Identification and in silico characterization of *cis*-acting elements of genes involved in carotenoid biosynthesis in tomato

Archana Koul¹ · Deepak Sharma¹ · Sanjana Kaul¹ · Manoj K. Dhar¹

Received: 18 December 2018 / Accepted: 6 June 2019 / Published online: 26 June 2019
© King Abdulaziz City for Science and Technology 2019

Abstract

Carotenoids, the widespread and structurally diverse class of pigments, accumulate in the fruits of tomato plants in a tissue specific manner. The carotenoid biosynthetic pathway genes have been cloned and characterized in tomato and other plants, however, its regulation is still obscure. We collected and analyzed forty different accessions of tomato for the present study. HPLC analysis revealed differential accumulation of major carotenoids (lycopene and β -carotene) in the ripe fruit tissue. In order to understand the underlying regulatory mechanisms in carotenoid biosynthesis and accumulation, we sequenced the *cis*-acting elements i.e. promoter, 5' and 3' untranslated regions of the carotenoid pathway genes, in all accessions, followed by their in silico validation. Major differences observed in the CAAT Box, Opaque-2 Box and L-box in the promoters of carotenoid isomerase and lycopene-beta cyclase genes, respectively, along with the variations in musashi binding element of 5' untranslated regions of the carotenoid isomerase gene, suggest their differential role in regulating the carotenogenesis process in tomato. The binding sites for various transcription factors namely RIN, AGAMOUS, CRY, RAP2.2 and PIF1 on the promoters of important carotenoid pathway genes were predicted in silico. We propose that expression of carotenoid genes and also the formation of protein product in ripe tomato fruits, is regulated efficiently by the binding of these transcription factors at selected sites in the promoter region. Finally, the differential expression of the above-mentioned genes in different developmental tissues supports the possible involvement of promoters and untranslated regions in carotenoid biosynthesis and accumulation process. The present study has generated significant information concerning regulatory players involved in the carotenoid biosynthesis in tomato.

Keywords *Cis*-acting elements · Carotenoids · BioMart · PlantCARE · Tomato · Gene regulation

Introduction

Tomato (*Solanum lycopersicum* L.) is the most important model system to understand the carotenoid biosynthetic pathway in plants (Dhar et al. 2014). The latter represents one of the most important secondary metabolic pathways in plants and the maximum accumulation of carotenoids occur in tomato fruits. Carotenoids especially, lycopene and β -carotene are beneficial for the humans as they act

as anti-oxidant, anti-cancerous, photoprotective, immune-stimulatory agents (Krinsky and Johnson 2005; Tanaka et al. 2012).

Genes that encode most of the enzymes of this pathway have been cloned and characterized in several plants such as in tomato, pepper, *Arabidopsis*, etc. (Chen et al. 2010; Ruiz-Sola and Rodriguez-Concepcion 2012; Su et al. 2015); however, reports on the characterization of *cis*-acting elements i.e. promoters and 5' and 3' untranslated regions (UTRs) present on the carotenoid genes is still in its infancy that needs to be elucidated. Therefore, it is important to understand the whole genomic biology/regulation of carotenoid biosynthetic pathway in order to develop a versatile and novel genotype of tomato with higher carotenoid content.

Plant growth and development involves temporal and spatial expression of the specific gene subsets which are influenced by various external and internal stimuli. Besides, the strength of *cis*-acting elements present either in the

Electronic supplementary material The online version of this article (<https://doi.org/10.1007/s13205-019-1798-1>) contains supplementary material, which is available to authorized users.

✉ Manoj K. Dhar
manojkdhar@rediffmail.com; md.jusbt@gmail.com

¹ Genome Research Laboratory, School of Biotechnology, University of Jammu, Jammu 180006, India

promoters or UTR's (5' and 3') regions, plays a significant role in controlling specific gene expression in all the developmental tissues of a plant. Gene promoters; their *cis* sequences and different copy number of their motifs, allow differential expression of pathway genes and ultimately differential protein/metabolite production in the plant system (Bilas et al. 2016). *Cis* regulatory sequences/motifs are located directly in the transcribed DNA strand i.e. promoters or may be active during post-transcriptional modifications like 5'cap, poly-A tail and signal sequences present in the UTRs (Vaughn et al. 2012). However, *cis* sequences mediate the gene regulation by forming active complexes with the *trans* elements (Venter and Botha 2010; Porto et al. 2014) mainly transcription factors (TFs) which enhances the gene transcription and its further expression in different plant organs. TFs bind to a specific sequence; near CAAT box of promoter and recruit RNA polymerase for transcription and its further processing. The activity of TFs depends upon the sequence pattern of the specific gene promoter. The TFs in collaboration with promoters have been shown to effect carotenoid pathway (Dhar et al. 2014).

