FUNCTIONAL CHARACTERIZATION OF AN ENDOSPERM SPECIFIC PROMOTER *p1062* FROM COMMON BUCKWHEAT (*FAGOPYRUM ESCULENTUM* MOENCH) FOR DRIVING TISSUE SPECIFIC GENE EXPRESSION

FUNKCIJSKE LASTNOSTI ENDOSPERMSKEGA PROMOTORJA *p1062* NAVADNE AJDE (*FAGOPYRUM ESCULENTUM* MOENCH) ZA OMOGOČANJE TKIVNO SPECIFIČNE EKSPRESIJE GENOV

Lashaihun DOHTDONG[1] & Nikhil Kumar CHRUNGOO[1]

http://dx.doi.org/10.3986/fbg0066

ABSTRACT

Functional characterization of an endosperm specific promoter *p1062* from common buckwheat (*Fagopyrum esculentum* Moench) for driving tissue specific gene expression

Seed storage proteins of grain crops meet the major dietary protein requirement of over half of the world population. PCR based genome walking the 5'UTR of the gene coding for a lysine rich legumin type protein amplified a 1.1kb DNA fragment representing the promoter region of the gene. Clustal alignment of this sequence with other sequences in the Genbank database clearly showed 100 percent complementary base match of 282 bases at the 3' end of the sequence, corresponding to nucleotide position 780-1062 with correspondingly similar number of bases on the 5' end of the 1.7kb *Bwleg* gene.We detected one prolamin box and three RY-repeat motifs in the sequence. Seven deletion fragments of the putative promoter were generated by 5' nested PCR and cloned in pCAMBIA1304 upstream of *GUS* gene after excising the CaMV 35S promoter from the vector. *Arabidopsis* plants plants harbouring the deletion construct *pBwlDF1* to *pBwlDF6* clearly showed seed specific expression of the reporter gene. Seeds harbouring the constructs *pBwlDF3*, *pBwlDF4* and *pBwlDF5* showed a nearly threefold decrease in GUS activity than those harbouring the construct with full length promoter.

Key words: buckwheat, DNA, promoter, constructs

IZVLEČEK

Funkcijske lastnosti endospermskega promotorja *p1062* navadne ajde (*Fagopyrum esculentum* Moench) za omogočanje tkivno specifične ekspresije genov

Založne beljakovine semen zrnastih poljščin ustrezajo glavnim potrebam po beljakovinah za več kot polovico svetovnega prebivalstva. S PCR in 5'UTR so za kodiranje kakovostnih beljakovin leguminskega tipa pomnožili odlomek 1,1 kb DNK, ki je promotorsko gensko območje. Vzporejanje te sekvence z drugimi sekvencami podatkovne baze genske banke jasno pokaže popolno komplementarnost 282 baz na 3' koncu sekvence, kar ustreza pozicijam 780-1062 z ustreznim številom baz na 5' koncu gena 1,7 kb *Bwleg.* V sekvenci smo odkrili eno prolaminsko škatljo in tri RY-ponovljene motive. Sedem delecijskih fragmentov putativnega promotorja smo generirali z 5' PCR kloniranjem pCAMBIA1304 navzgor od *GUS* gena po izločitvi promotorja CaMV 35S iz vektorja. Semena s konstrukti *pBwlDF3*, *pBwlDF4* in *pBwlDF5* so izražali skoraj trikratno zmanjšanje GUS aktivnosti v primerjavi s konstrukti, ki so vsebovali polne dolžine promotorjev.

Ključne besede: ajda, DNK, promotor, konstrukti

[1] Department of Botany, North-Eastern Hill University, Shillong, India; lashaihund24@gmail.com, nchrungoo@gmail.com

INTRODUCTION

Plant genetic resources, representing the entire generative and vegetative reproductive material of species with economic and /or social value for the agriculture of the present and the future, with special emphasis on nutritional plants are the most important gifts of nature to mankind. Of the total available genetic diversity, mankind has utilized only a few plants as major food sources with cereal grains and legume seeds being the major sources of vegetarian dietary proteins for human consumption. However, the nutritional quality of the proteins in both these crops do not match the WHO standards. While the major amino acid deficiency in legume seed proteins is their low content of sulphur containing amino acids cysteine and methionine, cereal proteins have low levels of lysine (Boulter 1981; Shotwell & Larkins 1989).

Strategies developed over the years to improve the level of essential amino acids in seed storage proteins of important crop plants through conventional breeding. However, in most cases these attempts have either led to a severe depletion in seed storage protein levels or abnormalities in seed development. The negative correlation between the seed protein content and the level of essential amino acids per unit protein has, therefore, come as a major handicap in improving the amino acid composition of seed proteins in crops. Due to such limitations in conventional breeding methodologies, molecular approaches have provided alternate strategies to conventional breeding programmes aimed at compensation of amino acid deficiencies in conventional crop plants. With the potentiality of genetic engineering being amply demonstrated a first step towards production of transgenic plants with improved amino acid composition entails the purification and characterization of the specific seed storage protein rich in essential amino acids followed by cloning a full length gene coding for the target protein. Such proteins and their genes have been isolated from soyabean (Hill & Breidenbach 1974), pea (Higgins et al. 1986; Hoffman et al. 1988), *Lupinus albus* (Melo et al. 1994), rice (Takaiwa et al. 1987; Krishnan & Pueppke 1993), oat (Shotwell et al. 1990), *Phaseolus vulgaris* (Goossens et al. 1994), rape seed (Coupe et al. 1993); field bean (Heim et al. 1994); chickpea (Khitha et al. 1995; Mandaokar, & Koundal 1996; Saha & Koundal 1998); *Chenopodium* (Dey et al. 1993); *Brassica* (Dasgupta & Mandal 1991; Utsumi et al. 1993; Dasgupta et al. 1995); grain amaranth (Raina & Datta 1992), buckwheat (Bharali, & Chrungoo 2003). Due to the balanced amino acid composition, high nutrient value and homology with seed storage proteins of leguminous group of plants, the nutritionally rich component of protein of common buckwheat (*Fagopyrum esculentum* Moench) which is a 26kDa basic subunit, having more than 6% lysine and nearly 2% methionine (Bharali, & Chrungoo 2003) could be an important candidate for compensation of limiting amino acid in plants deficient in such amino acids. The present paper describes the functional characterization of an endosperm specific promoter *p1028* from common buckwheat (*Fagopyrum esculentum* Moench) for driving tissue specific gene expression.

