horie, 2010; Apse, Blumwald, 2007), including HKT, NHX, and SOS1 transporters. These Na^+ transporters are involved in various processes that help to alleviate sodium stress and maintain ion homeostasis.

HKT-type transporters are involved in regulating Na^+ circulation within plants. They facilitate the controlled transport of Na^+ ions from the roots to the shoots, allowing the redistribution of Na^+ ions and minimizing their accumulation in sensitive tissues (Garciadeblas et al., 2003). Another important function is to regulate the balance of Na^+ and K^+ in the cytosol under salinity stress, which helps to prevent their accumulation in the cytoplasm and sensitive organelles. When Na^+ ions accumulate at a high concentration, they reduce the water absorption capacity of roots, leading to a drought-like condition. Thus, it is believed that the HKT family plays a crucial role in protecting rice plants against both drought and salinity stress (Shaban et al., 2023).

Five NHX homologs (OsNHX1 to OsNHX5) have been discovered in rice, and they play important roles in salt stress tolerance by controlling Na+ compartmentalization and maintaining cellular ion homeostasis. These homologs exhibit elevated expression levels under salinity, osmotic stress, and abscisic acid (ABA) treatment, as reported by Fukuda et al. (2011). Examination of their exon-intron structure and phylogenetic analysis of the encoded proteins reveal two distinct subgroups. The first group, comprising proteins OsNHX1 to OsNHX4, is primarily found in the vacuolar membrane and serves to sequester excess Na^+ into the vacuole, thereby removing it from the cytoplasm. In contrast, OsNHX5 is restricted to the Golgi apparatus, where it regulates endomembrane pH homeostasis, which is required for protein trafficking (e.g., proglutelins) and potentially other ion transport mechanisms (Farooq et al., 2021; Zhu et al., 2019).

Another antiporter gene, SOS1, which is involved in the transport of ions through the plasma membrane, was found in *Arabidopsis*. The expression of *SOS1* is regulated by the *SOS2/SOS3* pathway and was found in epidermal cells at the tips of roots and in parenchyma cells at the boundary between the xylem and symplast in roots, stems, and leaves (Shi et al., 2002).

In rice, the homolog of *AtSOS1* was discovered and named *OsSOS1 (*Martínez-Atienza et al., 2007; Shi et al., 2000). Studies conducted on *OsSOS1* have provided valuable insights into its function and regulation in response to drought and salinity stresses.

Some transcription factors (TFs), including Bzip (Yang et al., 2009), AP2/ERF (Mie et al., 1999), WRKY (Jiang, Deyholos, 2009), NAC (Tran et al., 2004), bHLH (Jiang et al., 2009), and MYB (Cui et al., 2013), are essential regulators in connecting drought and salt sensory pathways to plant tolerance responses. The promoter region of transporter genes, including *NHX*, *SOS*, and *HKT* gene families, harbors these transcription factors, which significantly contribute to plant development and salt and drought stress responses through ABA-mediated pathways (Basu and Roychoudhury, 2014; Liang et al., 2023).

Despite extensive studies on the transporter genes of rice under drought and salinity conditions, many aspects still require further exploration and research. In this study, the *cis*-elements of promoter regions of 5 genes (*OsNHX1*, *OsNHX5*, *OsHKT1;1*, *OsHKT2;1* and *OsSOS1*) are in silico compared and analyzed. Candidate genes involved in stress tolerance and adaptation could be prioritized by examining the abundance of stress-related motifs. The comparison of motif sequences in genes related to drought and salt stress would enhance our understanding of the regulatory mechanisms and molecular pathways underlying plant responses to these environmental challenges, as well as provide insights into the evolutionary conservation and divergence of stress-responsive genes.

MATERIALS AND METHODS

Data source

In this study, we obtained the sequences, chromosomal location, length (bp), exon number, and general protein information of five transporter genes, *OsNHX1*, *OsNHX5*, *OsHKT1;1*, *OsHKT2;1*, and *OsSOS1*, from the genome of the Nipponbare cultivar, which is available in the Phytozome v13 database (https://phytozome-next.jgi.doe.gov/). After identifying the 5' untranslated regions (5'UTR) of each gene, a 1.5 kb sequence located before the ATG codon was extracted for research.

Transcription starts site and promoter region identification

The transcription start site (TSS) and promoter region of each gene were determined using two promoter predictors: the Neural Network Promoter Prediction (NNPP) version 2.2 (https://www.fruitfly.org/seq_tools/promoter.html) and TSSPlant (http://www.softberry.com/berry.phtml?topic=tssplant&group=programs&subgroup=promoter).

