Journal of
Plant Development Sciences
Meerut, India

RESEARCH ARTICLE

RELATIONSHIP BETWEEN TFBS ABUNDANCE AND GENE EXPRESSION IN CARBOHYDRATE METABOLISM-RELATED GENES DURING FLOWERING IN RICE

M. Krishna Kumar[+], D.P. Syamaladevi*[+£], S.M. Balachandran and D. Subrahmanyan

ICAR-Indian Institute of Rice Research, Hyderabad, India. PIN 500 030
[£] Currently at: ICAR-Indian Institute of Spices Research, Kozhikode
Email: dpsdevi@gmail.com

Received-01.05.2023, Revised-12.05.2023, Accepted-23.05.2023

Abstract: The abundance of transcription factor binding sites (TFBSs) on promoter sequences differs from gene to gene. Understanding how this difference affects gene expression would enable us to develop superior lines through allele mining based genomic selection or genome editing. To understand the relationship between the TFBSs abundance and transcription in carbohydrate metabolism-related genes that are central to growth and grain yield in rice. For this, we first analysed the expression pattern of four carbohydrate metabolism-related enzymes - Sucrose Phosphate Synthase (SPS), PME (Pectin Methyl Esterase), Galacturonosyl transferase (GalAT) and Phosphoglucomutase (PGM). Then, we obtained a reliable set of TFBSs on the core and distal promoters of these genes following Support Vector Machine (SVM) method and stringent cut off of 90% TFBS motif similarity. Fold change in gene expression during phenology change from vegetative to flowering showed a linear relationship with the TFBS abundance in the case of an SBP family transcription factor, SPL12. Further, identification of CpG islands in the promoters explained the deviations in this linear relationship between gene expression and TFBS abundance. Using two DNA parameters - the TFBS abundance and the CpG islands - we provide predictive insights to the gene expression of carbohydrate metabolism-related genes in rice.

Keywords: Transcription factor binding sites, Gene expression, Carbohydrate metabolism, Phosphoglucomutase, Rice

INTRODUCTION

Gene expression is regulated precisely at multiple levels with transcriptional regulation at the bottom. Transcription is a complex function of various factors viz., amount of transcription factors (TFs), affinity of TFs to the respective transcription factor binding sites (TFBSs), methylation status of the upstream regulatory sequences as well as the availability of suitably exposed TFBS (Brignall *et al*. 2019; Suzuki & Bird 2008). The amount of TF is further affected by protein movement, mitotic protein dilution and miRNA-controlled transcript accumulation (Hofhuis and Heidstra 2018; Samad *et al*. 2017). Many genes are regulated through TF affinity as well, for instance, WUS regulates *CLV3* through binding site affinity in apical meristem (Hofhuis and Heidstra 2018). Methylation status of promoter regions is known to play critical role in transcriptional regulation (Suzuki & Bird 2008).

Though motif predictions identify numerous TFBS on promoter sequences, it is unknown if all of them are involved in TF binding. Therefore, identifying the functional TFBS from a set of predicted TFBS is a difficult task in the absence of TF binding assay experiments such as ChIP-Chip and ChIP-Seq (Robertson *et al*. 2007; Mikkelsen *et al*. 2007; Johnson *et al*. 2007). Here we are providing insights on the relationship between the abundance of transcription factor binding sites and gene expression in the transcriptional level, using bioinformatics and qPCR expression study.

For this analysis, we chose sucrose, starch and pectin metabolism-related genes due to their importance in growth and grain yield in rice (Fu *et al*. 2011; Gebbing and Schnyder 1999; Daynard *et al*. 1969; Deng *et al*. 2016; Peaucelle *et al*. 2011). Sucrose Phosphate Synthase (SPS) is a key enzyme in the biosynthesis of sucrose (Bruneau *et al*. 1991). Phosphoglucomutase (PGM) inter-converts G-1-P to G-6-P and positively impact the sucrose biosynthesis. PGM is attributed to plant growth, seed and root development (Malinova *et al*. 2014). Galacturonosyl transferase (GalAT) is the major component of pectin biosynthesis protein machinery and has key roles in plant growth and development (Orfila *et al*. 2005; Sterling *et al*. 2006; Godoy *et al*. 2013). Pectin Methyl Esterase (PME) is another important enzyme associated with cellular growth in plant and thereby affecting yield (Wen *et al*. 1999; Louvet *et al*. 2006; Fujita *et al*. 2010). It contributes to the stiffening of the cell wall by producing blocks of de-methyl

*Corresponding Author

esterified carboxyl groups that can interact with calcium ions forming a pectate gel (Bosch *et al*. 2005) and it was shown to have a role in determining the growth rate in the model plant *Characoralina* (Proseus *et al*. 2006)

Here we quantified TFBS abundance in the promoters of genes related to carbohydrate metabolism and correlated with the expression data from qPCR experiments during the mid-vegetative stage to flowering stage

MATERIALS AND METHODS

Plant material

Indica rice variety Gothra Bidhan-3 was planted in pots during Kharif season with three biological replications. Leaf samples were collected at mid vegetative stage (45 days after germination) and 3 days after panicle emergence. To understand the trend in gene expression during mid vegetative to the flowering stage, samples were collected at 45 days (mid) and 65 days (late) after sowing.

Real-time expression using qPCR

For conducting Real-time qPCR, we isolated total RNA from each leaf sample using Trizol method (Chomczynski and Mackey 1995). The extracted RNA was treated with DNase I (Thermoscientific) as per the manufacturer's instructions and quality and quantity were assessed using a spectrophotometer. This RNA was used for First-strand cDNA synthesis using oligo dT primers from Revert Aid First Strand cDNA Synthesis Kit (Thermo Scientific). The first-strand cDNA synthesis reaction was set up as per the manufacturer's instructions. Using the Quant prime online server (http://quant prime.mpimp-golm. mpg.de/), we designed the qPCR primers specific to target genes and the internal control genes (Table 2). Actin was used as an internal control. Real-time PCR was set up in Roche 96 light cycler machine using Takara Bio SYBR Green Master Mix following qPCR protocols given in Table 3. For each biological replicate, there were three experimental replicates PCRs were set up. Primer optimization titration experiments were conducted to avoid interference from primer dimers and non-specific amplification. After qPCR, we calculated the relative expression by the comparative Cq method and expressed as a fold change of expression. Statistical significance was tested using student t-test at confidence level 0.001.

Database search and promoter sequence extraction

The abundance of TFBSs was obtained by following a protocol as depicted in Fig. 1. The KEGG database (https://www.genome.jp/kegg/) (Kanehisa and Goto 2000) search for sucrose and starch metabolism-related genes retrieved a set of enzymes and their EC numbers. The enzyme names were obtained from the BRENDA database (https://www.brenda-enzymes. org/) (Jeske *et al*. 2019) and subsequent keyword search in RAP-DB (https://rapdb.dna. affrc.go.jp/) (Sakai et al. 2013) provided a set of rice enzymes involved in sucrose and starch metabolism. Pectin metabolism-related enzymes were added to this data set. All locus IDs thus obtained were used to search in PlantPAN 3.0 (http://plantpan.itps. ncku.edu.tw/index.html) (Chow *et al*. 2019) and obtained the 1000bp upstream promoter sequence of each gene.

Transcription factor binding site (TFBS) prediction

PlantPAN3.0 collects plant transcription factor binding profiles from PLACE, TRANSFAC (public release 7.0), AGRIS, and JASPER databases. The TF binding sites having more than 90% similarity were shortlisted as most likely TFBS and obtained the TF distribution map/network diagram through the 'Gene group analysis' option in PlantPAN3.0 within the rice genome. The abundance of each TFBS was obtained by counting the conjoined locations in the TF distribution map/network diagram.

CpG/CpNpG island prediction

In plants, DNA methylations are found on the cytosine of CpG and CpNpG islands (Pradhan *et al*. 1999; Cao and Jacobsen 2002; Lindroth *et al*. 2001). CpG/CpNpG islands prediction was based on Ponger and Mouchiroud's method (Ponger and Mouchiroud. 2002). CpG/CpNpG islands were defined as the DNA regions that are longer than 500 nucleotides, with a moving average C+C frequency of above 0.5 and a moving average CpG/CpNpG observed/ expected (o/e) ratio more than 0.6 (Ponger and Mouchiroud. 2002). The predictions were performed using CpG island prediction tool in PlantPan 3.0 (Chang *et al*. 2008; Chow *et al*. 2019)

RESULTS AND DISCUSSION

Transcript levels of SPS, PME, GalAT and PGM during phenology transition from mid-vegetative to flowering stage

There was a marked increase in gene expression of SPS and PME during flowering compared to the mid vegetative stage. The fold changes of expression in SPS and PME during the flowering stage were 3.03 and 1.76 folds respectively (Fig. 1). GalAT and PGM levels also increased during flowering but to a lower extent. The gene expressions of GalAT and PGM during flowering were 1.35 and 1.14 folds over the mid vegetative stage (Fig. 1).

