IDENTIFICATION AND FUNCTIONAL CHARACTERIZATION OF THE PROMOTER REGION OF ENDOSPERM SPECIFIC LEGUMIN TYPE SEED STORAGE PROTEIN GENE IN *FAGOPYRUM ESCULENTUM* (MOENCH)

By

SHINY CH. SANGMA

THESIS SUBMITTED
IN FULFILMENT OF THE DEGREE OF
DOCTOR OF PHILOSOPHY IN BOTANY

NORTH EASTERN HILL UNIVERSITY
SHILLONG-793022, INDIA
2009

Thesis

DS
583.57
SAN

CONTENTS

Page

Declaration ..... (I)

Acknowledgement..... (II-III)

Abbreviation ..... (IV-V)

List of Figures..... (VI-XII)

List of Tables..... (XIII)

CHAPTERS

1. INTRODUCTION..... 1-6

2. REVIEW OF LITERATURE..... 8-27

3. MATERIALS AND METHODS..... 28-54

   (i) Materials..... 28-29

   (ii) Protocols..... 29-50

   (iii) Chemicals and Solutions..... 50-54

4. AMPLIFICATION AND SEQUENCE ANALYSIS OF LEGUMIN GENE OF COMMON BUCKWHEAT

   (i) Experimental..... 55-56

   (ii) Results..... 56-62

   (iii) Discussion..... 62-68

5. AMPLIFICATION OF 5' UPSTREAM FLANKING REGION BY PCR BASED-GENOME WALKING.

(i) Experimentl..... 69-70

(ii) Results..... 70-82

(iii) Discussion..... 82-92

6. GENERAL SUMMARY AND DISCUSSION..... 93-114

7. REFERENCES..... 115-134

The North Eastern-Hill University

July, 2009

Declaration

I, Shiny Ch. Sangma, hereby declare that the subject matter of this thesis entitled, "Identification and functional characterization of the promoter region of endosperm specific legumin type seed storage protein gene in *Fagopyrum esculentum* (Moench)" is the record of work done by me. The contents of this thesis did not form basis of award of any previous degree to me or to the best of my knowledge to anybody else and that the thesis has not been submitted by me for any research degree in any other University/Institute

This is being submitted to the North Eastern Hill University for the award of the degree of Doctor of Philosophy in Botany

SHINY CH. SANGMA

17/8/09

Head of the Department

17/8/09

Supervisor

ACKNOWLEDGEMENTS

First and foremost I would like to express my utmost gratitude to my supervisor, Prof. N. K. Chrungoo, Botany Department, NEHU for the time and effort that he has put into my thesis work. He has inspired me and guided me and caused me to think like a researcher. A special thanks also goes to ever-charming, Madam Jyoti Narayan for keeping Sir's head cool in spite of all the headaches we, as students give him.

I am thankful to Prof. M. S. Dkhar, Prof. N. K. Chrungoo and Prof. A. K. Misra, the present and the past Heads, Department of Botany NEHU for providing the necessary facilities. I am thankful to all the faculty members of the Department for their inspiration and constant encouragement.

My sincere thanks to my labmates Anu, Sanga, Cressida, Tia, Nabanita, Felly, Neeta, Ganga, Geetahree, Saptadeepa and Lashai. I specially thank Nabanita, who is also my room mate, for being a supportive companion and helping me out with the references of my thesis. My fellow research scholars from Molecular Genetics Lab: Anindita, Mhathung, Manik, David, Meena, Mipesweri, Insanara, Jubanri, thank you all for your help and cooperation. I am very much grateful to my fellow research scholars from Botany Department and my fellow hostel-mates for being there for me whenever I needed them. My fellow scholars from sister Department, "thanks for lending me chemicals to keep my work going". Manoj deserves special thanks for being an ever-ready help at hand.

I owe a big thank you to the Non-teaching staffs of Botany Department, who are ever ready to help in whatever way they can.

My deepest gratitude is due to my mother, brothers, sisters, cousins and other well wishers for their prayers and support.

I thank CSIR for fellowship during my tenure as CSIR-JRF-SRF.

Most of all, I thank God for everything.

