Cytosine Methylation of Isoflavone Synthase Gene in the Genic Region Positively Regulates Its Expression and Isoflavone Biosynthesis in Soybean Seeds

Om Prakash Gupta,1,* Anil Dahuja,1 Archana Sachdev,1 Pradeep Kumar Jain,2 Sweta Kumari,1 Vinutha T.,1 and Shelly Praveen1

Plants, being sessile organisms, have evolved several dynamic mechanisms of gene regulation. Epigenetic modification especially cytosine methylation and demethylation actively regulates the expression of genes. To understand the role of cytosine methylation during isoflavonoid biosynthesis and accumulation, we performed cytosine methylation analysis in the coding region of two isoforms IFS1 and IFS2 gene, in two contrasting soybean genotypes differing in total isoflavone content (NRC37: high isoflavone; and NRC7: low isoflavone). The results indicated increased 5-mC in both the isoforms in NRC37 (∗20.51% in IFS2 and ∗85% in IFS1) compared with NRC7 (∗7.8% in IFS2 and ∗2.5% in IFS1) genotype, which signifies the positive role of 5-mC in the coding region of the gene leading to enhanced expression. In addition, temporal expression profiling [35 days after flowering (DAF), 45, 55, and 65 DAF] of both the isoforms showed increasing trend of accumulation in both the genotypes with maximum in NRC37 at 65 DAF. To further establish a correlation between methylation and expression of transcripts, we quantified the different isoforms of isoflavone in both the genotypes across all the stages. Therefore, the finding of this study would certainly increase our understanding of epigenetic regulation of isoflavone biosynthetic pathway mediated by the cytosine methylation that would assist molecular breeders to get high-performing soybean genotypes with better isoflavone yield.

Keywords: cytosine methylation, IFS1 and IFS2, gene regulation, Glycine max, epigenetics

Introduction

In addition to protein and vegetable oil, soybean is a premier source of vital bioactive compounds such as phytic acid, protease inhibitor, saponins, lecithin, and isoflavones. Soybean has attained the status of functional food owing to health promoting compounds especially isoflavone (Setchell et al., 2003). Isoflavone is produced via multistep phenylpropanoid pathway in the plants. Plants belonging to Papilionoideae family including soybean are major source of isoflavone (Wang and Murphy, 1994). The complex phenylpropanoid pathway leading to the isoflavone biosynthesis embraces numerous crucial enzymes such as phenylalanine ammonia lyase, cinnamic acid 4-hydroxylase, 4-coumarate CoA ligase, chalcone synthase, chalcone isomerase, and chalcone reductase. (Ralston et al., 2005).

In the recent past, several animal and human studies have shown the key role of the isoflavone during cardiovascular diseases, diabetes (Aerenhouts et al., 2010), cancer (breast, prostate, and colon) (Rose et al., 1986; Limer and Speirs, 2004), menopausal abnormalities (Clarkson, 2000), osteoporosis, loss of bone mass intensity (Rochfort and Panozzo, 2007), etc. Similarly, isoflavones play a vital role during plant’s life cycle, including significant biological processes such as inducer of nodulation genes during symbiosis, phytoalexins, fungal spore germination, and antifeedants (Ndakidemi and Dakora, 2003). Among all the enzymes of isoflavones biosynthesis, 2-hydroxyisoflavanone synthase (isoflavone synthase, IFS) is the most significant enzyme that potentially discriminates two plant types, that is, with or without isoflavone by deciding flux channeling toward isoflavone biosynthesis. Soybean genome encompasses two isoforms of IFS genes (IFS1 and IFS2) that differ by 14 amino acid only (Jung et al., 2000). The epigenetic regulatory mechanism underlying the varying degree of accumulation of different isoforms of isoflavones (daidzein, genistein, and glycitein) in developing soybean seeds is still a virgin field that requires immediate attention.

