BioTechnologia vol. 98(1) • pp. 41-51 • 2017
Journal of Biotechnology, Computational Biology and Bionanotechnology RESEARCH PAPERS

http://doi.org/10.5114/bta.2017.66616

In silico identification of transcription factors associated with the biosynthesis of carotenoids in corn (*Zea mays* L.)

ZAHRA ZINATI [1] *, LEYLA NAZARI [2], PAOLO BAGNARESI [3], RUDABE RAVASH [4]

[1] Department of Agroecology, College of Agriculture and Natural Resources of Darab, Shiraz University, Iran
[2] Seed and Plant Improvement Research Department, Fars Agricultural and Natural Resources Research Center, AREEO, Zarghan, Iran
[3] Consiglio per la Ricerca in Agricoltura e l'Analisi dell'Economia Agraria, Genomics Research Centre, Fiorenzuola d'Arda, Piacenza, Italy
[4] Department of Agronomy and Plant Breeding, Shahrekord University, Shahrekord, Iran

Abstract

Carotenoids, a diverse group of colorful pigments, contribute to the development, light harvesting and photoprotection in plants as well as human health. Due to the interesting properties of carotenoids, enhanced carotenoid biosynthesis has been of ongoing interest. Recent advances in computational biology and bioinformatics make it more feasible to understand the transcriptional regulatory network underlying carotenoid biosynthesis. Studies on carotenoid biosynthesis in corn (*Zea mays* L.) have indicated the pivotal role of the phytoene synthase gene *PSY1* (accession: GRMZM2G300348) in endosperm color and carotenoid accumulation in corn kernels. Computational approaches such as Genomatix, PlantPAN, PlantCARE, PlantTFDB and IGDE6 have been used for promoter prediction, regulatory features and transcription factor identification, as well as pairwise promoter comparisons. Four transcripts have been identified for the *PSY1* gene. Based on Genomatix and PlantPAN, the promoter predicted for GRMZM2G300348_T01 was different from that predicted for the other three transcripts (GRMZM2G300348_T02, GRMZM2G300348_T03 and GRMZM2G300348_T04). The results indicated that the promoter of GRMZM2G300348_T01 has more diverse motifs involved in hormonal/environmental stress responses. The most significant result obtained from this study is the discovery of two transcription factors belonging to the HB family that are co-expressed with all four transcripts of *PSY1* under environmental stresses. It is, therefore, likely that these transcription factors may act as critical regulators of *PSY1* gene expression in corn. Identification of the proteins acting upstream of *PSY1* within corn will shed light on the fine tuning of *PSY1* expression regulation. Such an understanding would also contribute to metabolic engineering aimed at enhanced carotenoid biosynthesis.

Key words: carotenoid biosynthesis, maize, transcription factors

Introduction

Carotenoids, a diverse family of isoprenoid pigments naturally found in plants, algae and certain species of fungi and bacteria, perform key roles in the development, photosynthesis and root-mycorrhizal interactions of phytohormones (such as abscisic acid, strigolactone or other signaling molecules) as well as their precursors (Cazzonelli et al., 2010). Furthermore, numerous research studies have shown the potential importance of carotenoids in the prevention and treatment of several kinds of chronic diseases, such as cardiovascular disease and certain types of cancer (Rao and Rao, 2007; Fiedor and Burda, 2014). These interesting properties of carotenoids have attracted considerable research effort aimed at carotenoid biosynthesis. Carotenoid biosynthesis is regulated during the life cycle of a plant by developmental and environmental signals (Welsch et al., 2000; Bramley, 2002; Li et al., 2008; Rodriguez-Villalon et al., 2009; Bou-Torrent et al., 2015). Phytoene synthase (PSY), encoded by the phytoene synthase 1 (*PSY1*) gene, is the first rate-limiting enzyme (Fu et al., 2010) and is considered the most important regulatory enzyme

* Corresponding author: Department of Agroecology, College of Agriculture and Natural Resources of Darab, Shiraz University; e-mail: zahrazinati@shirazu.ac.ir

in the carotenoid biosynthesis pathway (Cazzonelli et al., 2010). The transcript abundance of the phytoene synthase gene alters during fruit ripening (Tao et al., 2007), flower development (Zhu et al., 2002) or stress (Li et al., 2008; Arango et al., 2010), which leads to changes in the total carotenoid content (Cazzonelli et al., 2010).

In corn, there are three paralogous *PSY* genes which encode functional enzymes. Li and coworkers (2008) demonstrated a significant correlation between PSY1 expression level and total carotenoid content.

Other studies on carotenoid biosynthesis in corn (*Zea mays* L.) have indicated the pivotal role of the *PSY1* (accession: GRMZM2G300348) gene in endosperm color formation and carotenoid accumulation in corn kernels (Palaisa et al., 2003; Gallagher et al., 2004).

Recent studies on the regulation of phytoene synthase have provided evidence that post-transcriptional mechanisms play a major role in governing carotenoid biosynthesis. For example, *Arabidopsis* ORANGE (OR) proteins are significant post-transcriptional regulators of phytoene synthase in governing the biosynthesis of carotenoids (Zhou et al., 2015). In addition, in a recent study by Park and coworkers (2016) the sweet potato (*Ipomoea batatas*) *Orange* gene (*IbOr*) was found to play an essential role in the post-transcriptional regulation of IbPSY stability via its holdase chaperone activity, resulting in the accumulation of carotenoids and the development of abiotic stress tolerance in sweet potato (Park et al., 2016).

While structural genes involved in the carotenoid biosynthesis pathway are well characterized, little is known about the transcriptional regulators which control the expression of these structural genes in multiple plant systems (Sandmann et al., 2006). Identification of transcription factors regulating the expression of structural genes in the carotenoid biosynthesis pathway provides knowledge on gene expression for the potential enhancement of carotenoid production in plants via genetic engineering (Sagawa et al., 2015).