MATERIALS AND METHODS

Plant Materials: Grains of common buckwheat (*Fagopyrum esculentum* Moench) [Accession No. IC-188669] were procured from North Eastern Regional Station of National Bureau of Plant Genetic Resources, Shillong. (ii) Seeds of Arabidopsis col-0 (Columbia) ecotype were procured from the University of Nottingham, School of Biosciences through the European Arabidopsis Stock Centre.

DNA Isolation: Total genomic DNA was isolated from 14 days old etiolated seedlings of common buckwheat (acc. no. IC-188669) following a modified CTAB extraction protocol (Murray &Thompson 1980).

RNA Isolation: Grains of common buckwheat were harvested at the mid-maturation stage (16-20DAF) of development and Total RNA was isolated using TRIzol reagent (Invitrogen).

Isolation and *In-silico* analysis of the buckwheat legumin gene: PCR amplification of the legumin gene Bwleg from the cDNA template with primer pair BwlegF (5′GACTAGTATGTCAACTAAACTCATACT3′) and BwlegR (5′ACGCTAGATCTTTAGAAACGCTCCCTC3′). The nucleotide sequences were subjected to in-silico analysis. MOTIF SCAN software was used for motif search on amino acid sequence. The coding regions of the sequences were translated by the EXPASY tool to get the deduced amino acid sequences which were subsequently subjected to BLASTp to determine their homology with other known amino acid sequences in the

data bases. Physico-chemical properties like the Molecular weight and isoelectric point (pI) of the deduced amino acid sequences were obtained by ProtParam using EXPASY online tool.

Isolation and characterization of the 5'UTR of legumin gene *Bw*leg of common buckwheat: This study involved construction of genome walking adapter libraries from genomic DNA isolated from common buckwheat and amplification of the buckwheat legumin gene and its 5' upstream region (UTR) from the library. The approach was followed by amplification of the 5' UTR using primers designed from nucleotide sequences obtained during genome walking. After a series of in-silico analysis seven deletions of the putative promoter were created by 5' nested PCR. To further assess the tissue specificity and strength of promoter for driving reporter gene expressions, each deletion fragment was directionally cloned separately in pCAMBIA1304 plant expression vector upstream of GUS and GFP genes for reporter gene expression after excising the CaMV 35S promoter from the vector by digestion with BamHI and SpeI. For verifying the introgression of transgene into the nuclear DNA of transformed Arabidopsis, nuclear DNA was isolated from T1 and T2 generations of transformed Arabidopsis plants. PCR amplification of transgenes was carried out with genomic DNA isolated from transformed Arabidopsis plants as the template and transgene nucleotide sequence specific primers for amplification.

RESULTS AND DISCUSSION

PCR amplification of the legumin gene *Bwleg* from the cDNA template with primer pair BwlegF and BwlegR yielded a single amplicon corresponding to a molecular mass of 1.7kb (Fig.1). BLASTn analysis of the nucleotide sequence showed a maximum homology of 99%, 96% and 90% with a query coverage of 91%, 92% and 85% to common buckwheat legumin gene nucleotide sequences bearing accession numbers D87980, AF152003 and DQ849083, respectively. The sequence has been deposited in GenBank with accession number KM488332. The deduced amino acid sequence of the open reading frame derived from the nucleotide sequence bearing accession no. KM488332 represented a putative 64kDa pre-protein comprised of 565 amino acid residues with a theoretical pI of 5.63. Motif search on the deduced amino acid sequence identified a "Cupin 1" superfamily domain spanning from $P'_{49\text{-}275}$ and $P'_{390\text{-}539}$. Additionally, domain search on the deduced amino acid sequence using prosite identified an N-terminal signal sequence comprised of 18 amino acid residues represented by the sequence "MSTKLILSFSLCLMVLSC" highlighted in red in figure 2. Sequence analysis of the putative 64kDa pre-protein identified the ASN-GLY proteolytic cleavage site at P'_{377}. While the presence of the ASN-GLY proteolytic cleavage site indicated the presence of an α- and a β-subunit of the protein, the presence of cysteine residues at P´21 and P'381 shows the residues linking the α- and a β- subunit of the protein by a disulphide bond. ExPASy tool identified the α- subunit as a sequence of 376 residues with a predicted molecular mass of 43kDa and a theoretical pI of 5.23. The putative protein showed a lysine content of 1.9%. On the other hand, the β- subunit was identified as a sequence of 181 amino acid residues with a predicted molecular mass of 20kDa and a theoretical pI of 9.51. The subunit showed a lysine content of 5%. Analysis of the number of moles of different amino acids present in the α and β subunits of the putative 64kDa pre-protein revealed the presence of higher levels of lysine and leucine in the β subunit.