1Division of Biochemistry, ICAR-Indian Agricultural Research Institute, New Delhi, India.
2ICAR-National Research Centre on Plant Biotechnology, New Delhi, India.
*Present address: Division of Quality and Basic Sciences, ICAR-Indian Institute of Wheat and Barley Research, Karnal, India.
5-Methylcytosine (5-mC) is a hallmark of epigenetic gene modification and heterochromatin in both plants and mammals. In mammals, 5-mC is found exclusively at CG sites (often referred to as CpG sites) whereas plants methylate cytosines within CG, CHG, or CHH motifs (in which H is A, T, or C) (Wang et al., 2016). Plants contain relatively high levels of 5-mC, ranging from 6% to 25% of the total cytosine, depending on the species (Steward et al., 2002). Cytosine methylation also happens at differentially controlled promoters and within the protein-coding areas of highly expressed genes (Zilberman et al., 2007). Current literature has witnessed the dogma that most of the research work on epigenetic control of secondary metabolites comes from fungi (see review Cichewicz, 2010). Arase et al. (2012) established in planta assay systems to identify inhibition of cytosine methylation.

They reported that genistein, a major isoflavonoid, hinders cytosine methylation restoring transgene transcription. Similarly, genistein is also reported to restore the transcription of an epigenetically silenced endogenous gene in Arabidopsis plants. Overexpression of AtROSI increases the demethylation levels of both promoters as well as coding regions of genes involved in the flavonoid biosynthetic pathway and antioxidant pathway during control conditions that further increased during salt stress conditions. This is the first comprehensive study documenting the epigenetic regulation of flavonoid biosynthetic and antioxidant pathways during salt stress exposure of plants (Bharti et al., 2015). To date, there is no report on methylation status of IFS genes, one of immediate gene of isoflavone biosynthesis, in soybean genotypes that urges systematic study.

Abiding this knowledge, in this study, we have in silico characterized and checked the expression level of both the isoforms of IFS gene across the seed developmental stages followed by 5-mC analysis in soybean seeds in NRC37 and NRC7 soybean genotypes. Furthermore, we have profiled the differential accumulation pattern of different isoforms of isoflavones at four seed developmental stages (35, 45, 55, and 65 DAF) to establish a correlation with the level of miRNAs.

Materials and Methods

Plant materials, growth conditions, and sample preparation

In this study, we selected two extreme contrasting soybean genotypes (NRC7 and NRC37) differing in the total isoflavone content. The genotype NRC7 is a selection from Nanking Pb1 (selection from Nanking) and contains 490.4 and 1634.5 µg/g isoflavone content, respectively (Kumar et al., 2010). Both the genotypes were procured from ICAR-National Research Centre on Soybean, Indore, India. Before sowing, the seeds were initially treated with 1% sodium hypochlorite for 10 min and rinsed in distilled water for three times. Sterilized soybean seeds were then sown in 6” pot filled with autoclaved standard agrocoir peat growing media (Shaa Pith Media Company, India) in a 15 square feet growth chamber maintained at ICAR-National Phytotron Facility, Pusa Campus, ICAR-IARI, New Delhi-110012.

After little water sprinkling, pots were shifted to the growth chamber and maintained up to maturity with a growing condition of day/night temperature of 28/26°C, relative humidity ~75% with 16 h light duration. We selected four stages of soybean seed development that maximally contribute to the isoflavone accumulation, that is, 35, 45, 55, and 65 DAF. Each stage of each genotype was replicated three times. Seed samples were harvested at the mentioned stages and immediately frozen into liquid nitrogen and stored at −80°C till further use.

In silico analysis of IFS genes

Before going into the expression analysis and methylation studies, we first carried out bioinformatics analysis of IFS gene. Looking into the literature, we could identify two isoforms IFS1 and IFS2. Therefore, we first performed the sequence similarity analysis at both nucleotide and amino acid levels using BioEdit software version 5.09.04. After this, we carried out domain analysis of both the isoforms using SMART domain tool1 to check the name and functioning of domain. Furthermore, to know the presence and similarity of IFS1 and IFS2 in other crop species, we first carried out BLASTN analysis available at NCBI, and the most suitable nucleotide sequences were subsequently downloaded. These collected sequences were further used for clustalW analysis using BioEdit software version 5.09.04, to know the similarity and conserved nature of the domains across the species. In addition, we did phylogenetic analysis of the downloaded sequences to know the evolutionary relationship of IFS of other crops with that of soybean.

