Research article

Analysis of γ-Tocopherol methyl transferase3 promoter activity and study of methylation patterns of the promoter and its gene body

Khushboo Kumaria,b, Monika Prakash Raib, Navita Bansalb,b, G. Rama Prashatc, Sweta Kumaria, Veda Krishnan’a, Rohini Srivathsa,d, Anil Dahuja’a, Archana Sachdev’a, Shelly Praveena,b,*, T. Vinuthaa,∗∗

a Division of Biochemistry, IARI, New Delhi, 110012, India
b Amity University, Noida, Uttar Pradesh, 201313, India
c Division of Genetics, IARI, New Delhi, 110012, India
d National Research Centre for Plant Biotechnology, Pusa, New Delhi, 110012, India

ARTICLE INFO

Keywords: α-tocopherol Tocopherol biosynthesis cis-regulatory elements Promoter Methylation Gene body methylation Gene expression

ABSTRACT

Soybeans are known for its good source of protein (40%), oil (20%) and also serve as a source of nutraceutical compounds including tocopherols (toc). To know the molecular basis of differential α-toc accumulation in two contrasting soybean genotypes: DS74 (low α-toc - 1.36 μg/g and total-toc - 29.72 μg/g) and Bragg (high α-toc - 10.48 μg/g and total-toc 178.91 μg/g), the analysis of γ-TMT3 promoter activity and its methylation patterns were carried out. The sequencing results revealed nucleotide variation between Bragg:γ-TMT3-P and DS74:γ-TMT3-P, however none of the variations were found in core-promoter region or in cis-elements. The histochemical GUS assay revealed higher promoter activity of Bragg:γ-TMT3-P than that of DS74:γ-TMT3-P and correlated with significantly higher and lower (P < 0.05) expression of γ-TMT3 gene respectively. To know the molecular basis of differential accumulation of α-toc in these contrasting soybean genotypes, the DNA methylation pattern of γ-TMT3 gene body and its promoter was studied in both varieties. The results showed higher percentage (62.5%) of methylation in DS74:γ-TMT3-P than in Bragg:γ-TMT3-P (50%). Out of all the methylation sites in the promoter region, one of methylation site was found at CAAT box (-190 bp) of DS74:γ-TMT3-P. Further gene body methylation patterns revealed lowest % (40%) of CG methylation in DS74:γ-TMT3 gene as compared to Bragg:γ-TMT3 (64.2%). Thus our study revealed that, expression of γ-TMT3 gene was influenced by its promoter activity and methylation patterns in cis-elements of γ-TMT3 promoter and gene body. This study will help us to understand the possible role of methylation and promoter activity in determining the α-toc content in soybean seeds.