A comprehensive genome-wide transcription factor (TF) annotation and classification has been conducted in corn B73, leading to the prediction of 2538 genes as TFs, which were classified into 64 families (Lin et al., 2014). However, to the best of our knowledge, only a few transcription factors implicated in the regulation of accumulation of carotenoids in corn have been reported. As described in previous research (Toledo-Ortiz et al., 2010; Martel, 2011), changes in the transcript abundance of transcription factor genes coincide with changes in *PSY* mRNA transcripts and consequent alteration of carotenoid content. For instance, a survey conducted on the photostimulation of carotenoid gene expression in *Arabidopsis thaliana* indicated that phytochrome-interacting factor 1 (PIF1) down regulates the accumulation of carotenoids by binding to the *PSY* promoter and repressing *PSY* mRNA expression. Both *in vitro* and *in vivo* evidence demonstrated the repression of *PSY* expression by PIF1 binding (Toledo-Ortiz et al., 2010). Bou-Torrent et al. (2015) showed that PIF1 and related photolabile PIFs (apart from photostable PIF7) promote a shade-triggered decrease in carotenoid accumulation. Using a chromatin immunoprecipitation (ChIP) assay, Martel (2011) showed that in tomato the *RIPENING INHIBITOR* gene, a member of the MADS box family of transcription regulators, is directly involved in the regulation of phytoene synthase gene expression (*PSY1*).

The identification of cis-regulatory elements (CREs) which regulate gene transcription by functioning as binding sites for transcription factors can offer vital information on creating models in order to decipher the transcriptional regulatory network (Lee et al., 2002). CREs are specific short DNA motives, approximately 5-25 bp in length, that are recognized by TFs to regulate gene expression (Rani, 2007). Although most eukaryotic genes have a single promoter which lies near to the transcription start site, some genes contain alternative promoters activating transcription at different genome positions, often under particular conditions (Wittkopp and Kalay, 2011).

Recent developments in computational approaches to the prediction of promoters, their regulatory features (transcription factors binding sites and CpG/CpNpG islands) and the identification of transcription factors have made it more feasible to unravel the regulatory networks underlying the biosynthesis of carotenoids. In this regard, bioinformatics servers and databases are used to analyze the molecular mechanisms regulating *PSY1* gene expression and to identify co-expressed transcription factors which bind to the *PSY1* promoter regions.

Materials and methods

Identification of splice variants of the PSY1 gene in corn

In an attempt to identify the difference between all transcripts of *PSY1*, which result from alternative splic-

ing, we obtained all splice variants of *PSY1* and their sequence information from maizeGDB (http://www.maizegdb.org/) (Andorf et al., 2015).

Promoter prediction

In order to recognize the promoter regions, nucleotide sequences of 1.5 Kbp upstream from the translation start sites were identified using PlantPAN (Plant Promoter Analysis Navigator) (Chow et al., 2016) and Genomatix (http://www.genomatix.de/).

Tools for the identification of regulatory features (transcription factor binding sites and CpG/CpNpG islands)

The predicted promoter regions were scanned for the presence of transcription factor binding sites using PlantCARE, a web-based database of plant promoters and their Cis-Acting Regulatory Elements (http://bioinformatics.psb.ugent.be/webtools/plantcare/html/) (Lescot et al., 2002). For further studies, a variety of regulatory elements predicted using the PlantCARE database were categorized according to their functions as various types of elements: light cycle, hormonal/environmental response, site binding, metabolic, seed specific, and promoter core-related elements. Other elements were also categorized, including OBP-1 site and circadian elements, and elements with an unknown function. Based on the distribution of a number of cis-acting regulatory elements in each category, certain statistical methods were employed to carry out pairwise promoter comparisons; e.g. the Audic and Claverie (AC) test, the Fisher exact test, Chi-squared 2 × 2, Greller and Tobin, R of Stekel and Falciani and General Chi-squared available at IDEG6, a web tool for the identification of differentially expressed genes in multiple tag sampling experiments (http://telethon.bio.unipd.it/bioinfo/IDEG6/) (Romualdi et al., 2003).

Also, CpG/CpNpG islands were detected by CpGProD integrated into PlantPAN. The CpG/CpNpG islands were defined as short stretches of DNA that are longer than 500 nucleotides, with a moving average C+C frequency above 0.5 and a moving average CpG/CpNpG observed/expected (o/e) ratio of more than 0.6 (Ponger and Mouchiroud, 2002).

Identification of co-expressed transcription factors binding to the PSY1 promoter regions

Gene IDs for four transcripts including GRMZM2G300348_T01, GRMZM2G300348_T02, GRMZM2G300348_T03 and GRMZM2G300348_T04 were used as the input in PlantPAN to analyze the transcription factor binding sites and their corresponding transcription factors. Then, all transcription factors were analyzed to identify co-expressed transcription factors with each transcript under environmental (biotic and abiotic) stresses. The co-expression analysis was based on 24 microarray samples under acid soils, 24 microarray samples under drought, 16 microarray samples under waterlogging, 12 microarray samples under *Colletotrichum graminicola*, 12 microarray samples under *Meloidogyne incognita*, 12 microarray samples under *Phytophthora cinnamomi*, 20 microarray samples under *Phytophthora cinnamomi*, *Sporisorium reilianum f.sp. zeae* (*Kühn*) and 36 microarray samples under *Ustilago maydis* conditions. The Spearman's rank correlation coefficient (Spearman's rho) was used to discover the strength of the link between gene expressions. The threshold for the co-expression analysis was greater than or equal to 0.8. Then, PlantTFDB (Plant Transcription Factor Database) (http://plntfdb.bio.uni-potsdam.de/v3.0/) (Pérez-Rodríguez et al., 2010) was used to analyze co-expressed transcription factors in more detail. Moreover, to gain additional support for the results of the co-expression analysis in corn, co-expression analyzes were performed for the corresponding *PSY1* transcripts in both Arabidopsis and rice. To do this, the PlantPAN tool was first used to find the corresponding *PSY1* transcripts and co-expressed transcription factors in Arabidopsis and rice. Then, co-expression analyzes were performed, as described for corn.