Some promoters of the carotenoid biosynthetic pathway genes have been functionally validated in plants. For example, Dalal et al. (2010) isolate and characterize the chromoplastic *lyc-b* promoter from *Solanum habrochaites*. Deletion fragment assay and its expression *in fruto* as well as in stable transgenic tomato revealed the tissue specific regulation of *lyc-b* promoter and its up-regulation in chromoplast-rich flowers and fruits. Similarly, UTR mediated regulation of carotenoid pathway genes has been observed and verified in plants. UTRs are known to be post-transcriptional regulators which effect the expression of each gene of a pathway in a different manner. Alvarez et al. (2016) showed regulation of carotenogenesis by 5'-UTR mediated translation of phytoene synthase splice variants in *Arabidopsis thaliana*. Two alternative splice variants (ASV) were found in *A. thaliana* with different length and exon/intron retention of their 5'-UTRs. ASV1 was found to contain a long 5'-UTR and predominantly involved in developmentally regulated carotenogenesis process whereas ASV2 has a short 5'-UTR and is induced when an immediate increase in carotenoid flux is required like under various kinds of stress (salt, water etc.) or different light intensities (Alvarez et al. 2016).

In a recent study, our group has identified notable regulators i.e. microRNAs (miRNAs), that are known to play a very significant role in the regulation of carotenogenesis process in tomato (Koul et al. 2016). In the latter publication, we have reported miRNAs that mediate possible regulation of carotenoid pathway genes at different developmental stages of tomato; negative correlation of miRNA and carotenoid genes was observed, with some exceptions.

From the above studies it has become clear that within a cell, an interlinked network of several regulatory mechanisms

e.g. promoters and UTRs driven regulation, miRNA mediated gene silencing (PTGS), transcriptional expression and regulation by TFs exist, which contribute towards the carotenoid biosynthesis and accumulation in plants. The promoters and UTRs associated with the regulation of carotenogenesis process have not been characterized so far. The main aim of the present work was the *in silico* retrieval of the promoters and 5' and 3' UTR sequences, designing specific primers, sequencing and their characterization using various computational tools. Also, putative transcription factor binding sites on carotenoid gene promoters and their interactions was analysed. Therefore, the present study provides an insight into the coordinated role of promoters, UTRs and TFs in the regulation of carotenoid biosynthesis. These regulatory players can be employed to manipulate the pathway genes, through the process of metabolic engineering which may lead to the enhanced production of carotenoids specially lycopene and β -carotene in tomatoes.

Materials and methods

Plant material

Seeds of forty different tomato accessions (Table S1) were collected from various areas of Jammu and Kashmir (J&K) and some other parts of the country. The plants were maintained in the Greenhouse at School of Biotechnology, University of Jammu under controlled conditions (Temperature; Day: 26 °C $\pm$ 2 and night: 16 °C $\pm$ 2, Photoperiod: 16 h). Mature (red-ripe) fruits were harvested and stored at –80 °C for further use.

Profiling of carotenoids

For carotenoid profiling, tomato fruits were collected at red-ripe (RR) stage and plunged in liquid N₂ and stored at –80 °C for further extraction. Extraction of carotenoids was carried out in triplicates using the protocol of Nicolle et al. (2004). Separation and quantification of carotenoids (lycopene and β -carotene) was done by employing the protocol of Koul et al. (2016). Carotenoids were identified using their retention times and absorption spectra of authentic standards (Sigma, USA). The concentration of each carotenoid was expressed as microgram per gram fresh weight (FW) and the mean of the data from triplicate samples of the same tomato accession.

Amplification and in silico characterization of promoters

Out of different accessions, twelve were selected for the characterization purpose. Based on the HPLC data, these accessions were divided into three groups; maximum (SOL1, SOL21, SOL22, SOL26), intermediate (SOL6,

SOL7, SOL11, SOL18) and minimum (SOL4, SOL5, SOL15, SOL16) (Table S1) carotenoid containing accessions. As 5'-UTR lie adjacent to the promoter region; therefore, the primers were designed in such a way that it amplified both the promoter as well as 5'-UTRs. The carotenoid genes whose promoters and UTRs were studied during the present case have been given in Table 1. For the amplification of carotenoid gene promoter regions from selected tomato accessions, specific primers were designed using IDT primer quest tool (Table S2) on the basis of nucleotide sequences which were retrieved from the Ensembl Plants genome browser. The nucleotide start site of the promoters was determined using transcription start site (TSSP) program of the Softberry software (Shahmuradov et al. 2003).

Genomic DNA was isolated from the leaf tissues of the different accessions using Cetyl Trimethyl Ammonium Bromide (CTAB) protocol according to the protocol described by Doyle (1991). Extracted DNA was used for the amplification of the eight carotenoid gene promoter regions. Amplified PCR products were purified, sequenced and analysed. Analysis of sequencing data was done using various computational tools like BLAST and CLUSTAL W to identify the sequence similarities among different accessions and previously submitted sequences in the sequence databases. Two bioinformatic tools, PlantCARE and PLACE were used for the characterization of different regulatory factors present in carotenoid gene promoter regions in the studied tomato accessions (Higo et al. 1998; Lescot et al. 2002). Differences were calculated in terms of number and functionality of the motifs present in the promoter sequences.

Prediction of potential transcription factor binding sites (TFBS)

Potential binding sites of different TFs on the promoters of carotenoid genes were predicted using in-built weight matrices of MatInspector software (Cartharius et al. 2005).