The NNPP tool is a computational method that uses neural network algorithms to predict the location of the promoter region in a DNA sequence and the minimum promoter score (between 0 and 1) (Reese, 2001). A section of DNA sequence spanning 1.5 kilobases (kb) upstream of ATG was exacted from each gene sequence, then subjected to analysis using the NNPP program. According to Kuwano et al. (2011), the promoter region of a gene in rice is typically located upstream of the transcription start site (TSS) and extends approximately 1.5 kb before the start codon. Before running, a default cut-off value of 0.8 was applied for

eukaryotes. A cut-off value of 0.8 means that any predicted promoter region with a predictive score equal to or higher than 0.8 is considered a potential promoter. The program scanned the sequences and displayed the transcription start site prominently. In cases where multiple transcription start site regions were identified within a sequence, the site with the highest prediction score was considered the most reliable and accurate.

TSSPlant, a recently developed promoter prediction tool, can significantly improve accuracy by taking DNA sequences as input and analyzing them to identify potential promoter regions. It considers various sequence motifs, nucleotide compositions, and other characteristics associated with plant Pol II promoters (Shahmuradov et al., 2017). The upstream sequences of the start codon (1.5 kb) from the studied genes were also submitted to the TSSPlant program. A range of sequence-based features, including promoter motifs, nucleotide composition, and structural characteristics of the DNA sequence, which are closely associated with transcriptional initiation and promoter regions, were examined. Then, the potential TSS within the provided sequences was predicted based on the TSS score, which refers to the confidence score or probability. Typically, a higher TSS score indicates a more confident prediction, as well as a higher probability of accuracy for a predicted TSS.

Determination of motifs in transporter gene promoter regions

Based on a 1.5-kb sequence region upstream of each gene (in FASTA format), the cis-regulatory elements in the promoter regions of five transporter genes were analyzed using the New PLACE database (https://www.dna.affrc.go.jp/PLACE/?action=newplace). Consensus sequences of cREs were checked using PlantPAN 3.0 (http://plantpan.itps.ncku.edu.tw/plantpan3/index.html).

Cis-elements across the promoter sequences were then compared to obtain common motifs shared among the genes and consensus or similar sequences in the motif patterns. Subsequently, the functional significance of the identified cREs related to drought and salt tolerance was investigated. Available information on TFs binding to these motifs and their roles in gene regulation was searched based on literature or databases.

RESULTS AND DISCUSSION

Sequence retrieval

Gene sequences of 2 members of the OsNHX gene family (*OsNHX1* and *OsNHX5*), 2 members of the OsHKT gene family (*OsHKT1;1* and *OsHKT2;1*), and 1 member of the OsSOS gene family (*OsSOS1*) were retrieved from the Phytozome database. The characteristics of each transporter gene sequence, including gene name, transcript ID, genome size (bp), number of exons, and type of ion transporter, are shown in Table 1. The transporter gene sequences obtained exhibit different nucleotide lengths, ranging from 2286 to 14451 bp. They all have coding sequences interrupted by introns, and the number of exons varies from three to twenty-three.

Table 1. Characteristics of identified transporter genes.

Gene name	Transcript ID (from Phytozome)	Number of exons	Genome size (bp)	Type of ion transporter
OsNHX1	LOC_Os07g47100.2	14	4924	Na^+/H^+ antiporter
OsNHX5	LOC_Os09g11450.1	22	10377	Na^+/H^+ antiporter
OsHKT1;1	LOC_Os04g51820.2	3	2841	Na^+/K^+ symporter
OsHKT2;1	LOC_Os06g48810.1	3	2286	Na^+/K^+ symporter
OsSOS1	LOC_Os12g44360.2	23	14451	Na^+/H^+ antiporter

Determination of transcription start sites and promoter regions

The precise location of transcription start sites (TSSs) and the promoter region within a gene is essential for mapping *cis*-regulatory elements and investigating gene regulation mechanisms (Wang et al., 2007). In this study, TSSs were identified from 1.5 kb upstream of the start codon of each targeted gene sequence using the NNPP version 2.2 tool (standard predictive score = 0.8), and the TSSPlant tool. The analysis in Table 2 reveals the relative positions of all TSSs and a significant variation in the number of transcriptions start sites depending on the method used.

Table 2. Potential TSSs obtained by NNPP and TSSPlant analysis.

Promoter	NNPP version 2.2 tool			TSSPlant tool		
	Number of TSS	Distance between TSS and ATG (Predictive score*)	Distance between TATA-Box and ATG (Score)	Number of TSS	Distance from the start codon (TSS score)	Distance between TATA-Box and ATG (Score)
OsNHX1	4	-1122 (0.99)	Not predicted	3	-1267 (1.87)	absent
		-789 (0.89)	Not predicted		-728 (1.99)	absent
		-373 (1.00)	Not predicted		-322 (1.98)	-352 (5.09)
		-321 (0.94)	Not predicted			
OsNHX5	4	-1090 (1.00)	Not predicted	3	-1080 (1.83)	-1121 (7.78)

		-503 (0.90)	Not predicted		-489 (1.87)	absent
		-406 (0.89)	Not predicted		-182 (2.00)	absent
		-205 (0.99)	Not predicted			
OsHKT1;1		-1288 (0.83)	Not predicted		-1251 (2.00)	absent
	2	-539 (0.92)	Not predicted	3	-535 (1.76)	-568 (7.96)
					-179 (1.99)	-212 (5.94)
OsHKT2;1	2	-503 (0.93)	Not predicted	2	-382 (1.98)	absent
		-56 (0.95)	Not predicted		-53 (1.98)	-87 (8.19)
OsSOS1	3	-607 (0.99)	Not predicted	4	-1193 (1.87)	absent
		-309 (0.91)	Not predicted		-885 (2.00)	absent
		-171 (0.92)	Not predicted		-521 (1.97)	-554 (4.42)
					-168 (1.97)	-202 (6.90)

*****: Prediction score and TSS score are different, they are calculated from two different tools (NNPP and TSSPlant) and cannot be compared with each other.

