

IDENTIFICATION, ANALYSIS AND EVALUATION OF STARCH BRANCHING ENZYME GENE PROMOTER FROM *ORYZA SATIVA*

PAKEEZA RUBAB, HIRA MUBEEN*, AMMARA MASOOD,
JAVAID IQBAL WATTOO AND MUSHTAQ A. SALEEM

Department of Biotechnology, University of Central Punjab, Lahore, Pakistan.

**Corresponding author's email: hira_sh@hotmail.com*

Abstract

Promoters are regulatory elements that control's transcription and overall expression profile of genes. Plants consists of several enzymes, which are important for controlling metabolic activities. Starch branching enzyme (SBE), is one of an important enzyme involved in biosynthesis of starch in plants. It also plays a crucial role in determining the structure and physical properties of starch granules. SBE has two types, SBEI and SBEII. SBEII further divided in SBEIIa and SBEIIb. Both of these have different effects and size in different crops. Starch branching enzyme from *Oryza sativa* was selected for the current research. Rice is a staple food of 70% population of the world. The present research was focused on identification of SBE gene orthologues, analysis of SBE promoter sequence through High throughput genome sequencing (HTGS) database, screening of cis regulatory elements through Plant CARE database and further detection of putative protein domains through Conserved protein domain family (CDD) database within the promoter region. Several bioinformatics software's were used for this purpose. The study provides a deep insight on importance of designing constitutive novel promoters, which can be effectively substituted to get enhanced transgene expression in agricultural crops.

Key words: HTGS, CDD, Plant CARE, Domain, Transgene.

Introduction

Promoter is a specialized part of the DNA which normally occurs upstream of the coding region of genes and can act as the essential controlling component for level of gene expression and regulation. Promoter helps to regulate transcriptional activity of a gene. Many novel constitutive promoters have been identified. However, to increase the availability of useful promoters is necessary for effective constitutive gene expression. Furthermore, variety of novel promoter sequences needs to be explored for utilization in transgenic crop production (Naqvi *et al.*, 2017). Moreover, the major problem to use such promoters covers some intellectual property right (IPR) issues (Masood *et al.*, 2017).

Starch is the inexpensive thicker, water binder and gelly like agent with a complex structure and homopolymer of glucose (Gilbert *et al.*, 2011; Syahariza *et al.*, 2013). Starch is the vital molecule, produced during the process of photosynthetic fixation of carbon dioxide by the chloroplast. The starch is a basic component of many plants as a storage tissue. Moreover, it has two types including the transitory starch and storage starch both of these are composed of one linear polymer, amylose and the other branched polymer, amylopectin. However, slow digestion ability of starch leads towards decrease in frequency of the metabolic disease, specifically obesity and diabetes. This is also useful to improve the complications associated with complex diseases (Lehmann & Robin 2010). To improve the efficiency of the agricultural crops, the need of appropriate promoters is essential. Some common promoters like CaMV 35S, rbcS promoter, actin promoter from rice, maize and soyabean promoters are being used actively (Jani *et al.*, 2002; Chiera *et al.*, 2007). Production of transgenic crops by transferring modified genes is a useful process. Plants having transgenes can express foreign proteins with industrial and commercial value (Mubeen *et al.*, 2017).

In plants, several starch branching enzymes have been discovered. Two most common types of starch branching enzymes includes, the SBEI and SBEII. Whereas, the SBEII have further two types the SBEIIa and SBEIIb. Both of these have a different role for the synthesis of the starch. SBEI uses amylose as its substrate and SBEII uses amylopectin as its substrate. The SBEI can transfer the long glucan chains while the SBEII can transfer the short chains of glucan (Martin & smith 1995; Bhattacharyya *et al.*, 1990).

Some of the previously reported SBE's differ in size in different crops. Particularly, in wheat both SBEI and SBEII have a size of 88kD. Similarly, in maize, the SBEI is specifically branched with the amylose which gives more long chains than the SBEIIb (Takeda *et al.*, 1993). In dicots, like pea and potato, the only single form of SBEII is defined with only few known genes, namely SBEI and SBEII. However, these were nonfunctional in the pea plant having wrinkled shape (Bhattacharya *et al.*, 1990; Poulsen *et al.*, 1993; Burton *et al.*, 1995). In rice, there are two types of SBE II including IIa and IIb, both unique to the endosperms (Yamanouchi *et al.*, 1992). In barely, both of the SBEIIa and SBEIIb express activities which are comparable in the endosperm of different crops but only the IIb shows some expression in the endosperm of barley (Sun *et al.*, 1997). The biosynthesis of starch can be done by many enzymes like the ADP-glucose, pyrophosphorylase, the granule bound starch synthase, soluble starch synthase, the starch branching enzyme and the starch debranching enzymes. Some of these enzymes like (GBSS, SSS, SBE & DBE) have many structures. SBE is a glucosyl transferase, it can help in the catalysis and formation of α 1-6 linkage of the amylopectin. It can also produce starch chains, which are responsible for synthesis of amylopectin (Wu *et al.*, 2013). The SBE is not only useful for the catalysis of the new branches but it is also beneficial for adding up new non reducing ends in the molecule of starch and this synthesis can be continued

at the new non reducing ends so it can show the pattern of branching in the amylopectin and can affect the starch amount (Yandea *et al.*, 2011).

Oryza sativa consists of 430MB genome comprising of 12 chromosomes. Alpha-amylase is the principal catalyst in the pathway that converts starch into glucose. However, the direction of an amylase quality articulation may assume an imperative part in seedling improvement. Oat seedling life corresponds with the measure of an amylase movement in the developing seed (Tsuchiya *et al.*, 1992). Rice is one of the essential nourishing crop consumed by some portion of the total population every year. Up to some extent, 90% of the Rice is starch, which includes the glucose polymer amylose and the amylopectin. They can be polymerized with the help of 1,4 and 1,6 linkages amylopectin, which is the enormously spared polysaccharide having high level of the polymerization (DP). Both the amylose and amylopectin are merged by two different pathways. Its union require the dynamic granule-bound starch synthase (GBSS), while amylopectin is a result of an intricate pathway, which includes distinctive isoforms of starch synthase (SS), starch expanding compounds (SBEs), and starch-debranching proteins (SDBEs) (Ball *et al.*, 2003). In higher plants, the biosynthesis of starch is carried out by four classes of the chemicals; ADP-Glc pyrophosphorylase (AGPase), starch synthase (SS), starch branching enzymes (BE) and starch debranching enzymes (DBE) (Smith *et al.*, 1997). However, in case of amylopectin the starch capacity is almost 65% to 85% having a proper structure with connected bunches (Jenkins *et al.*, 1993). Some of the hereditary studies, have uncovered numerous qualities and quantitative characteristic loci for grain quality. However, still some of the grain quality attributes are unpredictable. Furthermore, some real qualities has been cloned in a particular pathway, for example, starch, protein, lipid, and flavonoids biosynthesis.

