In silico Promoter Analysis Reveals Rice *Wx* Promoter Could be a TATA-less Promoter with a Putative Pyrimidine Rich Region

K.R.D. Hemage* and V.A. Sumanasinghe[1]

Centre for Education, Research and Training on Kidney Diseases
Faculty of Medicine, University of Peradeniya
Sri Lanka

ABSTRACT: *Rice waxy gene is one of the major genes, which determines the amylose content of a rice grain. The expression of this gene is regulated by various stimuli such as temperature, light, drought and diseases. Promoter region of a gene is responsible for changing the level of expression with the help of transcription factors and their binding sites, cis-acting regulatory elements found in the promoter region ranging from 2000 bp upstream and 500 bp downstream relative to the transcription start site. Present study investigated the putative transcription factor binding sites in the promoter region and possible expression regulation stimuli using five Sri Lankan rice varieties, two glutinous rice varieties, Arabidopsis thaliana and potato. The promoter regions of the plants were extracted either after mapping selected varieties from rice 3000 genome project or from promoter databases. Signal scan was carried out to capture putative signal sequences, and kappa incidence of coincidence vs guanine cytosine content, guanine cytosine skew, and multiple sequence alignment were also performed. The results revealed that rice Wx promoter could be a TATA-less promoter with a putative pyrimidine patch, and categorized under ATCG middle class of promoters. The possible transcription regulation stimuli of waxy gene include dehydration, light, temperature, and hormones. There were signal sequences related to mesophyll, root nodule, seed and root specific regulation of expression.*

Keywords: *Rice Wx promoter, cis-acting regulatory elements, TATA-less promoters, ATCG middle class, GC skew*

INTRODUCTION

Rice grain is mainly composed of starch (90%) deposited in the endosperm. The amount of starch content varies with genetic and environmental factors. Digestion rate and time of ingested rice vary with the amylose: amylopectin ratio, grain processing, physicochemical properties, size of the particles and amount of lipid-amylose complex (Hu *et al*., 2004). Out of these, the most influential character is amylose: amylopectin ratio. Higher the ratio, lower the stickiness of cooked rice and lower the gelatinous nature of rice floor. Most importantly, if the ratio is higher, glucose intake to the blood becomes slower and time taken to empty the gastrointestinal tract is increased, helping to regulate glucose level in blood (Behall *et al*., 1988).

[1] Department of Agricultural Biology, Faculty of Agriculture, University of Peradeniya, Sri Lanka.

* Author for correspondence: rus2dil@gmail.com

Waxy or *Wx* (LOC_Os06g04200) is one of the major genes which determine the amylose content of the rice grain. There are four different transcripts of the *Wx* gene where glutinous rice varieties transcribe mature 2.3 kb mRNA, whereas non-glutinous rice varieties transcribe a 2.3 kb and a 3.3 kb erroneous mature mRNA with the first intron (Bligh *et al*., 1998). *Wx* gene is translated into ADP glucose starch tansferace (EC 2.4.1.2.1, 59 kDa) (Dry *et al*., 1992; Nelson and Pan, 1995), and there are mainly two alleles of *Wx* locus, Wx^a and Wx^b, which can be differentiated using the accumulated Wx protein (Hirano and Sano, 2000).

Although *Wx* gene expression is temperature dependent, some differences are observed due to a single base substitution at the splice site donor of the first intron in Wx^b (Hirano and Sano, 1998; Isshiki *et al*., 1998). If the temperature is low, the expression is higher (Hirano and Sano, 1998; Larkin and Park, 1999), resulting higher amylose contents in the mature seeds (Hirano and Sano, 2000). Usually, these types of temperature dependencies are regulated with the help of the promoter regions of the genes.

Promoter of a gene is a sequence of DNA found towards the 5' end of the gene. It defines the beginning of the transcription by RNA polymerase II in eukaryotic organisms, usually which can be spread from 2000 bp upstream to 500 bp downstream from the transcription start site (TSS). A typical promoter consists of a TSS, core promoter (region immediately flanking TSS), proximal promoter (approximately the region from -250 bp to TSS), and distal promoter (region beyond -250 bp, which still affects the transcription).

A typical core promoter consists of an initiator element (INR), TATA box, TFIIB recognition element (BRE) and downstream core promoter element (DPE) (Butler and Kadonaga, 2002) (Figure 1).

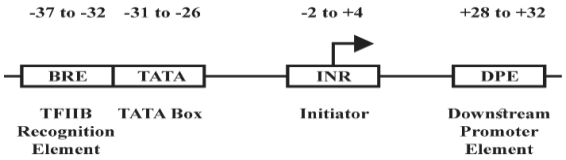

Figure 1. Core promoter region showing the basic *cis*-elements (adapted from Butler and Kadonaga, 2002)

The TATA box is found approximately 25 – 30 bp upstream to TSS, with a consensus sequence of TATAAA. However, not every core promoter contains TATA boxes (those are called TATA-less promoters), for example 43% of 205 core promoters in *Drosophila* lack TATA boxes (Kutach and Kadonaga, 2000). Transcription starts in the vicinity of INR in which TSS is located, and the consensus sequence is Py-Py(C)-A_{+1}-N-T/A-Py-Py (Bucher, 1990). DPE (consensus sequence – A/G_{+28}-G-A/T-C/T-G/A/C) is found exactly at +28 to +32 from A_{+1} in INR and mostly in TATA-less promoters (Butler and Kadonaga, 2002). TFIIB recognition element is found immediately upstream to TATA box and the consensus sequence is G/C-G/CG/A-C-G-C-C (Lagrange *et al*., 1998). Other than these elements, there are GC rich regions called CpG islands, CAAT boxes, downstream core element, and additional *cis*-acting elements that are recognized by the proteins activated under different environmental conditions such as light, cold and drought *etc*.