ABBREVIATIONS

Amp	: Ampicillin
BLAST	: Basic Local Alignment Search Tool
bp	: Base Pair
CTAB	: Cetyl Trimethyl Ammonium Bromide
dATP	: 2' Deoxyadenosine 5'-triphosphate
dCTP	: 2' Deoxycytosine 5'-triphosphate
dGTP	: 2' Deoxyguanosine 5'-triphosphate
DNA	: Deoxyribonucleic Acid
dNTP	: 2' Deoxynucleotide-triphosphate
dTTP	: 2' Deoxythymidine 5'-triphosphate
EDTA	: Ethylene Di-amine Tetra-acetic acid
GFP	: Green Fluorescent protein.
Kb	: Kilo base
LB	: Luria Bertani
ME	: β-Mercaptoethanol
MOPS	: Morpholinopropane Sulfonic acid
ORF	: Open Reading Frame
PCR	: Polymerase Chain Reaction
RNA	: Ribonucleic Acid
RPM	: Revolutions Per Minute
SDS	: Sodium Dodecyl sulphate

SSC	: Sodium Saline Citrate
SSP	: Seed Storage Protein
TBE	: Tris-borate-EDTA
Tm	: Melting temperature
TSS	: Transcription Start Site
UV	: Ultra Violet
μCi	: Micro Curie

LIST OF FIGURES

Fig 1.1a : Plants of common buckwheat (*Fagopyrum esculentum* Moench) under cultivation in the experimental garden of the Depatment of Botany, North Eastern Hill University Shillong.

Fig 1.1b : Close up of the plant of common buckwheat (*Fagopyrum esculentum* Moench) in flowering stage.

Fig 2.1 : Diagrammatic representation of the assembly of transcription machinery on a typical eukaryotic promoter.

Fig 3.1a : A typical three step cycle parameter for PCR amplification.

Fig 3.1 b : Two step cycle parameter for PCR-based genome walking with Universal GenomeWalker Kit (Clontech).

Fig 3.2 : Strategy for PCR-based gene walking using Universal GenomeWalker Kit from Clontech

Fig 3.3a : Map and features of pJET1.2/blunt vector (Fermentas) used for cloning of PCR products.

Fig 3.3b : Map and features of pGlow-TOPO (Invitrogen), a promoterless reporter vector with GFP as the reporter gene.

Fig. 4.1 : Total genomic DNA isolated from etiolated seedlings of buckwheat (*Fagopyrum esculentum* Moench.)

Fig. 4.2a : Restriction digestion profile of genomic DNA isolated from etiolated seedlings of buckwheat (*Fagopyrum esculentum* Moench) digested with *Eco*RI

Fig. 4.2b : Restriction digestion profile of genomic DNA isolated from etiolated seedlings of buckwheat (*Fagopyrum esculentum* Moench) digested with *Hind*III

Fig. 4.2c : Restriction digestion profile of genomic DNA isolated from etiolated seedlings of buckwheat (*Fagopyrum esculentum* Moench) digested with *Nco*I

Fig. 4.3 : PCR amplification profile with buckwheat genomic DNA as template and primer pair SS2F-SS3R.

Fig. 4.4 : Nucleotide sequence of 0.835 kb DNA fragment amplified with buckwheat genomic DNA as template and primer pair SS2F-SS3R

Fig. 4.5 : Clustal W (1.81) multiple alignment of 835 bp amplicon generated with buckwheat genomic DNA as template and primer pair SS2F-SS3R with nucleotide sequences of some *nad*1 gene sequences available in the database.

Fig. 4.6a : PCR amplification profile with buckwheat genomic DNA as template and primer pairs SS8F-SS1R and SS8F- SS3R

Fig. 4.6b : PCR amplification profile with buckwheat genomic DNA as template and primer pair SS4F-SS3R

Fig. 4.6c : PCR amplification profile with buckwheat genomic DNA as template and primer pair SS2F-SS1R.

Fig. 4.6d : Autoradiograph of the amplification products of primer pairs SS8F-SS1R and SS8F-SS3R hybridized with [α-^{32}P]-dATP labelled 0.835 kb DNA fragment amplified by primer pair SS4F-SS3R

Fig. 4.6e : Autoradiograph of the amplification products of primer pair SS2F-SS1R hybridized with [α-^{32}P]-dATP labelled 800 bp DNA fragment amplified by primer pair SS4F-SS3R

Fig. 4.7 : Diagramatic representation showing the relative positions of oligonucleotide primers SS2F, SS8F, SS4F, SS1R and SS3R on the target DNA.

Fig. 4.8 : Nucleotide sequence of 1613 bases for the 1.6 kb fragment amplified from buckwheat genomic DNA using the primer pair SS8F-SS3R.

Fig. 4.9 : Clustal W (1.81) multiple alignment of the nucleate sequence of 1613 bases of the 1.6 kb amplicon generated with buckwheat genomic DNA as template and primer pair SS8F-SS3R with nucleotide sequences of SSP genes available in the database

Fig. 4.10 : Diagramatic representation of the intron/exon architechture in the 1613 bp nucleotide sequence of 1.6 kb amplicon generated

buckwheat genomic DNA as the template and primer pair SS8F-SS3R.