The 5' upstream region of buckwheat legumin-type SSP generated was isolated by PCR-based genome walking using Universal Genome Walker Kit from Clontech (USA) which resulted in amplification of a 1.1 kb DNA fragment (Fig.3). The nucleotide sequence of 1062 bases for the amplicon has been deposited in Genbank with accession no. EU595873. It is known that the efficiency of ATG codon recognition is modulated by the context sequence of the codon. The context sequence of ATG at P'_{801} (TCCACCATGTCA) in the nucleotide sequence of p1062*Bwleg* matches the optimal context sequence CCACCATG(G) derived by Kozak (Kozak 1984; Kozak 1986). P**romoter prediction tool** (Neural Network Promoter Prediction) **identified three probable promoter regions** at $P'_{392\text{-}442}$, $_{473\text{-}523}$ and $_{721\text{-}771}$ in the sequence. 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 YR rule, i.e., pyrimidine (C/T) at position -1 and purine (A/G) at position +1 of the TSS by Yamamoto (Yamamoto et al. 2007) was suggested to be conserved in *Arabidopsis* and rice genes. The TSS at position 761 also follows the YR rule, having pyrimidine 'C at -1 and purine 'A' at+1 position ($C^{-1}A^{+1}$). Considering 'A' at position 761 (+1) as the predicted TSS and ATG at posi-

tion 801 (+40) as the initiating codon, the TATA at position $731_{(-62)}$ is therefore considered as the TATA box of the promoter. Apart from TATA box, the sequence revealed several other *cis*-elements, that are involved in the regulation of eukaryotic gene in general and seed-specific expression in particular. Online tool PLACE revealed presence of one prolamin box represented by the -171 core element 5'TGTAAAG3' and three RY-repeat motifs represented by -525, -138, -111 core elements "5'CATGCA3"(Fig 4). The core element 5'CATGCA3' also known as legumin box is considered to be the key element in regulating seed specific expression of genes (Chamberland et al. 1992; Ellerström et al. 1996; Reidt et al. 2000). This box is conserved in all the legumin genes (Shasany & Koundal 2000). The presence of other positive regulatory, enhancer like *cis* elements are only fully functional in conjunction with the core motif 5'CATGCA3' of the legumin box (Bäumlein et al. 1992). While the "P-box" ("TGTAAAG") is a -300 enhancer element present in SSP genes of cereals and several other dicots (Vickers et al. 2006), we detected a "P- box" as a -171 element in the buckwheat legumin gene *Bwleg* promoter p1062*Bwleg*. 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". Comparative search for regulatory motifs across other seed specific promoters also revealed the highly conserved nature of the 'TGTAAAG',' CANNTG', 'AAAG', 'CACA' and 'CANNTG' (MYC consensus box) motifs in SSP gene promoters of many dicots and monocots (Takaiwa et al. 1996; Sakata et al. 1997).

Seven deletion fragments of the putative promoter were amplified by 5' nested PCR using a common reverse primer *viz*. DLRl and seven forward primers *viz*. DLF1, DLF2, DLF3, DLF4, DLF5, DLF6 and DLF7. This generated a ladder of 7 amplicons with molecular mass of 790kb, 680kb, 560kb, 520kb, 380kb, 250kb and 180kb (Fig.5). Directional cloning of each deletion in pCAMBIA1304 plant expression vector upstream of *GUS* and *GFP* genes for reporter gene expression after excising the CaMV 35S promoter from the vector by digestion with *Bam*HI and *Spe*I generated constructs designated as *pBwlDF7, pBwlDF6, pBwlDF5, pBwlDF4. pBwlDF3, pBwlDF2* and *pBwlDF1*. These constructs were mobilized into *Agrobacterium tumefaciens* (LBA4404) for transformation of Columbia (Col-0) ecotype of *Arabidopsis thaliana*. While PCR amplification with genomic DNA of transformed *Arabidopsis* plants as template revealed amplification of each deletion fragment in conformity with the size of the transgene, no amplification was detected with genomic DNA isolated from untransformed plants of *Arabidopsis thaliana*. Similarly, amplification of GUS reporter gene was observed in the same pattern (Fig.6).

To determine the tissue specific expression and minimum effective promoter length, various plant parts (pod, flower and leaf) of all the seven confirmed homozygous T3 transgenic *Arabidopsis* harbouring buckwheat p1062*Bwleg* deletion construct (*pBwlDF1* to *pBwlDF7*) individually were subjected to GUS and GFP staining (Fig.7). Arabidopsis plants harbouring promoter deletion construct *pBwlDF1* to *pBwlDF6 clearly* showed seed specific expression of the reporter gene. However, wild type and transgenic *Arabidopsis* harbouring deletion construct *pBwlDF7* did not show any reporter gene activity in any tissue. Endosperm specificity of reporter gene expression by previously known seed specific promoters of monocots such as glutelin GluA-2 of rice (Wu et al. 1998), D hordein of barley (Horvath 2000), zein of maize (Russell & Fromm 1997) and dicots like lectin (Phillips et al. 1997), ß-conglycinin ά subunit from soybean (Nishizawa et al. 2003), ARCELIN5 (Goossens et al. 1999) and ß-phaseolin (Burow et al. 1992; van der Geest & Hall 1997) is well known. Reporter gene assay for p1062*Bwleg* clearly revealed the highest activity in plants harbouring the construct *pBwlDF1* followed by a marginally lower activity in plants harbouring the construct *pBwlDF2. S*eeds harbouring the constructs *pBwlDF3*, *pBwlDF4* and *pBwlDF5* showed a nearly threefold decrease in GUS activity than those harbouring the construct with full length promoter. We could not detect any significant GUS activity in plants harbouring the construct *pBwlDF7.* Thus, deletion of 5'CAAT3' $_{(-732)}$ motif, 5'CACA3' $_{(-774)}$ motif, 5'CANNTG3' $_{(-621)}$ motif and 5'AACA3' $_{(-761)}$ motifs had only a marginal effect on promoter activity. On the other hand, deletion of 5'CAAT3' elements at P'$_{-732, -570, -497, -451}$, CACA elements at P'$_{-774}$ and $_{-475}$, 5'CANNTG3' elements at P'$_{-621}$ and $_{-224}$, 5'AACA3' element at P'$_{-761}$ and 5'CATGCA3' element at P'$_{-525}$ caused a marked reduction in GUS expression in the seeds. CATGC, the core motif of legumin box, has been implicated as a key *cis* acting element for seed specific gene expression (Chamberland et al. 1992; Bäumlein et al. 1992). Deletion of this motif within the 2.4 kb *Le*B4 upstream sequence has been reported to lead to a drastic reduction in reporter gene expression, besides driving low level of expression in the leaves (Bäumlein et al. 1992). However, on the basis of their results on progressive deletions, leaving the CATGCATG motif intact in the *Le*B4 promoter, it was concluded that the