Table 1 Carotenoid biosynthetic pathway genes whose promoters and UTRs were studied in tomato

Gene name	Gene symbol
Geranyl geranyl pyrophosphate synthase 1	<i>ggpps1</i>
Phytoene synthase 1	<i>psyl</i>
Phytoene synthase 2	<i>psyl2</i>
Phytoene desaturase	<i>pds</i>
Zeta-carotene desaturase	<i>zds</i>
Carotenoid isomerase	<i>crtISO</i>
Lycopene-beta cyclase	<i>lyc-b</i>
Lycopene-epsilon cyclase	<i>lyc-e</i>

Amplification and in silico characterization of 5' and 3' UTRs

The whole genome of tomato was accessed using the Ensembl Plants Genome browser in order to retrieve the transcript ID of the carotenoid gene sequences. The transcript ID was further used for extracting the UTR sequences using a data mining tool; BioMart (Smedley et al. 2015). After getting full length UTR sequences, specific primers were designed (Table S3) and synthesized. Amplification of the UTRs was attempted from the genomic DNA of above selected accessions. Amplified PCR products were purified, sequenced and analysed. Analysis of both 5' and 3' UTRs sequences was done using BLAST and CLUSTAL W. UTRs were further analysed using UTRdb and UTRScan softwares (Mignone et al. 2005). UTRScan algorithm was used for signal/motif searches whereas description of the signals was retrieved using UTRsite tool of the UTRdb.

Gene expression studies using qRT-PCR

Total RNA was isolated from different developmental tissues of the SOL21 and SOL5 accessions (having maximum and minimum carotenoid content, respectively) using TRIzol® reagent (Life Technologies, Carlsbad, CA). cDNA synthesis was done using RT enzyme (Fermentas, Canada). qRT-PCR of important carotenoid genes i.e. *crtISO* and *lyc-β* gene was performed using SYBR green Applied Biosystems 7500 Sequence Detection System. Actin was used as an internal control. Reaction included specific forward and reverse primers (10 mM) (Table S4). Reaction was carried out at 95 °C for 10 min, followed by 30 cycles of 95 °C for 15 s and 60 °C for 1 min and then by a thermal denaturing step to generate dissociation curves. Three technical replicates were used during the study. The gene expression was calculated using $\Delta\Delta C_T$ method (Livak and Schmittgen 2001). Standard error was calculated from three technical replicates.

Results and discussion

Measurement of carotenoids

Lycopene and β -carotene content were determined by HPLC at red ripe (RR) stage of the tomato fruits. Values represent the mean of three replicates. Forty accessions used during the present study showed differential carotenoid accumulation during ripening stage (Table S1, Figure S1). Highest lycopene content of 40.8 ± 0.65 $\mu\text{g/g}$ of fresh weight (FW) was observed in accession SOL 21 (SKUAST-J), while the β -carotene content was 7.6 ± 0.91 $\mu\text{g/g}$ of FW. Similarly, least lycopene and β -carotene content of 5.52 ± 0.77 $\mu\text{g/g}$ and 0.76 ± 0.68 $\mu\text{g/g}$ of FW, respectively was observed in

accession SOL 15 (Nougron, Jammu). Three internationally known varieties namely ArkaVikas, PUSA Ruby and PUSA Dwarf were also used as reference varieties. Among reference varieties, Arka Vikas showed highest lycopene and β -carotene accumulation (58 ± 0.31 $\mu\text{g/g}$ of FW and 10.4 ± 0.11 $\mu\text{g/g}$ of FW, respectively). Gupta et al. (2015) also reported the similar content of lycopene and β -carotene in ArkaVikas where as the mutants having down regulated carotenoid genes showed very minute amount of the respective carotenoids in the fruit tissue. Similar study was carried out by Rosati et al. (2000) in wild-type and transgenic tomato fruit tissues. This indicates that the metabolic flux of carotenoids gets drastically altered as the spatio-temporal expression of the carotenoid biosynthetic genes and other regulatory molecules, change in response to various natural as well as metabolically engineered stimuli (Rosati et al. 2000).

Comparative analysis of promoters and their *cis*-sequences

Regulation of gene expression is a very complex process and can occur at various levels i.e. during transcription, mRNA processing followed by translation and post-translation modifications. However, regulation primarily occurs at the gene transcriptional level with the coordinated action of *cis*-acting motifs present in the promoters and *trans*-acting factors. The *cis*-acting motifs are DNA sequences in both the coding and non-coding regions which are present in the genome. Epigenomic studies have demonstrated the role of *cis*-acting elements in the chromatin remodelling and its modifications to generate an accessible region for binding of *trans*-factors in order to initiate transcription of a particular gene (Riechmann 2002). Organization of all the eukaryotic genomes is universally similar and almost all the regulatory elements are same. However, difference lies in the copy number of regulatory motifs present in the particular tissues/organs (Twyman 2003; Venter and Botha 2010). Plant promoters have attracted increasing attention because of their irreplaceable role in modulating the spatio-temporal expression of genes by interacting with transcription factors (TFs) (Liu et al. 2013). In total, promoters of eight carotenoid genes namely *ggpps1*, *psy1*, *psy2*, *pds*, *zds*, *crtISO*, *lyc-b* and *lyc-e* (Table 1), were amplified from the twelve accessions on the basis of HPLC profiles mentioned above. PCR resulted in the amplification of full length promoters (Figure S2). Promoter sequences of carotenoid genes were analyzed using PlantCARE and PLACE databases. Analysis showed that all promoters contain motifs necessary for their function, whereas some of the promoters also contain motifs essential for the carotenogenesis process to occur (Table S5). The promoter sequences were also characterized using BLAST and Clustal W software (Altschul et al. 1990; Larkin et al. 2007).