Total RNA isolation and cDNA synthesis and qPCR expression analysis

Total RNA was isolated from 100 mg seed tissue of four different developmental stages already mentioned in the Material and Methods section using Trizol (Invitrogen) following user’s manual. cDNA was synthesized by PrimeScript™ 1st strand cDNA Synthesis Kit (TaKaRa, Japan) at 42°C for 60 min following manufacturer’s instructions. The RNA and cDNA concentrations were quantified using NanoDrop spectrophotometer (NanoDrop technologies) and the A260/280 and A260/230 ratios were used to assess purity. Gene sequences of IFS were downloaded from GeneBank available at NCBI2 and soybean functional genomics database. Primers were designed using PrimerBlast3 available at which were further utilized to check for hairpin structure, homodimer, and heterodimer formation using4 and primers were made final if they were having ΔG less than −9 kcal/mol or more positive. We selected three reference genes as internal control (EF1α2α, cyclophilin, and Actin2/7) to find the most stable gene for data normalization. The primer sequence of genes is given in Table 1. All the primers were validated using thermal dissociation (Tm) curve of PCR amplicons that showed a single peak. Amplification of each gene was established with PCR with the following conditions of 94°C for 5 min, then 40

---

1http://smart.embl-heidelberg.de/
4http://eu.idtdna.com/calc/analyzer
cycles of 94°C for 15 s, 55°C for 30 s, and 72°C for 45 s. PCR products were resolved using 3% agarose gel electrophoresis with a size of ~150 bp. qPCR was carried out using Bio-Rad CFX96 machine in a reaction volume of 10 μL containing SYBR® Green JumpStart™ Taq Ready-Mix™ (Sigma), reverse primer (10 μM), forward primer (10 μM), and 10 ng/μL of cDNA using the mentioned PCR protocol. The relative expression level was calculated following the 2^ΔΔCt method\(^1\). Standard errors and standard deviations were calculated from replicates of three biological tissues.

### Isolation of genomic DNA and bisulfite treatment

Using standard protocol, isolation and recovery of DNA from mature and hard soybean seed are difficult. Therefore, we used a modified and efficient version of DNA isolation protocol that produced best recovery without losing integrity (Anonymous, 2007). The quality and quantity of the isolated DNA were confirmed by taking the ratio of 1.8 of A\(_{260}/A_{280}\) by the NanoDrop machine. The quality and integrity of DNA were further cross checked by running the DNA on 1% agarose gel, which showed no smear (Supplementary Fig. S1C). The agarose gel was freezed for 10 min at -20°C, which helps visualize DNA to some extent (Supplementary Fig. S1B).

#### PCR amplification of bisulfite-treated DNA, sequencing, and computational analysis

The bisulfite modified DNA of both the genotypes was subjected to PCR amplification using the primer mentioned in Table 1. The primers were designed using MethPrimer software available at Li’s Lab (www.urogene.org/methprimer/). The PCR was set up to 25 μL containing 5 μL of 5× GC buffer, 0.5 μL of 10 mM dNTP, 1.25 μL of 10 μM of forward primer, 1.25 μL of 10 μM of reverse primer, 0.25 μL of phusson polymerase (NEB), 3 μL of 15 ng/μL of modified DNA template, and 13.75 μL of sterile water. The reaction was run using a PCR program of 98°C for 1 min and 45 cycles of 98°C for 30 s, 55°C for 1 min, 72°C for 75 s, and a final extension of 10 min. The PCR run was resolved on 1.5% agarose gel, which confirmed the presence of ampli-con (Supplementary Fig. S1C).