The presence of several cis regulatory elements within the promoter sequence of gene plays a vital role in controlling overall expression of genes. Starch branching enzyme found to be highly rich with such functional motif and conserved protein domains. In the present study, we have identified the promoter sequence of SBE gene from *Oryza sativa* by using HTGS database and orthologues of SBE with higher similarity. We have further characterized the putative conserved protein domains and several cis-regulatory elements, which plays a crucial role in transcriptional regulation within the promoter region. The results are represented along with future utilization and benefits of promoter sequences.

Material and Methods

Bioinformatics approach for promoter analysis: The use of various bioinformatics tools and software's has made it possible to understand and analyze the gene expression patterns. The expression of genes can be studied by understanding the regulation process. However, a specific promoter present at 5' end of gene controls the gene expression regulation.

Sequence retrieval: The promoter sequence of SBE gene from *Oryza sativa* 2870bp was retrieved from NCBI-HTGS database. The selected gene was located at chromosome 6.

Analysis of orthologues: The SBE gene was searched in NCBI to find its genetic variants in different organisms. The orthologues predict similar functions of all SBE genes with subject SBE gene.

Analysis of Cis regulatory elements and conserved protein domains: The PlantCARE database was used for analysis of motif in SBE gene promoter. Protocomp software was used for prediction of localized proteins within promoter region. Further, the CDD software was used for prediction and analysis of conserved protein domains with structure and hierarchal representation of sequence clusters.

Results

Retrieval of SBE promoter sequence through high throughput genome sequencing: The promoter sequence of SBE gene from *Oryza sativa* was obtained from NCBI-High throughput genome sequencing (HTGS) database. The SBE gene was located at chromosome 6. The sequence is shown in Fig. 1 below:

Identification of SBE Motif: The SBE gene promoter sequence was searched in Plant CARE database for identification of putative motif present within the promoter sequence. Each motif has a different function and represented in a different color. Results are shown in Fig. 2 below:

Functions of SBE Motif: The SBE sequence analysis for identification of conserved region resulted in identification of several cis regulatory motif. Some of these were found useful in control of light responsiveness and some are involved in regulation of metabolism. Results are given in Table 1 below:

Identification of SBE gene orthologues: The subject SBE gene was matched with other SBE genes by using different bioinformatics software's for finding its orthologues. Results showed 41 different gene orthologues with percent similarity. All of these genes were located on different chromosomes. Few of these are given in Table 2 below:

Identification of proteins in SBE gene promoter: The selected SBE promoter sequence was searched against localized proteins by using softberry Protocomp 9.0 software for finding integral prediction of location of specialized proteins within the SBE promoter sequence. Results showed the presence of highest number of integral proteins in extracellular matrix and chloroplast region with value of 3.14-3.15 as given in Table 3 and Fig. 3 below:

```

>AF004004.1 Oryza sativa Japonica Group cultivar Nipponbare chromosome 6
clone OJ1165 E01, +++ SEQUENCING IN PROGRESS +++|
>ATGGTGATCTCAGCGCTTCGTTCCGCCAGGTGGAACAGCATCCTGTGATCAGCTGCATCTTCTTCAGTGTGTAGG
TCTCCTCTACCTCGTGAAGACGTAGTGCAGATCGGCCTTCTTGCCATTGAGTCCACCCGGCATTGTTCAATC
GATGGAGACGAGTACGACGACCCGCTTCGTATAGTCCCACGGCARTGAGGACTTGCTTTATGTACGCCATCTCAGT
ATAGCTCTCCTCGGTTTCAGTTGCGTTGCTCCTGTTCTTGCAATGATGATGATTTCATTGCTGGCTATGGGTGTGTAT
CAAATCTGACCAACTTTTCCGGCTCAGTTCTGCACAACATTGGACTCATTGTTTTACACTTCAGGTCTCAGTGAAA
CTCTGAAATGTAACGAACGACGAGAGAGGGCCGTAGTTTTTTTATACTGGCAAGTTGCAGTTCAGATCGCTGAAA
ATGGCGGGARCTGCTGCCTTTAACCCTCCACCGGGACACCGGGCTAGGAGGCAGCTCGGTGAAATCTCCTGAAAC
AATGCAGCARAGGAGTGAATCCATTGTACAAAAGAAAGTTATTCTGATCCTCTCAAGGTCAGGCCTGARATACAG
TTTGATTTTGACAAATGTAGTACAGTGCCTGCTGCAACGATGAAAGGGCTCAAARTTTCAGGAGGTGACGGTGG
TACATGCCGATCGCCATCTTGTGTGGGTGGAACCTGCCCCACGATGGCTTCAGTGACCAATGCGTCTTCTTCCCGA
ACAGCTTGTGGCTCTCGATCTCAAGCTGTTCCGAGATTTGACACTGTCCTCTCAAACGAACTGCTGCTGTTT
TTGCTCGCTGCTACTGCTGCACCTTCAGAAACCGCGTCTGAATCCCATGCCCTACCGCTGTCACACTATTATTGCTC
TGCAACAGGGCGACTTCGAGAGATCCCTGACGCCGGTGACAGAGAATGACCGGACTAGTGCCCGTGGCGACACGACGT
TGGCGTCAGCAAACTTGCATTCTCCTGTGTAGCTTTGGTCATCAARGCCAGAGCTGGAACCTGACTCTGACACTGAC
ACTGACAGAGCCATATCTGTCAACGACCTTATTCTTCTGCTTCTTCTGATTATTGGAGGAGAACAACTCATTGGA
CCCTTGATTGGTTCCACTGGTTCTTGAGGTGACACTCGTCCGACTGACGTGCTCCGGAGATGCATTCTTGCAGCAG
TAGCAGCAGCGTTGCGCTTCCGGTCACTCCTGAATGCTCTGAATGATTCTGATCGGGAGTACAGCCTCTTACCAAGA
TCCTGCCTTCTCACTGCTCTCTGATCTGCTTCGCGATGTCTCGAGCGATTTGCTTCGGGTGAGTAGTACCCAGCT
TCTCCTGACACCTCACTTTTGGACTTGAGCTTGAGCTCTTCTTGAGCCTCTCCTTCACTCCTCAAGAAACTCCT
CCATGTTCCCTTCCCTCTTCACTTGGATGAACTGAACAACGATTCTTGGTCACTGCCAACAAATGTCAAAAGTACTA
GGCTTGAGAGCACTATCCGTTTCTGGAAACGTGGTTTCCGAGACGTGCTTATACTGTTATTATGTTCACTCCTCAA
ATCTCTGAACCTGTCCGGCATCGTTGCAAGTGGCTCACCTTAARACCTTTATGACCTTATCTCCAGGAGATATGT
TAGATTCTGCTGCAGCTTTTGTGAGTARTCTCTGTATCACTACTATTATCTGTAGCATTCTTCAGGGTGTCTTTC
TCTGCCATGGAAATCGTGCAGCTTCCATAGCCAAATTTGCCATCTTTTCTTGTGCAGGTTCTCCTGTGCAAGAA
CTGAATGTACCTACCTTCTCCATGCGGTGTCTACTAGTAATAGTAGGACACCCCTGAAGCTGTAAAAACGGTTC
TTGCATTCTCAGTGCCTTGGATGTTGCCATGCCTGGAAATCCTCCCTGATCTTCTCAAGCAGCTCCTCCTGTGG
TGATGCTGACCGGAACGACTTCTTCGCGGTCTCATTTTCTTGAGACAATGCCGTGAAGTGTAAATCAACACTGCTGCT
ACCGTAGGAGATCAGGCTGCATTGATTGCTGAAATGTTGCAGAACGGGTGATGTGACAGTGCATTTCAGTT
TACTTGGCAGCTTCAGGTTGCTGACTTCTTCTGCATGGATCATCACTTCATCCCTTTTTGCAAGTACCGGGATTGTG
CCGATGCCCATCAGACGTGCATGACACTCGGCTGGCTGGGAAAATATCCTTGTGGATCACATTCTTTGCAGTGGT
TCTGCTTGAGGGGTGCTTCCCTTGGGTGGCTTCTGCCTCATCTAATATTTTATTAAAGAAAACAGTAAATACG
ATGGTACTAATTGATTGCATTAAACAAGAACTTGAGCATTAAAGCATAGGAGCAATTARGTACAACATACAGTACTAG
TCATGGTGATGGTGGTATGGTGCATAGTGGAGATCAAGGCTAATCCGGGGAGCTTCAGGGCTTCAACCAAACT
AGTAGACTCTCCATTAACCAATCAGAGGTCTGCAGATAATCCATAGGATTGCTATCTAAATAACCACTGTTTCCATG
TATTCTTCTTCCGATTTCCGCTGTTTTCTGGGACAGCTGCGTGTGTGTCATAGGATATCTGCCATGAGTTTCC
TTGAGRATTTAGTTCTTCTGCTAGCTGAAAGAGGGCAGCCCATATGTGTGGAGATTTAAGATGCAGTTCAGTAA
TTAAGAGGGAGGCCATARTTAATGTAGTAAAGCTGAACATCATGTCCGGTAGGATCAITTTAATTTTTGCATGTTTTT
CCTGACAAAGAAAAAGAGAACAGAATCTAACAAATGCCCACACTTTCATGCACAAATGTCTGTAAACTGTTTTT
TTTTTAAAAAGGATTTTTTGTATACTGCTAGTAAACAAAGTGTCTATTGCCGGGACCATCCTGAATATAGTCA
TTAGACTCATTACTGAAGAACATGTACATACTATAGCACAAATGCTGGAAGATGGAACTCTTAAGAGATCAAA
CTCTCAAGRGAAGAGAGAGTGCAGAACCTCCAGATTGTTGAAGAACAGCACAGCAATTGCATCATGCAGCTTCAG
AGATGAGATGCTTCAATTCCTTGGCAGGGAGAAGATATGTTGTGTTGATCGCCGAGGAAACAGGCAAATGGAAT
GGATGAAGACGAAGAGGAGTACTCCCAAGCACGAAGTGCACCGCAAGAAATGAAGCAACCTCAATCAGTTGAC
CAGATAAAATAGTAAGGGCAGAAATTTAAAGAGAGTGGCCTTTGATTTGTTTATCGCTGAAAGGGGACGACGTGC
TGGAGGGTCTAGGTCCTATGTTGGATTACCTCGTGCAGATCATCATCATCTCATTTTCTTCTGGGCTTTAT
AAGTATATGTTGACGAATGAATGAATTAATAAATGTTGATTAATTAGTCTGCAGGAGGTA
    
```