Results

Promoter analysis

The sequence information for 4 transcripts extracted for the *PSY1* gene is given in Table 1. Based on Genomatix and PlantPAN, the promoter region predicted for GRMZM2G300348_T01 was different from that predicted for the other three transcripts (GRMZM2G 300348 _T02, GRMZM2G300348_T03 and GRMZM2G300348 _T04).

Functional motives for both predicted promoter regions are given in Table 2. With respect to hormonal/environmental response-related elements, AuxRR-core (cis-acting regulatory element involved in auxin responsiveness), CGTCA-motif (cis-acting regulatory element involved in themethyl jasmonate (MeJA)-responsiveness),

Table 1. Splice variants of *PSY1* and their sequence information

Transcript ID	Transcript	UniProt	bp	PFAM ID	KEGG EC	KOG ID	KEGG Ortholog	Arabidopsis best hit	Rice best hit	Protein
GRMZM2G300348_T01	AFW75702	A0A096T3W8	1811	PF00494				AT5G17230.3	LOC_Os06g51290.1	204aa
GRMZM2G300348_T02	AFW75700	A0A096T3W9	2042	PF00494				AT5G17230.1	LOC_Os06g51290.1	414aa
GRMZM2G300348_T03	AFW75701	A0A096T3X0	2084	PF00494	2.5.1.32	KOG1459	K02291	AT5G17230.1	LOC_Os06g51290.1	407aa
GRMZM2G300348_T04	AFW75699	Q6EIC3	2093	PF00494	2.5.1.32	KOG1459	K02291	AT5G17230.1	LOC_Os06g51290.1	410aa

Table 2. Functional motif numbers of identified Cis-Acting Regulatory Elements

	CARE	Promoter region of GRMZM2G300348_T01	GRMZM2G300348_T02, T03, T04	Function
Light cycle – related element	AE-box	2	1	part of a module for light response
	ATCT-motif	2	1	part of a conserved DNA module involved in light responsiveness
	Box 4	1	1	part of a conserved DNA module involved in light responsiveness
	G-box	1	4	cis-acting regulatory element involved in light responsiveness
	ACE	0	1	cis-acting element involved in light responsiveness
	G-Box	0	2	cis-acting regulatory element involved in light responsiveness
	GA-motif	0	2	part of a light responsive element
	GAG-motif	0	1	part of a light responsive element
	I-box	0	1	part of a light responsive element
	Sp1	0	4	light responsive element
	TCT-motif	0	1	part of a light responsive element
	rbcS-CMA7a	0	1	part of a light responsive element
Hormonal/ environmental responses – related element	ABRE	0	3	cis-acting element involved in the abscisic acid responsiveness
	AuxRR-core	1	0	cis-acting regulatory element involved in auxin responsiveness
	CGTCA-motif	1	0	cis-acting regulatory element involved in the MeJA-responsiveness
	GC-motif	0	1	enhancer-like element involved in anoxic specific inducibility
	HSE	1	0	cis-acting element involved in heat stress responsiveness
	LTR	1	0	cis-acting element involved in low-temperature responsiveness

Hormonal/ environmental responses – related element	P-box	2	0	gibberellin-responsive element
	TC-rich repeats	2	1	cis-acting element involved in defense and stress responsiveness
	TCA-element	2	0	cis-acting element involved in salicylic acid responsiveness
	W box	1	0	pathogen-induced activity
	Box-W1	1	0	fungal elicitor responsive element
	TGACG-motif	1	0	cis-acting regulatory element involved in the MeJA-responsiveness
Site binding – related element	AT-rich element	0	1	binding site of AT-rich DNA binding protein (ATBP-1)
	MBS	3	1	MYB binding site involved in drought-inducibility
Metabolic – related element	O2-site	0	1	cis-acting regulatory element involved in zein metabolism regulation
Seed specific – related element	Skn-1_motif	4	1	cis-acting regulatory element required for endosperm expression
	GCN4_motif	2	0	cis-regulatory element involved in endosperm expression
	RY-element	1	1	cis-acting regulatory element involved in seed-specific regulation
Promoter core/ function element	CAAT-box	33	8	common cis-acting element in promoter and enhancer regions
	TATA-box	27	28	core promoter element around -30 of transcription start
	5UTRPy-rich stretch	0	3	cis-acting element conferring high transcription levels
Unknown function element	CTAG-motif	1	0	ACTAGCAGAA
	Unnamed__11	2	0	TCCACATAGA
	Unnamed__4	9	16	CTCC
	AC-II	0	1	(C/T)T(T/C)(C/T)(A/C)(A/C)C(A/C)A(A/C)C(C/A)(C/A)C
	Unnamed__1	0	2	CGTGG
	Unnamed__2	0	1	CCCCGG
	Unnamed__3	0	2	CGTGG
	box S	0	1	AGCCACC
	AAGAA-motif	0	1	GAAAGAA
	plant_AP-2-like	0	1	CGCGCCGG
Other element	OBP-1 site	1	0	cis-acting regulatory element
	circadian	3	0	cis-acting regulatory element involved in circadian control