Clustal W analysis showed that the promoters of carotenoid genes i.e. *ggpps1*, *psy2*, *pds*, *zds* amplified from the twelve accessions were 100% similar to each other, respectively and the reference sequences retrieved from various databases. This suggests that the promoters of these genes play uniform role in carotenogenesis process in different tomato accessions. Variations were observed in the promoters of *psy1*, *crtISO* and *lyc-b* genes. Promoter sequences of *psy1* gene showed 98–99% sequence similarity to one another however; the differences in terms of motifs and functionality were not significant. On the other hand, significant differences were observed in the promoter sequences of *crtISO* and *lyc-b* genes amplified from twelve accessions, which showed 97–98% and 98–99% sequence similarity to one another, respectively.

PLACE and Plant Care analysis of the promoters of *crtISO* showed differences in the number of CAAT box and Opaque-2 box (O2). For example, SOL 21 had 17 CAAT-box and one O2-box whereas; SOL 5 had 14 CAAT-box and no O2-box (Fig. 1). CAAT-Box is the essential enhancer element present in the promoter, which is one of the driving elements for enhancing the expression of a particular gene. CAAT box is required by the pathway genes for higher expression which ultimately results in the increased production of proteins inside the cell e.g. the CAAT box in association with neighboring elements and bending by NF-Y is considered to be a major regulatory mechanism for the transcription activation (Mantovani 1999). Dai et al. (1992) characterized the functional role of CAAT box present in the nopaline synthase (*nos*) promoter by mutational analysis. Point mutation in the CAAT box region reduced the promoter activity in the transgenic tobacco plants. Similarly, in our case, CAAT box represents one of the major regulatory motifs present in the promoter region that possibly enhances the transcription/expression of *crtISO* gene, which ultimately leads to the enhanced biosynthesis and accumulation of lycopene in the red-ripe fruits as confirmed by HPLC data. However, in vivo elucidation needs to be carried out in order to explore the actual involvement of the CAAT boxes in the enhancement of gene transcription. The second difference in the promoters of *crtISO* lied in the O2 (Opaque-2) box. Promoter of *crtISO* gene in SOL 21 has an O2 box whereas SOL 5 does not contain such motif. O2 box is a bZIP type transcriptional activator (Wu et al. 1998) which enhances transcription of a particular gene. Wu et al. (1998) explained that *o2* mutant had significantly reduced contents of the product as compared to the wild-type. Similarly, Aukerman et al. (1991) demonstrated that an arginine to lysine substitution by ethyl methanesulfonate (EMS) mutation in the *o2* mutant reduced the transcriptional expression of the storage protein, zein which shows the importance of this motif for the enhanced expression of a gene. Presence or absence of the O2 motif in the promoters of *crtISO*

Fig. 1 Depiction of presence of O2 motif in the promoter of *crtISO* gene amplified from SOL 21 (represented by red arrow) while as it is absent in SOL 5, detected using Plant Care algorithm

gene, may be responsible for the differential expression of this gene, and possibly differential amount of lycopene in the tomato fruits of various accessions. Similarly, differences were observed in the promoter sequences of *lyc-b* gene. Clustal W analysis of promoter of *lyc-b* revealed SNPs, which were further confirmed using PLANT CARE and PLACE software. For example, In SOL 21, extra light responsive motifs namely L-box, AAGAA and TCCACC T were present, which play essential roles in enhancing the carotenogenesis process (Fig. 2). The last two motifs are *cis*-elements involved in seed-specific expression and developmental processes. The major difference lied in L-box which is a light responsive element. In SOL 21, extra light responsive motif i.e. L-box was present which was absent from the promoter of *lyc-b* gene of SOL 5. It has been established that light has a major impact as a regulator of transcriptional control in higher plants (Meier et al. 2011), especially carotenoids. Similarly, Meier et al. (2011) showed that the expression of many carotenoid genes is suppressed in dark grown plants and activated by photoactivated phytochromes. Light induced synthesis of carotenoids is characterized by an increase in the expression of *psy* and *MEP* pathway genes.