CATGC motif was necessary, but not absolutely essential, for SSP for gene expression. *Arabidopsis* plants harbouring the construct *pBwlDF7* did not show any GUS expression in their seeds. This confirms the role of proximal elements like the 5'TGTAAG3' $_{(-171)}$ motif or P-box in regulating the expression of SSP genes. A similar profile of activity was observed for GFP activity in the seeds of transformed *Arabidopsis* plants. This functional analysis revealed the presence of the significant proximal and distal regulatory elements spanned in the biparietal organization of p*1062Bwleg* promoter. Successful introgression of the promoter into a heterologous system for driving spatial and temporal manner of expression of the reporter gene also accentuates its candidature in seed specific gene expression for nutritional enhancement programmes.

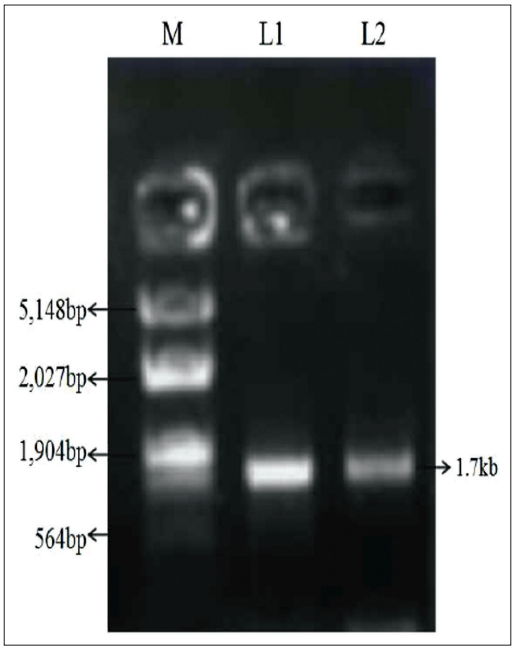

Fig 1: Electrophoresis profile of the legumin gene Bwleg using cDNA as the template with primer pair BlegF and BlegR. M: 500bp ladder and lane 1 shows the 1.7kb profile of the legumin gene.

```
10         20         30         40         50         60
MSTKLILSFS LCLMVLSCSA QLLPWQNGQR SRPHHGHQHI HHQCDITRLT ASEPSRKVRS

        70         80         90        100        110        120
EAGVTETRDN DTPEFRCAGF VAVRVVIQPG GLLLPSYSNA PYITFVEQGR GVQGVVVPGC

       130        140        150        160        170        180
PETFQSESEF EYPQSQRDQR SRQSESEESS RGDQRTRQSE SEEFSRGDQR TRQSESEEFS

       190        200        210        220        230        240
RGDQRTRQSE SEEFSRGDQR TRQSESEEFS RGDQHQKIFR IRDGDVIPSP AGVVQWTHND

       250        260        270        280        290        300
GDNDLISITL YDANSFQNQL DGNVRNFFLA GQSKQSREDR RSQRQTREEG SDRQSRESDD

       310        320        330        340        350        360
DEALLEANIL TGFQDEILQE IFRNVDQETI SKLRGDNDQR GFIVQARDLK LRVPEEYEEE

       370        380        390        400        410        420
LQRERGDRKR GGSGRSNGLE QAFCNLKFKQ NVNRPSRADV FNPRAGRINT VNSNNLPILE
```

Asn-Gly linkage

P7´

```
       430        440        450        460        470        480
FIQLSAQHVV LYKNAILGPR WNLNAHSALY VTRGEGRVQV VGDEGRSVFD DNVQRGQILV

       490        500        510        520        530        540
VPQGFAVVLK AGREGLEWVE LKNDDNAITS PIAGKTSVLR AIPVEVLANS YDISTKEAFR

       550        560
LKNGRQEVEV FLPFQSRDEK ERERF
```

Fig. 2 (a)Sequence analysis of the putative 64kDa pre-protein showing an N-terminal signal sequence comprised of 18 amino acid residues represented by the sequence "MST-KLILSFSLCLMVLSC" highlighted in red and the ASN-GLY proteolytic cleavage site at P'$_{377}$ indicating the presence of an α- and a β-subunit of the protein with the presence of cysteine residues at P´21 and P'381 linking the α- and a β- subunit of the protein by a disulphide bond.