1. Introduction

Soybean [Glycine max (L.) Merrill] is one of the globally important crop because of its exclusive nutrient composition [protein (40%) and oil (21%)] and it is consumed in wider range of forms - natto, tofu, soy flour, miso, soy sauce, soy protein and soymilk (Wohleser, 2007; DellaPenna, 2005). Soybean is considered as "functional food" as it contains number of health promoting compounds such as vitamin E (vit-E), phenolics including phenolic acids, flavonoids and isoflavonoid (Fara and Vasanthan, 2004; Bramley et al., 2000). Tocopherols (toc) are the powerful antioxidant which is synthesized by photosynthetic organisms like higher plants and green algae (Traber and Atkinson, 2007). Vit-E is the combined term given to tocochromanol compounds which include four toc (α, γ, β and δ) and tocotrienols, which are structurally differ from one another by the number and position of methyl groups on the chromanol ring (Shintani and DellaPenna, 1998; Asemifi-F aboard and Munne-Bosch, 2010). Among the four toc, α-toc is attributed to the highest vit-E activity (DellaPenna and Pogson, 2006). Soybean contains highest amount of γ-toc but the vit-E activity of γ-toc is about one tenth of that of α-toc. Therefore manipulating the toc pathway to convert the γ-toc to the most bioactive α-toc in soybean seeds has gained much attention in recent years (Eenennaam et al., 2003). γ-TMT (γ-tocopherol methyl transferase) is the rate-limiting enzyme of tocopherol biosynthesis pathway, catalyzes the synthesis of α-
toc using γ-toc as a substrate (d’Harlingue and Camara, 1985; Koch et al., 2003). Several approaches have been employed to enhance the content of α-toc from most inactive γ-toc through expression of the γ-TMT gene in soybean seeds. Increased α-toc content in soybean seed, Arabidopsis thaliana and lettuce leaves have previously been demonstrated by over expressing the γ-TMT gene (Shintani and DellaPenna, 1998; Eenennaam et al., 2003; Cho et al., 2005). Generally, the regulation of gene expression at spatial, temporal level as consequences of various biotic and abiotic stresses is mediated at the transcriptional, post-transcriptional and post-translational levels. The transcriptional control of gene expression is regulated largely by promoter of the gene and its cis-acting elements (Zou et al., 2011). Promoters of protein-coding genes contains a “core promoter” region (60 bp to 120 bp) which is positioned upstream to transcriptional start site (TSS) and comprises TATA and CAAT-boxes (Maston et al., 2006). These regions are recognized and bound by the basal RNA polymerase II transcription machinery (Hahn, 2004; Kadonaga, 2002). Higher promoter activity and its direct relation with higher expression of Methyl phytylhydroquinone methyltransferase (VTE3) was reported in tomato (Quadrana et al., 2014) and positive correlation between α-toc accumulation and γ-TMT expression was also reported in soybean seeds (Dwiyanti et al., 2007, 2011). In soybean, γ-TMT proteins are classified as three different isoforms viz, γ-TMT1, 2 and 3 based on the variation found in the amino acid sequences at N-terminal regions and showed highest expression of γ-TMT3 in soybean seeds (Dwiyanti et al., 2011; Vinutha et al., 2015). Recent evidences have shown the regulation of gene expression through epigenetic mechanism leading to modulation in gene expression and therefore leading to novel phenotypes even in the absence of genetic variation (Boyko and Kovalchuk, 2011; Peredo et al., 2006). DNA methylation is the common epigenetic phenomenon in higher plants which plays an important role in growth and developmental processes, virus defense, transposon silencing and gene imprinting (Meijon et al., 2010; Villar et al., 2009). The most common methylation of cytosines occurs at CG, CHG and CHH contexts in plants where H is any nucleotide apart from guanine (Cokus et al., 2008; Zemach et al., 2010; Feng et al., 2010; Takuno and Gaut, 2013). DNA methylation in promoter region is often associated with decrease in transcriptional activity and ultimately inhibits the gene expression (Rambani et al., 2015; Wang et al., 2013). Promoter DNA methylation patterns was also shown to influence gene expression due to the modification of transcription factor (TF) binding sites on promoters (Turker, 2002). It was reported that 33% of the active genes in Arabidopsis possess DNA methylation (Zilberman et al., 2007; Zhang et al., 2006) most commonly in the transcribed regions (referred to as “gene body DNA methylation”) (Suzuki and Bird, 2008). From many of the reports, it was evident that, gene body methylation can affect the gene expression (Gupta et al., 2019; Wang et al., 2013). Methylome analysis in different tissues of soybean revealed highest CG methylation in gene body of highly expressed genes and lowest expressed genes showed the highest CHG and CHH methylation levels in gene body (Song et al., 2013 and Zhang et al., 2006). A large number of genes possess CG DNA methylation (mCG) in the existing transcribed region (Zilberman et al., 2007) and are found to be evolutionarily conserved and constitutively expressed (Niederhuth et al., 2016). The promoter region of TaEXP1A-B (Triticum aestivum expansin gene) and EXP1A-B (expansin gene) were shown to have stronger influence on the silencing effect of these genes in leaf tissue of wheat (Hu et al., 2013). Hypomethylation at promoter region of FWA gene showed the stronger induction of its gene expression (Kinoshita et al., 2004). Since methylation at gene body and promoter may cause significant amount of changes in gene expression, identification of methylation target sites at key cis-acting regulatory elements (CAREs) and at gene body region are important for complete understanding the processes involved in plant development, gene expression and metabolite accumulation. The present study was carried out to know the effect of γ-TMT3 promoter activity and direct impact of methylation of γ-TMT3 gene body and its promoter on the differential expression of γ-TMT3 and α-toc accumulation in soybean genotypes namely DS74 (low α-toc - 1.36 μg/g and total-toc -29.72 μg/g) and Bragg (high α-toc - 10.48 μg/g and total-toc - 178.91 μg/g) (Vinutha et al., 2015; Vinutha et al., 2017). The GUS reporter-aided approach was used to identify the γ-TMT3 promoter activities and methylation patterns were studied using bisulfite sequencing method.

2. Material and method

2.1. Plant materials and growth conditions

Seeds of soybean (DS74 and Bragg) and Arabidopsis thaliana L. (Col-0) were received from Division of Genetics, IARI and National Research Centre for Plant Biotechnology respectively. Soybean seeds were sowed and grown under greenhouse conditions at national phytotron facility, Pusa campus, New Delhi, at 28°C and 70–75% relative humidity (RH). Arabidopsis plants were grown at 20–22°C and 70% RH under 16-hours light/8-hours dark conditions for about 1–3 months.

2.2. DNA extraction

1 g of soybean seeds collected after 50 DAF (days after flowering) were weighed and crushed with help of liquid nitrogen. 10 ml of CTAB buffer and 0.2% of β-mercapto-ethanol were added to each tube and mixed mix vigorously for 10 min followed by incubation at 65°C for 1 h. Tubes were subjected to cool down to room temperature. Equal volume of chloroform: isoamyl alcohol (24:1) was added and mixed gently for 10 min. The tubes were centrifuged at 10000 rpm for 10 min at 4°C. The resulting supernatant was then transferred into a fresh tube followed by addition of two volumes of absolute ethanol and incubated for 1 h at −20°C. Further the mixture was centrifuged for 10 min at 10000 rpm 4°C. The pellet was retained and 70% ethanol washing was given. Pellet was dried and suspended in TE buffer.