In the present study, rice *Wx* gene promoter was selected to identify putative regulatory elements and the respective stimuli, to direct future research on eating and cooking quality

improvements since rice is the staple food in Sri Lanka. The study was focused on characterizing TFBS in rice *Wx* promoter and comparing these elements with a similar promoter in rice (GBSSIb), and *Wx* promoters of a non-starchy plant (*Arabidopsis thaliana*) and a starchy crop (*Solanum tuberosum*).

MATERIALS AND METHODS

Extraction of the promoter region

Three glutinous rice varieties (*Caddit, Fukushima Mochi* and *Taisen Glutinous*) and five Sri Lankan rice varieties (*Godawel, Puttunellu, Hondarawalu, Pannithi* and *Karuthaseenati*) were selected for the extraction of rice *Wx* promoter. Selection of Sri Lankan rice varieties was based on the continuity (no gaps in the mapped sequence) of the *Wx* promoter region, of 48 Sri Lankan varieties in the rice 3000 genome project in 2014. Rice GBSSIb, potato, and *A. Thaliana Wx* promoter regions were extracted from the PlantPAN database (Chang *et al*., 2008) and *A. thaliana* TAIR10.

Mapping sequence reads, filtering and slicing the promoter region

Sequence reads (paired Illumina FASTQ) of all the Sri Lankan varieties and the three glutinous varieties from rice 3000 genome project were transferred to Galaxy server (https://usegalaxy.org/) (Giardine et al., 2005; Blankenberg *et al*., 2010; Li *et al*., 2014) from European Nucleotide Archive (https://www.ebi.ac.uk/ena) under the project number ERP005654. Soon after uploading sequence reads (adapter contamination removed) to the server, quality reports were generated using FastQC tool (Galaxy tool version 0.63) (Andrews, 2012). Sequence reads were converted to FastQ Sanger using FastQ Groomer version 1.0.4 (Blankenberg *et al*., 2010), and mapped to the reference genome (*Oryza sativa* Os-Nipponbare-Reference_IRGSP-1.0 version 7.0) (Kawahara *et al*., 2013) using Burrows-Wheeler Aligner (BWA) version 1.2.3 (Li *et al*., 2009) with default parameters. Output files (BAM – Binary Sequence Alignment/ Map format) from BWA was filtered (using map quality >= 20) using Filter tool version 0.0.1 (Barnett *et al*., 2011). Slice BAM version 0.0.1 (Li and Durbin, 2009; Blankenberg *et al*., 2010) was used to slice the BAM file using a BED file (Browser Extensible Data) with the coordinates of targeted region. Since all the varieties had at least six samples sequenced, all of them were mapped, filtered, sliced, and merged to one BAM file using Convert, Merge, Randomize tool version 0.0.1 (Barnett *et al*., 2011) to increase the depth of the final alignment. Consensus calling was done using Integrated Genome Viewer version 2.3 (Thorvaldsdottir *et al*., 2013) after visualizing to find gaps. First five Sri Lankan varieties that had no gaps in the targeted region were selected for further analysis (Figure 2).

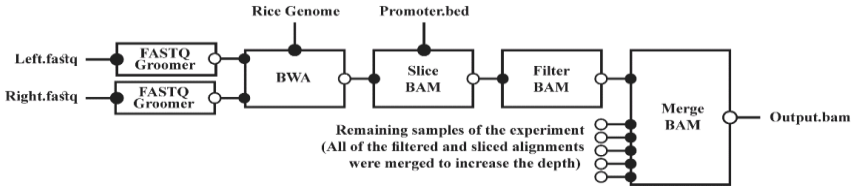

Figure 2.Workflow used in mapping of paired reads into the rice reference genome. Rectangles represent tools, black circles are inputs, white circles are outputs and each line in between represents the flow of data from input to output

Signal scan, motif mapping, GC skew plots, and other analyses

Signal scan was carried out using Plant *cis*-acting regulatory DNA elements (PLACE) database (Higo *et al*., 1999) using the consensus sequence. Statistical analyses were performed using R statistical language version 3.2.2 (R Core Team, 2017) in RStudio version 1.0.136 (RStudio Team, 2015). Motifs were mapped using gplot2package (Wickham, 2009). GC skew plots (as described in Tatarinova *et al*. 2003) were generated using a sliding window of 20 and the step size of five for the core promoter region. Kappa index of coincidence vs guanine-cytosine content plots were created to find out the promoter class. Frequency of TFBS along the promoter was calculated using a sliding window of 250 bp and step size of 10. Frequencies of TFBS of promoters were calculated. Multiple sequence alignments for full promoter regions were performed using ClustalW2 (Larkin *et al*., 2007). Dendrograms were generated using ggdendro package (De Vries and Ripley, 2016) using similarity coefficients and UPGMA method.