Prediction of potential transcription factor binding sites (TFBS)

Transcription factors represent the *trans*-factors which in coordination with *cis*-regulatory elements lead to the enhanced gene expression and ultimately enhanced protein/metabolite accumulation. Potential binding sites of different TFs on the promoters of carotenoid pathway genes were predicted using MatInspector software using in-built weight matrices. TFBS were found on the promoters of the three important carotenoid genes; *psy1*, *crtISO* and *lyc-b*. Three TFs; RAP2.2 (light responsive motif), PIF1 and RIN were predicted to bind the promoter of *psy1* gene (Fig. 3) which were assigned to the matrix family P\$LREM. Similarly, PIF1 and RIN were assigned to the matrix family P\$MYCL and P\$AHBP, respectively. TF AGAMOUS (MADS-box) was predicted to bind the promoter of *crtISO* gene (Fig. 3) which was assigned to the matrix family P\$MADS. Three TFs; AGAMOUS, CRY and RAP2.1 were predicted to bind the promoter of *lyc-b* gene (Fig. 3). CRY and RAP2.1 were observed to be assigned to the matrix families P\$IBOX and P\$DREB, respectively. These TFs are known to play

Fig. 2 Pictorial representation of important motifs present in the promoter of *lyc-b* gene detected using Plant Care software. In SOL 21, extra light responsive motif i.e. L-box, AAGAA and TCCACCT (represented by red arrows) are present which is absent from SOL 5

Motifs Found	SOL 21	Motifs Found	SOL 5
+	A-box	+	A-box
+	AAGAA-motif ←	+	Box I
+	Box I	+	Box-W1
+	Box-W1	+	CAAT-box
+	CAAT-box	+	CCGTCC-box
+	CCGTCC-box	+	ERE
+	ERE	+	GT1-motif
+	GT1-motif	+	GT1-motif
+	L-box ←	+	LTR
+	LTR	+	MBS
+	MBS	+	MNF1
+	MNF1	+	Sp1
+	Sp1	+	Sp1
+	TATA-box	+	TATA-box
+	TC-rich repeats	+	TC-rich repeat
+	TCCACCT-motif ←	+	Unnamed_4
+	Unnamed_4	+	W box
+	W box	+	chs-CMA2a
+	chs-CMA2a	+	circadian
+	circadian		

essential roles in modulating the expression of promoters belonging to the carotenoid gene family using various mechanisms in plants (Dhar et al. 2014). A transcription factor may regulate multitude of genes individually or in a coordinated manner by binding with *cis*-regulatory element for the precise expression patterns (Srivastava and Kumar 2019). Therefore, various individual and coordinated gene regulatory networks have to be designed using bioinformatics tools followed by in vivo elucidation, in order to unravel the regulatory roles of *cis*-acting elements along with *trans*-factors, which ultimately lead to the enhanced protein production inside the cell in a particular organ/tissue of a plant.

5' and 3' UTRs analysis

The UTR region represents a segment of mRNA transcript, encoded in DNA strand and acts during transcription and

not the translation process, however, it actually regulates the latter process (Barrett et al. 2012). In the present study, PCR amplification and in silico characterization of 5' and 3'-UTRs of carotenoid genes was conducted. PCR amplification of 5' and 3'-UTRs of carotenoid genes i.e. *ggpps1*, *psyl*, *psy2*, *pds*, *zds*, *crtISO*, *lyc-b* and *lyc-e* from twelve accessions was attempted using genomic DNA as template. PCR resulted in the amplification of full length UTRs (Figures S2, S3). 5'-UTRs were represented as the nucleotide sequence upstream of the start codon whereas 3'-UTRs as the nucleotide sequence downstream of the stop codon to the poly (A) tail according to Aoki et al. (2010). Further start and stop nucleotide sequence site of the promoter and 5'-UTR were differentiated using the TSSP and UTRdb softwares. Accession numbers of the promoter and UTR sequences submitted to the GenBank are from KY100933 to KY100975.

and RAP2.1 have been shown by the arrows and different colors assigned to the different matrix families P\$LREM, P\$MYCL and P\$AHBP, P\$MADS, P\$IBOX and P\$DREB, respectively

of a plant. UTRs of carotenoid genes showed differences in terms of number and hence strength of uORFs present within them. Highest number of uORFs was observed in the 3' UTR of *psy2* gene. *Psy2* is actively expressed in the green-tissues (mainly leaves) which results in the biosynthesis of lutein and has very less participation in lycopene accumulation. Therefore, maximum number of uORFs in this gene may lead to feedback regulation of carotenoids in leaves. 5' UTR region of *ggpps* and *crtISO* gene contains minimum number of uORF i.e. only one, confirming no feedback regulation and therefore, their active participation in the biosynthesis and accumulation of lycopene in the ripe fruit. Similarly, *lyc-b* and *lyc-e* (major catabolic genes) contain higher number of uORFs, indicating the feedback regulation of these genes, which possibly results in the increased accumulation of lycopene than its catabolic products i.e. β -carotene and lutein in the red-ripe fruit.