To understand the pattern of gene expression just before the flowering stage, we performed a qPCR experiment during the late vegetative stage and the fold changes were calculated over the mid vegetative stage (Fig. 2). SPS expression during the late vegetative phase over mid vegetative phase was 3.00 folds, comparable to that during flowering (3.03folds). This indicates that there was no significant induction of SPS expression during late vegetative to flowering stage. PME expression during the late vegetative phase was 1.05 folds,

which was lower than that during flowering. This suggests that there is a considerable induction of PME after the late vegetative stage (Fig. 2).

Unlike SPS and PME, GalAT and PGM gene expressions decreased drastically during the late-vegetative stage as indicated by the negative fold changes of -3.67 and -2.09 folds respectively (Fig. 2). This decrease was followed by a hike revealing the induction of GalAT and PGM genes during flowering. Both the genes are known to be involved in molecular processes during the reproductive stages of plant growth (Egli *et al*. 2010; Wang *et al*. 2013).

TFBS abundance of an SBP family transcription factor and gene expression

SBP family transcription factor SPL12is known to be induced during phenology transition from vegetative to the reproductive stage and is an important member regulating plant transition from juvenile to adult (Shikata *et al*. 2009; Xu *et al*. 2016, Wang *et al*. 2016). TFBS abundance was the highest in PME promoter with five SPL12 binding sites followed by GalAT and PGM (Fig. 3). The expression data of these three genes also followed the same trend - highest in PME, followed by GalAT and PGM (Fig. 1). SPS promoter showed the lowest SPL12 binding sites harboring just one site (Fig. 3). In line with this, change in expression of SPS gene also remained more or less the same with 3.00 and 3.03 folds during late vegetative and flowering stages respectively (Fig. 1, 2). However, the inducibility of GalAT and PGM appears to be more during the late vegetative stage to flowering (from -ve fold change to positive), even though they had a lower number of SPL12 binding sites compared to PME. This suggests the involvement of additional factors such as the position of the TFBS and methylation patterns in the regulation of gene expression during flowering.

TFBS position and CpG islands on 1Kb promoter

In the PME promoter, the SBP binding sites were mostly congregated in the core promoter, whereas in the case of SPS and PGM promoters they were in the distal promoter region (Fig. 4). In the case of GalAT, there was one site at the core promoter, remaining were at the distal promoter. Only PME promoter harbored most of the SPL12 binding sites on its core promoter region possibly contributing to its steady increase in expression during the mid-vegetative stage to flowering.

Chip-Chip experiments have shown that TFBS that contain CpG are involved in constitutive gene expression and some CpG containing sequences are also involved in inducible and tissue-specific gene regulation (Rozenberg *et al*. 2008; Luu *et al*. 2013). Our analysis identifies CpG islands in the entire 1Kb promoters (covering core and distal promoter regions) of SPS (Fig. 5a) which are indicative of possible methylation at these sites. The single SPL12 site identified at -500bp position is located within this CpG island. In PME only distal ~600bp (-1000 to -367bp) region was found to have CpG island sparing all the SPL12 binding sites in the core promoter (Fig. 5a). In GalAT, the promoter region between -893 to -253 (covering parts of the core and distal promoter) was detected to be GC rich, leaving just two SPL12 binding sites outside the CpG island (Fig. 5a). No CpG island was located in the PGM promoter. The G+C, G/C ratio and start-up score were highest in SPS followed by PME and GalAT (Fig. 5b). Separation of SPL12 TFBS to core promoter and the CpG islands to the distal promoter would have been playing a role in the steady induction of PME from mid vegetative stage to flowering. In the case of GalAT and PGM, two SPL12 TFBS have been available without interference from methylation (Fig. 4) yielding a slight induction in gene expression during flowering. In the case of SPS, there was no induction as the entire promoter was predicted to be a CPG island which encompassed the single SPL12 site it harboured. Taken together, it can be deduced that the presence of CpG islands and the TFBS abundance are contributory factors in the regulation of expression of these genes during flowering.

Though SPS induction during flowering was negligible, the fold change of gene expression of SPS remained higher than other genes. Similarly, there was down regulation of PGM and GalAT during the late vegetative stage. These deviations could be attributed to TF concentration or methylation like mechanisms of chromatin remodeling (Ballare *et al*. 2013).

Prediction of expression of carbohydrate metabolism-related genes

Based on this understanding of SPL12 binding sites and gene expression during flowering, we analysed 1Kb promoter regions of genes related carbohydrate metabolism following the schema in Fig. 6 in an attempt to predict expression levels of genes during flowering in rice. It is most likely that Sucrose phosphate phosphatase (SPP), which is having the maximum number of SPL12 binding sites in its promoter get maximally induced during flowering (Fig. 7a). SPP harboured nine SPL12 binding sites and no CpG island in its promoter making us predict higher induction during flowering (Fig. 7a, 7b). On the other hand, isoamylase, beta - fructofuranosidase and starch phosphorylase had the least number of SPL12 binding sites (Fig. 7a). Beta –fructofu-ranosidase and starch phosphorylase do not have any CpG island, whereas isoamylase has G/C rich regions (Fig. 7b, c) and therefore can be predicted to be poorly induced during flowering. In the case of NDP Glucose - starch Glucosyl transferase, Fructo-kinase and Sucrose phosphate phosphatase genes the TFBS were mostly in the core promoter regions

whereas in starch phosphorylase there was no TFBS present in the core promoter (Fig. 8).

TF families with abundant TFBSs

Further, to understand the expression of carbohydrate metabolism-related genes under different spatiotemporal and environmental conditions, we obtained the TFBS distribution pattern of seven different transcription factor families. TFBSs corresponding to three transcription factors, B3 (IDEF), NFYB and TCP families were the most abundant with an average number of TFBSs ranging from 18-25 (Fig. 9a). TFBSs of other transcription factor families were sparse with average number of sites ranging from 3-6 (Fig. 9b). The distribution maps of TFBSs of B3, TCP and NFYB reveal uniform distribution throughout the promoter (Fig. 10, 11, 12).

B3 family: B3 family transcription factor, IDEF1, the expression is known to be induced by Fe deficiency (Kobayashi 2007; Ogo *et al*. 2008; Kobayashi *et al*. 2010). The TFBSs of IDEF1 was highest in NDP-glucose-starch glucosyltransferase and isoamylase (Fig. 13). However, isoamylase expression could be reduced due to the presence of CPG island in its promoter (Fig. 7a) to a lower level than starch synthase in response to IDEF1. In contrast, the promoter of Glucose 6-Phosphate isomerase (GPI) harboured the least number of TFBSs for IDEF1 and predicted to have CpG island (Fig. 7b, c). This suggests a scanty expression of GPI during Iron deficient conditions. GPI is involved in starch accumulation in Mesophyll Cells, Growth and Photosynthetic Capacity and antioxidant metabolism mediated stress tolerance (Bahaji *et al*. 2015; Seong *et al*. 2013). This suggests that down regulation of GPI due to absence of IDEF1 binding sites in its promoter could be a contributing factor in growth retardation under iron deficient conditions. Allele mining or genome editing targeting large number of the IDEF1 binding sites on GPI promoter can lead to development of stress tolerant lines.

NFYB family: NFYB is a ubiquitous, drought-responsive transcription factor (Li *et al*. 2008). SPP, due to a large number of NFYB binding sites and the absence of CpG islands, is expected to be induced during drought stress. However, a lower starch synthase transcription is expected in response to drought owing to the lower NFYB binding sites as well as the presence of a CpG island in its promoter (Fig. 7b, c).

TCP family: TCP is directly involved in leaf development and growth hormones signalling (Li *et al*. 2015). TCP binding sites predominate in the core promoters of SPS and Fructo kinase genes (Fig. 13). SPS due to the presence of CpG island, would not be responding much to TCP family TFBS.

TF families with sparse TFBS

Relatively sparse TFBS with a 3-6 average number of sites were found for bHLH, bZip, Homeo and SBP transcription factors (Fig 9, Fig. 14). The bHLH sites were highest in starch synthase and isoamylase. SPP, GPI and PGM promoters harboured the least number of bHLH binding sites. The bZIP sites were in general lesser compared to other sites. Notably, Beta-fructofuranosidase, has only one site for bZIP family transcription factor.