The TSS score in the TSSPlant tool and the predictive score in the NNPP version 2.2 tool are valuable indicators to evaluate the likelihood of a position in a gene being a transcription start site (TSS). Although they are specific to their respective tools and may have different calculation methods and interpretations.

For TSS prediction of the *OsNHX1* gene, the two least different results obtained from the NNPP and TSSPlant tools are at the positions -321 and -322, respectively. Both results received high accuracy prediction scores compared with others, 0.94 (in NNPP) and 1.98 (in TSSPlant). Moreover, the TSSPlant tool predicts the location of the TATA box of *OsNHX1* (-352) with a score of 5.09.

In terms of the *OsNHX5* gene, the two least different results obtained from the NNPP and TSSPlant tools are at positions -205 and -182, respectively. They also received high accuracy prediction scores compared with others: 0.99 (in NNPP) and 2.00 (in TSSPlant). The TATA box was not predicted by TSSPlant. However, the TATA box position, TSS sequence, and some important *cis*-elements of the *OsNHX* genes could be obtained from the results of Fukuda et al. (2011). According to the author, the

TSS sequence positions of the *OsNHX1* and *OsNHX5* genes are -313 and -150, respectively, relative to the ATG position. This prediction is similar to our results.

In predicting the TSS of the *OsHKT1;1* gene, we obtained positions of -539 using NNPP and -535 using TSSPlant programs, with corresponding accuracy scores of 0.92 and 1.76, respectively. The prediction score of 1.76 in TSSPlant is not higher than other prediction scores we get from TSSPlant. However, the TATA box positioned reasonably at -568 with a high score of 7.96.

In the case of the *OsHKT 2;1* gene, the least different results obtained from the NNPP and TSSPlant tools were the positions at -56 and -53, respectively, with the higher prediction accuracy scores compared with others being 0.95 (in the NNPP) and 1.98 (in the TSSPlant). The position of the TATA box is predicted by the TSSPlant at -87 with a score of 8.19.

The closely related results from the NNPP and TSSPlant tools were also received for TSS prediction of the *OsSOS1* gene at positions -171 and -168, respectively, with relatively higher prediction accuracy scores compared with others -0.92 (in NNPP) and 1.97 (in TSSPlant). Furthermore, the TATA box is in a reasonable position at -202 and has a high score (6.90).

The ATG codon serves as the initiation site for translation, and the region between the TSS and the ATG is known as the 5' untranslated region (5' UTR). In eukaryotic genomes, the typical distance between the TSS and the ATG codon ranges from around 500 to 1500 nucleotides (de Medeiros Oliveira et al., 2021). However, this distance could be shorter or longer, depending on specific genes. Alternative splicing and TSSs can further contribute to the variable distance between the TSS and the ATG codon.

When testing long sequences of intergenic regions in plant genomes, Shahmuradov et al. (2017) evaluated and compared the TSSPlant with other tools, including NNPP version 2.2, Proscan, and EP3. Among promoters within 1100-base pair regions of 55 plant protein-coding genes, which had experimentally confirmed transcription start sites (TSS) positioned at 1001 base pairs (PlantProm DB database), TSSPlant demonstrated the highest accuracy among tested programs (Sensitivity (Sn) = 72%, a harmonic mean of precision and accuracy (F1) = 47%), followed by TSSP (Sn = 63.6%, F1 = 41.2%), NNPP (Sn = 51%, F1 = 31%), EP3 (Sn = 20%, F1 = 31%) and Proscan (Sn = 9%, F1 = 15%) (Shahmuradov et al., 2017). Therefore, the promoter region prediction of TSSPlant is highly appreciated compared to NNPP.

Thus, the potential TSS positions selected for the five genes *OsNHX1*, *OsNHX5*, *OsHKT1;1*, *OsHKT2;1* and *OsSOS1* are -322, -182, -535, -53, and -168, respectively.

Motifs in the promoter regions of genes encoding transporters

The promoter core elements (including TATA-box, G-box, and CAAT-box) play a crucial role in gene transcription regulation. Informative analysis of cREs in the promoter regions of the five studied genes revealed the presence of numerous factors involved in gene regulation. Under analysis by the NewPLACE tool and PlantPAN 3.0, *OsNHX1*, *OsNHX5*, *OsHKT1;1*, *OsHKT2;1* and *OsSOS1* promoters gained 92, 78, 95, 97, and 83 types of cREs, respectively. However, in this study, only core promoter elements relating to salt and drought tolerances, such

as ABRE, GT1GMSCAM4, MYBRS, MYCRS, NAC-binding site, W-box, and DRE, are considered.

