

Analysing Structural diversity of Seed storage protein gene promoters: Buckwheat a case study

Upasna Chettry, Lashaihun Dohtdong and N.K. Chrungoo
Dept. of Botany, North Eastern Hill University
Shillong-793002, Meghalaya, India.

E-mail: nchrungoo@gmail.com

DOI <https://doi.org/10.3986/fag0004>

Received November 20, 2017; accepted August 12, 2018

Keywords: buckwheat, cis-regulatory elements, phylogenetic profiling and nucleosome

ABSTRACT

Multiple sequence alignment of 5'UTR of SSP genes from accessions of *Fagopyrum esculentum* revealed the invariant nature of sequences with the transcription start site at P₇₆₁ and TATA box located -30bp upstream the TSS. Other *cis*-elements identified in the sequences included the legumin box (-581, -524, -184, -135, -91), the -131 prolamin box, DOF element (-718, -649, -540, -432, -272, -225, -128) and CAAT box (-692, -530, -475, -411, -282, -168, -54). Other elements identified included those involved in abscisic acid signalling *viz.*, ABI3 at P_{-470,-95,-68}, RAV1 at P₋₆₉₄ and ₋₅₄₃ and AGL15 at P₋₆₇₁. A comparative analysis of regulatory elements of SSP gene promoters of distantly related species the presence of five *cis*-regulatory elements *viz.* TATA BOX, E-BOX, RY- element, CAAT box and the Endosperm box, which interplay in seed specific SSP gene expression. Other

modulators influencing seed specific gene expression detected in the sequences included the ABA-responsive elements ABI3, RAV1 and AGL15 which play an integral role in seed maturation. Identification of potential nucleosome binding sites in SSP gene promoters of *Cicer arietinum*, *Brassica napus*, *B. campestris*, *Vicia faba*, and *Pisum sativum* at positions 78, 635, 195, 112 and 152 respectively surmises the spatial fine tuning of SSP gene transcriptional regulation in these species. On the other hand, absence of nucleosome binding sites in the promoters of *Fagopyrum esculentum*, *Zea mays*, *Avena sativa*, *Triticum aestivum* and *Oryza sativa* may indicate relatively easier access of transcription factors to the proximal promoter, thereby providing higher level of gene expression.

INTRODUCTION

Seed storage proteins are a major class of proteins that not only serve as a source of nutrition to germinating seedlings but also as an important source of dietary proteins for human consumption. Genes encoding such proteins are under tight spatial and temporal transcriptional control. The basic building blocks for promoters of such genes are regions of *cis*-regulatory DNA, which in eukaryotes often comprises clusters of *cis*-regulatory elements (CREs) that modulate gene expression through their interaction with *trans*-acting factors. Plant *cis*-regulatory motifs are often reported as consensus sequences which are commonly delineated by reporter gene expression assays (Guilfoyle, 1997). Nevertheless, PLACE database, a collection of experimentally characterized plant *cis*-regulatory elements sequences, remains an invaluable resource for annotating motifs discovered in sequences that have not been characterized experimentally (Higo et al., 1998). Majority of contemporary computational approaches for the discovery of *cis*-regulatory elements use the position weight matrix (PWM) motif model, based on the frequencies of nucleotides at each position in a collection of regulatory elements (GuhaThakurta, 2006). The exponential growth in development of bioinformatics tools to discover specific motifs in DNA or protein sequences and creation of genomic resources representing specialized databases of plant *cis*-acting elements have greatly facilitated *in silico* analysis of promoters. However, discovery of such elements is hindered by the variability within their sequences, which typically tolerate

nucleotide substitutions without loss of functionality. Further, in majority of the cases, and especially so in higher eukaryotes, TFs often regulate gene expression by binding to specific elements in the promoter regions of different genes independently or in synergy with other regulatory proteins. Different *cis*-elements of a given promoter are also known to interact with different parts of *cis*-regulatory module (CRM), where the relative positions of *cis*-elements and the distances between them are crucial. an overall regulatory complex (Arnone and Davidson, 1997).

Detailed analysis of expression of genes coding for seed storage proteins has revealed that the expression of SSP genes and accumulation of the proteins is limited to the endosperm/ embryos or cotyledons of the seeds (Perez-Grau and Goldberg, 1989; Fujino et al., 2001; Milisavljevic et al, 2004; Jain, 2004). Seed-specific expression has been shown to be conferred by the promoter regions of various storage protein genes (Devic et al, 1996; Lee et al, 2007; Moreno-Risueno et al, 2008). Signature *cis*-elements identified in the promoters of specific class of plant genes include the "legumin box" comprising of the core "RY motif " having the sequence 5'CATGCA3' (Baumlein et al., 1986, Dickinson et al, 1988, Forde et al, 1985) and the "vicilin box" having the core sequence 5'GCCACCTCAT3' in legumes (Vicente et al., 1997; Weschke et al., 1988) and the "prolamin box" (5'TGTAAAG3') or endosperm motif (E-motif) in cereals (Vicente et al., 1997; Shewry and Halford, 2003.). The promoter region of prolamin genes comprises of three CREs including the