2.3. 5′ RACE for mapping TSS

5′ RACE (Rapid Amplification of cDNA End) was employed to map TSS site of γ-TMT3 transcript using 5′ First Choice RLM-RACE kit (Ambion, USA). The soybean leaf sample (1 g) were ground with liquid nitrogen and total RNA was isolated and cDNA was synthesized using universal forward primer and outer reverse TMT3 primer specific to γ-TMT3 exon, and nested PCR was performed using gene specific primers TMT3 inner reverse and universal forward primer. The primer sequences are given in Supplementary Table 1.

2.4. Primer designing

Primers were designed from PRIMER3 software for full length amplification of γ-TMT3 promoter (~1.5 kb). For methylation study, the primers were designed using METH PRIMER tool.

2.5. Analysis of the γ-TMT3 promoter sequence

The γ-TMT3 promoter region (~1.5 kb) was amplified by PCR from soybean seed DNA (DS74 and Bragg) using γ-TMT3 specific primers (Supplementary Table 1) and cloned into pGEM-T Easy vector (Promega Corporation, USA). The clones were confirmed by restriction digestion and sequencing using Sanger’s dyeoxy method (Sequencer Tech Pvt. Ltd.). The sequencing results of γ-TMT3 were subjected to BLAST using NCBI BLAST tool and the sequences from both varieties showed 85% match with the γ-TMT3 promoter of soybean. The PLANTCARE software was used to analyze the CAREs in the promoter region (~1.5 kb) upstream to the defined TSS (Higo et al., 1999).
2.6. Development of γ-TMT3 promoter and GUS reporter gene fusion constructs

After confirming the γ-TMT3 promoter sequence cloned in pGEMT vector, γ-TMT3 promoter-GUS (β-glucuronidase) fusion constructs were developed using binary vector pORE-R2 (Coutou et al., 2007) containing GUS gene expression. The γ-TMT3 promoter fragments (~1.5 kb) from both varieties (DS74 and Bragg) were amplified by PCR. Each PCR fragment was further purified and digested with BamHI and Xho1 (New England Biolabs, UK). The digested products were ligated into pre-digested pORE-R2 vector with BamHI and Xho1 to create DS74:γ-TMT3-P::GUS and Bragg:γ-TMT3-P::GUS constructs. Clones were confirmed by restriction enzyme digestion followed by analysis on agarose gel (1%) and sequencing.

2.7. In planta Agrobacterium mediated transformation of Arabidopsis plants

Six-week-old Arabidopsis plants were used for transformation of γ-TMT3 constructs using floral dip method (Clough and Bent, 1998). These plants had more number of unopened buds and leaves were dipped into Agrobacterium suspension containing 5% sucrose and 0.05% Silwet-L77 solution and kept under dark for 12 h. Thereafter the transformed plants were washed with distilled water and allowed to grow till maturity. The harvested seeds from mature plants were sterilized and subjected to kanamycin (50 μg/ml) MS medium. The putative transgenic plants of Arabidopsis with two to six-leaf stages were then transferred from the antibiotic resistance plates to soil-rite containing pots. The presence of γ-TMT3 promoter in transgenic T2 lines was confirmed by PCR using specific primers.

2.8. GUS activity and assays

Two and six-week-old uniformly grown control and transgenic Arabidopsis seedlings were selected. For GUS activity, transgenic Arabidopsis leaves were crushed in protein extraction buffer (50 mM NaPO4 pH 7.0, 0.2% Triton X-100, 10 mM EDTA Na2, 10 mM β-mercaptoethanol) and centrifuged for 10 min at 10,000 g, 4°C. The Supernatant was then transferred to fresh tube and concentration of protein was determined using Bradford reagent with standard i.e BSA (bovine serum albumin). Further, supernatant was diluted from 0.1 to 1 mg/ml, reacted with MUG (1 mM 4-methylumbelliferyl- b-D-glucoside) and then incubated at 37°C for 2 h. The reaction was 10 times diluted with the stop buffer (0.2 M Na2CO3) and the measurements of enzymatic products was done by fluorometry at wavelength of 365/455 nm. GUS activity was expressed in 4-MU (pmoles 4-methyumbelliferone) produced per minute per milligram protein. To carry out GUS assay, leaves were collected and incubated in 50 mM sodium phosphate buffer (pH 7.0, 1.0 mM potassium ferro- and ferri-cyanide, 0.01M EDTA, pH 8.0, 20% methanol and 0.1% TritonX 100) containing 1 mM X-gluc (5-bromo-4-chloro-3-indolyl β-glucuronide) at 37 °C for overnight. After incubation, the chlorophyll was removed from leaves (bleaching) by using 70% ethanol. The results were recorded after repeated washing for 4-5 times with fresh 70% ethanol solution. The leaves from GUS positive plants were examined with a bright field microscope at a low magnification and images were taken by a digital camera.