Another major difference was observed in the Musashi binding elements (MBE) present in SOL1, SOL21, SOL22 and SOL26 (having maximum carotenoid accumulation) accessions. MBE represents one of the most significant UTR motif observed in the 5' UTRs of *crtISO* gene of different

accessions which possibly leads to the early translational activation and therefore, increases the gene expression (Charlesworth et al. 2006; Arumugam et al. 2010). Significant differences observed in the 5'-UTRs of *crtISO* gene in terms of number of MBE hints towards its differential role in regulating carotenoid biosynthesis (mainly lycopene) and accumulation. The tomato accessions having medium and lowest amounts of lycopene had one MBE whereas the accessions having highest amounts of carotenoids contained two MBE. Musashi is known to be one of the post-transcriptional regulators which act in different signal transduction pathways and enhance the gene expression and its activity (Zearfoss et al. 2014). However, an in-depth analysis of UTR regions needs to be done using transgenic and functional characterization approach in order to prove that whether the difference in the carotenoid content in different accessions is due to the difference in the sequences present in the UTR regions of the carotenoid genes. Different functional motifs present in 5' and 3' UTR's of carotenoid genes predicted by UTRScan software have been listed in Table S6. There are several reports on the role of 3' UTR in enhancing the gene expression (Papadakis et al. 2004; Vignesh et al. 2013) which has been ascribed to presence of several regulatory factors or coordinated role of 3' UTR and 5' UTR during loop formation facilitating the translation initiation process.

Also, many miRNAs are known to bind to the 3' UTR and facilitate the enhanced expression of the pathway genes which ultimately leads to the enhanced product accumulation in different tissues of plants (Barrett et al. 2012).

Expression profiling of *crtISO* and *lyc-b* gene

As the differences were observed in the promoters and UTRs of *crtISO* and *lyc-b* genes, transcript profiles of these genes were generated in both SOL 21 and SOL 5 in vegetative as well as reproductive tissues. Transcript profiling allowed us to identify the differential expression of *crtISO* and *lyc-b* among different developmental tissues of tomato: leaves and other green tissues (stem, IG and MG fruits), flower, ripening fruits (BR and RR fruits) and roots gene in SOL 21 and SOL 5 accessions. Differential expression was observed in various tissues of the tomato plant (Fig. 4). *CrtISO* gene showed highest expression at red ripe stage (RR) in SOL 21 as compared to SOL 5 followed by root, stem, breaker stage (BS), immature green (IG), flower, mature green (MG) while leaf had the least. The expression increased during the ripening process and about 1 to twofold increased transcript level was detected in RR with lowest in MG (Fig. 4). On the other hand, highest expression of *lyc-b* gene was observed in SOL 5 as compared to SOL 21 with higher expression

Fig. 4 Relative transcript levels quantified by qRT-PCR in relative to the expression of actin (internal control), expressed as $2^{-\Delta\Delta C_t}$. Relative gene expression profile of **a** *crtISO*, **b** *lyc-b* in SOL 21 and **c** *crtISO*, **d** *lyc-b* in SOL 5, respectively

in stem followed by flower, leaf and root and less expression in ripening tissues i.e. RR, IG, MG and least in BS (Fig. 4). *CrtISO* gene was highly expressed in ripening fruits (BR and RR) indicating its major role in the carotenogenesis, especially biosynthesis of lycopene in fruits. As for the chromoplastic-specific cyclase, *lcy-b* seems to be highly expressed in flower, green tissues and in roots while as less expression was observed in the ripening stages of the fruit. This up-regulation of *lcy-b* in flower and green tissues is responsible for the accumulation of β -carotene and various other downstream carotenoids like zeaxanthin and lutein (Stigliani et al. 2011).

A pictorial representation showing the possible regulation of carotenoid biosynthesis in tomato by promoters, UTRs, *cis* sequences and TFs is shown as Fig. 5. However, many aspects like copy number of *cis* sequences and their diversity within the plant species still remain unexplained and need further in silico and in vivo research (Liu and Bao 2009). Construction of the novel gene cassettes using the above regulatory players can be an excellent strategy to increase the expression of carotenoid genes and ultimately the production and accumulation of carotenoids in plants. For example, Liu and Bao (2009) explained that the availability of TFs is a limiting factor for the transcription process

to occur. Therefore, using the gene constructs with multiple motif copies definitely requires the sufficient quantity of TFs in the nucleus. Therefore, the major aim is to obtain stable, efficient and universal cassettes for driving the increased expression of the pathway genes.

Conclusion

The present study depicts the possible regulation of carotenoid biosynthetic pathway at the transcriptional as well as biochemical level. Carotenoid profiling of different accessions reveals differential accumulation of carotenoids (lycopene and β -carotene) at red-ripe stage of tomato. Characterization of promoters, UTRs, their *cis*-motifs, TFs indicates the possible regulation of this pathway via. differential expression/regulation of *crtISO* and *lcy-b* gene. This was also validation using Real-time PCR at different developmental stages. Overall, the possible inter-relationship between gene promoters, UTRs and carotenoid content was predicted. This study has provided us an initial path towards understanding the *cis*-regulatory elements (promoters and UTR's) of the carotenoid biosynthetic pathway which can

Fig. 5 A pictorial representation depicting the possible mechanisms responsible for the differential expression as well as differential accumulation of carotenoids in the different tomato accessions

be employed in future for the production of carotenoid rich tomatoes.