Table 3 presents the most analyzed cREs, with the exception of DPE, in the *OsNHX1* promoter. Meanwhile, the remaining genes lack more than one type of cRE. This suggests that *OsNHX1* may have a more diverse and complex expression mechanism than the other genes.

Table 3. Position of cREs in the promoter region relates to the TSS of the genes (NewPLACE & PlantPAN 3.0 source).

Gene name / **CRE name**	***OsNHX1***	***OsNHX5***	***OsHKT1;1***	***OsHKT2;1***	***OsSOS1***
Core region					
TATABOX (TBP binding site)	-31, -30(-) TTTATATA, TATATAA	Absent	-34, -35(-) TATAAAA, TTTATA	-32(-), -35 TTATTT, TATAAAT	-35 TATAAAT,
CAATBOX	-156(-), -156(-), -159, -63(-) CAAT, CCAAT, CAAAT	-157 CAAT	-121(-), -109, -110 CAAT, CCAAT	-144(-), -108, -103(-), -57, -45(-), -109, -58, -150 CAAT, CCAAT, CAAAT	-109(-), -95(-) CAAT, CAAAT
DPE	Absent	+28 AGTCC	Absent	Absent	Absent
G-box (recognized by bZIP)	-70(-), -70, -339, -339(-), -71(-),-72,-72(-) CACGTG, CACGTGG, CACGTGGC, MCACGTGGC	Absent	Absent	Absent	Absent
Y-Patch	-43 CCTCCCTC	-30 TTCTCC	-45(-) TTCTCTCT	-52, +15 CTCCTCC, CTTTCTC	-28(-), -50(-) CCTCTC, CTTTTC
Salinity, drought stress					
ABRE (ABA-responsive element)	-69, -70(-), -338, -339(-), -1210, -1211(-), -1489, -70, -71 (-), -339, -1211, -1212(-),-	Absent	-574, -298, -443(-), -575 ACGTG, MACGYGB	-1489, -1490 ACGTG, MACGYGB	Absent

	72(-),-72(-),-72, -73(-),-72(-),-72(-) ACGTG, MACGYGB, YACGTGGC, ACGTGKC, ACGTGGC, RYACGTGGYR, CACGTGGC, MCACGTGGC				
MYBRS (MBS)	-99, -163(-), -207(-),-265, -366, -1064, -648, -1289(-), -1312, -648(-), -1289, -1289 MACCWAMC, GGATA, WAACCA, CNGTTR, YAACKG, TAACTG	-166(-), -718(-), -1118(-), -274, -277(-), -719, -1403(-), -277, -277, -479, -502(-), -878(-),-565(-), -932, -1330(-) CCWACC, CNGTTR, YAACKG, TAACTG, GGATA, WAACCA	-90(-), -1079, -117(-), -129, -413, -981, -303, -1091, -1268(-), -504(-), -504, -413, -129(-), -504, -413(-) GGATA, AACGG, CNGTTR, WAACCA, CNGTTR, TAACTG, YAACKG	-154, -381, -1012(-), -1328, -381(-), -381(-), -521(-), -521(-), -544(-), 1228(-), 1476(-), 1025(-), 1087(-), 1315(-) WAACCA, CNGTTR, YAACKG, TAACTG, TAACAAA, TAACARA, GGATA, CCWACC, ACGG	-77, -1436(-), -296(-), -394, -394(-), -394(-), -1436 GGATA, CNGTTR, AACGG, YAACKG, TATCCAT
MYCRS (bHLH family binding site)	-70, -70(-), -172, -172(-), -339, -339(-), -387, -387(-), -616, -616(-), -648, -648(-), -671, -671(-), -1211, -1211(-), -387, -616, -671(-), -387(-), -616(-), -671, -542, -566 CANNTG, CACATG, CATGTG, AAATATTC, AAATATACT	-21, -21(-), -473, -532, -532(-), -738, -738(-), -904, -904(-), -1031, -1031(-), -1303, -1030(-), -473, -738, -738(-), -473(-) CANNTG, CACATG, CATGTG	-103, -103(-), -129, -129(-), -265(-), -370, -370(-), -388, -388(-), -606, -606(-), -722, -722(-), -811, -811(-), -1084, -1084(-), -265, -772, -811(-), -265(-), -722(-), -811, +261, +460 CANNTG, CACATG,	-189, -189(-), -510, -510(-), -863, -863(-), -888, -888(-), -1009, -1009(-), -1412, -1412(-), -344, -477 CANNTG, AAATATTCC, AAAGATGC	-163, -163(-), -178, -178(-), -254, -254(-), -310, -310(-), -1167, -1167(-), -1244, -1244(-), -254, -254(-), -254(-), -381(-) CANNTG, CACATG, CATGTG, AAGATTCT,