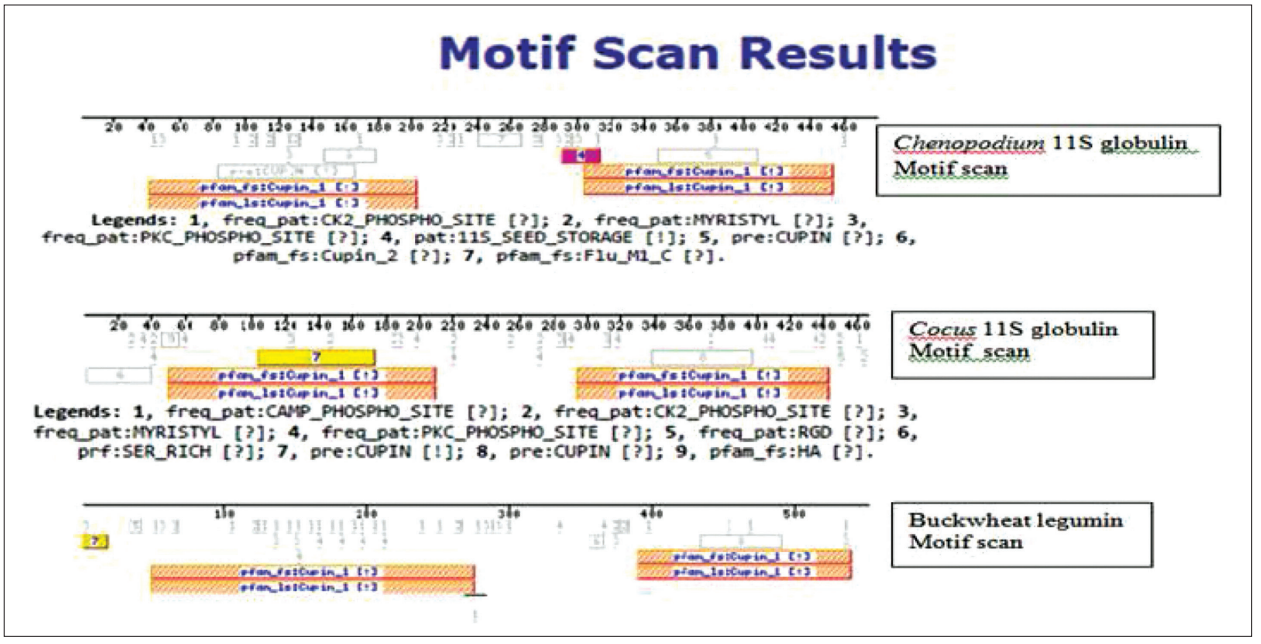

Fig2(b)Motif search on the deduced amino acid sequence showing a "Cupin 1" Superfamily domain spanning from P'$_{49-275}$ and P'$_{390-539}$.

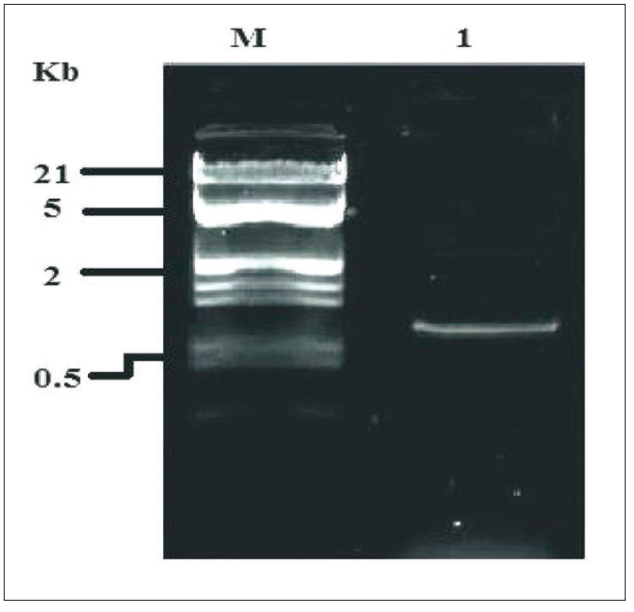

Fig. 3: Electrophoresis profile of the 5'UTR of the legumin like seed storage protein gene. M: EcoRI/HindIII double digest λ DNA, lane 1 shows the profile of the 5'UTR of the legumin like seed storage protein gene.