Acknowledgements This work was supported by the Department of Biotechnology (DBT), Govt. of India funded research project on tomato metabolomics. Authors thank Coordinator, Bioinformatics Centre (DBT BIF) for providing necessary facilities. Archana Koul is grateful to the Department of Science and Technology (DST), Govt. of India for INSPIRE Fellowship [xxxviii].

Compliance with ethical standards

Conflict of interest The authors declare that they have no conflict of interest.

References

- Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ (1990) Basic local alignment search tool. *J Mol Biol* 215:403–410
- Alvarez D, Vob B, Maass D, Wust F, Schaub P, Beyer P et al (2016) Carotenogenesis is regulated by 5' UTR-mediated translation of phytoene synthase Splice variants. *Plant Physiol* 172:2314–2326
- Aoki K, Yano K, Suzuki A (2010) Large-scale analysis of full-length cDNAs from the tomato (*Solanum lycopersicum*) cultivar Micro-Tom, a reference system for the Solanaceae genomics. *BMC Genomics* 11:210
- Arumugam K, Wang Y, Hardy LL, MacNicol MC, MacNicol AM (2010) Enforcing temporal control of maternal mRNA translation during oocyte cell cycle progression. *EMBO J* 29:387–397
- Aukerman MJ, Schmidt RJ, Burr B, Burr FA (1991) An arginine to lysine substitution in the bZIP domain of an opaque-2 mutant in maize abolishes specific DNA binding. *Genes Dev* 5:310–320
- Barrett LW, Fletcher S, Wilton SD (2012) Regulation of eukaryotic gene expression by untranslated gene regions and other non-coding elements. *Cell Mol Life Sci* 69:3613–3634
- Biłas R, Szafran K, Hnatuszko-Konka K, Kononowicz AK (2016) Cis-regulatory elements used to control gene expression in plants. *Plant Cell Tissue Organ Cult* 127:269–287
- Cartharius K, Frech K, Grote K, Klocke B, Haltmeier M, Klingenhoff A et al (2005) MatInspector and beyond: promoter analysis based on transcription factor binding sites. *Bioinformatics* 21:2933–2942
- Charlesworth A, Wilczynska A, Thampi P, Cox LL, MacNicol AM (2006) Musashi regulates the temporal order of mRNA translation during *Xenopus* oocyte maturation. *EMBO J* 25:2792–2801
- Chen Y, Li F, Wurtzel ET (2010) Isolation and characterization of the *Z-ISO* gene encoding a missing component of carotenoid biosynthesis in plants. *Plant Physiol* 153:66–79
- Dai Z, An K, Edward GE, An G (1992) Functional role of CAAT box element of the nopaline synthase (*nos*) promoter. *J Plant Biol* 42:181–185
- Dalal M, Chinnusamy V, Bansal KC (2010) Isolation and functional characterization of lycopene β -cyclase (CYC-B) promoter from *Solanum habrochaites*. *BMC Plant Biol* 10:61
- Dhar MK, Sharma R, Koul A, Kaul S (2014) Development of fruit color in Solanaceae: a story of two biosynthetic pathways. *Brief Funct Genomics* 14:199–212
- Doyle J (1991) DNA protocols for plants. Molecular techniques in taxonomy. Springer, Berlin, pp 283–293
- Gupta P, Sreelakshmi Y, Sharma R (2015) A rapid and sensitive method for determination of carotenoids in plant tissues by high performance liquid chromatography. *Plant Methods* 11:5
- Higo K, Ugawa Y, Iwamoto M, Higo H (1998) PLACE: a database of plant cis-acting regulatory DNA elements. *Nucleic Acids Res* 26:358–359
- Koul A, Yogindran S, Sharma D, Kaul S, Rajam MV, Dhar MK (2016) Carotenoid profiling, in silico analysis and transcript profiling of miRNAs targeting carotenoid biosynthetic pathway genes in different developmental tissues of tomato. *Plant Physiol Biochem* 108:412–421
- Krinsky NI, Johnson EJ (2005) Carotenoid actions and their relation to health and disease. *Mol Aspects Med* 26:459–516
- Laing WA, Martinez-Sanchez M, Wright MA, Bulley SM, Brewster D, Dare AP, Rassam M, Wang D, Storey R, Macknight RC, Hellen RP (2015) An upstream open reading frame is essential for feedback regulation of ascorbate biosynthesis in *Arabidopsis*. *Plant Cell* 27:772–786
- Larkin MA, Blackshields G, Brown NP, Chenna R, McGettigan PA, McWilliam H et al (2007) Clustal W and Clustal X version 20. *Bioinformatics* 23:2947–2948
- Lescot M, Dehais P, Thijs G, Marchal K, Moreau Y, Van de Peer Y et al (2002) PlantCARE, a database of plant cis-acting regulatory elements and a portal to tools for in silico analysis of promoter sequences. *Nucleic Acids Res* 30:325–327
- Liu S, Bao Y (2009) Effects of copy number of an octopine synthase enhancer element and its distance from the TATA box on heterologous expression in transgenic tobacco. *Acta Phys Plant* 31:705–710
- Liu Y, Yin J, Xiao M, Mason AS, Gao C, Liu H et al (2013) Characterization of structure, divergence and regulation patterns of plant promoters. *J Mol Biol Res* 3:23
- Livak KJ, Schmittgen TD (2001) Analysis of relative gene expression data using real time quantitative PCR and the $2^{-\Delta\Delta C_T}$ method. *Methods* 25:402–408
- Mantovani R (1999) The molecular biology of the CCAAT-binding factor NF-Y. *Gene* 239:15–27
- Meier S, Tzfadia O, Vallabhaneni R, Gehring C, Wurtzel ET (2011) A transcriptional analysis of carotenoid, chlorophyll and plastidial isoprenoid biosynthesis genes during development and osmotic stress responses in *Arabidopsis thaliana*. *BMC Syst Biol* 5:77
- Mignone F, Grillo G, Licciulli F, Iacono M, Liuni S, Kersey PJ et al (2005) UTRdb and UTRsite: a collection of sequences and regulatory motifs of the untranslated regions of eukaryotic mRNAs. *Nucleic Acids Res* 33:141–146
- Nicollé C, Simon G, Rock E, Amouroux P, Remesy C (2004) Genetic variability influences carotenoid, vitamin, phenolic, and mineral content in white, yellow, purple, orange, and dark-orange carrot cultivars. *J Am Soc Hortic Sci* 129:523–529
- Papadakis ED, Nicklin SA, Baker AH, White SJ (2004) Promoters and control elements: designing expression cassettes for gene therapy. *Curr Gene Ther* 4:89–113
- Porto MS, Pinheiro MPN, Batista VGL, Cavalcanti dos Santos R, de Albuquerque Melo Filho P, de Lima LM (2014) Plant promoters: an approach of structure and function. *Mol Biotechnol* 56:38–49
- Riechmann JL (2002) Transcriptional regulation: a genomic overview. *Arabidopsis Book* 1:e0085
- Rosati C, Aquilani R, Dharmapuri S, Pallara P, Marusic C, Tavazza R et al (2000) Metabolic engineering of beta-carotene and lycopene content in tomato fruit. *Plant J* 24:413–420
- Ruiz-Sola MA, Rodriguez- Concepcion M (2012) Carotenoid biosynthesis in arabidopsis: a colorful pathway. *Arabidopsis Book* 10:158
- Shahmuradov IA, Gammerman AJ, Hancock JM, Bramley PM, Solov'yev VV (2003) PlantProm: a database of plant promoter sequences. *Nucleic Acids Res* 31:114–117
- Smedley D, Haider S, Durinck S, Pandini L, Provero P (2015) The BioMart community portal: an innovative alternative to large, centralized data repositories. *Nucleic Acids Res* 43:589–598