GCN4-like (GLM) element (5'GRTGAGTCAT3'), the prolamin-box (5'TGTAAAGT3') and the AACA (5'AACAAACTCTATC3') element that respectively interact with bZIP, DOF and MYB family transcription factors (Fauteux and Strömviik, 2009). A comprehensive analysis of the *napA* gene promoter in rapeseed (*Brassica napus* L.) has revealed the presence of two regulatory complexes which include the B-box, that contains the distB element (5'GCCACTTGTC3') together with the proxB element (5'TCAAACACC3'), and the RY/G complex which contains two RY repeats (5'CATGCA3') and one G-box (5'CACGTG3') (Ezcurra et al., 1999; Chandrasekharan et al., 2003). G-box, CCAAAT box, E-box (5'CACCGT3') and RY elements have been demonstrated to have a strong role in mediating gene expression in embryos (Lindstrom et al., 1990). Motifs conferring seed-specific expression are known to lie in the proximal region of the promoter, often within 500

MATERIALS AND METHODS

Nucleotide sequences of the 5'UTR of legumin like seed storage proteins of ten accessions of common buckwheat viz. IC-107090, IC-107285, IC-107265, IC-108517, IC-79192, IC-16550, IC-188669, IC-324313, IC-18864 and IC-363973 were generated by nucleotide sequencing of the relevant amplicons. Nucleotide sequences of the promoter regions of seed storage protein genes of other distantly related species were retrieved from Genbank database of NCBI for comparative

bp upstream of the transcriptional start (Wu et al., 2000; Fujimori S et al., 2005).

Although the availability of *cis*-acting regulatory element database and tools of bioinformatics help to predict the transcriptional properties of new entry sequences with considerable accuracy, understanding the structural features of DNA, such as GC skew, bendability, topography, free energy, curvature and nucleosome positioning gives a better understanding of the regulatory landscape of such genes (Florquin et al., 2005; Kanhere and Bansal 2005b). The present study describes the profiling of the 5'UTR of legumin-like seed storage protein gene in ten accessions of common buckwheat *vis. a vis.* seed storage protein gene promoters from distantly related species. This uncovered the presence of specific conserved motifs in SSP gene promoters across plant species and moderate nucleosome binding potential in 5'UTR of buckwheat legumin genes.

analyses. The accession numbers of the sequences retrieved from Genbank data bases included EU595873 of *Fagopyrum esculentum*, AF420598 of *Brassica napus*, X67833.1 of *B. juncea*, Y13108 of *B. campestris*, Y13166 of *Cicer arietinum*, S60289.1 of *Vicia faba*, X02983.1 of *Pisum sativum*, X65064.1 of *Hordeum vulgare*, X65064.1 of *Oryza sativa*, EU189096.1 of *Triticum aestivum* and JQ241267 of *Zea mays*.

Sequence analysis

BLASTn analysis of the nucleotide sequences of the 5'UTR of legumin like seed storage proteins of ten accessions of common buckwheat was carried out using the BLAST tool of NCBI. The sequences were aligned using the multiple alignment tool MULTALIN (<http://multalin.toulouse.inra.fr/multalin/>). Distribution of *cis*-elements within the sequences

Nucleosome formation potential

Comparative analysis on the nucleosome formation potential of the representative sequences from each species was performed with Strong Nucleosome tool (<http://strn-nuc.haifa.ac.il:8080/mapping/home.jsf>).

Result and discussion

Profiling of the 5'UTR of buckwheat legumin gene

BLAST analysis of nucleotide sequences from all the accessions revealed more than 98% homology with 5'UTR of sequence bearing accession no. EU595873, the gene coding for legumin like seed storage protein gene of common buckwheat. Alignment of the sequences using MULTALIN clearly showed a highly conserved nature of the sequences (Fig. 1). Promoter prediction tool (Neural Network Promoter Prediction) identified three probable promoter regions between P'₃₉₂₋₄₄₂, 473-523 and 721-771 in the sequences. Out of the three predicted transcription start sites, the TSS at P'₇₆₁ was located closest to the predicted ATG start codon at P'₈₀₁. The TSS at P'₇₆₁ also followed the YR rule (C⁻¹A⁺¹), having the pyrimidine 'C' at -1 and the purine 'A' at +1 position (Yamamoto et al. 2007).

was identified out by PLACE (<http://www.dna.affrc.go.jp/PLACE/signalscan.html>) and AtPAN (<http://atpan.itps.ncku.edu.tw>). Neural Network Promoter Prediction tool (http://www.fruitfly.org/seq_tools/promoter.html) was used to identify the transcriptional start site in the target sequences.