Table 3. List of the significant categories including light cycle-, hormonal/environmental responses-, sites binding-, seed specific-, promoter core-related element, other elements including OBP-1 site and circadian elements, and elements with unknown function between the promoter regions of *PSY1* identified by Audic and Claverie (AC) test, Fisher exact test, Chi-squared 2×2, Greller and Tobin, *R* of Stekel and Falciani and General Chi-squared

Category	Lib1	Lib2	Lib1(norm)	Lib2(norm)	AC 1	Fisher 1 2	Chi2×2 1 2	GT	*R*	Chi
Light cycle-related element	6	20	571.4	2127.7	0.000802	0.001358	0.001146	–	0.001961	0.001146
Hormonal/ environmental responses-related element	13	5	1238.1	531.9	0.026119	0.090352	0.082935	–	0.090992	0.082935
Seed specific-related element	7	2	666.7	212.8	0.048255	0.175797	0.123969	–	0.119783	0.123969
Promoter core/function element	60	39	5714.3	4148.9	0.012574	0.033312	0.027465	–	0.115924	0.027465
Unknown function element	12	25	1142.9	2659.6	0.003273	0.009956	0.006041	–	0.012614	0.006041
Other element	4	0	380.10	0	0.040896	0.123514	0.055921	–	0.023696	0.055921

Table 4. Results of annotating CpG in two promoter sequences of *PSY1* transcripts

Gene ID	CpG island number	Length	G+C frequency	CpG o/e ratio	AT skew	CG skew	Start-p	Strand	Strand-p
GRMZM2G300348_T01	0	–	–	–	–	–	–	–	–
GRMZM2G300348_T02, T03, T04	1	1208	0.57	0.87	0.08	−0.07	0.49	minus	0.85

G+C frequency – G+C frequency of the CpG island; CpGo/e ratio – CpG o/e ratio of the CpG island; Start-p – predicted probability to be located over the transcription start site (range from 0 to 1); AT skew – AT skew value of the CpG island; GC skew – GC skew value of the CpG island; strand and (strand-p) – strand of the promoter and predicted probability to be located over this strand (range from 0.5 to 1)

HSE (cis-acting element involved in heat stress responsiveness), LTR (cis-acting element involved in low-temperature responsiveness), P-box (gibberellin-responsive element), TCA-element (cis-acting element involved in salicylic acid responsiveness), W box (pathogen-induced activity), Box-W1 (fungal elicitor responsive element), and TGACG-motif (cis-acting regulatory element involved in the MeJA-responsiveness) were only detected in the promoter region of GRMZM2G300348_T01, while ABRE (cis-acting element involved in the abscisic acid responsiveness), and GC-motif (enhancer-like element involved in anoxic specific inducibility) were only present in the promoter region of the other three transcripts. In addition, TC-rich repeat elements (cis-acting element involved in defense and stress responsiveness) were common to the two promoter sequences.

Significant differences were determined between the two promoters in terms of the number of repetitions of each category including light cycle-, hormonal/environmental responses-, site binding-, seed specific-, and promoter core-related elements, other elements including OBP-1 site and circadian elements, and elements with unknown function (Table 3).

CpG annotation in two promoter sequences of *PSY1* transcripts is shown in Table 4. No CpG island was found in the promoter of GRMZM2G300348_T01. On the other hand, in the promoter region of the three other transcripts, one possible CpG island was found covering almost all the promoter region (Table 4).

Co-expression analysis for PSY1 transcripts in corn

Co-expression analysis showed 4 transcription factors co-expressed with *PSY1* transcripts under environmental

stresses belonging to the HB and MYB families (Table 5).

Another significant result revealed from this study is the discovery of two transcription factors (accessions: GRMZM2G130442; Homeodomain leucine zipper family IV protein; Outer cell layer5a; Uncharacterized protein and GRMZM2G118063; Homeodomain leucine zipper family IV protein; Putative homeobox/lipid-binding domain family protein isoform 1; Putative homeobox/lipid-binding domain family protein isoform 2; Uncharacterized protein) belonging to the HB family that are co-expressed with all four transcripts of the *PSY1* gene (GRM ZM2G300348_T01, GRMZM2G300348_T02, GR MZM 2G300348_T03, GRMZM2G300348_T04) under environmental stresses. GRMZM2G130442 (HDZIV5 _OCL5) and GRMZM2G118063 (HDZIV10_OCL10) are homologous genes located on chromosome 4 and 10, respectively.

Co-expression analysis for PSY1 transcripts in Arabidopsis and rice

The corresponding *PSY1* transcripts and co-expressed transcription factors in Arabidopsis and rice are presented in Table 6. Since GRMZM2G130442 and RMZM 2G118063 are homologs, the corresponding genes for GRMZM2G130442 and GRMZM2G118063 were assigned to the same gene ID in both Arabidopsis and rice, as shown in Table 6. Co-expression analyzes revealed that in Arabidopsis AT4G04890 (Homeobox-leucine zipper protein PROTODERMALFACTOR2; PDF2) is co-expressed with AT5G17230 (*PSY1*) under environmental stresses with a Spearman's rho equal to 0.82. In rice, although the co-expression analysis based on the Spearman's rho method gave no results, Os08t0136100 (Homeobox-leucine zipper protein ROC7; ROC7) showed co-expression with Os06t0729000 (*PSY1*) under environmental stresses based on the Pearson correlation method with a correlation coefficient equal to 0.74.

Discussion

This study increases current understanding of the regulation of carotenogenesis in corn. Although the main carotenoid biosynthesis pathway has been elucidated (Cunningham and Gantt, 1998; Fraser and Bramley, 2004), the lack of a fundamental understanding of carotenogenesis regulation in plant cells is noticeable (Sandmann et al., 2006). The main factor influencing carotenoid production in different plant species is the transcriptional regulation of genes encoding phytoene synthase (PSY), which is the first and main rate-determining enzyme of the pathway (Fraser and Bramley, 2004; Sandmann et al., 2006; Li et al., 2008; Welsch et al., 2008; Maass et al., 2009; Rodríguez-Villalón et al., 2009).