Congregation of TFBS in core and distal promoters

In the case of TF families for which the TFBSs were abundant, they appeared in core as well as distal promoters. For instance, the average percentage of the three most abundant TFBs in the 300bp core promoter region ranges from 31-39% (Fig. 10, Table 1), leaving 61-69% of TFBS distributed over 700bp. In contrast, in the case of sparse TFBSs, the binding sites showed distinct positioning. For instance, SPL12 sites in PME were congregated at the core promoter (Fig. 4), whereas in the case of starch phosphorylase and PGM they were mostly distributed at the distal promoter (Fig. 8).

Variation in promoters

Sequence variations between promoter sequences of different alleles of the same gene from different cultivars or haplotypes would result in differences in TFBSs and therefore gene expression. Therefore, the indels and substitutions in the promoters of select genes of two indica cultivars (Shushui 498 and RPBio-226) were compared with that of the japonica cultivar Nipponbare. The promoters showed high identity (>98%) with few mutations (Table 4). The predicted TF binding sites of SPL12 did not overlap with the mutant sites. However, in GalAT, the proximal promoter was found to be highly variable and one SPL12 binding site on GalAT at position 957 (Table 4) was found to be within this variable region.

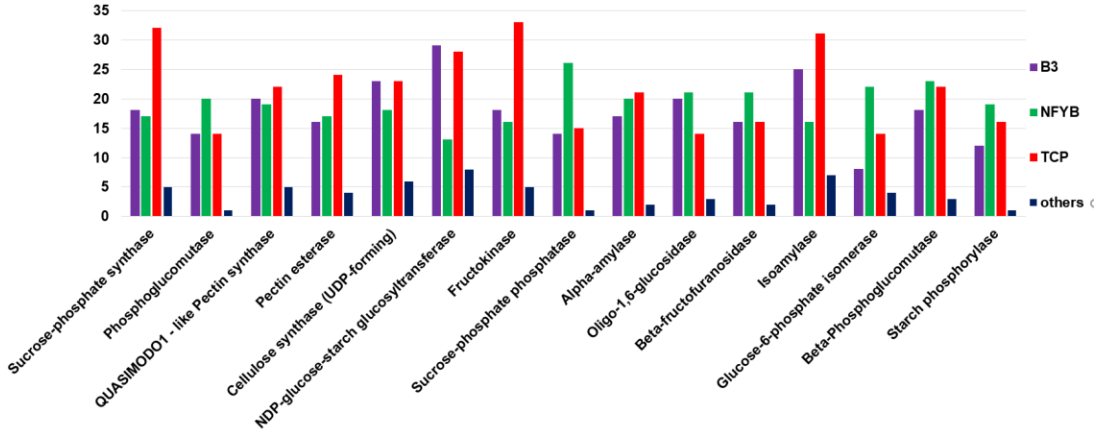

Fig. 1 qPCR expression of SPS, PME, GalAT and PGM during flowering stage compared to mid vegetative stage in indica variety GothraBidhan-3. Y-axis shows the fold change of expression and the bars shows the standard deviation of the mean of 3 biological replicate samples.

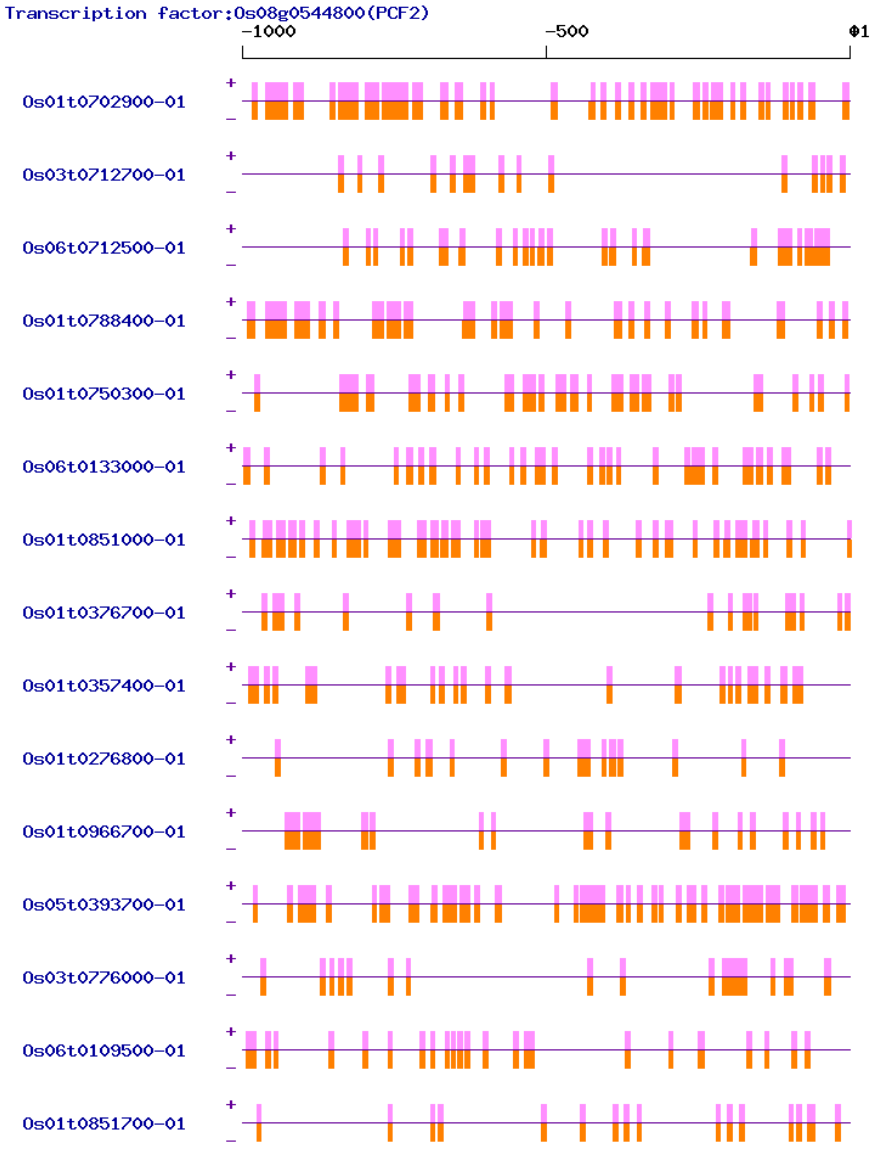

Fig. 2 Fold-change of gene expression from real-time qPCR experiment is shown in Y axis for SPS, PME, GalAT and PGM during late vegetative stage compared to mid vegetative stage in GothraBidhan-3. Error bars in the graph is the standard deviation of the mean fold change in gene expression for 3 biological replicates.