Sequences with statistical value of scoring peaks between 50-65 were used to determine the potential position of nucleosome along the DNA sequence.

Considering 'A' at position 761 (+1) as the predicted TSS and ATG at position 801 (+40) as the initiating codon, the TATA at position 731⁽⁻⁶²⁾ was identified as the TATA box of the promoter. Apart from TATA box, the sequences revealed several other *cis*-elements, that are involved in the regulation of eukaryotic gene expression in general and seed-specific expression in particular. The transcription start site predicted for the sequences of all the accessions followed the YR rule with the TATA box motif being localized at P'₋₃₀ relative to the TSS. Alignment of the context sequences around TATA, TSS and ATG-start codon of buckwheat seed storage protein gene with the corresponding regions of seed storage protein genes from other accessions clearly

established the high degree of conservation in spacing between these elements. Sequence analysis identified 3 legumin boxes comprised of the core sequence 5'CATGCA3' at P[']_{-470, -95, and -68}, a single prolamin box, comprising of the sequence 5'TGTAAAG3' at P[']₋₁₃₁ and 7 DOF motifs with the core sequence 5'AAAG3' at P[']_{-718, -649, -540, -432, -272, -225, and -128} with respect to TSS. Legumin box is considered to be the key element in regulating seed specific expression of genes coding for legumin type proteins (Bäumlein et al., 1992; Ellerström et al., 1996; Reidt et al., 2000). Destruction of the legumin box by a 6 bp deletion in an otherwise intact 2.4 kb 5'-noncoding upstream sequence of *Vicia faba* legumin gene LeB4 was shown to drastically reduce LeB4 expression in seeds (Baumlein et al., 1992). Similar observations were made by Ezcurra et al. (1999) for RY elements in the promoter region of napin gene. Baumlein et al. (1992) has shown that the enhancer-like *cis*-elements in 5'UTR were fully functional only in conjunction with the core motif 5'CATGCATG3' of the legumin box, thereby indicating a possible role of legumin box in modulating enhancer activity in promoter of SSP genes. RY motif has been shown to interact with the conservative B3-domain of the transcriptional activators VP1 of maize (McCarty, 1995) and *fus3* and *abi3* proteins of *Arabidopsis* (Ezcurra et al., 1999; Reidt et al., 2000). Analysis of several other seed specific promoters has confirmed the importance of RY elements for quantitative expression of seed specific genes as well as the potential of this motif in repression of gene expression in non-seed tissues (Mönke et al., 2004; Singh, 1998)

While the “P-box” (5'TGTAAAG3') is a -300 enhancer element present in SSP genes of cereals and several other dicots (Vickers et al., 2006), we detected the “P- box” as a -131 element in the buckwheat legumin gene promoter. This element has also been reported to be involved in quantitative regulation of gene expression in seeds (Wu et al., 2000; Chandrasekharan et al., 2003). In many cases the “P-box” and “GCN4” motifs are coupled with each other with only a few nucleotides separating them. This module has been named as “bifactorial endosperm box”. The CAAT box, noted as an enhancer element involved in quantitative regulation of gene expression, was located at positions -692,-530,-475,-411,-282,-168, and -54. Sequence analysis also revealed the presence of SEF1 binding motif, having the core motif 5'ATATTTATA3' at P[']₋₃₀₇. Lessard et al. (1990) have demonstrated strong interaction between SEF1 and A-T rich sequences present far upstream in genes coding for α' and β subunits of β -conglycinin. They suggested SEF1 recognizes its binding site with greater affinity than the other SEF factors and it may be involved in directing nucleosome phasing within the promoter region, analogous to mammalian high mobility group chromosomal proteins (HMG-1). Zhou et al. (2014) have reported that while deletion of SEF3 and SEF4 binding motifs from promoter of seed-specific allergen gene of *Arachis hypogaea* did not affect promoter activity, deletion of three E-boxes and one SEF1 motif caused a marked decrease in promoter activity. Their results suggest the possibility of a role for E-box and SEF1 binding motifs in regulating seed-specific expression of genes.

Seed storage protein gene promoters exhibit signature motifs

Comparative analysis using AtPan software generated the co-occurrence of *cis*-motifs in the promoters of *Cicer arietinum*, *Brassica.napus*, *B.campestris*, *Vicia faba*, *Pisum sativum*, *Fagopyrum esculentum*, *Zea mays*, *Avena sativa*, *Triticum aestivum* and *Oryza sativa* (Table 1). The overall consensus generated from the SSP gene promoters investigated in the present study can be broadly divided into five conserved composite motifs (CREs). These included CRE1, which is 20 to 30 bases upstream of the TSS. This element has been reported universally from all promoters (Joshi, 1987). CRE2 included a G-box-like and a CAAT motif, nested into an E-box (5'CANNTG3'). CRE3 comprised of the RY element (5'CATGCA3') with core motif CATG. This motif is known to be highly conserved in seed specific promoters of both dicots and monocots (Dickinson et al., 1988). CRE4 comprised of the CAAT box which has been suggested to act as an enhancer element involved in quantitative regulation of gene expression (Schirm et al, 1987; Wu et al., 2000). CRE5 included P-box or the endosperm box "5'TGTAAAG3' that interacts with the transcription factor DOF which plays a key role in activating the expression of prolamin genes in cereals. Forde et al. (1985) has suggested the presence of at least