RESULTS

According to the rice 3000 genome project, the average depth of sequencing was 14 with an average genome coverage of 94%. However, for the samples used for quality control, the average depth and coverage were 6 and 85%, respectively. PLACE results for the rice *Wx* promoter region (2.5 kb, -2 kb and + 0.5 kb relative to TSS), consisted of a total of 591 transcription factor binding sites (TFBS), out of which 283 in the forward strand and 308 in the reverse strand. Unique TFBS count was 126, of which 92 were in the forward strand and 34 in the reverse strand.

Core promoter region and common TFBS

There were 22 TFBS in the rice *Wx* core promoter with two INR elements at +10 and +88 positions and no TATA boxes were found from signal scan. A putative pyrimidine patch (or Y patch with 15 CT rich overlapping CTRMCAMV35S) was observed from +49 to +86 (Figure 3 (a)). Comparatively, rice *Wx* core promoter had less TFBS (22 motifs) than that of *A. thaliana* (44 motifs), potato (40 motifs), and rice GBSSIb (34 motifs). There were only four motifs found common to all four promoters. CACTFTPPCA1 (known as G box), NODCON2GM (which is one of the putative nodulin consensus sequences), OSEROOTNODULE (characteristic motif found in promoters activated in infected cells of root nodules), and SORLIP1AT (which is an over represented sequence in light induced promoters) are those (Figure 3).

Apart from these, rice *Wx* core and GBSSIb promoters shared six more motifs namely, CBFHV (binding sites of C-repeat binding factors), CGCGBOXAT (CGCG box recognized by AtSR1-6), CTRMCAM35S (CT rich motif), CURECORECR (Core of copper response element), PALBOXAPC (Box A), and SORLIP2AT (sequence over represented in light induced promoters). Moreover, INR was found in all three *Wx* promoters (+10 and +88 in rice *Wx*, +95 in *A. Thaliana Wx*, and -24 in potato *Wx*, relative to the TSS) and in potato in which a TATA box was found at +36. CAAT boxes were found in *A. Thaliana* (-56), and potato (+32 and +33), DOFCOREZM (Dof protein binding motif) and DPBFCOREDCDC3 (novel class of bZIP transcription factors, DPBF-1 and 2 binding site) were found in all three *Wx* core promoter regions.