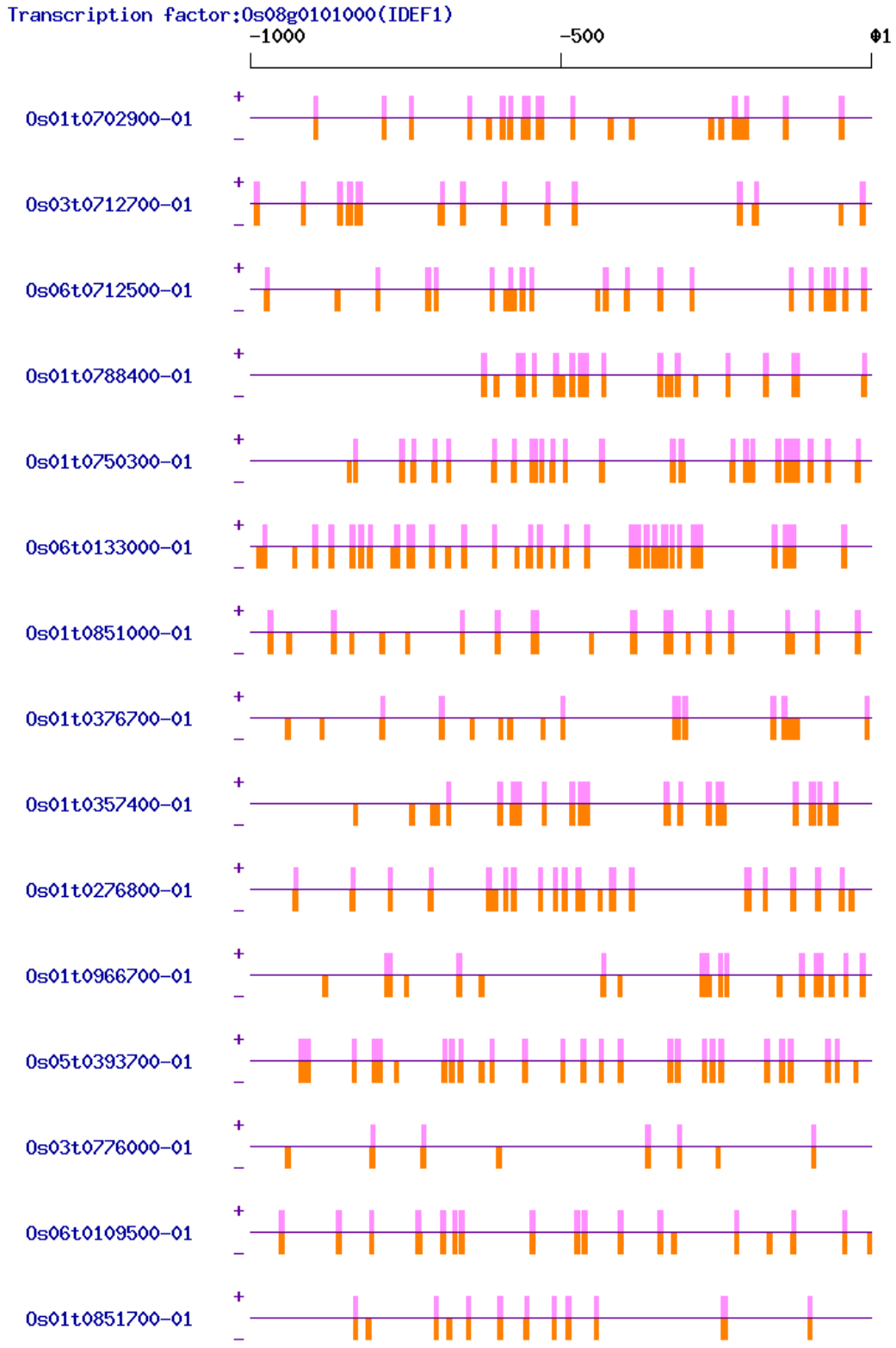

Fig. 3 The number of Transcription Factor Binding Sites (TFBSs) corresponding toSPL12 (an SBP family transcription factor) in 1Kb promoter sequences of SPS, PME, GalAT and PGM genes.

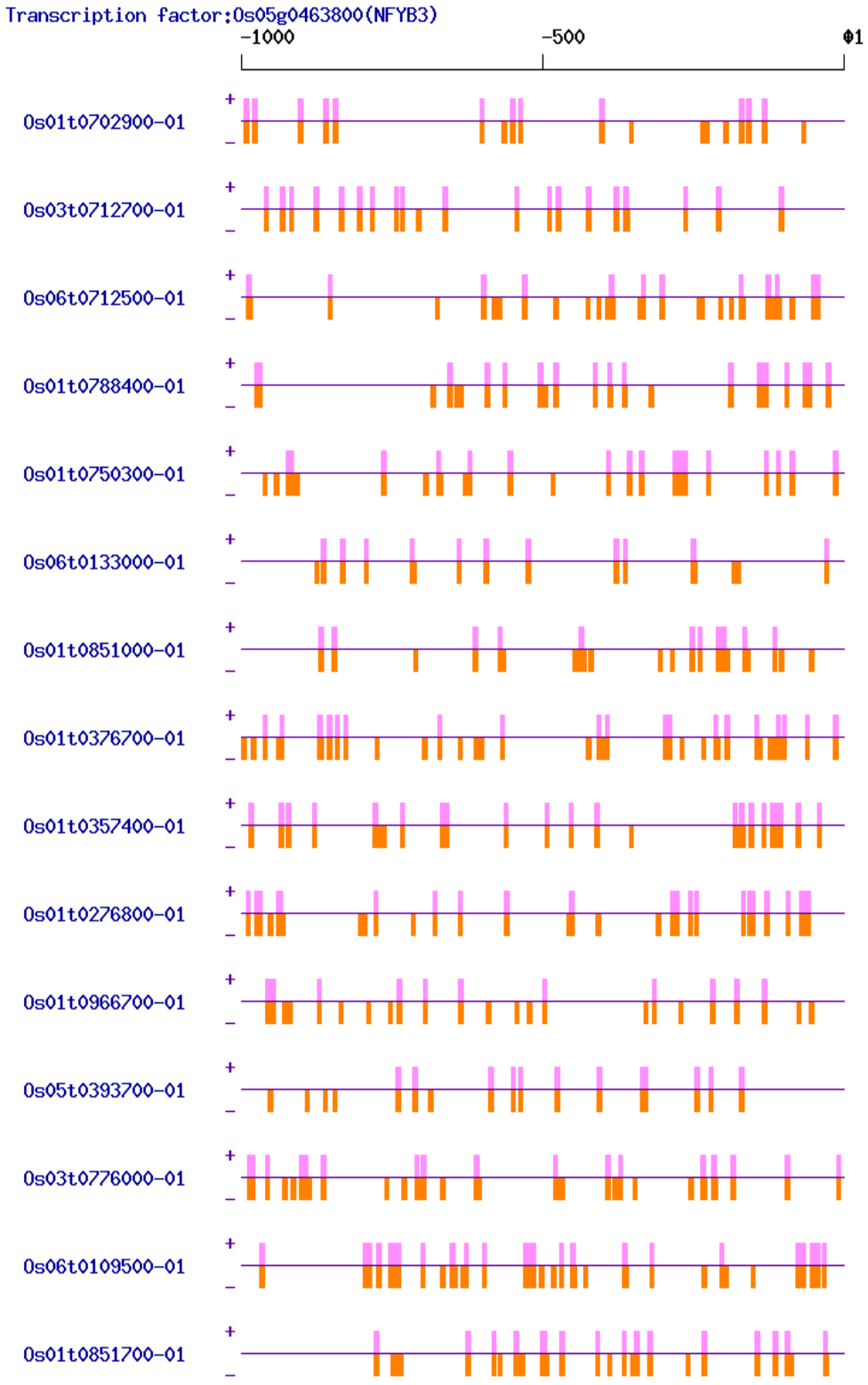

Fig. 4 The map showing distribution of TFBSs corresponding to SPL12 in 1Kb promoter sequences of SPS, PME,GalAT and PGM genes.

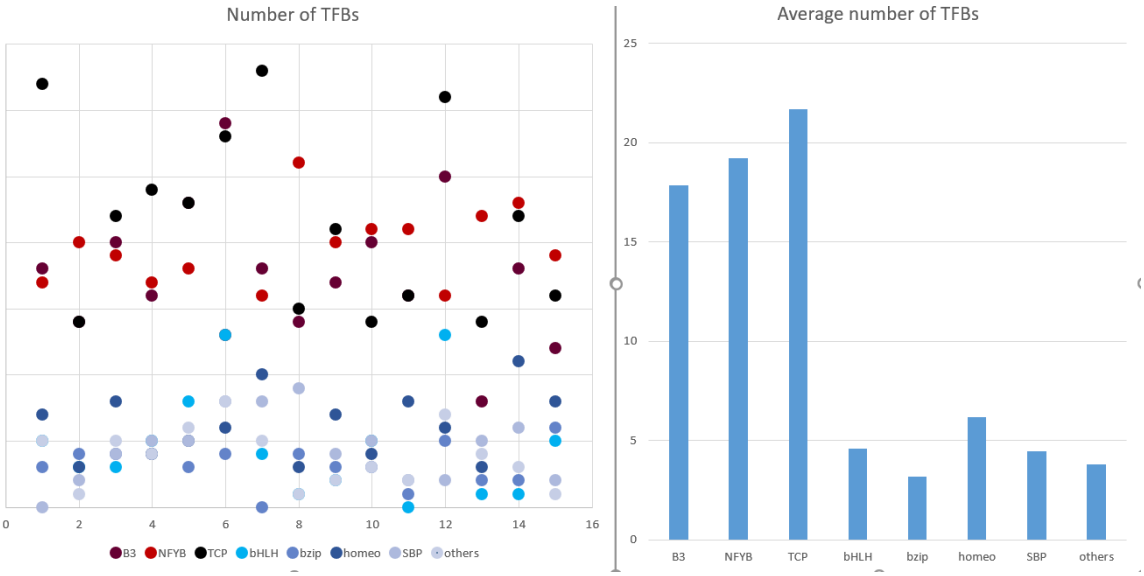

Fig. 5 a) Graph showing length (base pairs) of predicted CpG islands and b) CpG island features (C+G frequency, CpGo/e ratio, Startup) in 1 Kb promoter sequences of SPS, PME, GalAT and PGM genes