Ten independent PCR purified methylated sequences of both IFS1 and IFS2 were sequenced by Chromus Biotech (India). As the sequence length was ~220 bp, the sequencing was done from both the ends using forward and reverse primers to get full coverage of the sequence. The obtained sequences were used for further downstream analysis. For a comparative analysis of the bisulfite sequence, both the sequences of IFS1 and IFS2 for different contrasting soybean genotypes were aligned by using ClustalX software and visualized manually by using BioEdit graphical view. The methylation data were analyzed by using the Kismeth software, which allows the analysis of methylation at every particular site. The identification of the differentially methylated

### Table 1. List of Primers Used in Quantitative Polymerase Chain Reaction and Bisulfite Sequencing Analyses in Soybean

<table>
<thead>
<tr>
<th>Sr. No.</th>
<th>Name</th>
<th>Accession No.</th>
<th>Forward primer (5’-3’)</th>
<th>Reverse primer (5’-3’)</th>
<th>Amplicon size (bp)</th>
<th>Tm (°C)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>IFS1</td>
<td>AF195798</td>
<td>AGA ATT CCG TCC CGA GAG GTT</td>
<td>TGC CAT TCC TGA AGT AGC CAA</td>
<td>127 64</td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>IFS2</td>
<td>AF195799</td>
<td>AAT GTG CCC TGG AGT CAA TCT G</td>
<td>GGC GTC ACC ACC CTT CAA TAT</td>
<td>100 66</td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>EF1α2a</td>
<td>XM_003524541.1</td>
<td>GCT CTC ACA GAG GCT CTT CCC</td>
<td>ATG ATG ATA ACC TGG GCA GTG A</td>
<td>147 59</td>
<td></td>
</tr>
<tr>
<td>4</td>
<td>Cyclophilin</td>
<td>CF806591</td>
<td>GCA CGA GTC CTC TTC CTC AG</td>
<td>CTC GGA GTC ACG TCG GC</td>
<td>144 59</td>
<td></td>
</tr>
<tr>
<td>5</td>
<td>Actin2/7</td>
<td>BW677100</td>
<td>ACT TGC CCA TCA GGA AGC TC</td>
<td>TGT TCA CCA CCT CTG CCA AG</td>
<td>145 60</td>
<td></td>
</tr>
<tr>
<td>6</td>
<td>IFS1 (d)*</td>
<td>AF195798</td>
<td>TGG YYT TTG AAG TAT YTY AAG GTT G</td>
<td>CCT TRR TAA TTT TRA TCT CCA TRR TCT C</td>
<td>220 64.6</td>
<td></td>
</tr>
<tr>
<td>7</td>
<td>IFS2 (d)*</td>
<td>AF195799</td>
<td>TGG YYA TTG AAG YAT YTY AAG GTT GG</td>
<td>TRA TCT CCA TRR TCT CAT CCT CAR C</td>
<td>208 64.3</td>
<td></td>
</tr>
</tbody>
</table>

(d)* denotes degenerate.

qPCR, quantitative polymerase chain reaction.

\(2^\Delta\Delta C_t\) method

\(1^\text{st}\) PCR run was resolved on 1% agarose gel to visualize the constant smear as a result of DNA degradation by bisulfite treatment. This smeared DNA is now single stranded that does not allow EtBr to bind properly as later binds with dsDNA. Therefore, the agarose gel was freezed for 10 min at -20°C, which helps visualize DNA to some extent (Supplementary Fig. S1B).

\(\text{PCR amplification of bisulfite-treated DNA, sequencing, and computational analysis}\)

The bisulfite modified DNA of both the genotypes was subjected to PCR amplification using the primer mentioned in Table 1. The primers were designed using MethPrimer software available at Li’s Lab (www.urogene.org/methprimer/). The PCR was set up to 25 μL containing 5 μL of 5× GC buffer, 0.5 μL of 10 mM dNTP, 1.25 μL of 10 μM of forward primer, 1.25 μL of 10 μM of reverse primer, 0.25 μL of phusson polymerase (NEB), 3 μL of 15 ng/μL of modified DNA template, and 13.75 μL of sterile water. The reaction was run using a PCR program of 98°C for 1 min and 45 cycles of 98°C for 30 s, 55°C for 1 min, 72°C for 75 s, and a final extension of 10 min. The PCR run was resolved on 1.5% agarose gel, which confirmed the presence of ampli-con (Supplementary Fig. S1C).