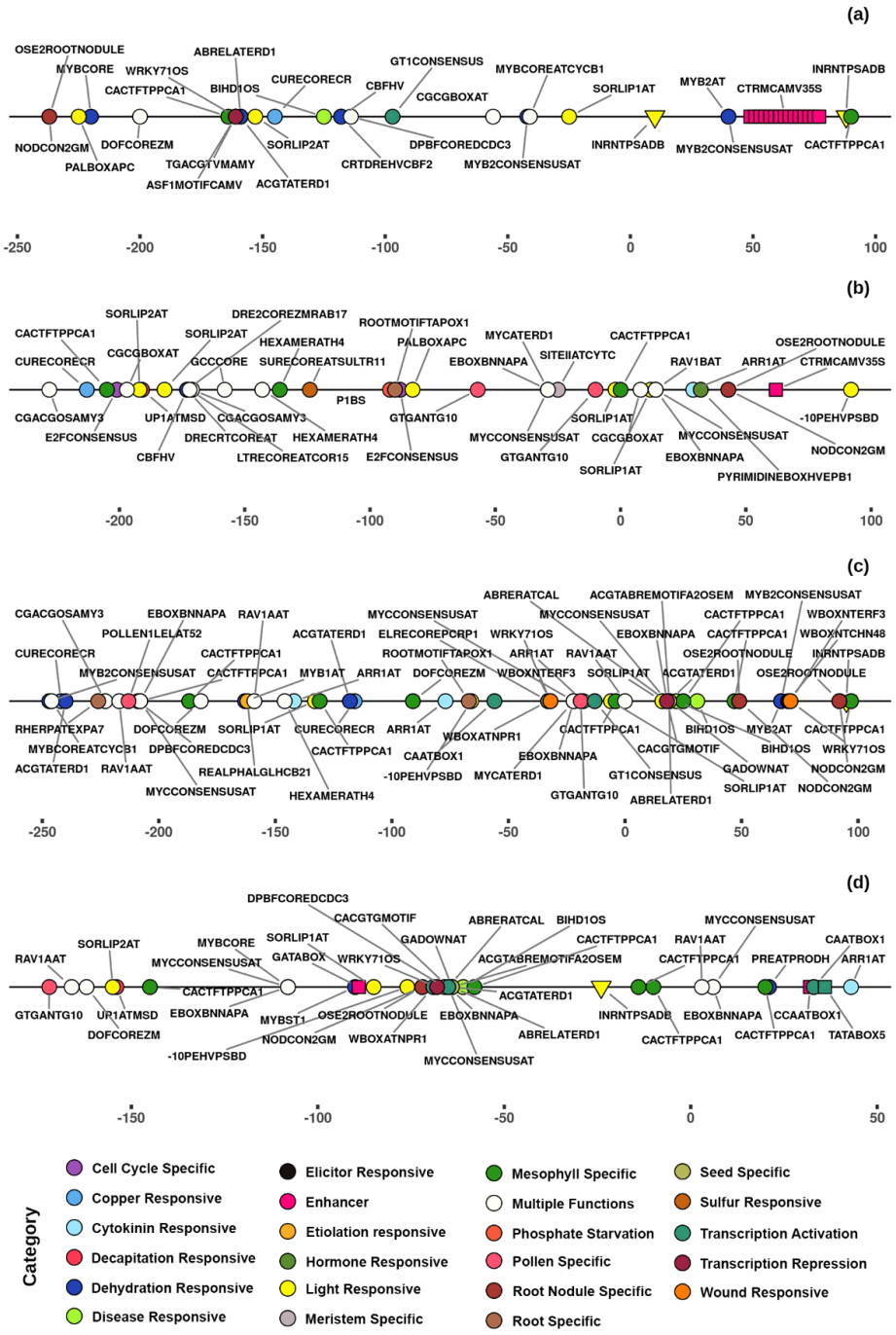

Figure 3.Core promoter region with TFBS for rice *Wx* (a), rice GBSSIb (b) *A. thaliana Wx* (c) and potato*Wx* (d) promoters. Categories show either the function / pathway / response / place of transcription regulation of the signal sequence. TFBS depicted as triangles are initiator elements, squares are enhancers and essential sequences, and circles represent all the other TFBS

GC skew plots

According to the GC skew plots (Figure 4), rice *Wx* core promoter showed a peak GC skew from -15 to -5, whereas GC skew of *A. thaliana Wx* promoter peaked at -160, -75, +5, and +50 suggesting multiple transcription start sites. Rice GBSSIb and potato *Wx* promoters displayed peaks at -20 and +25, and -5 to +10, respectively.

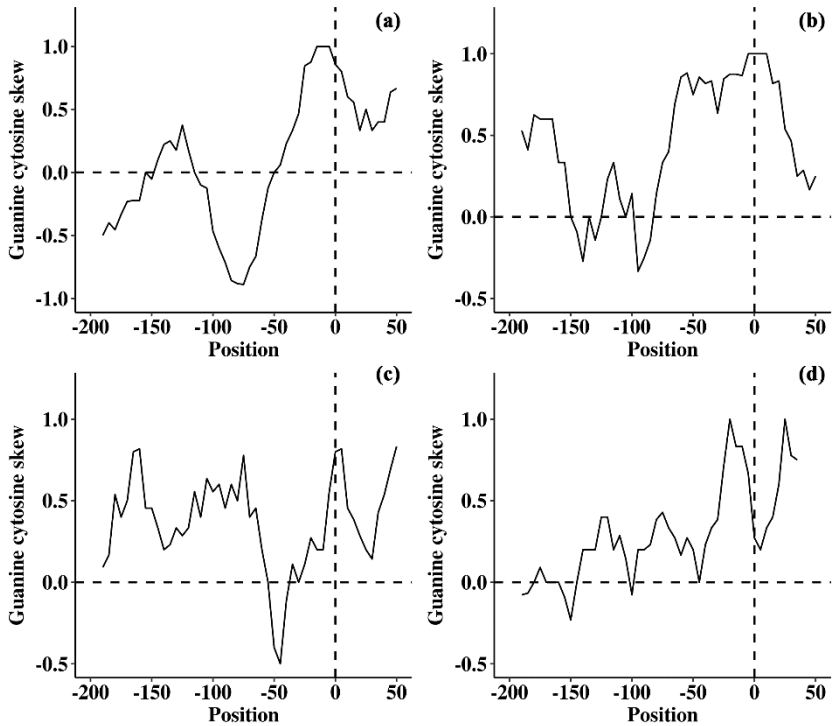

Figure 4. GC skew plots for rice *Wx* (a), rice GBSSIb (b), *A. thaliana Wx* (c), and potato *Wx* (d) promoters. All the promoters showed a peak of GC skew at the vicinity of TSS. Graphs of *A. thaliana* and potato suggest multiple TSS

Kappa index of coincidence (KIC) vs GC content

According to KIC vs GC content graphs for promoter classes, rice *Wx* and GBSSIb followed an ATCG middle and CG less patterns respectively whereas *A. Thaliana* showed an AT based pattern. Since potato *Wx* full promoter region had large sequence gaps, it was not considered for this analysis (Figure 5).

Frequencies of TFBS

TFBS with more than five occurrences were arranged according to a descending pattern for rice *Wx* and GBSSIb, and *A. thaliana Wx* promoters (since potato promoter region had gaps in the sequence, it has been omitted). ACGTATERD1 was the most frequent *cis*-element (18 times) in rice *Wx* promoter which was followed by CTRMCAMV35S (15 times) and CGCGBOXAT (13 times) (Table 1).

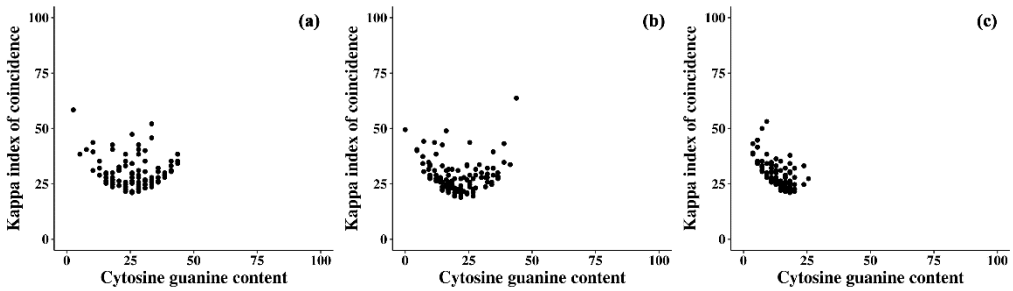

Figure 5. KIC vs GC content graphs for rice *Wx* (a), rice GBSSIb (b) and *A. thaliana Wx*(c) promoters. Rice *Wx* show ATCG middle pattern whereas rice GBSSIb and *A. thaliana Wx* follow a CG less and AT based patterns, respectively