2.9. Quantitative gene expression analysis

Total RNA was isolated from Arabidopsis leaves harboring both the constructs i.e DS74:γ-TMT3-P::GUS and Bragg:γ-TMT3-P::GUS via Trizol method. Total RNA was isolated and cDNA (250 ng) was prepared from developing seeds of soybean (30, 40 and 50 DAF) similarly as mentioned here for Arabidopsis leaves. Then γ-TMT3 expression analysis was carried out using γ-TMT3 specific primers. After DNase treatment (1 μg) total RNA was used to synthesize cDNA (cDNA synthesis kit). Equal concentration of cDNA (250 ng) was used to carry out Real Time PCR in a 10 μl reaction volume (Bio-Rad, USA) using SYBR Green PCR Master Mix (Takara, Japan). The GUS expression analysis in Arabidopsis plants harboring DS74:γ-TMT3-P and Bragg:γ-TMT3-P constructs were evaluated by using GUS-specific primers. The normalization of gene expression was done by using reference gene actin for Arabidopsis and EF1α-mRNA (Elongation factor 1 α-mRNA accession number- XM003553244, Glyma19g07240). The relative gene expression was calculated using ΔΔCt method (Schmittgen and Livak, 2008). All the primer sequences are listed in Supplementary Table 1.

2.10. Bi-sulfite treatment

This was done by using Epitect Bisulfite Qiagen kit (#59104). Bi-sulfite mixture was prepared using DNA (3 μg) isolated at 50 DAF stage from both varieties (DS74 and Bragg), 85 μl bisulfite mixture, 35 μl DNA protect buffer and volume made to 140 μl. Bi-sulfite DNA conversion was done in thermal cycler which was followed by washing and desulphonation of membrane bound DNA and elution of pure, converted DNA from spin column.

2.11. PCR and sequencing

Amplification was carried out using Bisulfite treated DNA of both varieties using PCR program of 35 cycles as follows: initial denaturation at 95 °C for 5 min, denaturation at 95 °C for 45 s, annealing at 58 °C for 50 s, extension at 72 °C for 90 s and final extension at 72 °C for 8 min. The PCR amplified products were separated on agarose gel (1%) by electrophoresis, purified and sequenced. Methylated cytosines were analyzed using BioEdit sequence alignment tool (Alzohairy, 2011).