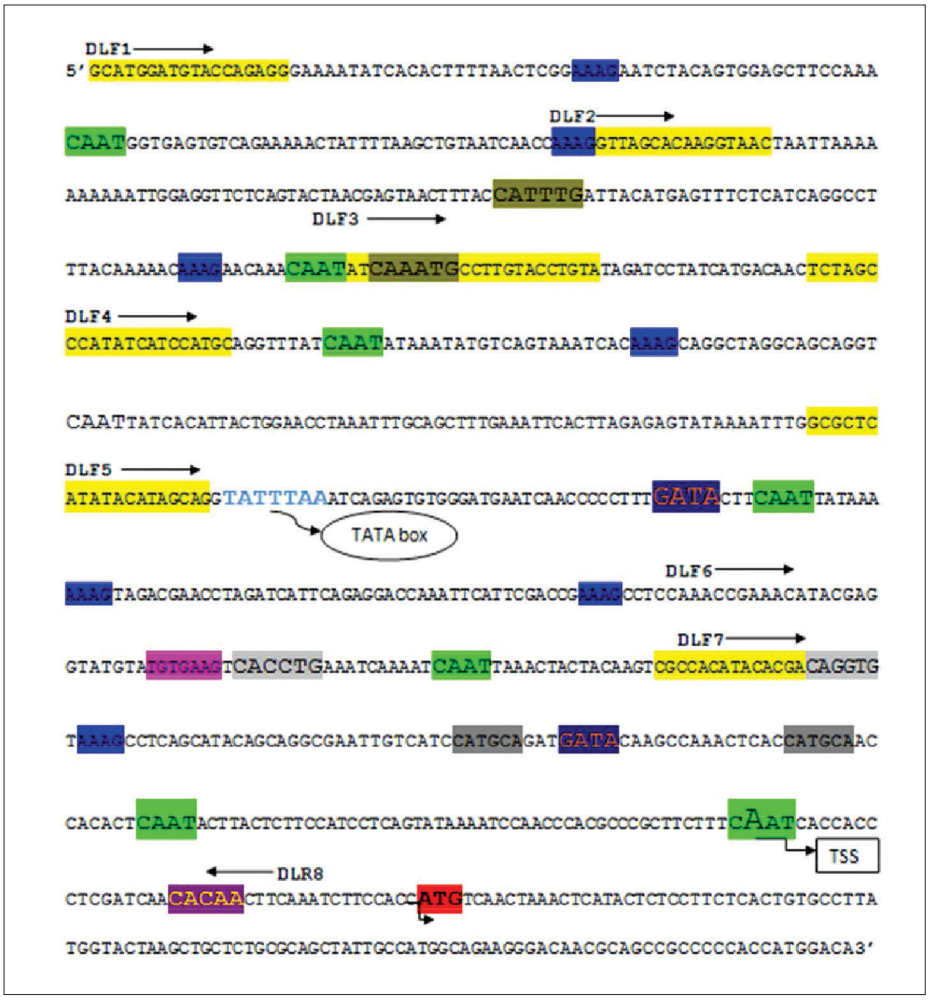

Fig.4: Nucleotide sequence of the the 5' UTR of legumin seed storage protein gene of common buckwheat showing position of primers for deletion analysis along with the position of various regulatory elements.

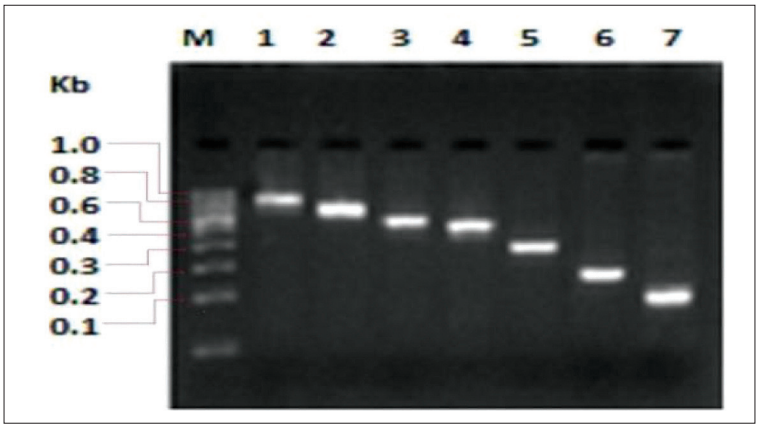

Fig.5: Electrophoresis profile of deletion fragments; Lane M- 100bp DNA ladder,1-DF1(784bp), 2-DF2(668bp), 3-DF3(541bp), 4- DF4(502bp), 5DF5(358bp), 6- DF6(224bp) and 7- DF7(149).

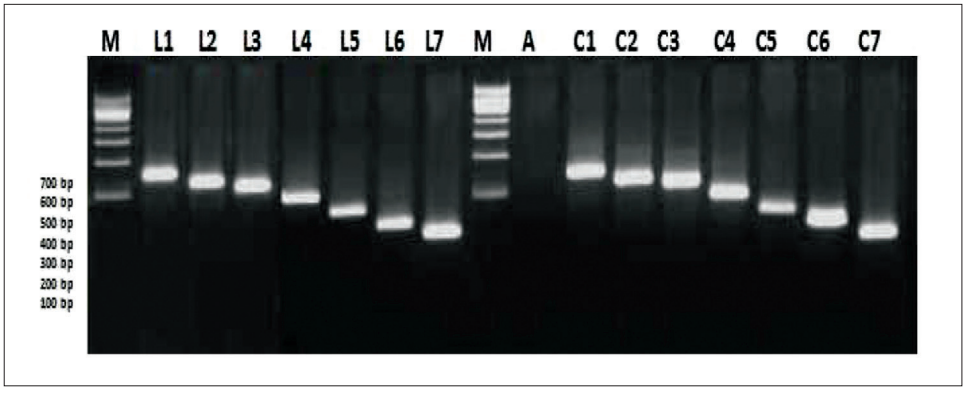

Fig.6: Electrophoresis profile of PCR amplification of (i)Deletion fragments (L1- L7) from transformed plants using genomic DNA isolated from transformed Arabidopsis plants (ii) Deletion fragments (C1- C7) from genomic DNA isolated from common buckwheat (iii) Deletion fragment from Arabidopsis A kept as control using genomic 20DNA isolated M-500bp DNA marker.