Fig. 1. Retrieval of promoter sequence from HTGS database.

```

>SBE 2870nt your sequence is >1500nt. For cpu reasons it was truncated to 1500nt
+ CCGTCTCATT TTCTTGAGAC AATGCCGTGTA AGTGTAAATCA ACACTGCTGC TACCGTAGGA GATCAGGCTG
- GCCAGAGTAA AAGAACTCTG TTACGGACAT TCACATTAGT TGTGACGACG ATGCCATCCT CTAGTCGGAC
+ CATTGGATT GCTGAATGT TCCAGAACGG GGTGATGTA CAGTGATCAT TTCCAGTTA CTTCGCAGCT
- GTAACCTAA CCACTTACA ACGTCTTGGC CCACTACACT GTCAGTAGTA AAGGTCAAAT GAACCGTCGA
+ TCAGGTTGCT GACTTCTTCT GCATGGATCA TCACTTCATC GCTTTTGGCA GATACCGGGA TTGTGGCGAT
- AGTCCAACGA CTGAGAAGA CGTACCTAGT AGTGAAGTAG GGAAAAAGCT CTATGGCCC AATACGGCTA
+ GCCCATCAGA CGTCCTTCA CACTGGGCTG GCTGGGAAA ATATCCTTGT GGATCACATT CTTTGGAGTG
- CGGGTAGTCT GCACTTACT GTGAGCCGAC GCAOCCITTT TATAGGAACA CTTAGTGTAA GAAAGTGCAC
+ GTCTGTCTT GAGGGTGTCT TCCCTTGGGT GCCTTCTGCC TCACTAAAT ATTTTATTA AAGAAAACA
- CAAGACAGAA CTCCCAACGA AGGGAACCA CCGAGAGAGG AGTAGATTA TAAAAATAAT TTCTTTTGT
+ GTAATAACA TGGTACTAAT TGATTGCAT AACAMGAAMC TTGAGCATT ACGATAGGAG CAATAGTA
- CATTATGCT ACCATGATA ACTAMGAAA TCTCTCTTG AAGCCCGAT TCGATCTCTC CTTAATGAT
+ CAACATACAG TACTAGTCAAT GGTGATGGTG GTATGCTGCT ACGTGGAGA GTCAGGCTA TTCCGGGAG
- CTGATCTCTC AAGAAAGTA GCACTAGCC CACTGGAGA TGTGAGCTCT CAGTTCGAT AAGGCTCTC
+ CTCAGGGCC TACAACAACA ACTAGTAGAG TCTOCATTA CCATTCAGAG GTCGCGAGAT AATCCATAGG
- GAGGCTGGG AATCTCTCTG TGAATAGCT AAGGTAATT GGTAACTCTC CAGACGCTA TTAGGTATCC
+ ATTCCTATCT AATACGCAC TGTTCGATG TATTCTTCTT GCCGATTTCC GCTGTCTTCT GGGCAGAGCT
- TAACGATAGA TTTATGGCTG ACAAMGTAC ATAGAGAA CCGCTAAAGG CGACAAAAGA CCGTCTCGA
+ GCGTGTGCT CCAAGGAT ATCTGCCATG AGTTTCTTG AGAATTTAGT TCTTCTCTCT AGCTTAACTC
- CGCAACACCA CGTTATCTTA TAGACGGTAC TCAAGGAMC TCTTAAATCA AGAAGGACGA TCGACTTCT
+ GCGAGCCCA TATGTGGA GATATTAAGA TCCAGTTCAA GTAATTAAGA GGAGCCATA ATTAATGTAG
- TCGTCCGGT ATACAGACT CTATAATTCT ACCTCAMGT CATTAAITCT CTTCCGGTAT TAATTACATC
+ TAAAGCTGA CATCATGTC GGTAGGATCA TTTTAAITTT TGCATGTTT TCTGACAAA GAAAAAGAG
- ATTCCACTT GATGACAGC CCATCTAGT AAAATTAAMA ACCTACAAA AGGACTCTT CTTTTTCTC
+ AACGAATCT AACAAAATC CCAGCATT CATCCAAA ATGTCTGTAA ACTGTTTTT TTAATAAAG
- TTGTCTAGA TTGTTTACG GGTGTGAAA GTAGGTGTT TACAGACATT TGACAAAAA AAATTTTTTC
+ GATTTTTGT ATACTGCTAG TACATAACA AAGTCTATT GCCGGGACCA TCTGAATAT AATGATAGA
- CTAATAACA TARGAGATC ATGTATGTT TCCAGATAA CGGCCCTGT AGGACTTATA TCGTAAITCT
+ CTCATTACTG AAGACATGT ACATACTATA GCACAAAATG CTGGAAGAAT GGAAGCTCTT AAGAGATCA
- GAGTAATGC TTCTGTACA TGTATGATAT CGTGTTTTAC GACCTTCTTA CTTCCGAGAA TTCTTCTAGT
+ AACTCTTCAA GAGAAGAGA GAGTCCGAAA OCTOCAGATT GTTGAAGAAC AGCACAGCAA TTGCATCATG
- TTGAGAAGTT CTCTTCTCT CTCACGCTT GGAGTCTAA CACTTCTTG TCGTGTGTT AAGTATGAC
+ CAGCTTCAGA GATGAGATGC TTCAATCT TGGCAGGAG AAGAATATGT TGTGTGATC GCGAGGAAA
- GTCGAAGTCT CTACTCTAG AAGTAAAGG ACCGTCCCTC TTCTTATAA ACAGACTAG CCGCTCCTT
+ CAAGCCAAA TGGATGGAT GAGAGGAAG AGGAGAGTA CTOCCAACA CGAAGTGAC CCAAGAAIT
- GTTCGGTTT ACCTTACCTA CTCTCTCTC TCTTCTCAT GAGGGTCTG GCTTCACTGC CGTTCTTIA
+ GAAGGAACA TAAACAGT GAGCAGAATA AAATAGTAG GCCAGAAAT TTAAGAGAG TGGCCTTGA
- CTTCCTTGGT AGTATGCAA CTGCTTAT TTTATCATT CCGTCTTAA AATTTCTCTC ACCGAAACT
+ TTTGTTTATC GCTGAAAGG GACGACGTG TGGAGGCTT AGGTCCCATG TTGATTACC CTGTGGAG
- AAACAAATAG CGACTTCC CTGCTGCAG ACCTCCAGA TCCAGGTAC AACCTAATGG GAGCAGCTC
+ TCAATCAT CATCTCAIT TTCTTCTGG GCTTTATAG TATATGTTGA CGAATGAAT GAATIAAAG
- ACTAGATA GTAGATAA AAGAAAGACC CGAATATC ATATACAAT CTTACTTAA CTTAATTTG
+ AAGTTGAT AATTAGCTG CAGGAGTT
- TTACAACIAA TTAATCAGAC GTCCTCAA

```

Fig. 2. Shows presence of motif at conserved regions within the SBE promoter sequence obtained from PlantCARE database.