3. Result and discussion

3.1. Sequence analysis of γ-TMT3 promoter

From our previous study, we showed genotypic variations in toc content as well as its composition in soybean germplasm collections and identified two contrasting genotypes (Bragg and DS74) for total toc and α-toc content. We also showed positive correlation between γ-TMT3 expression and α-toc accumulation in Bragg seeds as compared to DS74 (Vinutha et al., 2017). To know the molecular basis of α-toc variation, the complete CDS and promoter region including 5′ UTR sequences of γ-TMT3 gene were isolated from DS74 and Bragg seeds (Fig. 1A). The sequencing result showed no variation either in CDS region or in 5′ UTR sequences (gene sequence submitted to NCBI, accession number MK988566-DS74 and MK416205-Bragg), which instigated us to analyze the promoter region of γ-TMT3 gene to know the molecular basis for differential α-toc accumulation in these two contrasting soybean genotypes. To carry out this work, the TSS of the γ-TMT3 promoter was identified using 5′-RACE. More than five to eight 5′-RACE-PCR products were sequenced and identified one possible TSS at −28 bp (AAAAACAAAGAGAGCAAGGAGAGGAG). The γ-TMT3 promoter (~1.5 kb) upstream to TSS site was isolated by PCR amplification of genomic DNA from DS74 and Bragg seeds using specific primers (Supplementary Table 1), cloned and sequenced (NCBI Gene ID: MG855743-DS74 and MG855744-Bragg). The sequencing result was analyzed using PLANTCARE database to find CAREs of γ-TMT3 promoters. The analysis showed that, γ-TMT3 promoter region covers different classes of regulatory motifs. Eighteen different kinds of CAREs were found to be common in DS74:γ-TMT3-P and Bragg:γ-TMT3-P except for O2-SITE which was found to be unique in Bragg:γ-TMT3-P. The
core promoter region (60–120 bp) was found to be composed of only CAAT boxes which are similar in number (three CAAT boxes at −63,−64,−81(DS74) and at −62,−63,−80(Bragg)) in both DS74:γ-TMT3-P and Bragg:γ-TMT3-P. TATA-boxes were found to be the most abundant cis-acting elements i.e. eighty three and seventy three in numbers from Bragg:γ-TMT3-P and DS74:γ-TMT3-P respectively. The proximal sequences at TATA-boxes of γ-TMT3 promoter are relatively AT-rich. TATA-boxes are the key regulatory elements which contain binding sites for TF, enhancers and repressors required for spatiotemporal expression of the genes (Wittkopp and Kalay, 2012). CAT boxes like sequence (CAAT) are also present upstream to TATA-box at −21 bp (DS74:γ-TMT3-P) and at −22 bp (Bragg:γ-TMT3-P). A total of 16 CAAT-boxes are observed in γ-TMT3 promoters from both the genotypes. Seven different types of light responsive element (LRE) which includes ACE, Box4, Box1, I-Box, TCCC-motif, TCT-motif and Gbox were observed, all the six LREs were found to be present on the similar position in both DS74:γ-TMT3-P and Bragg:γ-TMT3-P, whereas the position of Box-1-LRE was found to be different i.e at −312 bp in DS74:γ-TMT3-P and at −294 bp Bragg:γ-TMT3-P. These LREs appears to be essential for light regulated transcriptional activity (Ochoa et al., 2007), however the presence of combinations of LREs in both DS74:γ-TMT3-P and Bragg:γ-TMT3-P acting as light responsive units (LRUs) which might be crucial to confer light responsiveness. Further we also observed the presence of ARE (anaerobic induction element) at −1425 bp in both DS74:γ-TMT3-P and Bragg:γ-TMT3-P. GC and GT are the two important motifs of AREs, in which GC is responsive to drought and low oxygen induced element (Dolferus et al., 2001). A HSEs (heat stress responsive elements) motif
<table>
<thead>
<tr>
<th>Name</th>
<th>Position in DS74</th>
<th>Position in Bragg</th>
<th>Number in DS74</th>
<th>Number in Bragg</th>
<th>Function</th>
</tr>
</thead>
<tbody>
<tr>
<td>AAGAA motif</td>
<td>(+) 673, 769</td>
<td>(+) 673, 769</td>
<td>2</td>
<td>2</td>
<td>Unknown</td>
</tr>
<tr>
<td>ACE</td>
<td>(−) 439</td>
<td>(−) 439</td>
<td>1</td>
<td>1</td>
<td>Light responsive element (LRE) part of a conserved DNA module involved in LRE</td>
</tr>
<tr>
<td>TCT-motif</td>
<td>(+) 1186</td>
<td>(−) 1186</td>
<td>3</td>
<td>3</td>
<td>LRE</td>
</tr>
<tr>
<td>ARE</td>
<td>(−) 1425</td>
<td>(−) 1425</td>
<td>1</td>
<td>1</td>
<td>Part of LRE</td>
</tr>
<tr>
<td>CAAT box</td>
<td>(−) 22, 52, 63, 64, 81, 190, 278, 287, 336, 342, (+) 372, 621, 786, 791, 1344, 1388</td>
<td>(−) 21, 51, 62, 63, 80, 190, 275, 284, 335, (+) 341, 372, 21, 786, 791, 1344, 1388</td>
<td>16</td>
<td>16</td>
<td>Enhancer regions</td>
</tr>
<tr>
<td>ERE</td>
<td>(+) 311</td>
<td>(−) 308</td>
<td>1</td>
<td>1</td>
<td>Ethylene responsive element</td>
</tr>
<tr>
<td>HSE</td>
<td>(−) 1059</td>
<td>(−) 1059</td>
<td>1</td>
<td>1</td>
<td>Element involved in heat stress responsiveness</td>
</tr>
<tr>
<td>MBSI</td>
<td>(+) 1244</td>
<td>(+) 1244</td>
<td>1</td>
<td>1</td>
<td>MYB binding site-Flavonoid biosynthesis required for endosperm expression</td>
</tr>
<tr>
<td>Skn-1 motif</td>
<td>(−) 1459, (+) 231</td>
<td>(−) 1459</td>
<td>2</td>
<td>2</td>
<td>Required for endosperm expression</td>
</tr>
</tbody>
</table>

(continued on next page)
was found at $-1059$ bp which was reported to be responding to elevated temperatures and regulate the gene expression levels (Storozenko et al., 1998). ABREs (abscisic acid responsiveness element) present at $-1056$ bp, found in both DS74-$\gamma$-TMT3-P and Bragg-$\gamma$-TMT3-P and are responsive to induce ABA-dependent and ABA-independent gene expression under osmotic and cold stress conditions (Yamaguchi-Shinozaki and Shinozaki, 2005). The Motif Skn-1 is specific to gene expression in endosperm was also detected at $-1459$ bp in both DS74-$\gamma$-TMT3-P and Bragg-$\gamma$-TMT3-P. However, O2-SITE (Opaque-2) present at core promoter region at $-228$ bp position present only in Bragg-$\gamma$-TMT3-P which has role in regulation of zein metabolism (Muth et al., 1996). From the available literature search, it was clear that O2 site is responsible for binding of O2 or a class of bZip proteins (Wu et al., 1998). The different type of cis-elements are grouped as O2 sites which are validated from different gene promoter; for example in maize -GCN4 motif-bound by either O2 or bZip TF (Wu et al., 1998) and regulates Glutelin gene expression in wheat; ACGT, ACAT and TCCACGT also reported as O2 binding motif for the regulation of $\alpha$-zein and $\beta$-zein genes in maize (Qiao et al., 2016; Neto et al., 1995). Thus we speculate that, presence of O2 site only in Bragg-$\gamma$-TMT3-P might not linked to the regulation of tocopherol biosynthesis. The functions of all these predicted cis-elements in both DS74-$\gamma$-TMT3-P and Bragg-$\gamma$-TMT3-P are listed in Table 1.