Table 1. Frequencies of TFBS in rice *Wx*, GBSSIb and *A. Thaliana Wx* promoters TFBS are sorted from highest to lowest frequency (F is frequency) Superscripts denote common motifs

Rice *Wx* promoter		Rice GBSSIb Promoter		*A. thaliana Wx* promoter	
Motif	**F**	**Motif**	**F**	**Motif**	**F**
ACGTATERD1[a]	18	CACTFTPPCA1[d]	19	CACTFTPPCA1[d]	20
CTRMCAM35S	15	GTGANTG10[l]	14	DOFCOREZM[h]	19
CGCGBOXAT	13	DOFCOREZM[h]	13	ARR1AT[m]	18
SORLIP1AT[b]	12	ACGTATERD1[a]	10	GT1CONSENSUS[c]	16
GT1CONSENSUS[c]	12	EBOXBNNAPA[f]	9	WRKY71OS[j]	15
CACTFTPPCA1[d]	9	MYCCONSENSUSAT[g]	9	GTGANTG10[l]	15
CURECORECR[e]	9	GT1CONSENSUS[c]	8	CAATBOX1	11
EBOXBNNAPA[f]	8	ARR1AT[m]	8	ACGTATERD1[a]	10
MYCCONSENSUSAT[g]	8	CURECORECR[e]	6	WBOXNTERF3	10
DOFCOREZM[h]	8	WRKY71OS[j]	6	EBOXBNNAPA[f]	9
GATABOX[i]	7	CGACGOSAMY3[k]	6	MYCCONSENSUSAT[g]	9
WRKY71OS[j]	7	SORLIP1AT[b]	6	POLLEN1LELAT52[n]	8
CGACGOSAMY3[k]	6	POLLEN1LELAT52[n]	6	GATABOX[i]	8
ABRELATERD1	6			WBOXHVISO1	8
				RAV1AAT	8
				BIHD1OS	6

Multiple sequence alignment

Multiple sequence alignment and grouping using hierarchical clustering method **(**UPGMA**)** separated *Nipponbare* and *Karuthaseenati* from the rest of the varieties where the latter was more distant from the rest **(**Figure 06**).**

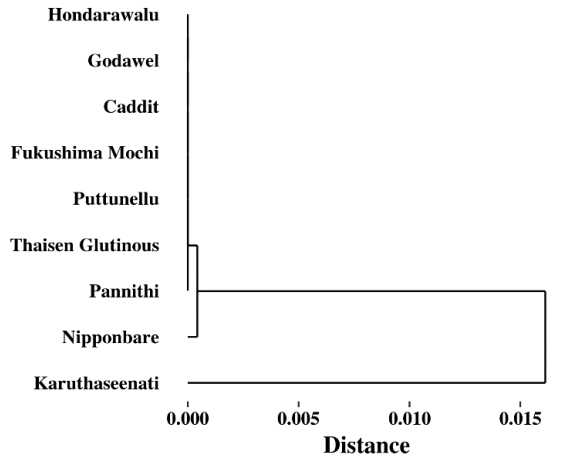

Figure 6. Dendrogram of rice varieties used in the study, constructed using the similarity coefficients obtained from the multiple sequence alignment of the full promoter regions (-2.0 kb to +0.5 kb) and UPGMA method

DISCUSSION

Rice *Wx* core promoter

KIC vs GC content graph showed that *Wx* promoter has an ATCG middle class pattern (which is a rare type), where it contains balanced A+T and G+C with higher than average Kappa IC values (Gagniuc and Ionescu-Tirgoviste, 2012). This pattern is mostly found in TATA promoters as against TATA-less promoters. Compared with rice GBSSIb and *A. thaliana Wx* promoters, which had rather common CG less (high frequency of AT repetitive sequences, usually found in rice) and AT based patterns (high A+T %, common in *A. thaliana*), respectively (Gagniuc and Ionescu-Tirgoviste, 2012), rice *Wx* promoter is unique.

GC skew plot of rice *Wx* confirms transcription start site where GC skew was at the peak of 1, in the region close to +1 position (-15 to -5) and the presence of INR elements (INRNTPSADB) (Nakamura *et al*., 2002). Having two INR elements at +10 and +88 suggest possible multiple transcription start sites. However, rice GBSSIb promoter showed a peak from -5 to +10 with no INR elements in the immediate vicinity which is true for *A. thaliana* and potato where both provide evidence for multiple transcription start sites.