Regulatory mechanisms of *PSY1* expression can be understood by identification of cis-acting regulatory elements, and transcription factors binding to *PSY1* promoter regions and co-expressed with the *PSY1* transcripts which regulate the expression of *PSY1* gene. Identification of transcription factors binding to *PSY1* promoters and co-expressed with *PSY1 may provide* valuable insights into an understanding of the regulatory networks of *PSY1* expression to develop improvements in strategies of carotenoid biosynthesis.

Therefore, in this study, certain of bioinformatics servers and databases such as PlantPAN, Plant TFDB and maizeGDB were used to analyze the *PSY1* gene in more detail and identify co-expressed transcription factors which bind to the *PSY1* promoter regions and which may be associated with carotenoid biosynthesis. Since identification of transcription factors that regulate structural genes involved in carotenoid biosynthesis is highly demanding, these rapid bioinformatic approaches were applied, and since the results may not precisely correlate with the experimental expression data, experimental studies are essential to validate the results of computational analyzes.

The striking result emerging from the comparison of promoters is the higher total occurrence of hormonal/environmental responses-related elements in the promoter region of GRMZM2G300348_T01. The results indicated that the promoter of GRMZM2G300348 _T01 has diverse motifs involved in hormonal/environmental responses. Therefore, a comparison was conducted between the regulatory regions of promoters. The presence of AuxRR-core, CGTCA-motif, TCA-element, and TGACG-motif in the promoter of GRMZM2G300348 _T01 indicatesauxin, MeJA, gibberellins, and salicylic acid-signaling roles in the regulation of its expression. Moreover, a promoter analysis shows the presence of two temperature associated motives, LTR and HSE, in the promoter region of GRMZM2G300348_T01. HSE is consistently conserved in the regulatory regions of many heat-induced genes (Larkindale and Vierling, 2008). The ABA-responsive element (ABRE; PyACGTG/TC), a well-

Table 5. Transcription factors that are co-expressed with *PSY1* transcripts under conditions of environmental stress

Gene ID	Description	Gene symbol	Family	GRMZM2G300348_T01	GRMZM2G300348_T02	GRMZM2G300348_T03	GRMZM2G300348_T04
GRMZM2G122897	homeodomain leucine zipper family IV protein; putative homeobox/lipid-binding domain family protein	HDZIV8_OCL8	HB		*	*	*
GRMZM2G130442	homeodomain leucine zipper family IV protein; outer cell layer5a; uncharacterized protein	HDZIV5_OCL5	HB	*	*	*	*
GRMZM2G118063	homeodomain leucine zipper family IV protein; putative homeobox/lipid-binding domain family protein isoform 1; putative homeobox/lipid-binding domain family protein isoform 2; uncharacterized protein	HDZIV10_OCL10	HB	*	*		*
GRMZM2G045748	uncharacterized protein		MYB	*			

Table 6. Corresponding *PSY1* transcripts and co-expressed transcription factors in Arabidopsis and rice

Gene ID in corn	GRMZM2G300348 (*PSY1*)	GRMZM2G130442 (HDZIV5_OCL5)	GRMZM2G118063 (HDZIV10_OCL10)
Similar gene ID in rice	Os06t0729000	Os08t0136100	Os08t0136100
Similar gene ID in Arabidopsis	AT5G17230	AT4G04890	AT4G04890

characterized cis-element, is involved in ABA-induced gene expressions (Hattori et al., 2002; Fujita et al., 2011). In this study, the ABREs located in the promoter of GRMZM2G300348_T02, T03, T04 indicated a major role for ABA-signaling in the regulation of their expressions. Therefore, studying *PSY1* promoters could provide useful information about the gene and signaling networks underlying gene expression. The presence of HSE in the regulatory region of *PSY1* supports previous research showing the essential role of *PSY1* in heat stress-induced biosynthesis of carotenoid in corn (Li et al., 2008). Due to the presence of diverse motifs involved in hormonal/environmental stress responses in the promoter regions, we conclude that abiotic and biotic stresses can alter *PSY1* transcript levels. This finding also accords with Cazzonelli and Pogson' (2010) report that salt, drought, ABA, temperature, high levels of light, and photoperiod influence *PSY* expression.

Cytosine DNA methylation in plants is found primarily in transposable elements, CpG/CpNpG islands and repetitive DNA sequences (Bender, 2004; Tran et al., 2005). As CpG islands are highly correlated with gene regulation, they play a critical role in the genome. CpG-rich regions inactivate DNA through methylation, which in turn is frequently linked to heterochromatin, gene silencing, and pathogen control (Jeddeloh et al., 1998; Kooter et al., 1999; Vaucheret and Fagard, 2001; Rombauts et al., 2003). DNA methylation can be found on CpNpG islands and non-symmetrical trinucleotides in addition to cytosine of CpG islands (Pradhan et al., 1999; Lindroth et al., 2001; Cao and Jacobsen, 2002). Abiotic stress-induced methylation is thought to be associated with various biochemical pathways involved in plant adaptation and stress responses (Uthup et al., 2011). As demonstrated in this study, one CpG island is present in the promoter of GRMZM2G300348_T02, T03, T04

covering almost all promoter regions, suggesting an important role for DNA methylation in the transcriptional regulation of the expression of these transcripts.

Given the presence of diverse motifs involved in environmental responses in the corresponding promoter of GRMZM2G300348_T01 as well as the shorter length of the transcript, we hypothesize that under stress conditions and as an adaptive response to minimize the biosynthetic cost, the expression of GRMZM2G300348_T01 is higher than that of other transcripts. Since synthesizing larger proteins composed of longer amino-acid chains is metabolically more expensive than synthesizing smaller proteins, this could imply repression of abundant and large proteins and upregulation of small proteins for most stress responses. This idea has been supported by Vilaprinyo et al. (2010).