two types of controls operating on prolamin gene expression. While one was responsible for coordinating induction of genes during endosperm development, the other regulated subsequent rates of prolamin accumulation. It was suggested that these two controls have the ability to act differentially on subsets of prolamin genes. The two control systems were together named as the endosperm box which is a bipartite motif consisting of the prolamin box and the GCN4 like motif. The GCN4 motif has been reported to be a target of basic leucine zipper transcription factor that belongs to maize Opaque-2 (O2)-like protein family which is also known as *RISBZ* in rice. Yamamoto et al. (2006) have demonstrated that the prolamin binding factor transactivated several storage protein genes via an AAAG target sequence located within the promoters of such genes. They observed a synergism between RPBF and *RISBZ1* in recognizing the GCN4 motif (TGA(G/C)TCA) for inducing expression of SSP genes. It was suggested that RPBF gene, which predominately expressed in maturing endosperm and coordinately expressed with seed storage protein genes, was involved in quantitative regulation of genes expressed in the endosperm in cooperation with *RISBZ1*.

Nucleosome Mapping determines potential nucleosome binding site for SSP promoters

Variations in position of different *cis*-elements in the promoter sequences are expected to affect gene expression either through their interaction with transcription factors or through differences in nucleosome favouring and/or nucleosome excluding sequences (Tirosh et al.,2008). Besides the knowledge of *cis*-acting regulatory elements for prediction of transcriptional control of genes, information about structural features of DNA, such as GC skew, bendability, topography, free energy, curvature and nucleosome positioning would give a better understanding of the regulatory landscape. Therefore mapping of promoters for potential nucleosome binding sites would generate a deeper insight into long range interactions that may not be evident from sequence variations alone. Nucleosome positioning demarcates the promoter region and transcription start site. While promoters which confer ubiquitous gene expression are essentially free of nucleosomes, Levitsky et.al (2001) have suggested that promoters conferring tissue specific expression of genes display higher nucleosome formation potential. Nucleosome positioning map of each seed storage gene promoter, generated using Strong Nucleosome Mapping tool, revealed the highest scoring peak between 45-51. While a a scoring peak > 65 is considered to be statistically significant in determining potential position of the nucleosome along the DNA sequence, values between 50-60

indicate a moderate affinity towards accommodating a nucleosome. This is due to involvement of determinants such as CpG islands or epigenetic regulation in modulating nucleosome positioning. With a scoring peak of >50, the promoters of seed storage proteins of *C. arietinum*, *B. napus*, *B. campestris*, *V. faba*, and *P. sativum* showed potential nucleosome binding sites at positions P'_{78, 635, 195, 112 and 152} respectively. The sites were located at positions -100 to -300 with respect to the TSS. On the other hand, analysis of nucleotide sequences of 5'UTR of genes coding for seed storage proteins in *Fagopyrum esculentum*, *Zea mays*, *Avena sativa*, *Triticum aestivum* and *Oryza sativa* by Strong Nucleosome tool revealed a score of <45, thereby indicating absence of nucleosome binding sites in the promoters of SSP genes of these crops. Jiang and Pugh (2009) has suggested that in comparison to the transcribed region of DNA, the UTR was essentially free of nucleosome binding potential. Our results indicate that compared to the promoters of genes coding for SSPs in *C. arietinum*, *B. napus*, *B. campestris*, *V. faba*, and *P. sativum*, those of *Fagopyrum esculentum*, *Zea mays*, *Avena sativa*, *Triticum aestivum* and *Oryza sativa* have lower accessibility to nucleosome, thereby ensuring easy access to transcription factors and a consequent higher level of gene expression.

ACKNOWLEDGMENTS:

Financial support received from Department of Biotechnology, Govt. of India, New Delhi for undertaking the work under DBT Biotech Hub project vide grant no. BT/04/NE/2009 to NKC and

INSPIRE fellowship from Department of Science & Technology, Govt. of India to UC is gratefully acknowledged.