Table 1. Shows presence of cis regulatory elements in different organisms.

| Motif name | Function | Organisms |
|-----------------|---|----------------------------------|
| AAGAA-motif | GAAAGAA | <i>Avena sativa</i> |
| AE-box | AGAAACTT part of a module for light response | <i>Arabidopsis thaliana</i> |
| Box 4 | ATTAAT part of a conserved DNA module involved in light responsiveness | <i>Petroselinum crispum</i> |
| CAAT-box | CAAT common cis-acting element in promoter and enhancer regions | <i>Hordeum vulgare</i> |
| CAT-box | GCCACT cis-acting regulatory element related to meristem expression | <i>Arabidopsis thaliana</i> |
| CGTCA-motif | CGTCA cis-acting regulatory element involved in the MeJA-responsiveness | <i>Hordeum vulgare</i> |
| CTAG-motif | ACTAGCAGAA | <i>Avena sativa</i> |
| G-box | CACGTC cis-acting regulatory element involved in light responsiveness | <i>Zea mays</i> |
| GAG-motif | AGAGATG part of a light responsive element | <i>Spinacia oleracea</i> |
| I-box | CTCTTATGCT part of a light responsive element | <i>Nicotiana plumbaginifolia</i> |
| O2-site | GATGA(C/T)(A/G)TG(A/G) cis-acting regulatory element involved in zein metabolism regulation | <i>Zea mays</i> |
| Pc-CMA2c | GCCCACACA part of a light responsive element | <i>Spinacia oleracea</i> |
| Skn-1_l_motif | GTCAT cis-acting regulatory element required for endosperm expression | <i>Oryza sativa</i> |
| Sp1 | CC(G/A)CClight responsive element | <i>Zea mays</i> |
| TATA-box | TAATA core promoter element around -30 of transcription start | <i>Glycine max</i> |
| TATA-box | TTTAA core promoter element around -30 of transcription start | <i>Oryza sativa</i> |
| TC-rich repeats | ATTTTCTTCA cis-acting element involved in defense and stress responsiveness | <i>Nicotiana tabacum</i> |

Table 2. Shows the SBE gene orthologues.

| Sr. No. | Name gene | Description | Chromosome location |
|---------|--|---|---------------------|
| 1. | SBE2.2 ID: 831769 | <i>Arabidopsis thaliana</i> | Chromosome 5 |
| 2. | SBE1 ae1 ID: 542238 | <i>Zea mays</i> | Chromosome 5 |
| 3. | LOC4342117 ID: 4342117 SBE1 | <i>Oryza sativa</i> var. <i>japonica</i> Group (Japanese rice) | Chromosome 6 |
| 4. | LOC102596498 ID:102596498 SBE II | <i>Solanum tuberosum</i> | Chromosome 4 |
| 5. | LOC106762405 ID:106762405 SBELLb | <i>Vigna radiate</i> | Chromosome 5 |
| 6. | LOC4329532 ID: 4329532 SBE3 | <i>Oryza sativa</i> var. <i>japonica</i> Group (Japanese rice) | Chromosome 2 |
| 7. | LOC103440187 ID:103440187 SBEI | [<i>Malus domestica</i> (Apple)] | Chromosome 8 |
| 8. | LOC7487520 ID: 7487520 | <i>Populus trichocarpa</i> | Chromosome 1 |
| 9. | LOC8059621 ID: 8059621 Sbellb | <i>Sorghum bicolor</i> (Sorghum) | Chromosome 4 |
| 10. | Bathy06g02210 ID: 19015201 | <i>Bathycoccus prasinus</i> | Chromosome 6 |

Table 3. Shows the location of proteins within the SBE promoter region.

| Location weights: | LocDB / | PotLocDB / | Neural Nets / | Pentamers / | Integral |
|-------------------|---------|------------|---------------|-------------|----------|
| Nuclear | 0.0 / | 0.0 / | 0.00 / | 0.00 / | 0.00 |
| Plasma membrane | 0.0 / | 0.0 / | 1.00 / | 0.00 / | 0.00 |
| Extracellular | 0.0 / | 0.0 / | 1.00 / | 0.87 / | 3.15 |
| Cytoplasmic | 0.0 / | 0.0 / | 0.00 / | 1.51 / | 0.00 |
| Mitochondrial | 0.0 / | 0.0 / | 0.00 / | 0.99 / | 0.77 |
| Endoplasm. retic. | 0.0 / | 0.0 / | 0.00 / | 1.38 / | 0.15 |
| Peroxisomal | 0.0 / | 0.0 / | 1.00 / | 0.00 / | 2.79 |
| Golgi | 0.0 / | 0.0 / | 0.00 / | 1.31 / | 0.00 |
| Chloroplast | 0.0 / | 0.0 / | 0.00 / | 0.00 / | 3.14 |
| Vacuolar | 0.0 / | 0.0 / | 0.00 / | 0.02 / | 0.00 |