Carotenoid production can be elevated through genetic engineering by the identification of transcription factors regulating the expression of structural genes in carotenoid biosynthesis (Sagawa et al., 2015). As shown in Table 5, GRMZM2G045748, a transcription factor belonging to the MYB family, is co-expressed only with GRMZM2G300348_T01 under conditions of environmental stress. On the other hand, a promoter analysis showed that three MBS motifs are located in the promoter of GRMZM2G300348_T01, while there is only one motif in the promoters of GRMZM2G300348_T02, 3, 4. Therefore, the findings from co-expression analyzes are in agreement with those of promoter analyzes.

Transcription factors of MYB, one of the largest transcription factor families, are involved in a wide variety of plant biological processes. Ge and coworkers (2016) reported a transcription factor-mediated regulatory network based on gene functional annotations and stated that *ZmMYB138*, a member of the MYB transcription factor family, promotes embryonic callus formation in the corn embryo through a GA signal transduction. Several studies have revealed the importance of MYB transcription factors in the biosynthesis of anthocyanin (Kobayashi et al., 2004; Takos et al., 2006; Espley et al., 2007). In this regard, *ZmMYBC1* (Paz-Ares et al. 1987) and *ZmMYBP1* (Grotewold et al., 1994) have been reported to play a role in anthocyanin synthesis.

The most significant result obtained in this study can be summarized as the discovery of two transcription factors (accessions: GRMZM2G130442; Homeodomain leucine zipper family IV protein; Outer cell layer5a; Uncharacterized protein and GRMZM2G118063; Homeodomain leucine zipper family IV protein; Putative homeobox/lipid-binding domain family protein isoform 1; Putative homeobox/lipid-binding domain family protein isoform 2; Uncharacterized protein) belonging to the HB family that are co-expressed with the four identified transcripts (GRMZM2G300348_T01, GRMZM2G300348_T02, GRMZM2G300348_T03, GRMZM2G300348_T04) of the *PSY1* gene under conditions of environmental stress. The evidence for this finding was substantially strengthened by co-expression analyzes for the corresponding *PSY1* transcripts in both Arabidopsis and rice. In this way, co-expression of these transcription factors with *PSY1* not only in corn but also in Arabidopsis and rice lends further support to the idea that these transcription factors may act as critical regulators of *PSY1* gene expression. Homeodomain leucine zipper (HD-Zip) proteins are plant specific transcription factors and have been categorized into four distinct classes (HD-Zip I-IV) based on a sequence analysis (Ariel et al., 2007). The corn HD-ZIP TF family comprises approximately 65 members (Lin et al., 2014). According to Kubo and coworkers (1999), HD-Zip IV proteins play significant roles during anthocyanin accumulation, differentiation of epidermal cells, trichome formation and root development in Arabidopsis (Kubo et al., 1999). Thus, the identification of GRMZM2G130442 and GRMZM2G118063 as belonging to HD-Zip IV also supports the idea of the role of HD-Zip IV in the synthesis of carotenoids.

Conclusion

This study focused on *PSY1* genes and their non-coding regulatory regions studied by a computational analysis. Comparisons between regulatory regions of two promoter regions were conducted and significant differences in categories were determined, including light cycle-, hormonal/environmental responses-, site binding-, seed specific-, promoter core-related elements, other elements including OBP-1 site and circadian elements, as well as elements with unknown function.

Based on the analyzes of promoters, specific roles can be suggested for auxin, MeJA, gibberellins, and salicylic acid-signaling in GRMZM2G300348_T01 expression and the role of other three transcripts, including GRMZM2G300348_T02, GRMZM2G300348_T03 and GRMZM2G300348_T04, for ABA-signaling in expression regulation.

Candidate co-expressed transcription factors reported here will provide a basis for the identification and elucidation of their function in carotenoid biosynthesis in corn that could contribute to the enhanced production of carotenoids. However, due to the presence of false positive results in bioinformatics prediction, experimental studies are crucial in order to confirm the results of computational analyzes.

References

Andorf C.M., Cannon E.K., Portwood J.L., Gardiner J.M., Harper L.C., Schaeffer M.L., Braun B.L., Campbell D.A., Vinnakota A.G., Sribalusu V.V. et al. (2015) *MaizeGDB 2015: new tools, data, and interface for the maize model organism database.* Nucl. Acids Res. 44(D1): D1195-1201.

Ariel F.D., Manavella P.A., Dezar C.A., Chan R.L. (2007) *The true story of the HD-Zip family.* Trends Plant Sci. 12: 419-426.

Arango J., Wüst F., Beyer P., Welsch R. (2010) *Characterization of phytoene synthase from cassava and their involvement in abiotic stress-mediated responses.* Planta 232: 1251-1262.

Bender J. (2004) *DNA methylation and epigenetics.* Annu. Rev. Plant Biol. 55: 41-68.

Bou-Torrent J., Toledo-Ortiz G., Ortiz-Alcaide M., Cifuentes-Esquivel N., Halliday K.J., Martinez-García J.F., Rodriguez-Concepcion M. (2015) *Regulation of carotenoid biosynthesis by shade relies on specific subsets of antagonistic transcription factors and cofactors.* Plant Physiol. 169(3): 1584-1594.

Bramley P.M. (2002) *Regulation of carotenoid formation during tomato fruit ripening and development.* J. Exp. Bot. 53: 2107-2113.

Cao X., Jacobsen S.E. (2002) *Locus-specific control of asymmetric and CpNpG methylation by the DRM and CMT3 methyltransferase genes.* Proc. Natl. Acad. Sci. U.S.A. 99(Suppl4): 16491-16498.