References

- Arnone, M.I. and Davidson, E.H., 1997. The hardwiring of development: organization and function of genomic regulatory systems. *Development*, 124(10): 1851-1864.
- Baud, S., Dubreucq, B., Miquel, M., Rochat, C. and Lepiniec, L., 2008. Storage reserve accumulation in *Arabidopsis*: metabolic and developmental control of seed filling. *The Arabidopsis Book*.: e0113.
- Bäumlein, H., Wobus, U., Pustell, J. and Kafatos, F.C., 1986. The legumin gene family: structure of a B type gene of *Vicia faba* and a possible legumin gene specific regulatory element. *Nucleic Acids Res.* 14: 2707-2720.
- Chandrasekharan, M.B., Bishop, K.J. and Hall, T.C., 2003. Module-specific regulation of the β -phaseolin promoter during embryogenesis. *The Plant J.* 33(5): 853-866.
- Devic, M., Albert, S. and Delseny, M., 1996. Induction and expression of seed specific promoters in *Arabidopsis* embryo-defective mutants. *Plant J.* 9: 205-215.
- Dickinson C.D., Evans R.P. and Nielsen N.C., 1988. RY repeats are conserved in the 5'-flanking regions of legume seed protein genes. *Nucleic Acids Res.* 16: 371.
- Ellerström, M., Ståhlberg, K., Ezcurra, I. and Rask, L., 1996. Functional dissection of a napin gene promoter: identification of promoter elements required for embryo and endosperm-specific transcription. *Plant Mol. Biol.* 32(6): 1019-1027.
- Ezcurra, I., Ellerström, M., Wycliffe, P., Ståhlberg, K. and Rask, L., 1999. Interaction between composite elements in the napA promoter: both the B-box ABA-responsive complex and the RY/G complex are necessary for seed-specific expression. *Plant Mol. Biol.* 40(4): 699-709.
- Fauteux, F. and Strömvik, M.V., 2009. Seed storage protein gene promoters contain conserved DNA motifs in Brassicaceae, Fabaceae and Poaceae. *BMC Plant Biol.* 9(1): 126.
- Florquin, K., Saeys, Y., Degroeve, S., Rouze, P. and Van de Peer, Y., 2005. Large-scale structural analysis of the core promoter in mammalian and plant genomes. *Nucleic Acids Res* 33: 4255–4264.
- Forde, B.G., Heyworth, A., Pywell, J. and Kreis, M., 1985. Nucleotide sequence of a B1 hordein gene and the identification of possible upstream regulatory elements in endosperm storage protein genes from barley, wheat and maize. *Nucl. Acids Res.* 13: 7327-7337.
- Fujino, K., Funatsuki, H., Inada, M., Shimono, Y. and Kikuta, Y., 2001. Expression, cloning, and immunological analysis of buckwheat (*Fagopyrum esculentum* Moench) seed storage proteins.
- GuhaThakurta, D., 2006. Computational identification of transcriptional regulatory elements in DNA sequence. *Nucleic Acids Res.* 34(12): 3585-3598.
- Guilfoyle, T.J., 1997. The structure of plant gene promoters. *Genet Eng Springer US.* 15-47