			CATGTG, AAATATGCT, AAATATGCC		
NAC-binding site	-387(-), -616(-), -671,-778, -779(-) CATGTG, AATGCATT, ATGCATTA	-473(-), -738(-), -1291, -591(-), -895(-) CATGTG, AATGCATT, ACGCATTG, ATGCAATG	-265(-), -722(-), -811, -555(-), -470(-) CATGTG, TACGTAA, ACGCATTG	-174(-), -961(-), -1465(-) ATGCAATG, TGTGCGTT, TGGCGTGT	-254 CATGTG
ACGTATERD1	-69, -69(-), -338, -338(-), -338(-), -957, -957(-), -1210, -1210(-), -1489, -1489(-) ACGT	Absent	-574, -574(-), -1022, -1022(-) ACGT	-1489, -1489(-) ACGT	Absent
GT1GMSCAM4	-1052(-), -998, -992, -633(-), -595(-), -576(-), -528, -454(-), -421, -312 GAAAAA	-460(-), -189, -59, -16 GAAAAA	-1483, -1010, -972, -957, -899, -661, -579, -225 GAAAAA	-528(-) GAAAAA	-1187(-), -1008, -950(-), -264(-), -173, -91 GAAAAA
W-box (WRKY binding site)	-751, -966, -1136(-), -1345(-), -751, -966, -1136(-), -1345(-), -1135(-), -967 TGACY, TGACT, TTGAC	-441, -626(-), -694(-), -821(-), -900, -907(-), -1324, -906(-), -901, -693(-), -442 TGACY, TTGAC	-545, -544, -695, -717, -544, -1374, -431(-) CTGACY, TGACY, TGACT, TTGAC	-697(-), -1121(-), -1161, -275(-), -432(-), -697(-), -788, -1121(-), -1160, -1168, -275(-), -432(-), -1160, -764, -408 CTGACY, TGACY, TGACT, TTGAC	-303, -1170(-), -1389(-), -115, -222, -236(-), -301, -449(-), -730(-), -862, -1170(-), -1389(-), -115, -222, -236(-), -449(-), -862, -1170(-), -1389(-), -863, -729(-), -510 CTGACY, TGACY, TGACT, TTGAC

DRE	-317 ACCGAGA	-1195, -1195 ACCGAC, RCCGAC	Absent	-1125(-) ACCGAGA	Absent

Table 4. Number of cREs related to drought and salinity tolerance in promoter sequence.

***Cis*-element name**	***OsNHX1***	***OsNHX5***	***OsHKT1;1***	***OsHKT2;1***	***OsSOS1***	**Putative binding TF or Function**
ABRE	18	0	4	2	0	CRE involved in the ABA responsiveness
MYBRS	12	15	15	14	7	CRE involved in regulation of water stress
MYCRS	24	17	25	14	16	Binding site of dehydration-resposive gene
NAC-binding sites	5	5	5	3	1	CRE involved in early responsive to dehydration
ACGTATERD1	11	0	4	2	0	CRE involved in early responsive to dehydration
GT1GMSCAM4	10	4	8	1	6	CRE involved in regulation of salinity stress
W-Box	10	11	7	15	22	CRE involved in regulation of stress responses
DRE	1	2	0	1	0	CRE involved in treating with ABA and drought

The TATA box is known to play a role in enhancing gene expression by binding to proteins involved in transcriptional activity. TATA-box ensures its function when it is located in the region [-35, -25] (upstream of TSS). Of the five genes studied, *OsNHX1*, *OsHKT1;1*, *OsHKT2;1*, and *OsSOS1* genes have motifs in this proximal region, except the *OsNHX5* gene (Table 3). Although the *OsNHX5* gene did not gain the support of TATA-box, the DPE motif was found in its core promoter region at +28 (downstream of

TSS). This DPE element is also involved in the transcription initiation of the *OsNHX5* gene.

With the function of enhancing the binding of the TFs and the RNA polymerase complex to the gene's promoter region, the CAAT box element helps to form the pre-initiation complex, which is essential for transcription initiation. Our analysis found that the number of CAAT-box elements ranges from 8 motifs for the *OsHKT 2;1* promoter region to 4 motifs for *OsNHX1*, 3 motifs for *OsHKT1;1* and 2 motifs for *OsSOS1*. The *OsNHX5* promoter has only one motif in the NewPLACE database. Specifically, the CAAT box element is recognized by a TF called NF-Y (Nuclear Factor Y). The binding of NF-Y to the CAAT box helps recruit other TFs and the RNA polymerase complex (Nardone *et al*., 2017).

In the core promoter regions of 5 genes, Y patches were found near TSS within a range of approximately -50 to +20. A previous study suggested that the Y Patch motif regulates gene expression when a gene lacks the TATA box (Yamamoto *et al*., 2007). Although the TATA box is a well-known TF binding site, it is not always present in all core promoters of rice plants. Alternatively, Y patches with T/C-rich decamer motifs are more prevalent in rice promoters. Approximately 19% of rice genes possess the TATA box, while around 50% contain one or more Y patches in their core promoters (Civan, Svec, 2009).