In the present study, nucleotide sequence analysis of $\gamma$-TMT3 promoter from both DS74-$\gamma$-TMT3-P and Bragg-$\gamma$-TMT3-P showed various additions, deletions and replacement of the nucleotides (Fig. 1B). The addition of nucleotides was found to be present at 14 and 17 different positions of Bragg-$\gamma$-TMT3-P and DS74-$\gamma$-TMT3-P respectively (Table 2). Analysis of CAREs between DS74-$\gamma$-TMT3-P and Bragg-$\gamma$-TMT3-P showed the difference only in the number of TATA boxes present; suggesting that different number of these TATA boxes might also contribute to their differential promoter activity (Wang et al., 2015). However in our study it was not explored further as presence of TATA boxes is quite common feature of a promoter and the number of TATA boxes is quite common feature of a promoter and the number of TATA boxes might also contribute to their differential promoter activity (Wang et al., 2015). However in our study it was not explored further as presence of TATA boxes is quite common feature of a promoter and the number of TATA boxes present in the core promoter region of DS74-$\gamma$-TMT3-P and Bragg-$\gamma$-TMT3-P are similar (two each in number). In both DS74-$\gamma$-TMT3-P and Bragg-$\gamma$-TMT3-P, the position of nucleotide addition were found to be similar except at $-1095$ bp (C), $-1130$ bp (T) and 1150 bp (T) in DS74-$\gamma$-TMT3-P. A stretch of 19 bp deletions were observed in both the genotypes at position from $-1256$ bp to $-1238$ bp. Single nucleotide deletions were found at $-1283$ bp (T) and 1300 bp (T), $-1299$ bp (G) in DS74-$\gamma$-TMT3-P, whereas in Bragg-$\gamma$-TMT3-P the single nucleotide deletion was found only at $-1236$ bp (T) (Fig. 1B). Replacement of nucleotides was also observed at 4 different positions i.e. at $-1160$ bp (A to T), $-1280$bp (T to A), $-1279$bp (C to G) and at $-1287$ bp (G to C) in DS74-$\gamma$-TMT3-P whereas only one type replacement was found at $-1160$ bp (A to T) in Bragg-$\gamma$-TMT3-P (Table 2). The sequence polymorphism analysis showed the presence of addition and deletion type of mutation at different position of DS74-$\gamma$-TMT3-P and Bragg-$\gamma$-TMT3-P, whereas no such variation were observed either in cis-elements or CDS region of the gene.

### 3.2. Validation of the promoter for $\gamma$-TMT3 gene expression in contrast varieties of soybean

In order to know whether the variation in the nucleotides is responsible for differential activity of the promoter, GUS histochemical activity and GUS expression analysis was carried out. In the present study, $\gamma$-TMT3 promoter from both varieties of soybeans was fused with promoter less GUS reporter gene. The pORE-R2 construct carrying the $\gamma$-TMT3::GUS reporter gene was prepared and used for transforming Arabidopsis plants. Plants harboring promoter less GUS construct and 3SS::GUS were used as negative and positive control respectively (Fig. 1C (i) and 1C (ii)). The successful transformation of pORE-R2-$\gamma$-TMT3::GUS construct in transgenic Arabidopsis plants (T2) was confirmed by GUS activity followed by GUS gene expression (Fig. 1D and

---

**Table 1 (continued)**

<table>
<thead>
<tr>
<th>Name</th>
<th>Number in Function</th>
</tr>
</thead>
<tbody>
<tr>
<td>DS74</td>
<td>0</td>
</tr>
<tr>
<td>Bragg</td>
<td>1</td>
</tr>
<tr>
<td>O2-SITE</td>
<td>(−) 228</td>
</tr>
<tr>
<td>TCA-element</td>
<td>(−) 1276</td>
</tr>
<tr>
<td>TC-rich repeats</td>
<td>(−) 56</td>
</tr>
</tbody>
</table>

**Table 1**

<table>
<thead>
<tr>
<th>Name</th>
<th>Number in Function</th>
</tr>
</thead>
<tbody>
<tr>
<td>DS74</td>
<td>0</td>
</tr>
<tr>
<td>Bragg</td>
<td>1</td>
</tr>
<tr>
<td>O2-SITE</td>
<td>(−) 228</td>
</tr>
<tr>
<td>TCA-element</td>
<td>(−) 1276</td>
</tr>
<tr>
<td>TC-rich repeats</td>
<td>(−) 56</td>
</tr>
</tbody>
</table>

...
γ-TMT3 showed stronger in T2 generation. The γ-TMT3 GUS bearing DS74- P::γ-TMT3::GUS (1.7 folds) in the leaves harboring Bragg:γ-TMT3::GUS and Bragg:γ-TMT3::GUS was found to be restricted to petiole, midrib and no GUS activity was observed in leaf apex, leaf margin as well as in whole leaf area whereas GUS activity driven by harboring Bragg:γ-TMT3::GUS promoter was observed in whole leaf area (Fig. 1C (iii) and 1C (iv)).