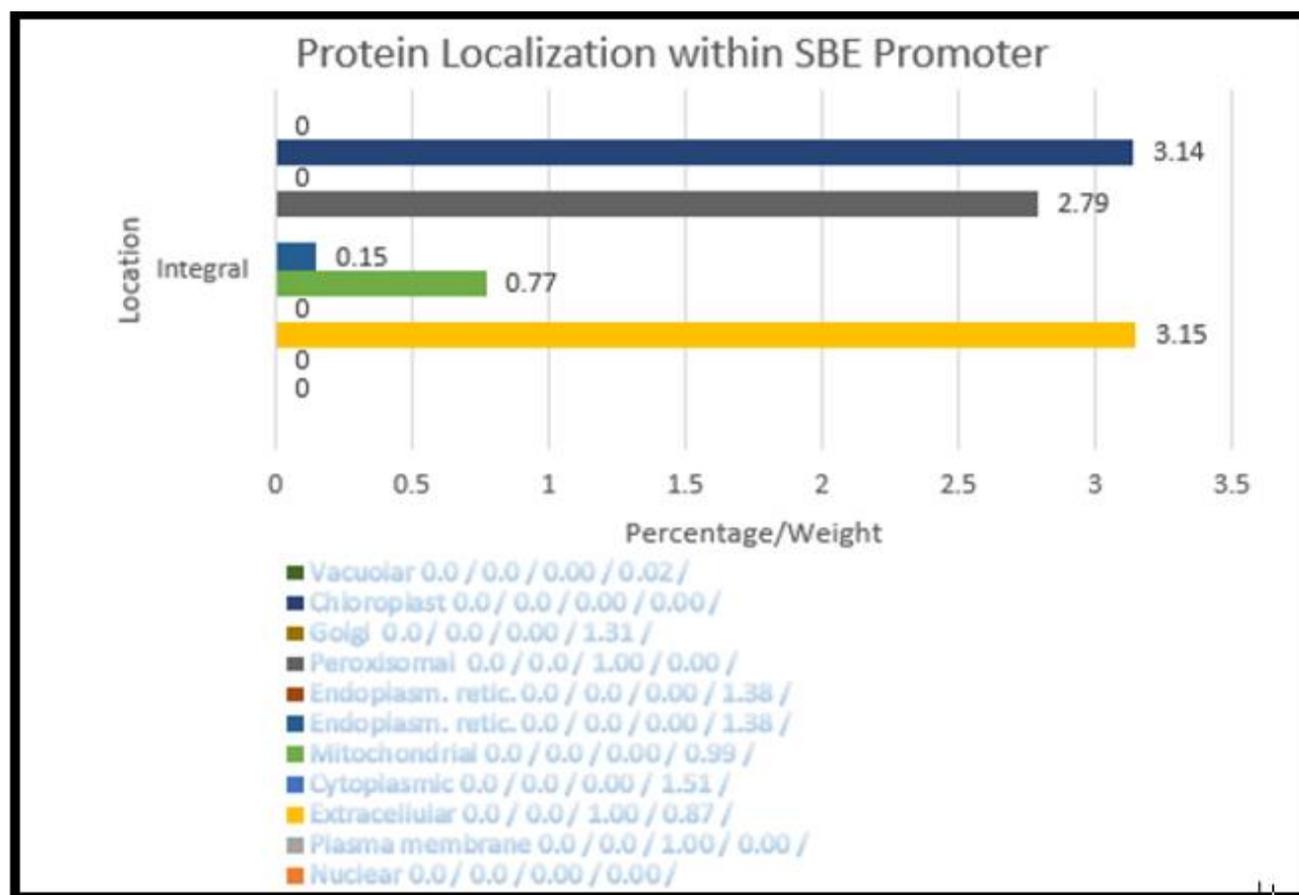


Fig. 3. Shows the diagrammatic representation of localized proteins of SBE.

Analysis of protein domains: The SBE promoter sequence was explored further by applying computational approach. The NCBI CDD data repository was used to find the putative conserved domains. Out of these, most of the domains were responsible for catalytic activity of proteins at the N-terminal end. Chitinases are important components necessary for hydrolyzing the abundant natural biopolymer chitin, which produce smaller chito-oligosaccharides. Several domains were identified with different catalytic activity. The sequence cluster (cd02848) was associated with chitinase catalytic domain. Furthermore, another N-terminal domain (cd11234) was identified and associated with glycogen debranching enzyme. Glycogen debranching enzymes exhibit two types of activities, including 4-alpha-glucanotransferase and amylo-1,6-glucosidase activity. The cluster tree for both of these domains were obtained from cluster protein domain family (CDD) and represented in Fig. 4(a, b) below:

Discussion

SBEs have various isoforms in growing stockpiling tissues of maize, rice, pea, wheat, potato and *Arabidopsis*. SBEs are isolated based on two gatherings made on auxiliary and the synergist properties; the principal aggregate is the SBE family II (Martin & Smith, 1995). Another SBE family I or B contains SBEI from maize, wheat, potato, cassava and SBEII from pea (Koporann *et al.*, 1991; Salehuzzaman *et al.*, 1992; Sun *et al.*, 1997; Larsson *et al.*, 1996; Fisher *et al.*, 1996; Burton *et al.*, 1995).

In *Arabidopsis*, this has been revealed that SBEII can be additionally subdivided into two kinds, generally delegated SBEIIa and SBEIIb, that vary marginally in synergist properties (Gao *et al.*, 1997). The requirement for numerous isoforms of SBE in plants is not easily comprehended and differentiates forcefully to the single glycogen-stretching compound found in microbes and warm-blooded animals. No doubt, plants require distinctive expanding exercises in various tissues and amid various formative phases of sink tissues. Moreover, it was observed that a move from SBEII to SBEI action amid pea incipient organism advancement was associated with changes in the amylopectin structure (Burton *et al.*, 1995). The fine structure of amylopectin is known to vary contingent upon plant species and different plant tissues, and this variety impacts the particular properties of the individual starch. Starch fanning protein (BE) catalyzes the development of α -1,6-glucosidic linkages of amylopectin amid starch biosynthesis, and hence BE has a critical impact in the arrangement of a particular fine structure of amylopectin.

Genome wide characterization studies revealed the discovery of 46 SBE's. Out of these, three were rich in *Arabidopsis thaliana*, *Zea mays*, barley and potato (Tetlow & Emes, 2014), four in *Oryza sativa* and seven in *Triticum aestivum* (Tyagi *et al.*, 2017) and six in cassava (Pei *et al.*, 2015). In addition to this, several SBE give rise some specialized modifying effects, which plays an important role in many other biological processes including germination, seedling growth and development.