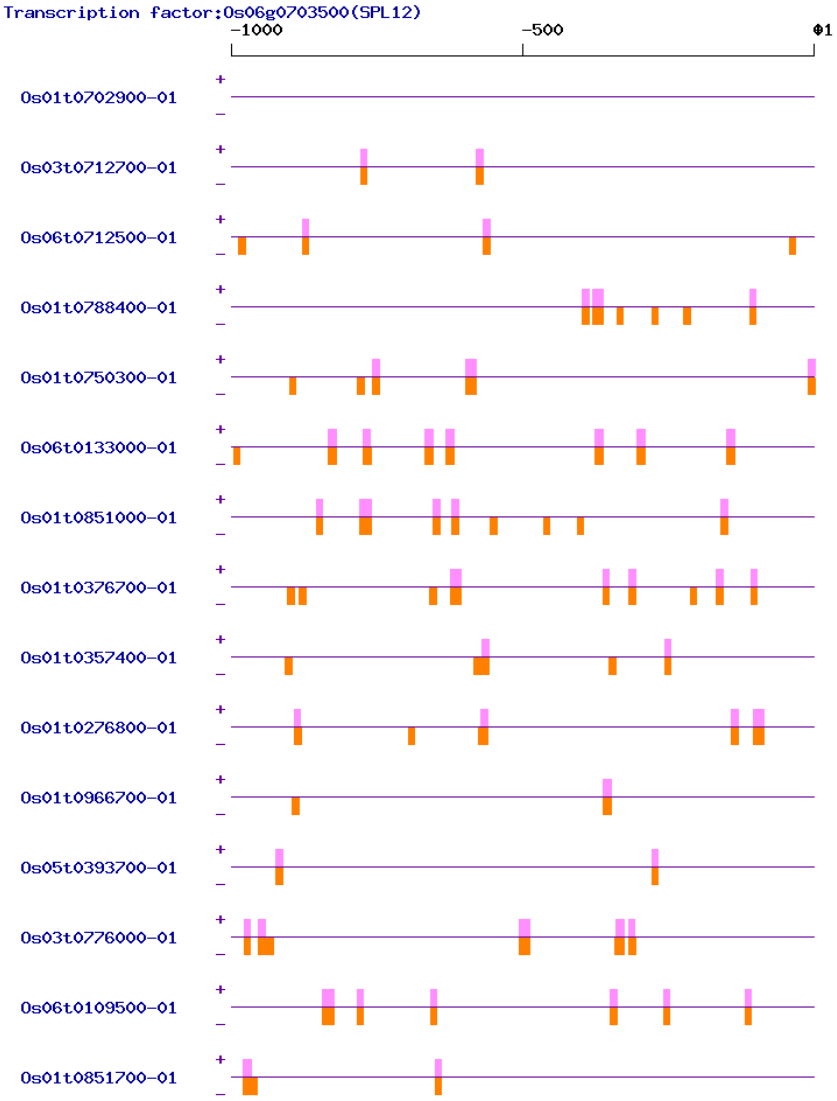

Fig. 6 Bioinformatics database search schema used for obtaining the TFBS distribution on 1Kb promoter sequences

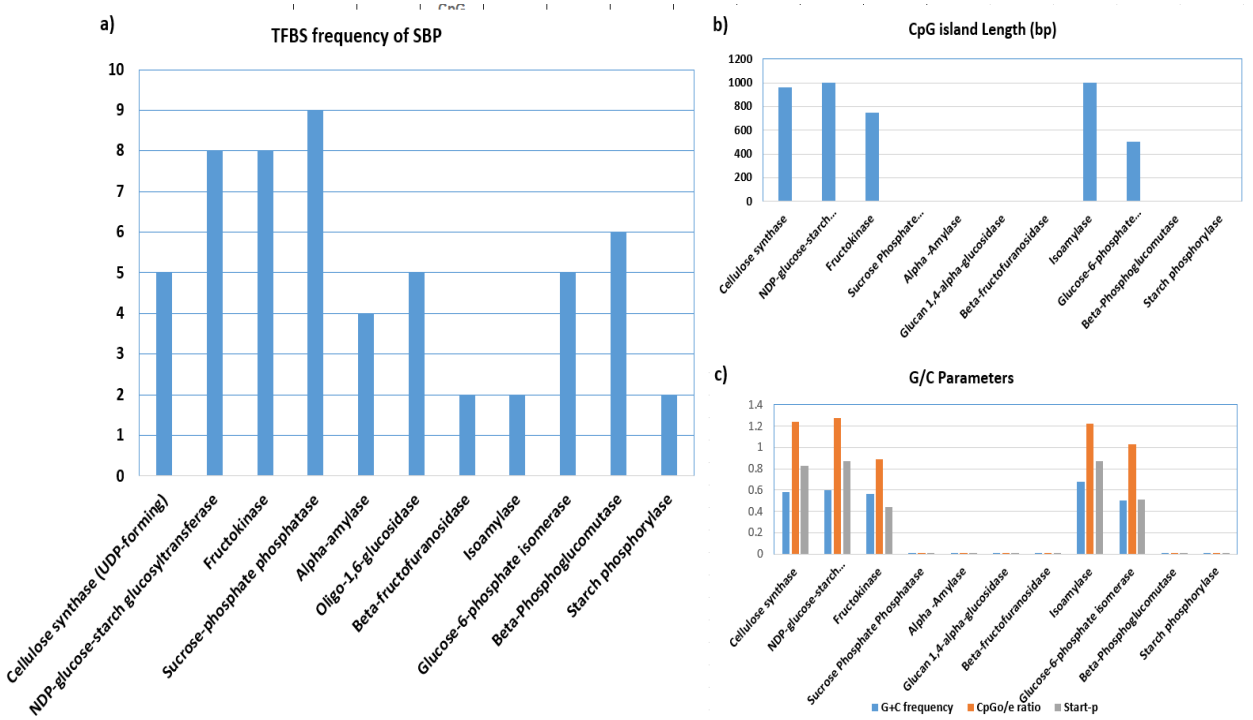

Fig. 7 a) TFBS frequency of SBP family transcription factor b) length of predicted CpG island (in base pairs) and c) CpG island features (C+G frequency, CpGo/e ratio, Startup) of 1 Kb promoter region of select carbohydrate metabolism related genes

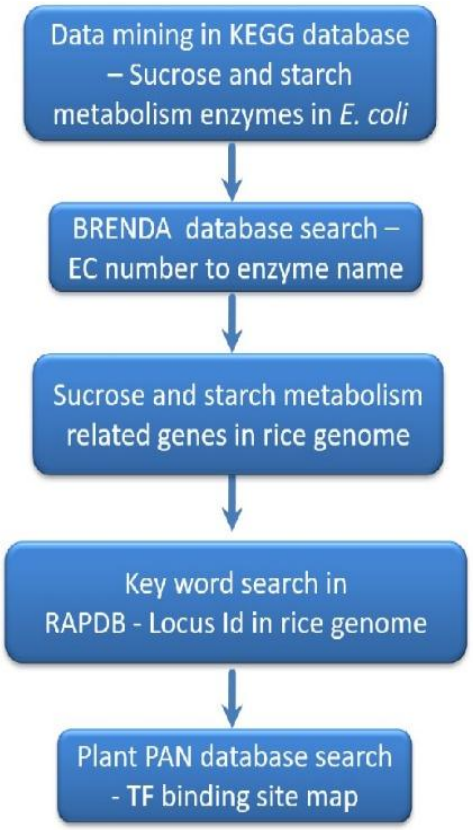

Fig. 8 Map showing the positions of TFBSs corresponding to SPL12 on 1Kb promoter region of the select carbohydrate metabolism related genes

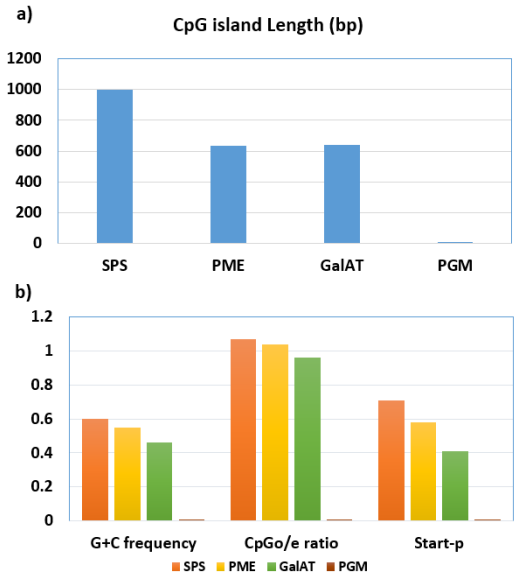

Fig. 9 a) Graph showing the distribution pattern TFBSs corresponding to 8 different transcription factor family proteins on the 1Kb promoter sequences of the select carbohydrate metabolism related genes b) Average number of TFBSs over the different carbohydrate metabolism related genes