Cazzonelli C.I., Pogson B.J. (2010) *Source to sink: regulation of carotenoid biosynthesis in plants.* Trends Plant Sci. 15: 266-274.

Chow C.N., Zheng H.Q., Wu N.Y., Chien C.H., Huang H.D., Lee T.Y., Chiang-Hsieh Y.F., Hou P.F., Yang T.Y., Chang W.C. (2016) *PlantPAN 2.0: an update of plant promoter analysis navigator for reconstructing transcriptional regulatory networks in plants.* Nucl. Acids Res. 44: 1154-1160.

Cunningham F.X., Gantt E. (1998) *Genes and enzymes of carotenoid biosynthesis in plants.* Annu. Rev. Plant Physiol. Plant Mol. Biol. 49: 557-583.

Espley R.V., Hellens R.P., Putterill J., Stevenson D.E., Kutty-amma S., Allan A.C. (2007) *Red colouration in apple fruit is due to the activity of the MYB transcription factor, MdMYB10.* Plant J. 49: 414-427.

Fiedor J., Burda K. (2014) *Potential role of carotenoids as antioxidants in human health and disease.* Nutrients 6: 466-488.

Fraser P.D., Bramley P.M. (2004) *The biosynthesis and nutritional uses of carotenoids.* Prog. Lipid Res. 43: 228-265.

Fu Z., Yan J., Zheng Y., Warburton M.L., Crouch J.H, Li J.S. (2010) *Nucleotide diversity and molecular evolution of the PSY1 gene in Zea mays compared to some other grass species.* Theor. Appl. Genet. 120: 709-720.

Fujita Y., Fujita M., Shinozaki K., Yamaguchi-Shinozaki K. (2011) *ABA-mediated transcriptional regulation in response to osmotic stress in plants.* J. Plant Res. 124: 509-525.

Gallagher C.E., Matthews P.D., Li F., Wurtzel E.T. (2004) *Gene duplication in the carotenoid biosynthetic pathway preceded evolution of the grasses.* Plant Physiol. 135(3): 1776-1783.

Ge F., Luo X., Huang X., Zhang Y., He X., Liu M., Lin H., Peng H., Li L., Zhang Z.H., Pan G., Shen Y. (2016) *Genome-wide analysis of transcription factors involved in maize embryonic callus formation.* Physiol. Plant. doi: 10.1111/ppl.12470.

Grotewold E., Drummond B.J., Bowen B., Peterson T. (1994) *The myb-homologous P gene controls phlobaphene pigmentation in maize floral organs by directly activating a flavonoid biosynthetic gene subset.* Cell 76: 543-553.

Hattori T., Totsuka M., Hobo T., Kagaya Y., Yamamoto-Toyoda A. (2002) *Experimentally determined sequence requirement of ACGT-containing abscisic acid response element.* Plant Cell Physiol. 43: 136-140.

Jeddeloh J.A., Bender J., Richards E.J. (1998) *The DNA methylation locus DDM1 is required for maintenance of gene silencing in Arabidopsis.* Genes Dev. 12(11): 1714-1725.

Kobayashi S., Goto-Yamamoto N., Hirochika H. (2004) *Retrotransposon induced mutations in grape skin colour.* Science 304: 982.

Kooter J.M., Matzke M.A., Meyer P. (1999) *Listening to the silent genes: transgene silencing, gene regulation and pathogen control.* Trends Plant Sci. 4(9): 340-347.

Kubo H., Peeters A.J., AartsM.G., Pereira A., Koornneef M. (1999) *ANTHOCYANINLESS2, a homeobox gene affecting anthocyanin distribution and root development in Arabidopsis.* Plant Cell 11: 1217-1226.

Larkindale J., Vierling E. (2008) *Core genome responses involved in acclimation to high temperature.* Plant Physiol. 146: 748-761.

Lee T.I., Rinaldi N.J., Robert F., Odom D.T., Bar-Joseph Z., Gerber G.K., Hannett N.M., Harbison C.T., Thompson C.M., Simon I. et al. (2002) *Transcriptional regulatory networks in Saccharomyces cerevisiae.* Science 298: 799-804.

Lescot M., Déhais P., Thijs G., Marchal K., Moreau Y., Van de Peer Y., Rouzé P. Rombauts S. (2002) *PlantCARE, a database of plant cis-acting regulatory elements and a portal to tools for in silico analysis of promoter sequences.* Nucl. Acids Res. 30(1): 325-327.

Li F.,Vallabhaneni R., Yu J., Rocheford T., Wurtzel E.T. (2008) *The maize phytoene synthase gene family: overlapping roles for carotenogenesis in endosperm, photomorphogenesis, and thermal stress tolerance.* Plant Physiol. 147: 1334-1346.

Lin J.J., Yu C.P., Chang Y.M., Chen S.C., Li W.H. (2014) *Maize and millet transcription factors annotated using comparative genomic and transcriptomic data.* BMC Genomics 15: 818.

Lindroth A.M., Cao X., Jackson J.P., Zilberman D., McCallum C.M., Henikoff S., Jacobsen S.E. (2001) *Requirement of CHROMOMETHYLASE3 for maintenance of CpXpG methylation.* Science 292(5524): 2077-2080.

Maass D., Arango J., Wüst F., Beyer P., Welsch R. (2009) *Carotenoid crystal formation in Arabidopsis and carrot roots caused by increased phytoene synthase protein levels.* PLoS ONE 4: e6373.

Martel C., Vrebalov J., Tafelmeyer P., Giovannoni J.J. (2011) *The tomato MADS box transcription factor RIPENING INHIBITOR interacts with promoters involved in numerous ripening processes in a COLORLESS NONRIPENING-dependent manner.* Plant Physiol. 157: 1568-1579.