Similar kind of results were shown in transgenic Arabidopsis leaves harboring GUS gene under the control of γ-TMT3 promoter from KAS (high α-toc content) showed higher GUS activity (Dwiyananti et al., 2011). The expression analysis of promoter region of atToc343-GUS (component of chloroplast protein import machinery) seems to be restricted to vascular cylinder whereas atToc34-GUS showed strong expression in meristematic centre of seedlings which correlate with strong accumulation of atToc34-GUS mRNA and overall decrease in the expression in the older tissues as compared to younger aerial tissue (Gutensohn et al., 2000). It was shown that polymorphism in OsZh2P3 (basic leucine zipper (bZIP) transcription factor) promoter between selected drought-tolerant and drought-sensitive rice genotype leads to different expression patterns of these genes with lowest accumulation of VTE mQTL (VTE3 (1)) alleles correlated with difference in DNA methylation of the promoter sequence and further noticed that that abundance of mRNA of 2-methyl-6-phytylquinol methyltransferase (VTE3 (1)) inversely corresponds to level of methylation of the proximal promoter region of VTE3 in tomato. Further, in tomato population, it was proved that higher accumulation of VTE3 (1) transcript levels in the population line of IL5-2-6 and IL5-2-6-1 was due to differences in the transcriptional promoter activity between domesticated and wild alleles of VTE3 (1). As it was evident from γ-TMT3 promoter sequence analysis from both the genotypes that, no variation was found in the sequences of known CAREs (Tables 1 and 2) except the presence of O2-SITE only in Bragg:γ-TMT3-P which has no role in tocopherol pathway genes expression as discussed earlier, thus the present study of methylation results suggested that, epigenetic regulation of α-toc synthesis was prominent primarily through lower CG methylation implying that hypomethylation of γ-TMT3 promoter might require for higher γ-TMT3 gene expression. A site specific methylation patterns were studied for promoter regions of several genes of mevalonate pathway and observed CG methylation in promoter regions of FDP (Farnesyl diphosphate synthase), REF (Rubber elongation factor) and Col1 (coronatine-insensitive1) genes as dominant methylation patterns correlated with down regulated expression patterns of these genes with lowest accumulation of latex in rubber (Uthup et al., 2011). Kinoshita et al. (2004) and Soppe et al. (2000) found that in late flowering phenotypes which were named as fwa-1 Arabidopsis mutants did not have change in the FWA nucleotide sequence; instead they found that, the phenotype was due to ec-topic expression of FWA gene associated with heritable hypomethylation around TSS sites. Epigenomic analysis of DMRs (differentially methylated genes of the CDS region) of several genes from soybean recombinant inbred lines (RILs) showed the positive correlation between CG gene-body methylation and gene expression levels (Schmitz et al., 2013). Hayashi and Yoshida (2009) showed that, the low level of methylation at transcribed region of Ptk59 gene exhibited lower gene expression which conferred resistance to blast disease in rice.

To know whether DNA methylation is present within the CAREs or in its close proximity in promoter sequence, we have searched for cytosine methylation sites within CAREs. The results showed that CAAT-box at −190 bp from ATG site in DS74:γ-TMT3-P was found to be

<table>
<thead>
<tr>
<th>Name of genotypes</th>
<th>Variations in nucleotide sequence in γ-TMT3 Promoter- position of nt from TSS</th>
</tr>
</thead>
<tbody>
<tr>
<td>Addition</td>
<td>Deletion</td>
</tr>
<tr>
<td>Bragg</td>
<td>−386 = A, −385 = G</td>
</tr>
<tr>
<td></td>
<td>−800 to 809 = GAGAGAT</td>
</tr>
<tr>
<td></td>
<td>−1163 = T, −1162 = A</td>
</tr>
<tr>
<td></td>
<td>−1959 = C1161,</td>
</tr>
<tr>
<td></td>
<td>−1150−1163 = T, −1162 = A</td>
</tr>
<tr>
<td>DS74</td>
<td>−386 = A, −385 = G</td>
</tr>
<tr>
<td></td>
<td>−800 to 809 = GAGAGAT</td>
</tr>
<tr>
<td></td>
<td>−1959 = C1161,</td>
</tr>
<tr>
<td></td>
<td>−1150−1163 = T, −1162 = A</td>
</tr>
</tbody>
</table>