According to the PLACE scan results, none of the promoters except *A. thaliana Wx,* contains TATA boxes between -200 to +50. There were no TATA boxes found within -1000 to +500 in rice *Wx* promoter in the signal scan results, although PlantPromDB categorized rice *Wx*promoter (promoter ID – PLPR0351) under TATA promoters (Shahmuradov *et al*., 2003). According to this database, a putative TATA box sequence (TACAAATA) has been identified between -35 to -27 (Shahmuradov *et al*., 2003). Signal scan in PLACE was not able to identify this sequence, because the TATA box sequences in PLACE database (CTATAAATAC, TATAAAT, TATTAAT, TATATAA, and TTATTT) are experimentally validated and TACAAATA sequence was not recognized as a TATA sequence (Higo *et al*., 1999). Further experiments are required to confirm whether this putative sequence (TACAAATA) is a TATA box and also to confirm whether rice *Wx* promoter is a TATA-less promoter because there were two INR elements (INRNTPSADB) which are found in

light responsive TATA-less promoters (Nakamura *et al*., 2002). A putative pyrimidine rich region with multiple CRTMCAMV35S (15 sites) was found between +49 to +86 which could be functioning as a pyrimidine patch (Y patch) although it does not follow consensus (CYTCYYCCYC) which is described elsewhere (Francki *et al*., 2009). Usually Y patches are observed from -49 to +50 in rice genes, and this region is known to be found in higher plants and the biochemical function is unknown (Yamamoto *et al*., 2007).

Transcription regulation

Signal scan results reveal that rice *Wx* promoter contains 8 dehydration responsive elements, 6 light responsive elements, and 4 multi-functional elements. It can be inferred that the promoter may be highly regulated under above varying conditions. Compared with other promoters, dehydration responsive elements are the most abundant elements, which is followed by light responsive, multi-functional and hormone responsive elements. All the *Wx* promoters and GBSSIb promoter had elements related to mesophyll, root nodule, seed and root specific regulatory elements. Additionally, several enhancers such as CRTMCAMV35S, and transcription activators and repressors can be observed. Further research needed to determine the exact mechanism and regulatory pathways of these elements.

If TFBS frequency of the full promoter region is considered, ACGTATERD1 is the single most abundantly present element throughout the region. This element is responsible for up-regulating early responses to dehydration by activating erd1 genes (Busk *et al*., 1997; Simpson *et al*., 2003; Suzuki *et al*., 2005). It is evident that this gene could be highly regulated by dehydration. Furthermore, the elements like CRTMCAMV35S, CGCGBOXAT, SORLIP1AT, and GT1CONSENSUS are present frequently, suggesting enhanced and light responsive transcription (Kaplan *et al*., 2006; Yang and Poovaiah, 2002).

According to Sano and Hirano (1998), rice *Wx* gene shows a high expression in cold temperatures showing that the gene is up-regulated by low temperature in mature seeds producing high amounts of amylose, and the expression is regulated organ specifically (Hirano and Sano, 2000). Results of the present study show the presence of CBFHV (at -118) and CRTDREHVCBF2 (at -118) which are binding sites for CBF1 (C-repeat binding factor 1) and CBF2, and core CRT/DRE motif respectively; these are known to be involved in cold regulated transcription activation (Gilmour *et al*., 1998; Jaglo-Ottosen *et al*., 1998; Dubouzet *et al*., 2003; Qin *et al*., 2004; Xue, 2003; Skinner *et al*., 2005). Apart from these, rice *Wx* and some other promoters contain disease resistant elements such as BIHD10S – binding site of OsBIHD1 (Luo *et al*., 2005), copper responsive elements like CURECORECR which areresponsible for transcription activation under copper deficiency (Quinn *et al*., 2002), and calcium and other ion responsive elements.

Identifying these motifs and their respective regulatory pathways are important because, manipulations such as deletions, additions and changing the position of the motifs could have differential effects on the expression *Wx* gene, changing the amylose content of the rice grain, and will have the benefit for future eating and cooking quality improvement research of rice.

CONCLUSIONS

The rice *Wx* promoter is an ATCG middle class, putative TATA-less promoter with INR element and putative pyrimidine patch (Y patch). Transcription could be mainly regulated by stimuli such as dehydration, light, hormones, and temperature. Evidence of organ specific

expression regulation of rice *Wx* gene in mesophyll, root nodule, seed and root, could be observed due to the presence of related *cis*-acting elements and further research needed to confirm the validity of this finding.

REFERENCES

Andrews, S. (2012). FastQC: A quality control tool for high throughput sequence data. [online]. [Accessed on 22.07.2015]. Available at http://www.bioinformatics.babraham.ac.uk/projects/fastqc/

Barnett, D.W., Garrison, E.K., Quinlan, A.R., Strömberg, M.P. and Marth, G.T. (2011). BamTools: A C++ API and toolkit for analyzing and managing BAM files. Bioinformatics. *27*(12), 1691-1692.

Behall, K.M., Scholfield, D.J. and Canary, J. (1988). Effect of starch structure on glucose and insulin responses in adults. Am. J. Clin. Nutr. *47*(3), 428-432.

Blankenberg, D., Gordon, A., Von Kuster, G., Coraor, N., Taylor, J., Nekrutenko, A. and Galaxy Team, the G. (2010). Manipulation of FASTQ data with Galaxy. Bioinformatics. *26*(14), 1783-1785.

Bligh, H.F.J., Larkin, P.D., Roach, P.S., Jones, C.A., Fu, H. and Park, W.D. (1998). Use of alternate splice sites in granule-bound starch synthase mRNA from low-amylose rice varieties. Plant Mol. Biol. *38*(3), 407-415.

Bucher, P. (1990). Weight matrix descriptions of four eukaryotic RNA polymerase II promoter elements derived from 502 unrelated promoter sequences. J. Mol. Biol., *212*(4), 563-578.

Busk, P.K., Jensen, A.B. and Pagès, M. (1997). Regulatory elements in vivo in the promoter of the abscisic acid responsive gene rab17 from maize. Plant J. *11*(6), 1285-1295.