Ten independent PCR purified methylated sequences of both IFS1 and IFS2 were sequenced by Chromus Biotech (India). As the sequence length was ~220 bp, the sequencing was done from both the ends using forward and reverse primers to get full coverage of the sequence. The obtained sequences were used for further downstream analysis. For a comparative analysis of the bisulfite sequence, both the sequences of IFS1 and IFS2 for different contrasting soybean genotypes were aligned by using ClustalX software and visualized manually by using BioEdit graphical view. The methylation data were analyzed by using the Kismeth software, which allows the analysis of methylation at every particular site. The identification of the differentially methylated
region (DMR) was carried out in every 100-bp window with a step size of 50 bp by comparing bisulfite sequences for different contrasting genotypes using the Fisher exact test with a p-value cutoff of 0.05 (Wang et al., 2016). A window with ≥3 differentially 5-mCs and a ≥1.5-fold change in DNA methylation level was considered DMR.

Quantification of different isoforms of isoflavone content in soybean developing seeds

The methodologies for quantification and data analysis of different isoforms of isoflavone were adopted from our previous publication (Gupta et al., 2017).

Statistical analysis

The data were analyzed by using one-way analysis of variance (ANOVA) by using statistical software (SPSS 19.0). Duncan’s multiple-range tests were performed to determine the significant difference between means at a significance level of p < 0.05.

Results

Epigenetics have emerged as a crucial regulatory circuit of gene expression in plants during various stages of growth and developments. Cytosine methylation is one of the key components that regulate the expression of genes in plants temporally and spatially. In this study, we have demonstrated the essential role of cytosine methylation in regulating the expression of IFS1 and IFS2 in the coding region. Starting with in silico analysis (sequence similarity at nucleotide and amino acid level, clustalW across the species, phylogenetic analysis and domain prediction etc.), we have carried out the expression pattern analysis of IFS1 and IFS2 at four seed developmental stages (35, 45, 55, and 65 DAF), followed by cytosine methylation analysis.

In silico analysis of both the isoforms IFS1 and IFS2

Nucleotide sequences of IFS1 (AF195798) and IFS2 (AF195799) were downloaded from NCBI. The total length of IFS1 and IFS2 were 1774 bp and 1824 bp, respectively. Similarity analysis showed ~84.8% and 97.8% identity at nucleotide and protein level, respectively (Supplementary Figs. S2 and S3). In addition, domain analysis of IFS1 and IFS2 showed presence of two domains, that is, signal peptide domain at N-terminal region and p-450 domain (Supplementary Figs. S4 and S5). In addition, similarity matrix analysis of IFS1 and IFS2 showed a range from 72.5% to 99% (Supplementary Table S1) and 48.2% to 99.2% (Supplementary Table S2), respectively. Furthermore, we carried out phylogenetic analysis to see the most closely related crop species for both IFS1 and IFS2. The result showed that both the isoforms of G. soja were most closely related to G. max (Supplementary Fig. S6).

Temporal expression analysis of IFS1 and IFS2 in NRC7 and NRC37 genotypes

To know the most crucial stage of total isoflavone accumulation in soybean seeds for cytosine methylation analysis, we carried out the expression profiling of both the isoforms, IFS1 and IFS2, at four seed developmental stages, that is, 35, 45, 55, and 65 DAF in NRC7 and NRC37 genotypes. The expression data suggested gradual increase in total accumulation of both the isoforms in both the genotypes across all the stages of seed development with maximum accumulation at 65 (Fig. 2A, B). The net accumulation of both the isoforms was higher in NRC37 genotypes than in NRC7 (Fig. 2A, B). Compared with IFS2, the expression of IFS1 was more deferring in both genotypes with maximum differences in NRC7. In NRC7, the expression of IFS1 was ~2.3, ~4.59, and ~6.59-fold and IFS2 was ~2.3, ~3.19, and ~5.29-fold at 45, 55, and 65 DAF, respectively (Fig. 2A). Similarly, in NRC37, the expression of IFS1 was ~28.5, ~31.15, and ~35.03-fold and IFS2 was ~27.29, ~30.32, and ~34.48-fold at 45, 55, and 65 DAF, respectively (Fig. 2B). The expression of IFS1 and IFS2 at 35 DAF was taken as control in both the genotypes and was considered as 1 (Fig. 2A, B).