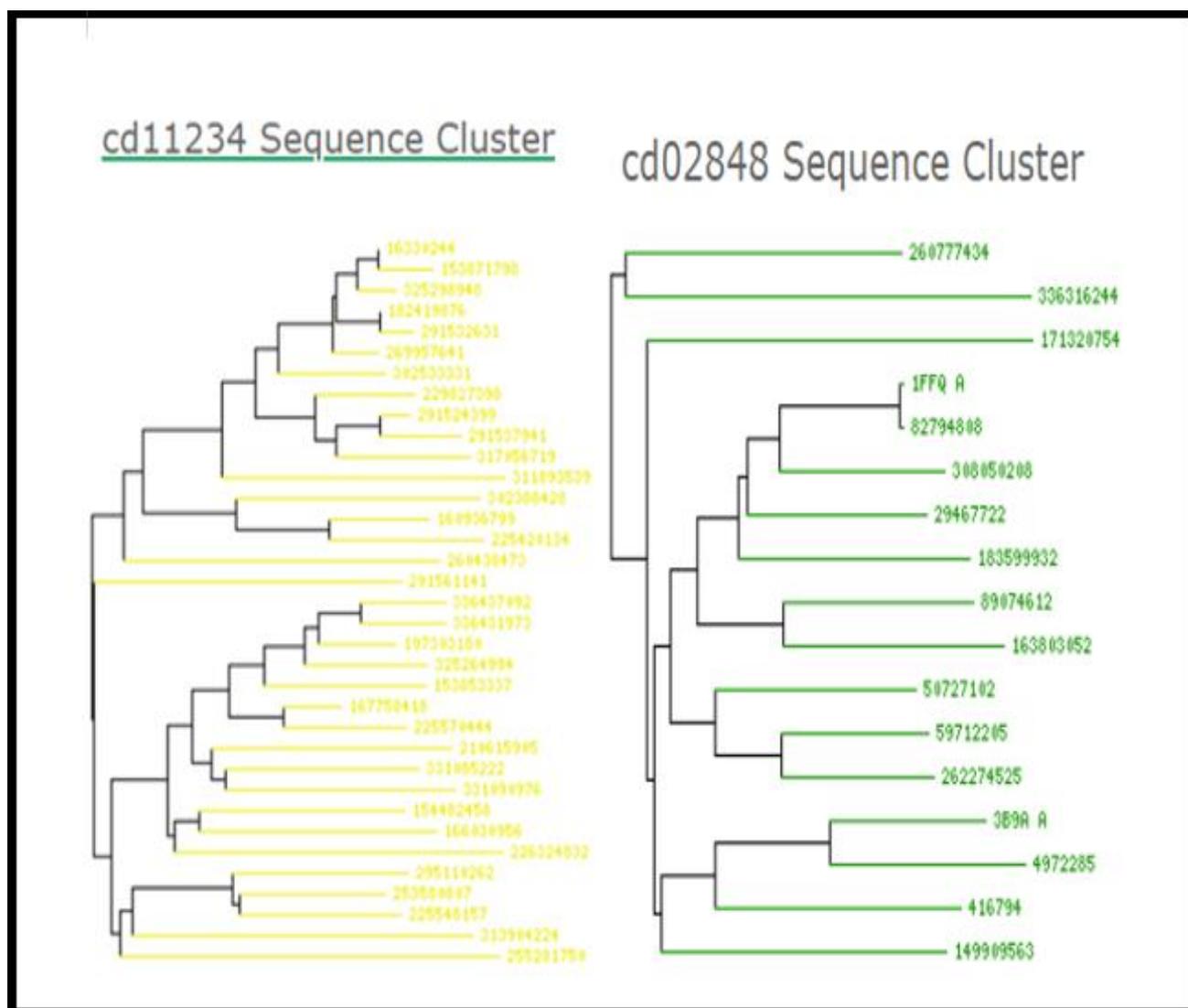


Fig. 4. (a). Shows the related sequences for (cd11234) in a cluster obtained from CDD. (b). Shows the related sequences for (cd2848) in a cluster obtained from CDD.

In this study, we have identified SBE gene promoter from *Oryza sativa* and studied its orthologues, by comparing ancestors using phylogenetic approach. Further, the identification of some cytoplasmic, localized and extracellular proteins is useful to identify expression patterns. Furthermore, the identification of several cis acting regulatory elements sheds light on underlying mechanism of targeted gene expression, development and genomic pathways.

The study of SBE gene promoter sequence revealed its importance as a key player of transcriptional control activities. The presence of several motifs, each linked with specific function revealed the importance of SBE. For instance, the motif CAAT box signals the binding site for RNA transcription factor appears within a conserved consensus sequence on enhancer region. CAT-box functions as a cis-acting regulatory element related to meristem expression, G-box is involved in light responsiveness, Skn-1_motif acts as a regulatory element required for endosperm expression. Similarly, TATA-box acts as a core promoter element around -30 of transcription start and TC-rich repeats as a common cis-

acting element involved in defense and stress responsiveness. Moreover, identification of protein domains within the SBE rice promoter also showed characteristics of some activities associated with N-terminal domain for catalytic activity of chitinase and glycogen debranching enzyme.

Conclusion

To understand the behavior and gene expression patterns, it is important to predict cis acting regulatory elements. Use of high throughput genomic methods have provided great success in identification and analysis of such motif and protein domains. Most of these domains are associated with different types of catalytic domains at any one terminal, either the N-terminal or C-terminal end, involved in different interactions like homodimeric, tetrameric and dodecameric interactions. These findings highlight the usefulness of SBE promoter, as one of the useful promoter for studying active gene regulation, metabolic pathways and expression of genes involved in starch synthesis.