Butler, J.E.F. and Kadonaga, J.T. (2002). The RNA polymerase II core promoter: a key component in the regulation of gene expression. Genes Dev. *16*(20), 2583-2592.

Chang, W.C., Lee, T.Y., Huang, H.D., Huang, H.Y. and Pan, R.L. (2008). PlantPAN: Plant promoter analysis navigator, for identifying combinatorial cis-regulatory elements with distance constraint in plant gene groups. BMC Genomics. *9*(1), 561.

De Vries, A. and Ripley, B.D. (2016). ggdendro: Create dendrograms and tree diagrams using 'ggplot2' [online]. [Accessed on 22.07.2016]. Available at https://CRAN.R-project.org/package=ggdendro.

Dry, I., Smith, A., Edwards, A., Bhattacharyya, M., Dunn, P. and Martin, C. (1992). Characterization of cDNAs encoding two isoforms of granule-bound starch synthase which show differential expression in developing storage organs of pea and potato. Plant J. *2*(2), 193-202.

Dubouzet, J.G., Sakuma, Y., Ito, Y., Kasuga, M., Dubouzet, E.G., Miura, S., Seki, M., Shinozaki, K. and 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*(4), 751-763.

Francki, M., Civáň, P. and Švec, M. (2009). Genome-wide analysis of rice (*Oryza sativa* L. subsp. *japonica*) TATA box and Y Patch promoter elements. Genome. *52*(3), 294-297.

Gagniuc, P. and Ionescu-Tirgoviste, C. (2012). Eukaryotic genomes may exhibit up to 10 generic classes of gene promoters. BMC Genomics. *13*(1), 512.

Giardine, B., Riemer, C., Hardison, R.C., Burhans, R., Elnitski, L., Shah, P., Zhang, Y., Blankenberg, D., Albert, I., Taylor, J., Miller, W., Kent, W. J. and Nekrutenko, A. (2005). Galaxy: A platform for interactive large-scale genome analysis. Genome Res. *15*(10), 1451-1455.

Gilmour, S.J., Zarka, D.G., Stockinger, E.J., Salazar, M.P., Houghton, J.M. and Thomashow, M.F. (1998). Low temperature regulation of the Arabidopsis CBF family of AP2 transcriptional activators as an early step in cold-induced COR gene expression. Plant J. *16*(4), 433-442.

Higo, K., Ugawa, Y., Iwamoto, M. and Korenaga, T. (1999). Plant *cis*-acting regulatory DNA elements (PLACE) database: 1999. Nucleic Acids Res. *27*(1), 297-300.

Hirano, H.Y. and Sano, Y. (1998). Enhancement of Wx gene expression and the accumulation of amylose in response to cool temperatures during seed development in rice. Plant Cell Physiol. *39*(8), 807-812.

Hirano, H.Y. and Sano, Y. (2000). Comparison of Waxy gene regulation in the endosperm and pollen in *Oryza sativa* L. Genes Genet. Syst. *75*(5), 245-249.

Hu, P., Zhao, H., Duan, Z., Linlin, Z. and Wu, D. (2004). Starch digestibility and the estimated glycemic score of different types of rice differing in amylose contents. J. Cereal Sci. *40*(3), 231-237.

Isshiki, M., Morino, K., Nakajima, M., Okagaki, R.J., Wessler, S.R., Izawa, T. and Shimamoto, K. (1998). A naturally occurring functional allele of the rice waxy locus has a GT to TT mutation at the 5' splice site of the first intron. Plant J. *15*(1), 133-138.

Jaglo-Ottosen, K.R., Gilmour, S.J., Zarka, D.G., Schabenberger, O. and Thomashow, M.F. (1998). *Arabidopsis* CBF1 over expression induces COR genes and enhances freezing tolerance. Science. *280*(5360), 104-106.

Kaplan, B., Davydov, O., Knight, H., Galon, Y., Knight, M. R., Fluhr, R. and Fromm, H. (2006). Rapid transcriptome changes induced by cytosolic Ca^{2+}transients reveal ABRE-related sequences as Ca^{2+}-responsive cis elements in Arabidopsis. Plant Cell. *18*(10), 2733–2748.

Kawahara, Y., de la Bastide, M., Hamilton, J.P., Kanamori, H., McCombie, W.R., Ouyang, S., Schwartz, D.C., Tanaka, T., Wu, J., Zhou, S., Childs, K.L., Davidson, R.M., Lin, H., Quesada-Ocampo, L., Vaillancourt, B., Sakai, H., Lee, S. S., Kim, J., Numa, H., Itoh, T.,

Buell, C.R. and Matsumoto, T. (2013). Improvement of the *Oryza sativa* Nipponbare reference genome using next generation sequence and optical map data. Rice. *6*(1), 4.

Kutach, A.K. and Kadonaga, J.T. (2000). The downstream promoter element DPE appears to be as widely used as the TATA box in *Drosophila* core promoters. Mol. Cell. Biol. *20*(13), 4754-4764.

Lagrange, T., Kapanidis, A.N., Tang, H., Reinberg, D. and Ebright, R.H. (1998). New core promoter element in RNA polymerase II-dependent transcription: Sequence-specific DNA binding by transcription factor IIB. Genes Dev. *12*(1), 34-44.