Analysis of variation in cytosine methylation pattern in the coding region of IFS1 and IFS2 in NRC7 and NRC37 genotypes

The coding region of IFS1 (~721–968 bp, near to middle of the gene body) of both the genotypes, that is, NRC7 and NRC37, was used for quantitative analysis of 5-mC...
The coding region of *IFS1* under study was found to contain 12 (30%) cytosines in the CG context, 5 (12.5%) cytosines in the CHG context, and 23 (57.5%) cytosines in the CHH context. This primary information was exploited for methylation studies of *IFS1* in both the contrasting soybean genotypes at 65 DAF. The detailed statistics of analysis is given in Supplementary Table S3. Nevertheless, alignment of representative bisulfite-treated *IFS1* sequence with the reference sequence (unmodified *IFS1*) was carried out in both the genotypes. Alignment result clearly indicated that NRC37 (8 cytosines) has more 5-mC methylation than NRC7 (3 cytosines) genotype (Fig. 4A, B).

Furthermore, quantitative analysis indicated context-specific variation in cytosine methylation in both the genotypes (Figs. 3B and 4C). An increase in the 5-mC level in the entire context, that is, CG, CHG, and CHH, was observed in the NRC37 soybean genotype (Fig. 3C). By contrast, increase in 5-mC level was observed in CG context only in NRC7 genotype (Fig. 3B). Out of 12 (30%) cytosines in the CG context, 5 (12.5%) cytosines in the CHG context, and 23 (57.5%) cytosines in the CHH context, we observed that 1 (8.33%) cytosine in CG context, 0 (0%) cytosine in CHG context, and 0 (0%) cytosine in CHH context in NRC7 soybean genotypes were methylated (Fig. 3C).

The coding region of *IFS2* (~708–940 bp, near to middle of the gene body) of both the genotypes, that is, NRC7 and NRC37, was used for quantitative analysis of 5-mC (Fig. 5A). The coding region of *IFS2* under study was found to contain 11 (28.2%) cytosines in the CG context, 5 (12.82%) cytosines in the CHG context, and 23 (58.97%) cytosines in the CHH context. This primary information was further exploited for the methylation studies of *IFS2* in both the contrasting soybean genotypes at 65 DAF. The detailed statistics of analysis is given in Supplementary Table S3. Nevertheless, alignment of representative bisulfite-treated *IFS2* sequence with the reference sequence (unmodified *IFS2*) was carried out in both the genotypes. Alignment result clearly indicated that NRC37 (~8 cytosines) has more 5-mC methylation than NRC7 (~3 cytosines) genotype (Fig. 6A, B).

Furthermore, quantitative analysis indicated context-specific variation in cytosine methylation in both the genotypes (Fig. 5B, C). An increase in the 5-mC level in the entire context, that is, CG, CHG, and CHH, was observed in the NRC37 soybean genotype (Fig. 5C). By contrast, increase in 5-mC level was observed in CG and CHG contexts only in NRC7 genotype (Fig. 5B). Out of 11 (28.2%) cytosines in the CG context, 5 (12.82%) cytosines in the CHG context, and 23 (58.97%) cytosines in the CHH context, we observed that 2 (18.18%) cytosines in CG context, 1 (20%) cytosine in CHG contexts, and 0 (0%) cytosine in CHH context of NRC7 genotype were methylated (Fig. 5C).
context in NRC7 (Fig. 5B), whereas 4 (36.36%) cytosines in CG context, 1 (20%) cytosine in CHG context, and 3 (13.04%) cytosines in CHH context in NRC37 genotypes were methylated (Fig. 5C).