Palaisa K.A., Morgante M., Williams M., Rafalski A. (2003) *Contrasting effects of selection on sequence diversity and linkage disequilibrium at two phytoene synthase loci.* Plant Cell 15: 1795-1806.

Park S., Kim H.S, Jung Y.J., Kim S.H., Ji C.Y., Wang Z., Jeong J.C., Lee H.S., Lee S.Y., Kwak S.S. (2016) *Orange protein has a role in phytoene synthase stabilization in sweetpotato.* Sci. Rep. 6: 33563.

Paz-Ares J., Ghosal D., Wienand U., Peterson P.A., Saedler H. (1987) *The regulatory c1 locus of Zea mays encodes a protein with homology to myb proto-oncogene products and with structural similarities to transcriptional activators.* EMBO J. 6: 3553-3558.

Pérez-Rodríguez P., Riańo-Pachón D.M., Corręa L.G.G., Rensing S.A., Kersten B., Mueller- Roeber B. (2010) *PlnTFDB: updated content and new features of the plant transcription factor database.* Nucl. Acids Res. 38: D822-D827.

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

Pradhan S., Urwin N.A., Jenkins G.I., Adams R.L. (1999) *Effect of CWG methylation on expression of plant genes.* Biochem. J. 341(Pt3): 473-476.

Rani V. (2007) *Computational methods to dissect cis-regulatory transcriptional networks.* J. Biosci. 32: 1325-1330.

Rao A.V., Rao L.G. (2007) *Carotenoids and human health.* Pharmacol Res. 55: 207-216.

Rodríguez-VillalónA., Gas E., Rodríguez-Concepción M. (2009) *Phytoene synthase activity controls the biosynthesis of carotenoids and the supply of their metabolic precursors in dark-grown Arabidopsis seedlings.* Plant J. 60: 424-435.

Rombauts S., Florquin K., Lescot M., Marchal K., Rouze P., van de Peer Y. (2003) *Computational approaches to identify promoters and cis-regulatory elements in plant genomes.* Plant Physiol. 132(3): 1162-1176.

Romualdi C., Bortolzuzz S., D'Alessi F., Danieli G.A. (2003) *IDEG6: a web tool for detection of differentially expressed genes in multiple tag sampling experiments.* Physiol. Genomics 12: 159-162.

Sagawa J.M., Stanley L.E., LaFountain A.M., Frank H.A., Liu C., Yuan Y.W. (2015) *An R2R3-MYB transcription factor regulates carotenoid pigmentation in Mimulus lewisii flowers.* New phytol. 209(3): 1049-1057.

Sandmann G., Römer S., Fraser P.D. (2006) *Understanding carotenoid metabolism as a necessity for genetic engineering of crop plants.* Metab. Eng. 8: 291-302.

Takos A.M., Jaffe F.W., Jacob S.R., Bogs J., Robinson S.P., Walker A.R. (2006) *Light-induced expression of a MYB gene regulates anthocyanin biosynthesis in red apples.* Plant Physiol. 142:1216-1232.

Tao N., Hu Z., Liu Q., Xu J., Cheng Y., Guo L., Guo W., Deng X. (2007) *Expression of phytoene synthase gene (Psy) is enhanced during fruit ripening of Cara Cara navel orange (Citrus sinensis Osbeck).* Plant Cell Rep. 26(6): 837-843.

Toledo-Ortiz G., Huq E., Rodríguez-Concepción M. (2010) *Direct regulation of phytoene synthase gene expression and carotenoid biosynthesis by phytochrome-interacting factors.* Proc. Natl. Acad. Sci. USA 22; 107(25): 11626-11631.

Tran R.K., Henikoff J.G., Zilberman D., Ditt R.F., Jacobsen S.E., Henikoff S. (2005) *DNA methylation profiling identifies CG methylation clusters in Arabidopsis genes.* Curr. Biol. 15(2): 154-159.

Uthup T.K., Ravindran M., Bini K., Thakurdas S. (2011) *Divergent DNA methylation patterns associated with abiotic stress in Hevea brasiliensis.* Mol. Plant 4: 996-1013.

Vaucheret H., Fagard M. (2001) *Transcriptional gene silencing in plants: targets, inducers and regulators.* Trends Genet. 17(1): 29-35.

Vilaprinyo E., Alves R., Sorribas A. (2010) *Minimization of biosynthetic costs in adaptive gene expression responses of yeast to environmental changes.* PLoS Comput. Biol. 6(2): e1000674.

Welsch R., Beyer P., Hugueney P., Kleinig H., von Lintig J. (2000) *Regulation and activation of phytoene synthase, a key enzyme in carotenoid biosynthesis, during photomorphogenesis.* Planta 211: 846-854.

Welsch R., Wüst F., Bär C., Al-Babili S., Beyer P. (2008) *A third phytoene synthase is devoted to abiotic stress-induced abscisic acid formation in rice and defines functional diversification of phytoene synthase genes.* Plant Physiol. 147: 367-380.

Wittkopp P.J., Kalay G. (2011) *Cis-regulatory elements: molecular mechanisms and evolutionary processes underlying divergence.* Nat. Rev. Genet. 6, 13(1): 59-69.

Zhou X., Welsch R., Yang Y., Álvarez D., Riediger M., Yuan H., et al. (2015) *Arabidopsis OR proteins are the major posttranscriptional regulators of phytoene synthase in controlling carotenoid biosynthesis.* Proc. Natl. Acad. Sci. USA 112(11): 3558-3563.

Zhu C., Yamamura S., Koiwa H., Nishihara M., Sandmann G. (2002) *cDNA cloning and expression of carotenogenic genes during flower development in Gentiana lutea.* Plant Mol. Biol. 48: 277-285.