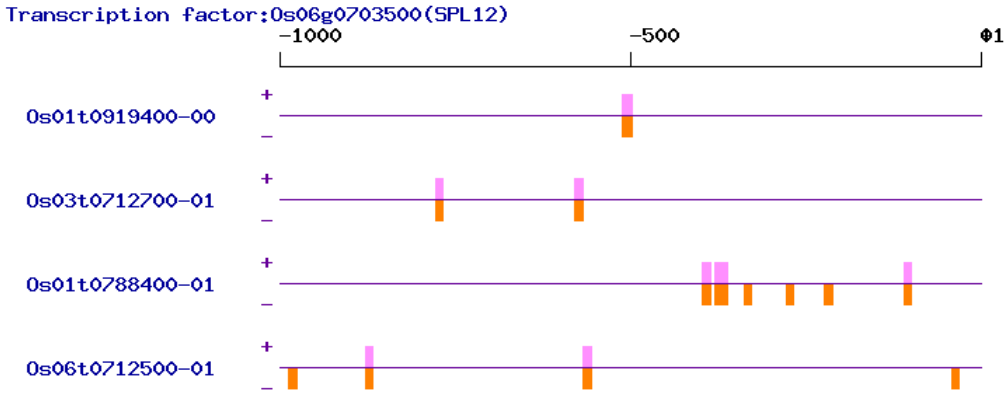

Fig. 10 Map showing the distribution of TFBS of B3 transcription factor family on 1Kb promoter region of select carbohydrate metabolism related genes

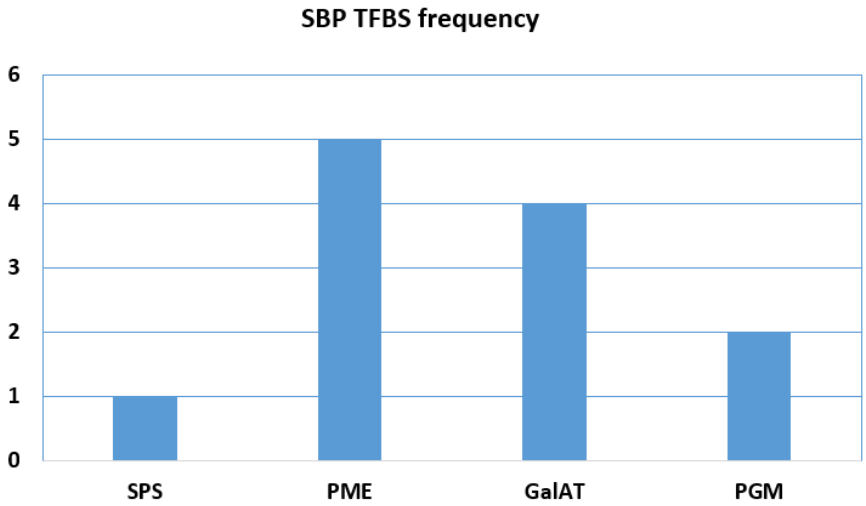

Fig. 11 Map showing the distribution of TFBS of NFYB transcription factor family on 1Kb promoter region of select carbohydrate metabolism related genes

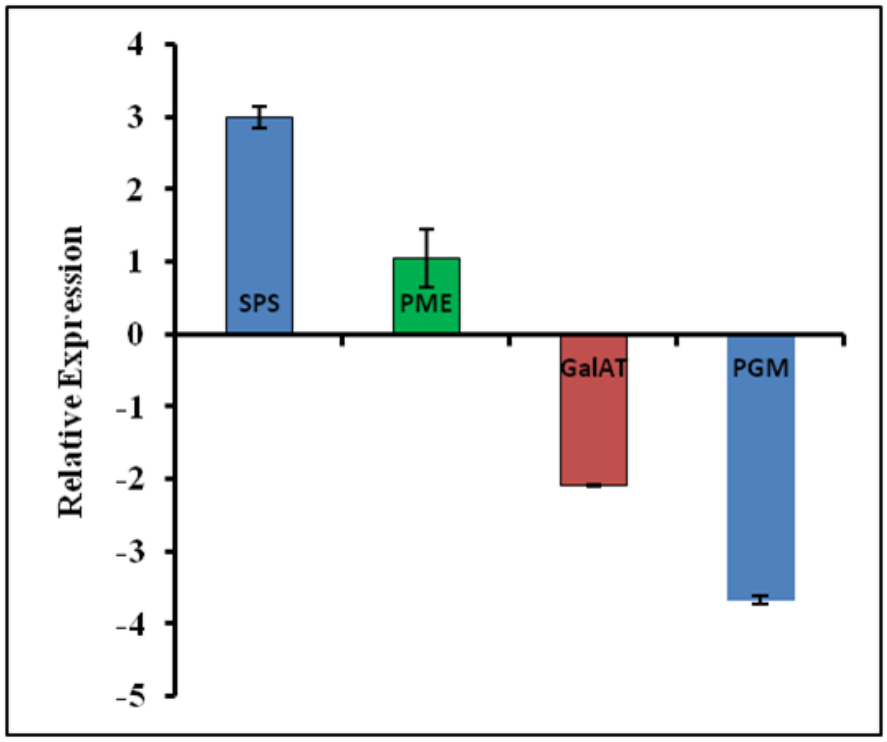

Fig. 12 Map showing the distribution of TFBS of TCP transcription factor family on 1Kb promoter region of select carbohydrate metabolism related genes

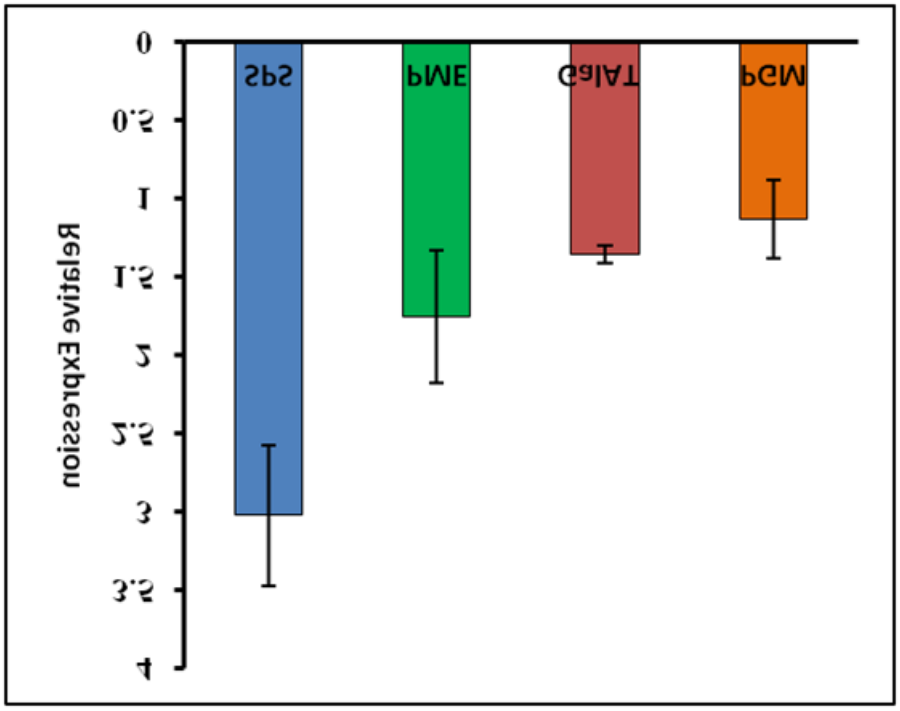

Fig. 13 Graph showing the TFBS frequency of a) B3, b) NFYB c) TCP and d) others on 1Kb promoter of select carbohydrate metabolism related genes