- Gutierrez, L., Van Wuytswinkel, O., Castelain, M. and Bellini, C., 2007. Combined networks regulating seed maturation. *Trends Plant Sci.* 12(7): 294-300.
- Higo, K., Ugawa, Y., Iwamoto, M., Higo, H., 1998. PLACE: a database of plant cis-acting regulatory DNA elements. *Nucleic Acids Res.* 26: 358-359.
- Jain, M., Tyagi, A.K. and Khurana, J.P., 2006. Molecular characterization and differential expression of cytokinin-responsive type-A response regulators in rice (*Oryza sativa*). *BMC Plant Biol.* 6(1): 1.
- Jiang, C. and Pugh, B.F., 2009. Nucleosome positioning and gene regulation: advances through genomics. *Nature Rev Genet.* 10.3 161-172.
- Joshi, C.P., 1987. An inspection of the domain between putative TATA box and translation start site in 79 plant genes. *Nucleic Acids Res.* 15(16): 6643-6653.
- Kanhere, A. and Bansal, M., 2005b. Structural properties of promoters: similarities and differences between prokaryotes and eukaryotes. *Nucleic Acids Res.* 33: 3165–3175.
- Lee, S.K., Hwang, S.K., Han, M., Eom, J.S., Kang, H.G., Han, Y., Choi, S.B., Cho, M.H., Bhoo, S.H., An, G. and Hahn, T.R., 2007. Identification of the ADP-glucose pyrophosphorylase isoforms essential for starch synthesis in the leaf and seed endosperm of rice (*Oryza sativa* L.). *Plant Mol. Biol.* 65(4): 531-546.
- Lessard, P.A., Allen, R.D., Bernier, F., Crispino, J.D., Fujiwara, T. and Beachy, R.N., 1991. Multiple nuclear factors interact with upstream sequences of differentially regulated β -conglycinin genes. *Plant Mol. Biol.* 16(3): 397-413.
- Levitsky, V.G., Podkolodnaya, O.A., Kolchanov, N.A. and Podkolodny, N.L., 2001. Nucleosome formation potential of eukaryotic DNA: calculation and promoters analysis. *Bioinformatics.* 17(11): 998-1010.
- Lindstrom, J.T., Vodkin, L.O., Harding, R.W. and Goeken, R.M., 1990. Expression of soybean lectin gene deletions in tobacco. *Dev Genet.* 11(2): 160-167.
- McCarty, D.R., 1995. Genetic control and integration of maturation and germination pathways in seed development. *Annu. Rev. Plant Biol.* 46(1): 71-93.
- Milislavljević, M.D., Timotijević, G.S., Radović, S.R., Brklja. ić, J.M., Konstantinović, M.M. and Maksimović, V.R., 2004. Vicilin-like storage globulin from buckwheat (*Fagopyrum esculentum* Moench) seeds. *J. Agric. Food Chem.* 52(16): 5258-5262.
- Mönke, G., Altschmied, L., Tewes, A., Reidt, W., Mock, H.P., Bäumlein, H. and Conrad, U., 2004. Seed-specific transcription factors ABI3 and FUS3: molecular interaction with DNA. *Planta*, 219(1): 158-166.
- Moreno-Risueno, M.Á., Gonzalez, N., Díaz, I., Parcy, F., Carbonero, P. and Vicente-Carbajosa, J., 2008. FUSCA3 from barley unveils a common transcriptional regulation of seed-specific genes between cereals and *Arabidopsis*. *Plant J.* 53: 882-894
- Perez-Grau, L. and Goldberg, R.B., 1989. Soybean seed protein genes are regulated spatially during embryogenesis. *The Plant Cell*, 1(11): 1095-1109.
- Reidt, W., Wohlfarth, T., Ellerström, M., Czihal, A., Tewes, A., Ezcurra, I., Rask, L. and Bäumlein, H., 2000. Gene regulation during late embryogenesis: the RY motif of maturation-specific gene promoters is a direct target of the FUS3 gene product. *The Plant J.* 21(5):401-408.
- Schirm, S., Jiricny, J. and Schaffner, W., 1987. The SV40 enhancer can be dissected into multiple segments, each with a different cell type specificity. *Genes Dev.* 1(1): 65-74.

- Shewry, P.R. and Halford, N.G., 2002. Cereal seed storage proteins: structures, properties and role in grain utilization. *J. Exp. Bot.* 53(370): 947-958.
- Singh, K. B., 1998. Transcriptional regulation in plants: the importance of combinatorial control. *Plant Physiol.* 118(4): 1111-1120
- Tirosh, I. and Naama B., 2008. Two strategies for gene regulation by promoter nucleosomes. *Genome Res.* 18(7): 1084-1091.
- Vicente-Carbajosa, J., Moose, S.P., Parsons, R.L. and Schmidt, R.J., 1997. A maize zinc-finger protein binds the prolamin box in zein gene promoters and interacts with the basic leucine zipper transcriptional activator Opaque2. *Proc. Natl. Acad. Sci.* 94(14): 7685-7690.
- Vickers, C.E., Xue, G. and Gresshoff, P.M., 2006. A novel cis-acting element, ESP, contributes to high-level endosperm-specific expression in an oat globulin promoter. *Plant molecular biology*, 62(1), pp.195-214.
- Weschke, W., Bassüner, R., Van Hai, N., Czihal, A., Bäumlein, H. and Wobus, U., 1988. The structure of a *Vicia faba* vicilin gene. *Biochem. Physiol. Pflanz.* 183(2-3): 233-242.
- Wu, C.Y., Washida, H., Onodera, Y., Harada, K. and Takaiwa, F., 2000. Quantitative nature of the prolamin-box, ACGT and AACA motifs in a rice glutelin gene promoter: minimal cis-element requirements for endosperm-specific gene expression. *The Plant J.*, 23(3): 415-421.
- Yamamoto, M.P., Onodera, Y., Touno, S.M. and Takaiwa, F., 2006. Synergism between RPBF Dof and RISBZ1 bZIP activators in the regulation of rice seed expression genes. *Plant Physiol.* 141(4):1694-1707.
- Yamamoto, Y.Y., Ichida, H., Matsui, M., Obokata, J., Sakurai, T., Satou, M., Seki, M., Shinozaki, K. and Abe, T., 2007. Identification of plant promoter constituents by analysis of local distribution of short sequences. *BMC Genomics*, 8(1): 67
- Zhou, Y., Yang, P., Zhang, F., Luo, X. and Xie, J., 2014. Analysis of promoter activity of the Peanut (*Arachis hypogaea* L.) seed-specific allergen gene Ara h 2.02 in transgenic *Arabidopsis*. *Bothalia Journal*, 44(12):80-97