Quantification of different isoforms of isoflavone in developing seed of soybean

By employing HPLC technique, we quantified the different isoforms (diadzein, genistein, and glycitein) of isoflavone in both the contrasting genotypes, that is, NRC7 and NRC37. The results significantly showed increasing trend of accumulation of total isoflavone across all the four stages with maximum content at 65 DAF (Figs. 7, 3D, E and 5D, E). Compared with NRC7, the increasing trend of content variation was more significant in NRC37 across all the stages (Fig. 7). Nevertheless, among all the isoforms, accumulation of diadzein was more significant followed by genistein and glycitein (Fig. 7).

Discussion

Owing to the significant role of isoflavone in plants and animals, its regulatory mechanism of biosynthesis at transcription level is crucial. Despite the potent regulatory mechanism of cytosine methylation during biotic and abiotic stress, plant growth, and development, its role during secondary metabolite biosynthesis especially isoflavone is highly unexplored.

Current literature have witnessed the facts that most of the research work on epigenetic control of secondary metabolites comes from fungi (Cichewicz, 2010). But recently, reports have started accumulating the knowledge on regulatory role of epigenetics during secondary metabolite biosynthesis and accumulation in the plants as well. Owing to several layers of regulation such as DNA methylation, histone modification, and chromatin remodeling, epigenomic control of secondary metabolite biosynthesis and accumulation in the plants could even be more complex as evidenced by the complex
studies done in fungi. In this study, we have demonstrated the essential role of cytosine methylation in regulating the expression of IFS1 and IFS2 in the coding region. The expression data of IFS1 and IFS2 suggested gradual increase in total accumulation of both the isoforms in both the genotypes across all the stages of seed development with maximum accumulation at 65 DAF. This trend of IFS expression is in agreement with the previous reports, where they have shown the increasing trend of accumulation with the advancement of the stages of seed development (Dhaubhadel et al., 2003). The net accumulation of both the isoforms was higher in NRC37 genotypes than in NRC7 genotypes delineating more active role leading to higher isoflavone content.

Temporal expression profiling of IFS1 and IFS2 led us to select 65 DAF as the most crucial stage, for cytosine methylation pattern analysis. 5-mC is a hallmark of epigenetic gene modification and heterochromatin in both plants and mammals. In the nuclei of differentiated mammalian cells, 5-mC is found exclusively at CG sites (often referred to as CpG sites), whereas plants methylate cytosines within CG, CHG, or CHH motifs (in which H is A, T, or C) (Zilberman et al., 2007). The designated coding region of both the isoforms showed significant difference in context-specific

FIG. 4. ClustalW analysis of bisulfite-treated IFS1 sequence along with the reference (unmodified) DNA sequence in soybean genotypes (A) NRC7; (B) NRC37. Each sample was sequenced four times (4×coverage) and representing sequence is depicted here in both modified and unmodified DNA.
cytosine methylation. The coding region of IFS1 showed differential context-specific 5-mC methylation pattern in both the contrasting soybean genotypes. The net context-specific cytosine methylation and total 5-mC methylation were more significant in NRC37 than in NRC7. Result showed an increase in total 5-mC (∼85%) in NRC37 than in NRC7 (∼2.5%). Similarly, the coding region of IFS2 showed an increase in 5-mC in NRC37 (∼20.51%) compared with ∼7.89% in NRC7. We also observed an increase in 5-mC in all contexts (CG and CHH) except CHG wherein it has shown the same pattern. Similarly, we observed higher compositional variation and total isoflavone content in NRC37 than in NRC7 genotypes (Fig. 7).

DNA methylation has been reported as one of the key factors that modulate gene expression at transcription level. Numerous studies have revealed the central role of DNA methylation in different plant developmental processes, such as seed/embryo development and gametophyte development (Hsieh et al., 2009; Zemach et al., 2010). Gene body methylation is not associated with repression. It is quite evident from the old literature that gene body methylation is a feature of transcribed genes especially in constitutive genes in angiosperms (Wolf et al., 1984; Bewick and Schmitz, 2017). Furthermore, widespread positive correlations between active transcription and gene body methylation have recently been established on the active X chromosome (Hellman and Chess, 2007) and by shotgun bisulfite sequencing of plant and animal genomes (Cokus et al., 2008; Lister et al., 2009).