Larkin, M.A., Blackshields, G., Brown, N.P., Chenna, R., McGettigan, P.A., McWilliam, H., Valentin, F., Wallace, I.M., Wilm, A., Lopez, R., Thompson, J.D., Gibson, T.J. and Higgins, D. G. (2007). Clustal W and Clustal X version 2.0. Bioinformatics. *23*(21), 2947-2948.

Larkin, P.D. and Park, W.D. (1999). Transcript accumulation and utilization of alternate and non-consensus splice sites in rice granule-bound starch synthase are temperature-sensitive and controlled by a single-nucleotide polymorphism. Plant Mol. Biol. *40*(4), 719-727.

Li, H. and Durbin, R. (2009). Fast and accurate short read alignment with Burrows-Wheeler transform. Bioinformatics. *25*(14), 1754-1760.

Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., Marth, G., Abecasis, G. And Durbin, R. (2009). The sequence alignment/map format and SAMtools. Bioinformatics. *25*(16), 2078-2079.

Li, J.Y., Wang, J. and Zeigler, R.S. (2014). The 3000 Rice genome project: New opportunities and challenges for future rice research. Gigascience, *3*:8. doi: 10.1186/2047-217X-3-8. eCollection 2014.

Luo, H., Song, F., Goodman, R.M. and Zheng, Z. (2005). Up-regulation of OsBIHD1, a rice gene encoding BELL homeodomain transcriptional factor, in disease resistance responses. Plant Biol. *7*(5), 459-468.

Nakamura, M., Tsunoda, T. and Obokata, J. (2002). Photosynthesis nuclear genes generally lack TATA-boxes: A tobacco photosystem I gene responds to light through an initiator. Plant J. *29*(1), 1-10.

Nelson, O. and Pan, D. (1995). Starch synthesis in maize endosperms. Annu. Rev. Plant Physiol. Plant Mol. Biol. *46*(1), 475-496.

Qin, F., Sakuma, Y., Li, J., Liu, Q., Li, Y.Q., Shinozaki, K. and Yamaguchi-Shinozaki, K. (2004). Cloning and functional analysis of a novel DREB1/CBF transcription factor involved in cold-responsive gene expression in *Zea mays* L. Plant Cell Physiol. *45*(8), 1042-1052.

Quinn, J.M., Eriksson, M., Moseley, J.L. and Merchant, S. (2002). Oxygen deficiency responsive gene expression in *Chlamydomonas reinhardtii* through a copper-sensing signal transduction pathway. Plant Physiol. *128*(2), 463-471.

R Core Team. (2017). R: A language and environment for statistical computing [online]. [Accessed on 22.10.2017]. Available at https://cran.r-project.org/.

RStudio Team. (2015). RStudio: Integrated development for R. [online]. [Accessed on 22.10.2017]. Available at https://www.rstudio.com/.

Shahmuradov, I.A., Gammerman, A.J., Hancock, J.M., Bramley, P.M. and Solovyev, V.V. (2003). PlantProm: A database of plant promoter sequences. Nucleic Acids Res. *31*(1), 114-117.

Simpson, S.D., Nakashima, K., Narusaka, Y., Seki, M., Shinozaki, K. and Yamaguchi-Shinozaki, K. (2003). Two different novel *cis*-acting elements of *erd1*, a *clpA* homologous *Arabidopsis* gene function in induction by dehydration stress and dark-induced senescence. Plant J. *33*(2), 259-270.

Skinner, J.S., von Zitzewitz, J., Szűcs, P., Marquez-Cedillo, L., Filichkin, T., Amundsen, K., Stockinger, E.J., Thomashow, M.F., Chen, T.H.H. and Hayes, P. M. (2005). Structural, functional, and phylogenetic characterization of a large CBF gene family in Barley. Plant Mol. Biol. *59*(4), 533-551.

Suzuki, M., Ketterling, M.G. and McCarty, D.R. (2005). Quantitative statistical analysis of *cis*-regulatory sequences in ABA/VP1- and CBF/DREB1-regulated genes of *Arabidopsis*. Plant Physiol. *139*(1), 437–447.

Tatarinova, T., Brover, V., Troukhan, M. and Alexandrov, N. (2003). Skew in CG content near the transcription start site in *Arabidopsis thaliana*. Bioinformatics. *19 Suppl. 1*, 313-314.

Thorvaldsdottir, H., Robinson, J.T. and Mesirov, J.P. (2013). Integrative genomics viewer (IGV): High-performance genomics data visualization and exploration. Brief. Bioinform. *14(2)*, 178-192.

Wickham, H. (2009). ggplot2: Elegant graphics for data analysis [online]. [Accessed on 22.10.2017]. Available at https://CRAN.R-project.org/package=ggplot2.

Xue, G.P. (2003). The DNA-binding activity of an AP2 transcriptional activator HvCBF2 involved in regulation of low-temperature responsive genes in barley is modulated by temperature. Plant J. *33*(2), 373-383.

Yamamoto, Y.Y., Ichida, H., Abe, T., Suzuki, Y., Sugano, S. and Obokata, J. (2007). Differentiation of core promoter architecture between plants and mammals revealed by LDSS analysis. Nucleic Acids Res. *35*(18), 6219-6226.

Yang, T. and Poovaiah, B.W. (2002). A Calmodulin-binding/CGCG box DNA-binding protein family involved in multiple signaling pathways in plants. J. Biol. Chem. *277*(47), 45049-45058.