Nucleosome binding potential of SSP gene promoters

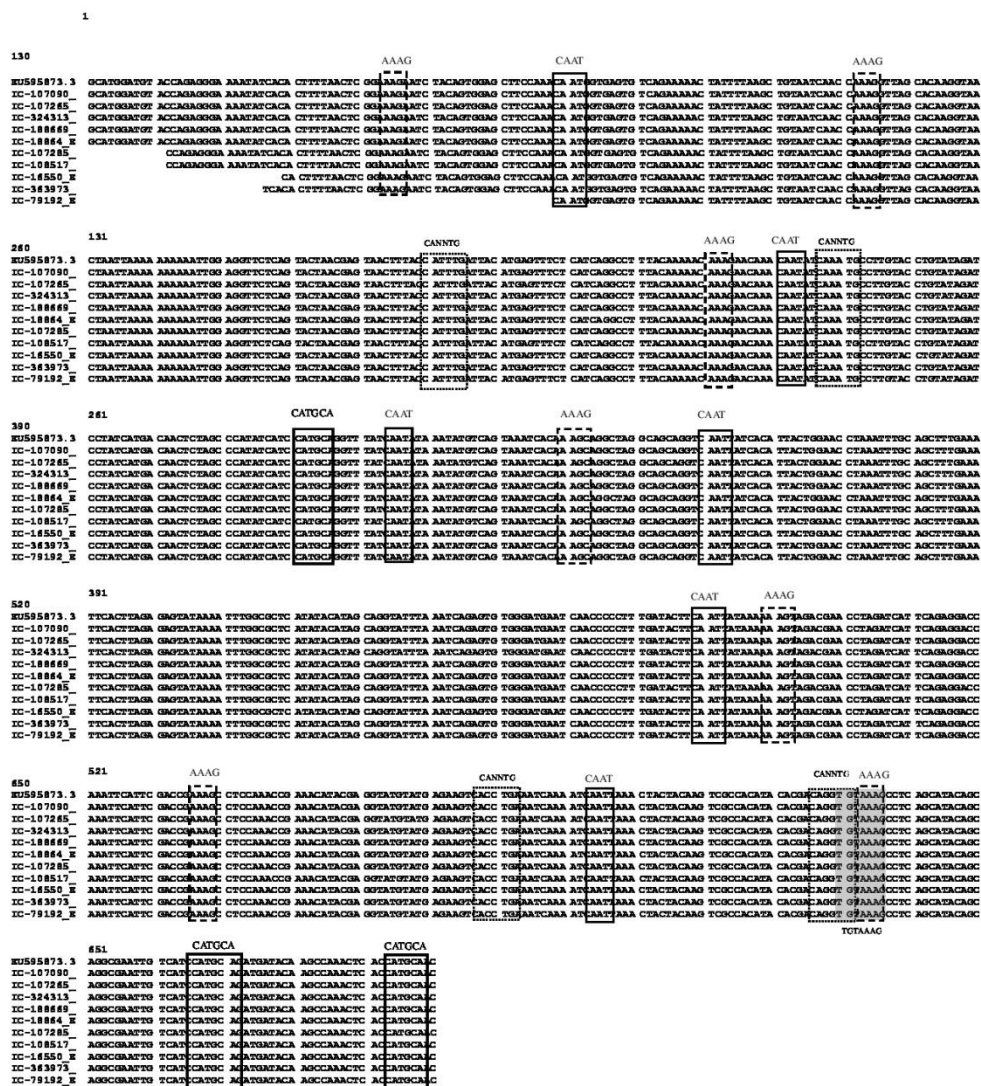


Fig 1: Multiple sequence alignment of 5'UTR of seed storage protein gene from ten accessions of *Fagopyrum esculentum* (Moench)