Furthermore, certain genes with an increased CG methylation were found to be more highly expressed but increased non-CG (CHG and CHH) methylation reduced their expression in Arabidopsis (Schmitz et al., 2013). Although the role of non-CG methylation in regulating gene expression in TEs through pre- and post-transcriptional silencing is well established, its effect on gene expression in plants remains underexplored. Owing to the all-time expression behavior of IFS gene, it is regarded as constitutive type gene.
Therefore, increased context-specific cytosine methylation as well as total 5-mC methylation in the coding region of ORFs of IFS gene in NRC37 could be linked to more expression of both the isoforms in NRC37, which is quite evident from our result leading to more isoflavone accumulation. In addition, there might be other factors contributing differential accumulation pattern of isoflavone such as the level of other genes of the pathway, genetic background of the cultivars, and methylation status of promoter of these genes, which needs to be further investigated in future.

FIG. 6. ClustalW analysis of bisulfite-treated IFS2 sequence along with the reference (unmodified) DNA sequence in soybean genotypes (A) NRC7; (B) NRC37. Each sample was sequenced four times (4x coverage) and representing sequence is depicted here in both modified and unmodified DNA.

FIG. 7. Compositional variation of different isoforms of isoflavone in developing soybean seeds. Data are the means of measurements from three independent biological replicates. Bars represent the SE, \( p \leq 0.05 \). Source: Gupta et al., 2017.
CYTOSINE METHYLPATION REGULATES ISOFLAVONE BIOSYNTHESIS IN SOYBEAN

Conclusion

To conclude, cytosine methylation is a key regulatory mechanism deciding the fate of gene expression in plants. Isoflavones present in soybean are very crucial compounds having several health benefits, but their content in the seed is not significant. Therefore, to design better performing soybean genotypes with significant isoflavone content, the regulatory mechanism of its biosynthesis should be understood. Therefore, to enhance our further molecular understanding at epigenetic level, we established the correlation of IFS gene expression and its cytosine methylation pattern on isoflavone accumulation. Nevertheless, other genes of the pathway might also be involved in deciding the accumulation of isoflavones, which could be a future line of action. This study highlights the important regulatory role of cytosine methylation during isoflavone accumulation in soybean, which could help molecular breeders to develop isoflavone-enriched soybean genotypes in future.

Acknowledgments

Authors thankfully acknowledge the financial support of the Department of Biotechnology (DBT), Government of India, New Delhi, for Grant-in-Aid project No. 24–469 and grateful to anonymous reviewers for their helpful suggestions. The authors are also thankful to Drs. Ranjeet Ranjan Kumar and Suresh Kumar.

Authors’ Contribution

O.P.G., A.D., A.S., and P.K.J. conceived and designed the experiments; O.P.G. carried out the experiments, analyzed the data, and drafted the article; O.P.G. and A.D. performed literature survey and statistical analysis; O.P.G., A.D., A.S., P.K.J., S.K., and S.P. revised and finalized the article. All authors have read and approved the article.

Disclosure Statement

The authors declare that there are no conflicts of interest.

Supplementary Material

- Supplementary Figure S1
- Supplementary Figure S2
- Supplementary Figure S3
- Supplementary Figure S4
- Supplementary Figure S5
- Supplementary Figure S6
- Supplementary Table S1
- Supplementary Table S2
- Supplementary Table S3

References


Wang, X., Li, Q., Yuan, W., Cao, Z., Qi, B., Kumar, S., et al. (2016). The cytosolic Fe-S cluster assembly component MET18 is required for the full enzymatic activity of ROS1 in active DNA demethylation. Sci Rep 6, 26443.


Address correspondence to:
Anil Dahuja, PhD
Division of Biochemistry
ICAR-Indian Agricultural Research Institute
Pusa Campus
New Delhi 110 012
India
E-mail: dahuja.anil@gmail.com

Received for publication December 14, 2018; received in revised form February 5, 2019; accepted February 24, 2019.