- Srivastava R, Kumar R (2019) The expanding roles of APETALA2/ethylene responsive factors and their potential applications in crop improvement. *Brief Funct Genomics* 00:1–15
- Stigliani AL, Giorio G, D'Ambrosio C (2011) Characterization of P450 carotenoid β - and ϵ -hydroxylases of tomato and transcriptional regulation of xanthophyll biosynthesis in root, leaf, petal and fruit. *Plant Cell Physiol* 52:851–865
- Su L, Diletto G, Purgatto E, Danoun S, Zouine M, Li Z et al (2015) Carotenoid accumulation during tomato fruit ripening is modulated by the auxin-ethylene balance. *BMC Plant Biol* 15:114
- Tanaka T, Shnimizu M, Moriwaki H (2012) Cancer chemoprevention by carotenoids. *Molecules* 17:3202–3242
- Twyman RM (2003) Growth and development: control of gene expression, regulation of transcription. In: Thomas B, Murphy DJ, Murray B (eds) *Encyclopedia of applied plant sciences*. Elsevier Science, London, pp 558–567
- Vaughn JN, Ellingson SR, Mignone F, von Arnim AG (2012) Known and novel post-transcriptional regulatory sequences are conserved across plant families. *RNA* 18:368–384
- Venter M, Botha FC (2010) Synthetic promoter engineering. In: Pua EC, Davey MR (eds) *Plant developmental biology—biotechnological perspectives*. Springer, Berlin, pp 393–414
- Vignesh M, Nepolean T, Hossain F, Singh AK, Gupta HS (2013) Sequence variation in 3'UTR region of crtRB1 gene and its effect on β -carotene accumulation in maize kernel. *J Plant Biochem Biotechnol* 22:401–408
- von Arnim AG, Jia Q, Vaughn JN (2014) Regulation of plant translation by upstream open reading frames. *Plant Sci* 214:1–12
- Wu CY, Suzuki A, Washida H, Takaiwa F (1998) The GCN4 motif in a rice glutelin gene is essential for endosperm-specific gene expression and is activated by Opaque-2 in transgenic rice plants. *Plant J* 14:673–683
- Zearfoss NR, Deveau LM, Clingman CC, Schmidt E, Johnson ES, Massi F, Ryder SP (2014) A conserved three-nucleotide core motif defines Musashi RNA-binding specificity. *J Biol Chem* 289:35530–35541