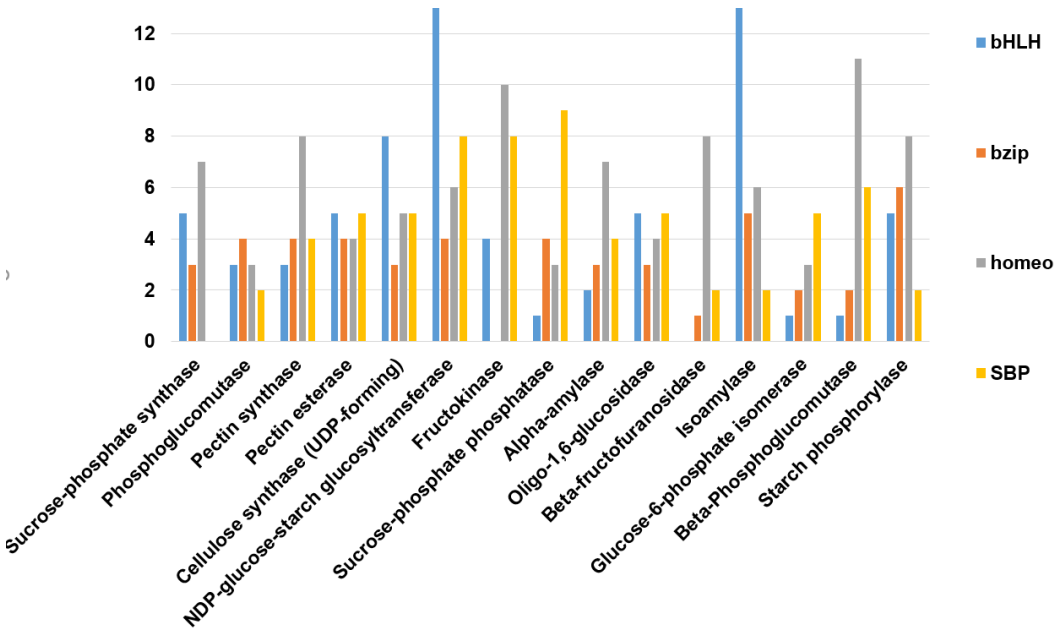

Fig. 14 Graph showing the TFBS frequency of transcription factor families -bHLH, bZip, Homeo and SBP.

Table 1. List of qPCR primers used in this study

ID	Locus	Description		Primer	Size(bp)
Os01t0919400-00	LOC_Os01g69030	Sucrose-phosphate synthase (EC 2.4.1.14).	Forward	5'CGAGGTCACAAGCACTCAGGTATC 3'	125
			Reverse	CGTCTTGTGTAACCCTGACAGC	
Os03t0712700-01	LOC_Os03g50480	Phosphoglucomutase (EC 5.4.2.2)	Forward	GACAACCTTGGAGGCGATAAGC	120
			Reverse	AGCAGCTCCAGCATCAACATTCTC	
Os01t0788400-01	LOC_Os01g57854	Pectin methyl esterase (EC 3.1.1.11)	Forward	CGACGGCTCCACAACTTTCAAC	83
			Reverse	TGTTCTCCACCTTGAGGTCCTG	
Os06t0712500-01	LOC_Os06g49810	GalAT (EC 2.4.1.-)	Forward	TGATGCTGAACTTCCCAAGAGTGC	130
			Reverse	AGCATCGCTCTTAACCGCTGAG	
Os04g57210	LOC_Os04g57210.1	actin	forward	CCGGTGGATCTTCATGCTTACCTGG	213
		actin	reverse	CGACGAGTCTTCTGGCGAAACTGC	

Table 2. List of enzymes related to starch and sucrose metabolism and percentage of TFBSs falling in core promoter region in 1Kb promoter sequences of their genes

Sl No	Enzyme name	Id No	% of TFBS in core promoter		
			B3	NFYB	TCP
1	Sucrose-phosphate synthase	Os01t0702900-01	33.33	35.29	43.75
2	Phosphoglucomutase	Os03t0712700-01	28.57	15	35.71
3	4-alpha-galacturonosyltransferase (Similar to QUASIMODO1)	Os06t0712500-01	30.00	47.36	27.27
4	Pectin methyl esterase	Os01t0788400-01	31.25	35.29	29.17
5	Cellulosesynthase (UDP-forming)	Os01t0750300-01	39.13	38.89	30.44
6	NDP-glucose-starch glucosyltransferase	Os06t0133000-01	17.24	30.77	35.71
7	Fructokinase	Os01t0851000-01	33.33	56.25	30.30
8	Sucrose-phosphate phosphatase	Os01t0376700-01	28.57	42.31	60.00
9	Alpha-amylase	Os01t0357400-01	41.18	40	42.86
10	Glucan 1,4-alpha-glucosidase	Os01t0276800-01	30.00	38.10	21.43
11	Beta-fructofuranosidase	Os01t0966700-01	56.25	28.57	50.00
12	Isoamylase	Os05t0393700-01	36.00	18.75	51.61

13	Glucose-6-phosphate isomerase	Os03t0776000-01	25.00	27.27	57.14
14	Beta-Phosphoglucomutase	Os06t0109500-01	27.78	30.44	27.27
15	Starch phosphorylase	Os01t0851700-01	16.67	31.58	43.75
Average % of TFBS in core promoter			**31.62**	**34.39**	**39.10**

Table 3. Real-time qPCR protocol

Process	**Time**
Pre-incubation	
95°c	600 seconds
Two step	
95°c	15 seconds
58°c	15 seconds
Melting curve	
95°c	10 seconds
65°c	60 seconds
97°c	60 seconds
Cooling	
37°c	60 seconds

Table 4. Single nucleotide Substitutions and indels in comparison with Nipponbare sequence

Promoter	**Cultivar**	**% ID**	**Hit**	**Substitutions/ Indels**	**Position of SPL12 binding site Nipponbare**
SPS	Shuhui498	98.80	CP018157.1	C13G, A293G, C347T, A350G, G361A, C408T, A426G, G436A, C515T, T553C, T578-, C920A	488-490
	RP Bio-226	98.80	CP012609.1	C13G, A293G, C347T, A350G, G361A, C408T, A426G, G436A	
PGM	Shuhui498	98.80	CP018159.1	A15G, G109A, G111A, A112G, C135T, A490G, A568C, G623T	221, 420
	RP Bio-226	99.20	CP012611.1	A15G, G109A, G111A, A112G, C135T, A490G, A568C, G623T, A708T, A709-, -743C, C785-	
PME	Shuhui498	99.20	CP018157.1	T63C, T90C, -263G, -263G, -263A, G740A, C822T, C880A	602, 619, 626,661, 721,776,889
	RP Bio-226	99.40	CP012609.1	T90C, -263G, -263G, -263A, C822T, C880A	
GalAT	Shuhui498	98.14	CP018162.1	A97G, A107-,A108T, A255G, G444A, A514G, T626A, T627-,T655C, A716G, T735C, -757T, -758G, -759T, -760A,T800G, T844G, 912-1000 highly variable	12,121, 432, 957
	RP Bio-226	98.03	CP012614.1 CP012614.1	A97G, A107-, A108T, A255G, G444A, A514G, T625A, T626A, T627-, T655C, A716G, T735C, -757T, -758G, -759T, -760A, T800G, T844G, 912-1000 highly variable	

CONCLUSION

In this study, we tried to analyse the relationship between promoter sequence features and gene expression in carbohydrate metabolism-related genes. The abundance of TFBS and the presence of CpG islands are two important aspects of gene expression regulation. Our analysis and qPCR expression study could show how these two aspects are correlated to gene expression in SPS, PME, GalAT and PGM. Our analysis also revealed abundant and sparse transcription factor binding sites in 1 Kb promoter region of carbohydrate metabolism-related genes. This study forms the basis for developing tools that can predict the functional TFBS from a large number of predicted TFBS on the promoter sequences. However additional regulatory pathways at post-transcriptional levels need to be investigated through further experimentation to

understand the complete mechanism of gene regulation.

ACKNOWLEDGEMENTS

Authors thank SERB (Grant No.YSS/2015/001916) and ICAR-IIRR for funding and providing facilities for experiments.

REFERENCES

Bahaji *et al.* (2015). Plastidic phosphoglucose isomerase is an important determinant of starch accumulation in mesophyll cells, growth, photosynthetic capacity, and biosynthesis of plastidic cytokinins in Arabidopsis PLoS ONE **10**(3), e0119641.

Google Scholar

Ballare *et al.* (2013). Nucleosome-driven transcription factor binding and gene regulation. *Molecular Cell*, **49**, 67–79.

Google Scholar

Bosch *et al* (2005) Pectin methyl esterases and pectin dynamics in pollen tubes. *Plant Cell*, **17**(12), 3219–3226.

Google Scholar

Brignall *et al.* (2019) Considering abundance, affinity, and binding site availability in the nf-κb target selection puzzle. *Frontiers in immunology*, **10**:609.