References

- Ball, S.G. and M.K. Morell. 2003. From bacterial glycogen to starch: understanding the biogenesis of the plant starch granule. *Annual Review of Plant Biology*, 54(1): 207-233.
- Bhattacharyya, M.K., A.M. Smith, T.N. Ellis, C. Hedley and C. Martin. 1990. The wrinkled-seed character of pea described by Mendel is caused by a transposon-like insertion in a gene encoding starch-branching enzyme. *Cell*, 60(1): 115-122.
- Burton, R.A., J.D. Bewley, A.M. Smith, M.K. Bhattacharyya, H. Tatge, S. Ring and C. Martin. 1995. Starch branching enzymes belonging to distinct enzyme families are differentially expressed during pea embryo development. *The Plant J.*, 7(1): 3-15.
- Chiera, J.M., R.A. Bouchard, S.L. Dorsey, E. Park, M.T. Buenrostro-Nava, P.P. Ling and J.J. Finer. 2007. Isolation of two highly active soybean (*Glycine max* L. Merr.) promoters and their characterization using a new automated image collection and analysis system. *Plant Cell Rep.*, 26(9): 1501-1509.
- Fisher, D.K., C.D. Boyer and L.C. Hannah. 1993. Starch branching enzyme II from maize endosperm. *Plant Physiol.*, 102 (3): 1045.
- Gao, M., D.K. Fisher, K.N. Kim, J.C. Shannon and M.J. Guiltinan. 1997. Independent genetic control of maize starch-branching enzymes IIa and IIb (isolation and characterization of a Sbe2a cDNA). *Plant Physiol.*, 114(1): 69-78.
- Gilbert, R.G. 2011. Size-separation characterization of starch and glycogen for biosynthesis–structure–property relationships. *Anal. & Bioanal. Chem.*, 399(4): 1425-1438.
- Syahariza, Z.A., S. Sar, J. Hasjim, M.J. Tizzotti and R.G. Gilbert. 2013. The importance of amylose and amylopectin fine structures for starch digestibility in cooked rice grains. *Food Chemistry*, 136(2): 742-749.
- Jani, D., L.S. Meena, Q.M. Rizwan-ul-Haq, Y. Singh, A.K. Sharma and A.K. Tyagi. 2002. Expression of cholera toxin B subunit in transgenic tomato plants. *Transg. Res.*, 11(5): 447-454.
- Jenkins, P.J., R.E. Cameron and A.M. Donald. 1993. A universal feature in the structure of starch granules from different botanical sources. *Starch-Stärke*, 45(12): 417-420.
- Koßmann, J., R.G. Visser, B. Müller-Röber, L. Willmitzer and U. Sonnewald. 1991. Cloning and expression analysis of a potato cDNA that encodes branching enzyme evidence for co-expression of starch biosynthetic genes. *Mol. & General Genetics MGG*, 230(1-2): 39-44.
- Larsson, C.T., P. Hofvander, J. Khoshnoodi, B. Ek, L. Rask and H. Larsson. 1996. Three isoforms of starch synthase and two isoforms of branching enzyme are present in potato tuber starch. *Plant Sci.*, 117(1-2): 9-16.
- Lehmann, U. and F. Robin. 2007. Slowly digestible starch—its structure and health implications: a review. *Trends in Food Sci. & Technol.*, 18(7): 346-355.
- Martin, C. and A.M. Smith. 1995. Starch biosynthesis. *The Plant Cell*, 7(7): 971.
- Masood, A., N. Iqbal, H. Mubeen, R.Z. Naqvi, A. Khatoon and A. Bashir. 2017. Cloning and expression analysis of d-hordein hybrid promoter isolated from barley (*Hordeum vulgare* L.). *Pak. J. Bot.*, 49(3): 1085-1095.
- Mubeen, H., A. Bashir, A. Ameen, A. Masood and S. Raza. 2017. Histochemical GUS expression of beta tubulin promoter in transgenic tobacco. *Afr. J. Biotechnol.*, 16(17): 945-952.
- Naqvi, R.Z., H. Mubeen, A. Masood, A. Khatoon and A. Bashir. 2017. Identification, isolation and evaluation of a constitutive sucrose phosphate synthase gene promoter from tomato. *Pak. J. Bot.*, 49(3): 1105-1112.
- Pei, J., H. Wang, Z. Xia, C. Liu, X. Chen, P. Ma and W. Wang. 2015. Phylogeny and expression pattern of starch branching enzyme family genes in cassava (*Manihot esculenta* Crantz) under diverse environments. *Mol. & Cellular Biochem.*, 406(1-2): 273-284.
- Poulsen, P. and J.D. Kreiberg. 1993. Starch branching enzyme cDNA from *Solanum tuberosum*. *Plant Physiol.*, 102(3): 1053.
- Salehuzzaman, S.N.I.M., R. Jacobsen and R.G.F. Visser. 1992. Cloning, partial sequencing and expression of a cDNA coding for branching enzyme in cassava. *Plant Mol. Biol.*, 20(5): 809-819.
- Smith, A.M., K. Denyer and C. Martin. 1997. The synthesis of the starch granule. *Ann. Review of Plant Biol.*, 48(1): 67-87.
- Sun, C., P. Sathish, S. Ahlandsberg, A. Deiber and C. Jansson. 1997. Identification of four starch-branching enzymes in barley endosperm: partial purification of forms I, IIa and IIb. *The New Phytologist*, 137(2): 215-222.
- Sun, Z. and C.A. Henson. 1991. A quantitative assessment of the importance of barley seed α -amylase, β -amylase, debranching enzyme, and α -glucosidase in starch degradation. *Arch. Biochem. & Biophys.*, 284(2): 298-305.
- Takeda, Y., H.P. Guan and J. Preiss. 1993. Branching of amylose by the branching isoenzymes of maize endosperm. *Carbohydr. Res.*, 240: 253-263.
- Tetlow, I.J. and M.J. Emes. 2014. A review of starch-branching enzymes and their role in amylopectin biosynthesis. *IUBMB life*, 66(8): 546-558.
- Tsuchiya, K., S. Tada, K. Gomi, K. Kitamoto, C. Kumagai and G. Tamura. 1992. Deletion analysis of the Taka-amylase A gene promoter using a homologous transformation system in *Aspergillus oryzae*. *Biosci. Biotechnol. & Biochem.*, 56(11): 1849-1853.
- Tyagi, R., A. Tiwari, V.K. Garg and S. Gupta. 2017. Transcriptome wide identification and characterization of starch 488 branching enzyme in finger millet. *Bioinformatics*, 13: 179-184 DOI 10.6026/97320630013179.
- Wu, A.C., M.K. Morell and R.G. Gilbert. 2013. A parameterized model of amylopectin synthesis provides key insights into the synthesis of granular starch. *PLoS One*, 8(6): e65768.
- Yamanouchi, H. and Y. Nakamura. 1992. Organ specificity of isoforms of starch branching enzyme (Q-enzyme) in rice. *Plant & Cell Physiol.*, 33(7): 985-991.
- Yandeau-Nelson, M.D., L. Laurens, Z. Shi, H. Xia, A.M. Smith M.J. Guiltinan. 2011. Starch-branching enzyme IIa is required for proper diurnal cycling of starch in leaves of maize. *Plant Physiol.*, 156(2): 479-490.

(Received for publication 10 May 2018)