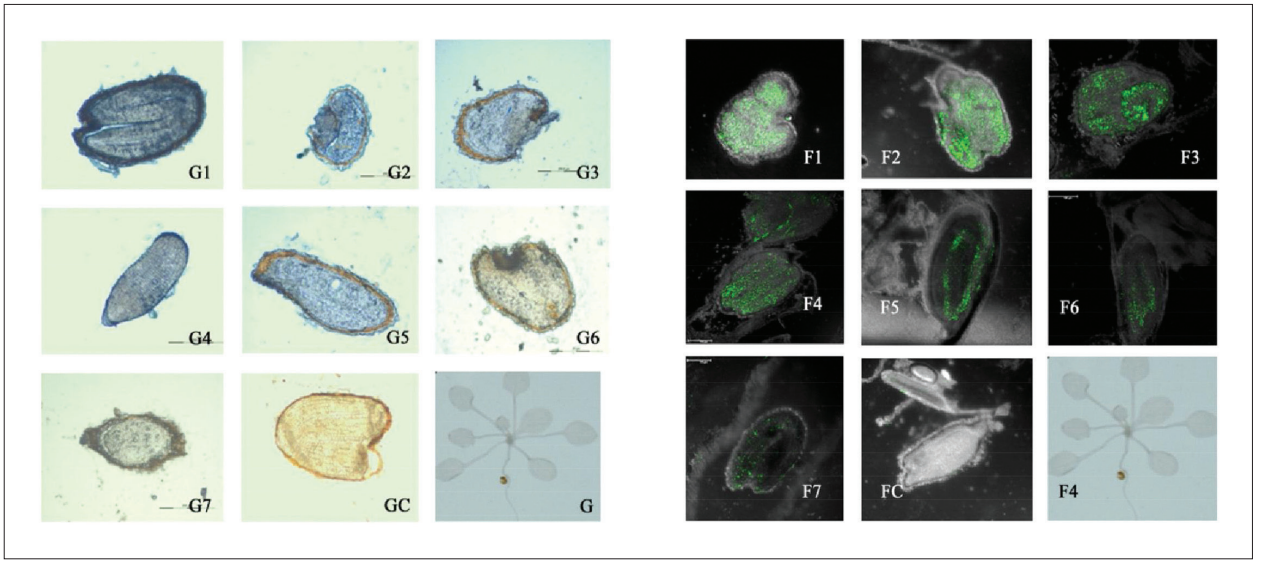

Fig.7: Reporter gene expression profile showing GUS gene (G1-G7) and GFP (F1-F7) activity in transformed Arabidopsis seed sections (transformed with constructs pBwlDF1, pBwlDF2, pBwlDF3, pBwlDF4. pBwlDF5, pBwlDF6 and pBwlDF7)and no activity in GC and FC seed section from control Arabidopsis seed section. leaf and stem of transformed plants (G) also showed no activity.

REFERENCES

Bäumlein, H., Villarroel, R., Inzé, D., & Wobus, U. (1992). Cis-analysis of a seed protein gene promoter: the conservative RY repeat CATGCATG within the legumin box is essential for tissue-specific expression of a legumin gene. Plant J., 2(2): 233-239.

Bharali, S., & Chrungoo, N. K. (2003). Amino acid sequence of the 26 kDa subunit of legumin-type seed storage protein of common buckwheat (*Fagopyrum esculentum* Moench), molecular characterization and phylogenetic analysis. Phytochem., 63(1): 1-5.

Boulter, D. (1981). Biochemistry of storage protein synthesis and deposition in the developing legume seed. Adv. Bot. Res., 9: 1-31.

Burow, M. D., Sen, P., Chlan, C. A., & Murai, N. (1992). Developmental control of the β-phaseolin gene requires positive, negative, and temporal seed-specific transcriptional regulatory elements and a negative element for stem and root expression. Plant J., 2(4): 537-548.

Chamberland, S., Daigle, N., & Bernier, F. (1992). The legumin boxes and the 3' part of a soybean β-conglycinin promoter are involved in seed gene expression in transgenic tobacco plants. Pl. Mol. Bio., 19: 937-949.

Chandrasekharan, M. B., Bisho K. J., & Hall, T. C. (2003). Module-specific regulation of the β-phaseolin promoter during embryogenesis. Plant J., 33(5): 853-866.

Coupe, S. A., Taylor, J. E., Isaac, P. G., & Roberts, J. A. (1993). Identification and characterization of a proline-rich mRNA that accumulates during pod development in oilseed rape (*Brassica napus* L.). Pl. Mol. Bio., 23(6): 1223-1232.

Dasgupta, J., Dasgupta, S., Ghosh, S., Roy, B., & Mandal, R. K. (1995). Deduced amino acid sequence of 2S storage protein from Brassica species and their conserved structural features. Ind. J. Biochem.and Biophys., 32: 378-384.

Dasgupta, S., & Mandal, R. K. (1991). Characterization of 2S seed storage protein of *Brassica campestris* and its antigenic homology with seed proteins of other Cruciferae. Biochem. Intl., 25: 409-417.

Dey, N. R. I.Singha., & Mandal, R. K. (1993). Characterisation of 2S albumin with nutritionally balanced aminoacid composition from the seeds of *Chenopodium album* and its antigenic homology with seed proteins of some Chenopodiaceae and Amaranthaceae species. Biochem. Mol. Biol. Int., 30(1): 149-157.

Ellerström, M., Stålberg, K., Ezcurra, I., & Rask, L. (1996). Functional dissection of a napin gene promoter: identification of promoter elements required for embryo and endosperm-specific transcription. Pl. Mol. Bio., 32(6): 1019-1027.

Goossens, A., Geremia, R., Bauw, G., Montagu, M., & Angenon, G. (1994). Isolation and characterisation of arcelin–5 proteins and cDNAs. FEBS. J., 225(3): 787-795.

Goossens, M., Rahtz, S. P., & Rahtz, S. (1999). The LaTeX Web Companion: Integrating TeX, HTML, and XML (Vol. 4). Addison-Wesley Professional.

Heim, U., Bäumlein, H., & Wobus, U. (1994). The legumin gene family: a reconstructed Vicia faba legumin gene encoding a high-molecular-weight subunit is related to type B genes. Pl. Mol. Bio., 25(1): 131-135.

Higgins, T.J., Chandler, P.M., Randall, P.J., Spencer, D., Beach, L.R., Blagrove, R.J., Kortt, A.A. & Inglis, A.S., 1986. Gene structure, protein structure, and regulation of the synthesis of a sulfur-rich protein in pea seeds. J. Biol. Chem., 261(24): 11124-11130.

Hill, J. E., & Breidenbach, R. W. (1974). Proteins of soybean seeds I. Isolation and characterization of the major components. Pl. Physiol., 53(5): 742-746.

Hoffman, L. M., Donaldson, D. D., & Herman, E. M. (1988). A modified storage protein is synthesized, processed, and degraded in the seeds of transgenic plants. Pl. Mol. Bio., 11(6): 717-729.