Another *cis*-element, the G-box, commonly found in the promoter regions of many plant genes, was identified in the *OsNHX1* promoter region (7 motifs) and absent in the other four. The G-box contains a consensus sequence of CACGTG and is recognized by TFs known as bZIP (basic leucine zipper) proteins. BZIP proteins are of the TP groups that regulate gene expression in response to environmental stimuli, including abiotic stress (Yu *et al*., 2020).

ABRE elements (ABA-responsive elements), functionally identified as cis regulation elements in the promoters of various ABA responsive genes, had been recognized in *OsNHX1* at 18 different positions. On the other hand, the number of ABRE elements was none or only a few in the other studied promoters (Table 4). The ABRE element is recognized by ABF/AREB proteins, which are activated by phosphorylation through the action of SnRK2 protein kinases (Kobayashi *et al.*, 2005). These TFs bind to the ABRE sequence in the promoter region of target genes, thereby regulating their expression. Activation of ABA-responsive genes leads to various physiological changes in plants, including stomatal closure (reducing water loss), accumulation of osmoprotectants (compounds that help maintain cellular osmotic balance), and the induction of stress tolerance mechanisms (Feng *et al.*, 2019).

MYBRS, MYCRS, and NAC binding-sites are sequence motifs that serve as recognition sites for MYB, MYC, and NAC TFs, respectively. The binding of these TFs to their respective motifs in gene promoters plays a vital role in the regulation of gene expression in plants, especially in response to dehydration (Joshi *et al.*, 2016). In our study, the total number of motif positions of these types in the genes *OsNHX1*, *OsNHX5*, *OsHKT1;1*, *OsHKT2;1* and *OsSOS1* were 41, 37, 45, 31, and 24, respectively (Table 4). Therefore, we could hypothesize that the activity of the *OsNHX1* and *OsHKT1;1* gene would be stronger than that of other genes under dehydration stress.

Furthermore, several factors indirectly involved in salt stress response were also found in the promoter region of studied genes, such as ACGTATERD1, DRE (Freitas *et al.*, 2019), and GT1GMSCAM4 (Liang and Jiang, 2017), which are known to be involved in the processes of response to salinity or drought stress. According to data presented in Table 4, the number of *cis*-elements belonging to these factors is significantly higher in the *OsNHX1* gene promoter region, thereby reinforcing our above hypothesis.

CONCLUSION

In conclusion, the analysis of *cis* elements associated with drought and salinity tolerance within the promoter regions of *OsNHX1*, *OsNHX5*, *OsHKT1;1*, *OsHKT2;1*, and *OsSOS1* using bioinformatics tools strongly supported our hypothesis. Potential transcription start sites for these genes were predicted at positions of -322, -182, -535, -53, and -168 for *OsNHX1*, *OsNHX5*, *OsHKT1;1*, *OsHKT2;1*, and *OsSOS1*, respectively. By comparing the number and positional information of these cREs, the findings suggest that, among the five studied genes, *OsNHX1* and *OsHKT1;1* may possess regulatory mechanisms that allow them to respond more robustly to environmental stresses. However, further experimental validation such as promoter activity assays, gene mutation studies, or ChIP (chromatin immunoprecipitation) should be carried out to confirm the association of specific TFs with *cis* elements and their impact on gene expression. This knowledge would provide suggestions for crop improvement strategies that enhance stress tolerance in agricultural systems.

ACKNOWLEDGEMENTS

This study was funded by project "Analysing the diversity, conservation and function of the NHX protein family in rice (*Oryza sativa* L.) by *in silico* method "code MHN 2023-01.03 from Ha Noi Open University.

CONFLICT OF INTEREST

The authors declare that there is no conflict of interests.

REFERENCES

Apse MP, Blumwald, E (2007) Na+ transport in plants. *FEBS Letters* 581: 2247-2254. https://doi.org/10.1016/j.febslet.2007.04.014.

Basu S, Roychoudhury A (2014) Expression profiling of abiotic stress-inducible genes in response to multiple stresses in rice (*Oryza sativa* L.) varieties with contrasting level of stress tolerance. *Biomed Res Int* 2014: 706890. https://doi.org/10.1155/2014/706890.

Civan P, Svec M (2009) Genome-wide analysis of rice (*Oryza sativa* L. *subsp. japonica*) TATA box and Y Patch promoter elements. *Genome* 52: 294-297. https://doi.org/10.1139/g09-001.

Cui MH, Yoo KS, Hyoung S, Nguyen HT, Kim YY, Kim HJ, Shin JS (2013) An Arabidopsis R2R3-MYB transcription factor, AtMYB20, negatively regulates type 2C serine/threonine protein phosphatases to enhance salt tolerance. *FEBS Lett* 587: 1773-1778. https://doi.org/10.1016/j.febslet.2013.04.028.

de Medeiros Oliveira M, Bonadio I, Lie de Melo A, Mendes Souza G, Durham AM (2021) TSSFinder-fast and accurate ab initio prediction of the core promoter in eukaryotic genomes. *Brief bioinfor* 22: bbab198. https://doi.org/10.1093/bib/bbab198.