Google Scholar

Bruneau *et al.* (1991) Sucrose phosphate synthase, a key enzyme for sucrose biosynthesis in plants: protein purification from corn leaves and immunological detection. *Plant Physiol*., **96**(2), 473–478.

Google Scholar

Cao and Jacobsen (2002) Locus-specific control of asymmetric and CpNpG methylation by the DRM and CMT3 methyl transferase genes. *Proc Natl Acad Sci USA*, **99**, 16491–16498.

Google Scholar

Chang *et al.* (2008) PlantPAN: Plant promoter analysis navigator, for identifying combinatorial cis-regulatory elements with distance constraint in plant gene groups BMC. *Genomics*, **9**, 561.

Google Scholar

Chomczynski, P. and Mackey, K. (1995) Modification of the TRI reagent procedure for isolation of RNA from polysaccharide- and proteoglycan-rich sources, *Biotechniques*,**19(6)**, 942-5.

Google Scholar

Chow *et al.* (2019) PlantPAN3.0: a new and updated resource for reconstructing transcriptional regulatory networks from ChIP-seq experiments in plants. *Nucleic Acids Res*., **47**, D1155–D1163.

Google Scholar

Daynard *et al.* (1969) Contribution of stalk soluble carbohydrates to grain yield in corn (*Zea mays* L.). *Crop Sci*., **9**, 831–834.

Google Scholar

Deng *et al.* (2016) Polyaspartate urea and nitrogen management affect nonstructural carbohydrates and yield of rice. *Crop Sci*., **56**, 3272–3285.

Google Scholar

Egli *et al.* (2010) Loss of cytosolic phospho-glucomutase compromises gametophyte development in Arabidopsis. *Plant Physiology*, **154**, 659–1671.

Google Scholar

Fu *et al.* (2011) Pre-anthesis non-structural carbohydrate reserve in the stem enhances the sink strength of inferior spikelets during grain filling of rice. *Field Crops Res*.,**123**, 170–182.

Google Scholar

Fujita *et al.* (2010) Rice expression atlas in reproductive development. *Plant and Cell Physiology*, **51**(12), 2060–2081.

Google Scholar

Gebbing and Schnyder (1999) Pre-anthesis reserve utilization for protein and carbohydrate synthesis in grains of wheat. *Plant Physiol*. **121**, 871–878.

Google Scholar

Godoy *et al.* (2013) Galacturonosyltransferase 4 silencing alters pectin composition and carbon partitioning in tomato. *Journal of Experimental Botany*, **64**(8), 2449–2466.

Google Scholar

Hofhuis and Heidstra (2018) Transcription factor dosage: more or less sufficient for growth Current Opinion in Plant Biology 45(A), 50-58.

Google Scholar

Jeske *et al* (2019) BRENDA in 2019: a European ELIXIR core data resource. *Nucleic Acids Res*., **47**(D1), D542-D549.

Google Scholar

Johnson *et al.* (2007) Genome-wide mapping of in vivo protein-DNA interactions *Science*, **316**, 1497–1502.

Google Scholar

Kobayashi *et al.* (2010) The spatial expression and regulation of transcription factors IDEF1 and IDEF2. *Ann Bot*., **105**(7), 1109–1117.

Google Scholar

Kobayashi *et al.* (2007) The transcription factor IDEF1 regulates the response to and tolerance of iron deficiency in plants Proceedings of *the National Academy of Sciences*, **104**(48), 19150-19155.

Google Scholar

Kanehisa and Goto (2000) KEGG: Kyoto Encyclopedia of Genes and Genomes Nucleic Acids Res., **28**, 27-30.

Google Scholar

Li *et al.* (2008) The *Arabidopsis* NFYA5 transcription factor is regulated transcriptionally and post transcriptionally to promote drought resistance Plant Cell 20(8), 2238-51.

Google Scholar

Li *et al.* (2015) The *Arabidopsis thaliana* TCP transcription factors: A broadening horizon beyond development Plant Signal Behav 10(7), e1044192.
Google Scholar

Lindroth *et al.* (2001) Requirement of Chromomethylase3 for maintenance of CpXpG methylation. *Science*, **292**, 2077–2080.
Google Scholar

Louvet *et al.* (2006) Comprehensive expression profiling of the pectin methylesterase gene family during silique development in Arabidopsis thaliana. *Planta*, **224**(4), 782-91.
Google Scholar

Luu *et al.* (2013) Disclosing the crosstalk among DNA methylation, transcription factors, and histone marks in human pluripotent cells through discovery of DNA methylation motifs. *Genome Res*., **23**, 2013-2029.
Google Scholar

Malinova *et al.* (2014) Reduction of the cytosolic Phosphoglucomutase in Arabidopsis reveals impact on plant growth, seed and root development, and carbohydrate partitioning PLoS ONE **9**(11): e112468.
Google Scholar

Mikkelsen *et al.* (2007) Genome-wide maps of chromatin state in pluripotent and lineage-committed cells. *Nature*, **448**, 553–560.
Google Scholar

Ogo *et al.* (2008) A Novel NAC Transcription Factor, IDEF2, That Recognizes the Iron Deficiency-responsive Element 2 Regulates the Genes Involved in Iron Homeostasis in Plants. *Journal of Biological Chemistry*, **283**(19), 13407-17.
Google Scholar

Orfila *et al* (2005) QUASIMODO1 is expressed in vascular tissue of *Arabidopsis thaliana* inflorescence stems, and affects homogalacturonan and xylan biosynthesis. *Planta*, **222**, 613–622.
Google Scholar

Peaucelle *et al* (2011) Pectin-induced changes in cell wall mechanics underlie organ initiation in Arabidopsis. *Curr Biol*., **21**, 1720–1726.
Google Scholar

Ponger and Mouchiroud (2002) CpGProD: identifying CpG islands associated with transcription start sites in large genomic mammalian sequences. **Bioinformatics**, **18** (4), 631-633.
Google Scholar

Pradhan *et al* (1999) Effect of CWG methylation on expression of plant genes. *Biochem J*. **341**, 473–476.
Google Scholar

Proseus *et al* (2006) Calcium pectate chemistry controls growth rate of Characorallina. *Journal of experimental botany*, **57**(15), 3989–4002.
Google Scholar

Robertson *et al* (2007) Genome-wide profiles of STAT1 DNA association using chromatin immunoprecipitation and massively parallel sequencing. *Nat Methods*, **4**, 651–657.
Google Scholar

Rozenberg *et al* (2008) All and only CpG containing sequences are enriched in promoters abundantly bound by RNA polymerase II in multiple tissues BMC *Genomics*, **9**, 67.
Google Scholar

Sakai *et al* (2013) Rice Annotation Project Database (RAP-DB): An integrative and interactive database for rice genomics. *Plant & Cell Physiol*., **54**(2):e6.
Google Scholar

Seong *et al* (2013) Transient overexpression of the Miscanthussinensis glucose-6-phosphate isomerase gene (MsGPI) in Nicotianabenthamiana enhances expression of genes related to antioxidant metabolism POJ **6**(6), 408-414.
Google Scholar

Shikata *et al* (2009) Arabidopsis SBP-box genes SPL10, SPL11 and SPL2 control morphological change in association with shoot maturation in the reproductive phase. *Plant Cell Physiol*., **50**(12), 2133-45.
Google Scholar

Sterling *et al* (2006) Functional identification of an *Arabidopsis* pectin biosynthetic homogalacturo nangalacturonosyl transferase Proceedings of *the National Academy of Sciences, USA* **103**, 5236–5241.
Google Scholar

Suzuki and Bird (2008) DNA methylation landscapes: Provocative insights from epigenomics. *Nature Reviews Genetics*, **9**, 465–476.
Google Scholar

Wang *et al* (2016) Squamosa promoter binding Protein-Like 2 controls floral organ development and plant fertility by activating ASYMMETRIC LEAVES 2 in Arabidopsis thaliana. *Plant Mol Biol* **92**(6), 661-674.
Google Scholar

Wang *et al* (2013) Arabidopsis Galacturonosyl transferase (GAUT) 13 and GAUT14 have redundant functions in pollen tube growth. *Molecular Plant*, **6**(4), 1131-1148.
Google Scholar

Wen *et al* (1999) Effect of pectin methylesterase gene expression on pea root development. *Plant Cell*, **11**, 1129–1140.
Google Scholar

Xu *et al* (2016) Developmental functions of miR156-regulated Squamosa Promoter binding protein-like (spl) Genes in *Arabidopsis thaliana* PLoS Genet **12**(8), e1006263.
Google Scholar