Horvath, C. M. (2000). STAT proteins and transcriptional responses to extracellular signals. TIBS., 25(10): 496-502. Jiang and Pugh, 2009

Khitha, S., Kansal, R., & Koundal, K. R. (1995). Isolation and Characterization of Genomic DNA of Chick Pea. Pl. Physiol. & Biochem.,-NEW DELHI-, 22: 1-5.

Kozak, M. (1984). Compilation and analysis of sequences upstream from the translational start site in eukaryotic mRNAs. Nucl Acids Res., 12(2): 857-872.

Kozak, M. (1986). Point mutations define a sequence flanking the AUG initiator codon that modulates translation by eukaryotic ribosomes. Cell, 44(2): 283-292.

Krishnan, H. B., & Pueppke, S. G. (1993). Nucleotide sequence of an abundant rice seed globulin: homology with the high molecular weight glutelins of wheat, rye and triticale. Biochem. Biophys. Res. Commun.,193(1): 460-466.

Mandaokar, A. D., & Koundal, K. R. (1996). Construction and screening of subgenomic library of chickpea (*Cicer arietinum* L.) for legumin genes and their analysis. Indian J. Exp Biol., 34(6): 496-501.

Melo, T. S., Ferreira, R. B., & Teixeira, A. N. (1994). The seed storage proteins from *Lupinus albus.* Phytochem., 37(3), 641-648.

Murray, M. G., & Thompson, W. F. (1980). Rapid isolation of high molecular weight plant DNA. Nucl Acids Res., 8(19), 4321-4326.

Nishizawa, K., Maruyama, N., Satoh, R., Fuchikami, Y., Higasa, T., & Utsumi, S. (2003). AC-terminal sequence of soybean β-conglycinin α' subunit acts as a vacuolar sorting determinant in seed cells. Plant J., 34(5): 647-659.

Phillips, J., Artsaenko, O., Fiedler, U., Horstmann, C., Mock, H. P., Müntz, K., & Conrad, U. (1997). Seed-specific immunomodulation of abscisic acid activity induces a developmental switch. EMBO. J., 16(15): 4489-4496.

Raina, A., & Datta, A. (1992). Molecular cloning of a gene encoding a seed-specific protein with nutritionally balanced amino acid composition from *Amaranthus*. Proc. Natl. Acad. Sci., 89(24): 11774-11778.

Reidt, W., Wohlfarth, T., Ellerström, M., Czihal, A., Tewes, A., Ezcurra, I., & 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. Plant J., 21(5): 401-408.

Russell, D. A., & Fromm, M. E. (1997). Tissue-specific expression in transgenic maize of four endosperm promoters from maize and rice. Transgenic Res., 6(2): 157-168.

Saha, S., & Koundal, K. R. (1998). Isolation and characterization of cDNAs encoding storage proteins of chickpea (*Cicer arietinum* L.). J. Biosci.,23(3): 213-223.

Sakata, Y., Chiba, Y., Fukushima, H., Matsubara, N., Habu, Y., Naito, S., & Ohno, T. (1997). The RY sequence is necessary but not sufficient for the transcription activation of a winged bean chymotrypsin inhibitor gene in developing seeds. Pl. Mol. Bio., 34(2): 191-197.

Shasany, A. K., & Koundal, K. R. (2000). Isolation and characterisation of legumin promoter sequence from chickpea (*Cicer arietinum* L.). Indian J. Exp. Biol. 38(4): 363-72.

Shotwell, M. A., & Larkins, B. A. (1989). The biochemistry and molecular biology of seed storage proteins. The Biochemistry of plants: a comprehensive treatise (USA). Purdue University, West Lafayette, IN.

Shotwell, M. A., Boyer, S. K., Chesnut, R. S., & Larkins, B. A. (1990). Analysis of seed storage protein genes of oats. J. Biol. Chem., 265(17): 9652-9658.

Takaiwa, F., Kikuchi, S., & Oono, K. (1987). A rice glutelin gene family—a major type of glutelin mRNAs can be divided into two classes. Mol. Gen. Genet., 208(1): 15-22.

Takaiwa, F., Yamanouchi, U., Yoshihara, T., Washida, H., Tanabe, F., Kato, A., & Yamada, K. (1996). Characterization of common cis-regulatory elements responsible for the endosperm-specific expression of members of the rice glutelin multigene family. Pl. Mol. Bio., 30(6): 1207-1221.

Utsumi, S., Kitagawa, S., Katsube, T., Kang, I. J., Gidamis, A. B., Takaiwa, F., & Kito, M. (1993). Synthesis, processing and accumulation of modified glycinins of soybean in the seeds, leaves and stems of transgenic tobacco. Plant Sci., 92(2), 191-202.

van der Geest, A. H., & Hall, T. C. (1997). The β-phaseolin 5' matrix attachment region acts as an enhancer facilitator. Pl. Mol. Bio., 33(3): 553-557.

Vickers, C. E., Xue, G., & Gresshoff, P. M. (2006). A novel cis-acting element, ES contributes to high-level endosperm-specific expression in an oat globulin promoter. Pl. Mol. Bio., 62(1): 195-214.

Wu, C. Y., Suzuki, A., Washida, H., & Takaiwa, F. (1998). The GCN4 motif in a rice glutelin gene is essential for endosperm-specific gene expression and is activated by Opaque-2 in transgenic rice plants. Plant J., 14(6): 673-683.

Wu, C. Y., Washida, H., Onodera, Y., Harada, K., & 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. Plant J., 23(3): 415-421.

Yamamoto, Y.Y., Ichida, H., Matsui, M., Obokata, J., Sakurai, T., Satou, M., Seki, M., Shinozaki, K. & Abe, T., (2007). Identification of plant promoter constituents by analysis of local distribution of short sequences. BMC Genomics, 8(1): p.67.