Farooq M, Park JR, Jang YH, Kim EG, Kim KM (2021) Rice Cultivars Under Salt Stress Show Differential Expression of Genes Related to the

Regulation of Na+/K+ Balance. *Front Plant Sci* 12: 680131. https://doi.org/10.3389/fpls.2021.680131.

Feng RJ, Ren MY, Lu LF, Peng M, Guan X, Zhou DB, Zhang MY, Qi DF, Li K, Tang W, Yun TY, Chen YF, Wang F, Zhang D, Shen Q, Liang P, Zhang YD, Xie JH (2019) Involvement of abscisic acid-responsive element-binding factors in cassava (Manihot esculenta) dehydration stress response. *Sci Rep* 9(1): 12661 https://doi.org/10.1038/s41598-019-49083-3.

Freitas EO, Melo BP, Lourenço-Tessutti IT, Arraes FBM, Amorim RM, Lisei-de-Sá ME, Costa JA, Leite AGB, Faheem M, Ferreira MA, Morgante CV, Fontes EPB, Grossi-de-Sa MF (2019) Identification and characterization of the GmRD26 soybean promoter in response to abiotic stresses: potential tool for biotechnological application. *BMC Biotechnol* 19(1): 79. https://doi.org/10.1186/s12896-019-0561-3.

Fukuda A, Nakamura A, Hara N, Toki S, Tanaka Y (2011) Molecular and functional analyses of rice NHX-type Na+/H+ antiporter genes. *Planta* 233: 175-188. https://doi.org/10.1007/s00425-010-1289-4.

Garciadeblas B, Senn ME, Banuelos MA, Rodriguez-Navarro A (2003) Sodium transport and HKT transporters: the rice model. *Plant J* 34: 788-801. https://doi.org/10.1046/j.1365-313x.2003.01764.x.

Hauser F, Horie T (2010) A conserved primary salt tolerance mechanism mediated by HKT transporters: a mechanism for sodium exclusion and maintenance of high K(+)/Na(+) ratio in leaves during salinity stress. *Plant Cell Environ* 33: 552-565. https://doi.org/10.1111/j.1365-3040.2009.02056.x.

Jiang Y, Deyholos MK (2009) Functional characterization of Arabidopsis NaCl-inducible WRKY25 and WRKY33 transcription factors in abiotic stresses. *Plant Mol Biol* 69: 91-105. https://doi.org/10.1007/s11103-008-9408-3.

Jiang Y, Yang B, Deyholos MK (2009) Functional characterization of the Arabidopsis bHLH92 transcription factor in abiotic stress. *Mol Genet Genomics* 282: 503-516. https://doi.org/10.1007/s00438-009-0481-3.

Joshi R, Wani SH, Singh B, Bohra A, Dar ZA, Lone AA, Pareek A, Singla-Pareek SL (2016) Transcription Factors and Plants Response to Drought Stress: Current Understanding and Future Directions. *Front Plant Sci* 7: 1029. https://doi.org/10.3389/fpls.2016.01029.

Kasuga M, Liu Q, Miura S, Yamaguchi-Shinozaki K, Shinozaki K (1999) Improving plant drought, salt, and freezing tolerance by gene transfer of a single stress-inducible transcription factor. *Nat Biotechnol* 17: 287-291. https://doi.org/10.1038/7036.

Kobayashi Y, Murata M, Minami H, Yamamoto S, Kagaya Y, Hobo T, Yamamoto A, Hattori T (2005) Abscisic acid-activated SNRK2 protein kinases function in the gene-regulation pathway of ABA signal transduction by phosphorylating ABA response element-binding factors. *Plant J* 44(6), 939–949. https://doi.org/10.1111/j.1365-313x.2005.02583.x.

Kuwano M, Masumura T, Yoshida KT (2011) A novel endosperm transfer cell-containing region-specific gene and its promoter in rice. *Plant Mol Biol* 76: 47-56. https://doi.org/10.1007/s11103-011-9765-1.

Liang L, Guo L, Zhai Y, Hou Z, Wu W, Zhang X, Wu Y, Liu X, Guo S, Gao G, Liu W (2023) Genome-wide characterization of SOS1 gene family in potato (Solanum tuberosum) and expression analyses under salt and hormone stress. *Front Plant Sci*, 14: 1201730. https://doi.org/10.3389/fpls.2023.1201730.

Liang M H, Jiang JG (2017) Analysis of carotenogenic genes promoters and WRKY transcription factors in response to salt stress in Dunaliella bardawil. *Sci Rep* 7: 37025. https://doi.org/10.1038/srep37025.

Liu C, Mao B, Yuan D, Chu C, Duan M (2022) Salt tolerance in rice: Physiological responses and molecular mechanisms. The Crop Journal 10: 13-25. https://doi.org/10.1016/j.cj.2021.02.010.