3.3. Impact of methylation on promoter and gene body region of γ-TMT3

To understand the possible role of methylation in γ-TMT3 gene body and its promoter in differential accumulation of α-toc, soybean seeds from DS74 and Bragg were collected at 50 DAF and total DNA was isolated. Bisulfite treatment was given to total DNA to assess the impact of DNA methylation. The GC rich regions in the form of a typical CpG island from −20 to −256 bp upstream for γ-TMT3 promoter and +1 bp to +237 bp downstream to ATG region for γ-TMT3 gene (Figs. 2C and 3A) were selected. The CpG region of γ-TMT3 promoter and its gene body were amplified through PCR using specific primers (Supplementary Table 1, Fig. 2A and B). The PCR products were cloned into pGEMT-easy vector and more than 15 clones were sequenced.
methylated whereas the similar site at −190 bp was not methylated in Bragg:γ-TMT3-P (Fig. 2F and G) implying that, methylation at CAAT boxes of core promoter region might have larger effect on lowering the expression of γ-TMT3 genes which was further confirmed by analyzing expression of γ-TMT3 gene which showed significantly (p < 0.05) lower expression in DS74 seeds as compared to Bragg seeds (Fig. 3D and E). CAAT-box sequences are the major CAREs which are also called as enhancers - known for their target site regulation and affect the transcriptional initiation frequency (Romier et al., 2003 and Dolfini et al., 2009). Methylation in such regulatory elements may severely affect transcription of gene (Song et al., 1998; Uthup et al., 2011; Yong-Villalobos et al., 2016). Similar studies were observed by Yong-Villalobos et al. (2016), where they reported that under low Pi-conditions (phosphate), the cis-regulatory elements within Pi-methDEG (phosphate-methylated differentially expressed genes) promoter were selectively methylated and leads to lower in expression of low Pi-responsive gene through affecting the binding capacity of TF to CAREs. Methylation at CAAT boxes of HMGR1 (3-hydroxy-3-methylglutaryl-CoA reductase1) promoter was found to be prominent and observed the significant down regulation of HMGR1 gene which encodes HMGR1 protein—the rate limiting enzyme of latex biosynthesis in rubber plants (Uthup et al., 2011). Further, DNA methylation patterns of γ-TMT3 gene body from both the genotypes were analyzed. In general, highest % of gene body methylation was observed at CHG sites followed by CHH and CG in both genotypes. Significantly highest% of CHG and CHH methylation (66.6%, 66%) was found in DS74:γ-TMT3-P as compared to CHG and CHH methylation (62.5%, 58.82%) in Bragg:γ-TMT3-P (Fig. 3B and C). However, the percentage of CG methylation (64.2%) was highest in case of Bragg:γ-TMT3-P as compared CG methylation (40%) in DS74:γ-TMT3-P (Fig. 3B, C, 5A and 5B). Similar study on gene body methylation of DMR associated genes was carried out through methylome analysis among different tissues of soybean and showed that the highest CG methylation pattern in gene body of highly expressed genes was observed and lowest expressed genes exhibited the highest CHG and CHH methylation levels in gene body (Song et al., 2013 and Zhang et al., 2006). Several other studies revealed the positive
Fig. 3. A) Diagrammatic representation highlighting the CpG region in the γ-TMT3 gene body. Cytosine methylation % in γ-TMT3 gene body isolated from DS74 B) and Bragg C) Expression pattern of γ-TMT3 gene was analyzed in soybean seeds from DS74 D) and Bragg E) at 30, 40 and 50 DAF by real-time qRT-PCR. Data are means ± SD of two biological and three technical replicates. Transcript levels were normalized with the expression of reference gene- EF1α-mRNA. Statistical analysis for gene expression was done through Duncan's test and P values < 0.005 were considered to represent significant differences.

Fig. 4. ClustalW analysis of bisulfite treated TMT3 promoter along with control (unmodified) γ-TMT3 DNA sequence encoding promoter region in soybean genotype DS74 A) Bragg B) showing methylated (presence of “C” nucleotide) & non-methylated cytosines (presence of “T” in place of “C” nucleotide).
correlation between gene body methylation at CHG and CHH sites and inhibition of gene expression, in spite of absence of methylation at promoter region (Busslinger et al., 1983; Barry et al., 1993; Hohn et al., 1996). However, we focused on the differential DNA methylation of γ-TMT3 promoter and its gene body from DS74 and Bragg and identified that γ-TMT3 gene expression was negatively correlated with methylation in promoter region and positively correlated with CG methylation in gene body. Promoter region of DS74-γ-TMT3 showed methylation within CAAT-box, whereas in Bragg-γ-TMT3-P showed no methylation at CAAT-box.

4. Conclusion

The γ-TMT3 promoter from high α-toc containing variety- Bragg showed higher promoter activity than the low α-toc containing variety- DS74. Although variation in the nucleotide sequence of cis-elements between Bragg-γ-TMT3-P and DS74-γ-TMT3-P was not observed, methylation of CAAT-box at core promoter region (−190 bp) was observed in DS74-γ-TMT3-P. Further higher and lower % of total methylation at promoter and CDS region of DS74-γ-TMT3 gene was observed respectively. DNA methylation pattern of gene body revealed, the highest CG methylation percentage in Bragg-γ-TMT3 CDS. Highest and lowest methylation rate at DS74-γ-TMT3 promoter and gene body respectively was correlated with significantly lowest expression of γ-TMT3 in DS74 as compared to Bragg. The differential accumulation of α-toc in contrasting genotypes thus found to be dependent on the enhanced γ-TMT3 gene expression due to higher promoter activity mediated through sequence diversity and methylation pattern at γ-TMT3 promoter and its gene in Bragg. The information generated in the present study will help us to develop promoter based molecular markers for breeding to enrich the α-toc in soybean. However, further studies are required to know deep insights on whether influence of methylation or sequence polymorphism or combination of both are required to enhance γ-TMT3 gene expression.

Ethical approval

This article does not contain any studies with human participants or animals performed by any of the authors.
pORE: a modular binary vector series suited for both monocot and dicot plant transformation. Transgenic Res. 16 (6), 771–781.