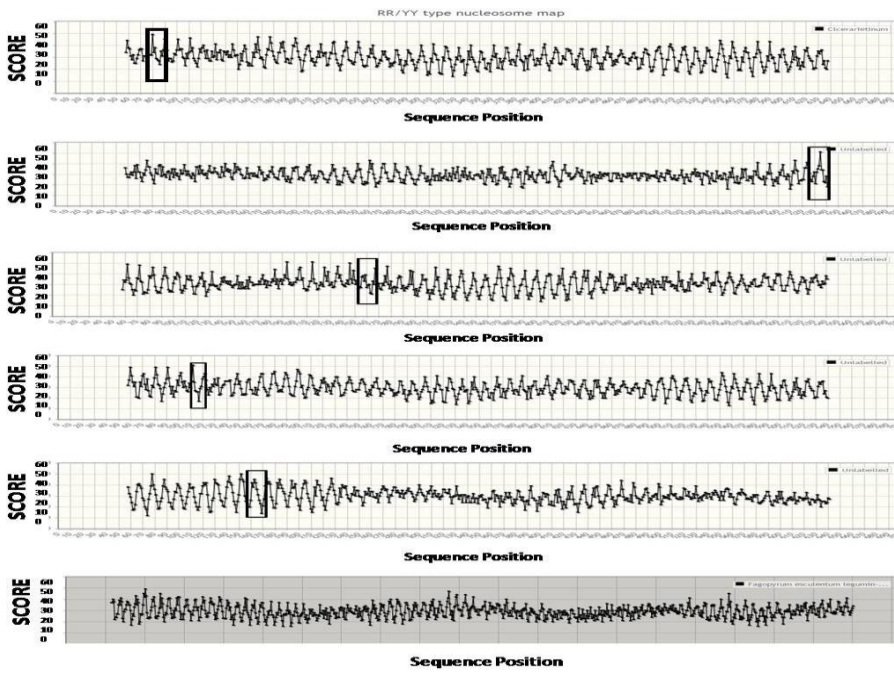


Fig. 2: Nucleosome positioning map highlighting the scoring peak at the probable nucleosome binding site in nucleotide sequences of 5'UTR of legumin genes of (A) *C. arietinum* [acc. no. Y13166] ; (B) *B. napus* [acc. no. X67833.1], (C) *B. campestris* [acc. no. Y13108], (D) *V. faba* [acc. no. X02983.1], and (E) *P. sativum* [acc. no. X02983.1]. A representative image of non existence of nucleosome binding site in *Fagopyrum esculentum*, *Zea mays* [acc. no. JQ241267], *Avena sativa* [acc. no. EU595873], *Triticum aestivum* [acc. no. (EU189096.1] and *Oryza sativa* [acc. no. (X65064.1] is given in (F). The nucleosome binding site for the respective accessions is boxed.

Table1 :Tabular representation of the significant regulatory elements present in SSP promoter across different species.

SITE	MO TIF	Oryza sativa	Zea mays	Hordeum vulgare	Triticum aestivum	Pisicum sativum	Vicia Faba	Cicer arrietum	Brassica napus	Brassica campestris	Brassica juncea	Fagopyrum esculentum	SIGNIFICANCE
PBF	TGTAAG	-34,-156,-161,-224,-320,-345,-454,-466,-476	-66,-136,-158,-173,-315,-540,-571	-45,-53,-65,-164,-273,-374	-50,-58,-69,-145,-169,-271,-332,-336	-124,-236,-260,-294,-308,-430,-480	--157,-259,-285,-438,-452,-576,-911,-927,-993,-1135,-1143,-1159	-155,-289,-369,-429,-786,-903,-1052,-1203,-1224,-1539,-2050,-2117,-2192	-1064,-225	--161,-259,-327,-611,-659,-726,-827,-1056,-1139,-1319,-1386	-225,-1068	-131	Core site required for binding of Dof proteins in maize
E-BOX	CANN TG	--391,-409	-804,-503,-263,-517	-355,-379	-182	-407,-824,-1156,-1124,-1251,-1285	-58,-126,-567,-849,-901	-517,-635,-747,1060,-1344,-2226	-912,-887,-594,-523	-107,-360,-560,631,-920,-1143,-1299,-1336	-58,-80,-138,-598,-651,-841,-917	-581,-524,-184,-135,-91	E-box of napA storage-protein gene of Brassica napus
CAATBOX I	CAAT	-10,-247,-402,-411	-196,-329,-415,-425,-504,-547,-870,-880,-940	-72,-109,-147	-77,-152	-62,-83,-493,-649,-715,-798,-904,-979,-1109	-370,-549,-676,-757,-938,-1104,-1130	-198,-473,-541,-951,-1010,-1100,-1613,-2023,-2122	-926,-829,-715,-345,-86	371,-902,-913,-1095,-1258,-1409	-46,-347,-405,-720,-827,-897,-924	-692,-530,-457,-411,-282,-168,54	"CAAT promoter consensus sequence" found in legA gene of pea;
DOF	AAAG	-34,-156,-161,-224,-320,-345,-454,-466,-476	-66,-135,-158,-174,-258,-315,-540,-572	-45,-53,-65,-164,-273,-374	169,-271,-332,-336	-124,-236,-260,-294,-308,-430,-480	--157,-259,-285,-438,-452,-576,-911,-927,-993,-1135,-1143,-1159	-155,-289,-369,-429,-786,-903,-1052,-1203,-1224,-1539,-2050,-2117,-2192	-225,-1064	--161,-259,-327,-611,-659,-726,-827,-1056,-1139,-1319,-1386	-225,-1068	-131	Core site required for binding of Dof proteins in maize
TATABOX X	TTATTT	-23	-23	-20	-24	23	-25	-23	-24	-21	-24	-30	TATA box elements are critical for accurate initiation
OPAQUE-2	TGAGTCA	-210											GNC4 motif is the recognition site for Opaque-2 (O2)-like proteins
RISBZI	TGAGTCA	-210											Required for the expression of GNC4 motif
SEF1						-1136	-803					-307	SEF1 binding motif; sequence found upstream region of soybean β-conglucinin gene