Martínez-Atienza J, Jiang X, Garciadeblas B, Mendoza I, Zhu JK, Pardo JM, Quintero FJ (2007) Conservation of the salt overly sensitive pathway in rice. *Plant Physiol* 143: 1001–1012. https://doi.org/10.1104/pp.106.092635.

Miyoshi SI, Sasaki T, Kaku N, Inoue T, Uozumi N, Maehara Y, Nakao H (2010) Assimilation of Metal Ions Bound to Porphyrins or Porphyrin-Peptides by Vibrio vulnificus, a Human Pathogen Inhabiting Estuarine and Marine Environments. *Biocontrol Science* 15: 1-6. https://doi.org/10.4265/bio.15.1.

Nardone V, Chaves-Sanjuan A, Nardini M (2017) Structural determinants for NF-Y/DNA interaction at the CCAAT box. *Biochim Biophys Acta Gene Regul Mech* 1860(5): 571–580. https://doi.org/10.1016/j.bbagrm.2016.09.006.

Reese MG (2001) Application of a time-delay neural network to promoter annotation in the Drosophila melanogaster genome. *Comput Chem* 26: 51-56. https://doi.org/10.1016/s0097-8485(01)00099-7.

Shaban AS, Safhi FA, Fakhr M A, Pruthi R, Abozahra MS, El-Tahan AM, Subudhi PK (2023) Comparison of the Morpho-Physiological and Molecular Responses to Salinity and Alkalinity Stresses in Rice. *Plants* (Basel, Switzerland) 13(1): 60. https://doi.org/10.3390/plants13010060.

Shahmuradov IA, Umarov RK, Solovyev VV (2017) TSSPlant: a new tool for prediction of plant Pol II promoters. *Nucleic Acids Res* 45: e65.https://doi.org/10.1093%2Fnar%2Fgkw1353.

Shi H, Ishitani M, Kim C, Zhu JK (2000) The Arabidopsis thaliana salt tolerance gene SOS1 encodes a putative Na+/H+ antiporter. Proc Natl *Acad Sci* 97: 6896–6901. https://doi.org/10.1073/pnas.120170197.

Shi H, Quintero FJ, Pardo JM, Zhu JK (2002) The putative plasma membrane Na(+)/H (+) antiporter SOS1 controls long-distance Na(+) transport in plants. *Plant Cell* 14: 465-477. https://doi.org/10.1105/tpc.010371.

Tran LS, Nakashima K, Sakuma Y, Simpson SD, Fujita Y, Maruyama K, Yamaguchi-Shinozaki K (2004) Isolation and functional analysis of Arabidopsis stress-inducible NAC transcription factors that bind to a drought-responsive cis-element in the early responsive to dehydration stress 1 promoter. *Plant Cell* 16: 2481-2498. https://doi.org/10.1105/tpc.104.022699.

Van Nguyen N, Ferrero A (2006) Meeting the challenges of global rice production. *Paddy Water Environ* 4: 1-9. https://doi.org/10.1007/s10333-005-0031-5.

Wang X, Bandyopadhyay S, Xuan Z, Zhao X, Zhang MQ, Zhang X (2007) Prediction of transcription start sites based on feature selection using AMOSA. *Proc LSS Comput Syst Bioinform Conf* 6: 183–193.

Yamaguchi T, Hamamoto S, Uozumi N (2013) Sodium transport system in plant cells *Front Plant Sci* 4: 410. https://doi.org/10.3389/fpls.2013.00410.

Yamamoto YY, Ichida H, Matsui M, Obokata J, Sakurai T, Satou M, Abe T (2007) Identification of plant promoter constituents by analysis of local distribution of short sequences. *BMC Genomics* 8: 67. https://doi.org/10.1186/1471-2164-8-67.

Yang O, Popova OV, Suthoff U, Luking I, Dietz KJ, Golldack D (2009) The Arabidopsis basic leucine zipper transcription factor AtbZIP24 regulates complex transcriptional networks involved in abiotic stress resistance. *Gene* 436: 45-55. https://doi.org/10.1016/j.gene.2009.02.010.

Yu, Y., Qian, Y., Jiang, M., Xu, J., Yang, J., Zhang, T., Gou, L., & Pi, E. (2020). Regulation Mechanisms of Plant Basic Leucine Zippers to Various Abiotic Stresses. *Front Plant Sci* 11:

1258. https://doi.org/10.3389%2Ffpls.2020.01258.

Zhu J, Ren Y, Wang Y, Liu F, Teng X, Zhang Y, Duan E, Wu M, Zhong M, Hao Y, Zhu X, Lei J, Wang Y, Yu Y, Pan T, Bao Y, Wang Y, Wan J (2019) OsNHX5-mediated pH homeostasis is required for post-Golgi trafficking of seed storage proteins in rice endosperm cells. *BMC plant biol*, 19(1): 295. https://doi.org/10.1186/s12870-019-1911-y.