Fig. 4.11 : Deduced amino acid sequence of the coding region of 1613 bp nucleotide sequence of the 1.6 kb amplicon generated buckwheat genomic DNA as the template and primer pair SS8F-SS3R

Fig. 4.12 : Clustal W (1.81) multiple alignment of the deduced amino acid sequence of the 1613 bases amplified DNA fragment using primer pair SS8F-SS3R with amino acid sequences of legumin-type SSPs available in the database

Fig. 4.13 : Phylogenetic tree based on alignment matrix of the deduced amino acid sequence of the 1.6 kb DNA fragment amplified with buckwheat genomic DNA as the template and primer pair SS8F-SS3R with the amino acid sequences of legumin-type proteins from various angiosperms and gymnosperms.

Fig. 4.14a : PCR amplification profile with buckwheat genomic DNA as template and primer SS8F-SS6R.

Fig. 4.14b : Autoradiograph of the 1.4 kb PCR product hybridized with [α-^{32}P]-dATPb labelled buckwheat legumin gene specific probe (DQ200889).

Fig. 4.15 : Partial nucleotide sequence of 564 bases for the 1.4 kb PCR product amplified from buckwheat genomic DNA with primer pair SS8F-SS6R

Fig.5.1 : Electrophoresis profile of genomic DNA isolated from etiolated seedlings of common buckwheat and digested with *Dra*I, *Eco*RV, *Pvu*II and *Stu*I.

Fig. 5.2 : Electrophoresis profile of amplification products of primary PCR in *Dra*I and *Eco*RV genomewalker libraries.

Fig. 5.3 : Electrophoresis profile of amplification products of nested PCR with products of primary PCR with *Dra*I adapter library and primer pair AP1- SS6R as the template and oligonucleotide primer pair AP2-SS10R

Fig. 5.4 : Electrophoresis profile of amplification products of nested PCR with products of primary PCR with *Eco*RV adapter library and primer pair AP1-SS11R as the template and oligonucleotide primer pair AP2-SS12R, *Eco*RV adapter library and primer pair AP1- SS13R as the template and oligonucleotide primer pair AP2-SS13R, *Dra*I adapter library and primer pair AP1-SS13R as the template and oligonucleotide primer pair AP2-SS13R.

Fig. 5.5 : Electrophoresis profile of amplification products of nested PCR with products of primary PCR with *Dra*I adapter library and primer pair AP1-SS6R as the template and oligonucleotide primer pair AP2-SS10R along with corresponding autoradiograph showing the hybridization of the amplification product of the nested PCR with [α-^{32}P]-dATP labelled buckwheat legumin gene specific probe (DQ200889).

Fig. 5.6 : Partial nucleotide sequence of 646 bases for the amplicon generated by nested PCR with products of primary PCR with *Dra*I adapter library and primer pair AP1- SS6R as the template and oligonucleotide primer pair AP2-SS10R.

Fig. 5.7 : Electrophoresis profile of amplification products of nested secondary PCR with products of primary PCR with *Eco*RV adapter library and primer pair AP1- SS11R as the template and oligonucleotide primer pair AP2-SS12R, *Eco*RV adapter library and primer pair AP1- SS13R as the template and oligonucleotide primer pair AP2-SS13R, *Dra*I adapter library and primer pair AP1- SS13R as the template and oligonucleotide primer pair AP2- SS13R, and corresponding autoradiograph showing the hybridization of the amplification products of the nested PCRs with [α-^{32}P]-dATP labelled amplicon generated by nested PCR with *Dra*I adapter library and primer pair AP1- SS13R as the template and oligonucleotide primer pair AP2-SS13R.

Fig. 5.8 : Nucleotide sequence of 1127 bases for the amplicon generated by nested PCR with *Eco*RV adapter library and primer pair AP1- SS11R as the template and oligonucleotide primer pair AP2-SS12R

Fig. 5.9 : Clustal W (1.81) multiple alignment of 1127 bp nucleotide sequence of Bw*Eco*RV lib.GWP3 with nucleotide sequences of some SSP genes of other plants.

Fig. 5.10a : Diagramatic representation of the intron/exon architechture in the 1127 bp nucleotide sequence of the amplicon generated with buckwheat *Eco*RV adapter library and primer pair AP1- SS11R as the template and oligonucleotide primer pair AP2-SS12R

Fig. 5.10b : Screen shot of the BLAST of the 1127 bp sequence of the amplicon Bw*Eco*RV lib.GWP3 generated with buckwheat *Eco*RV adapter library and primer pair AP1- SS11R as the template and oligonucleotide primer pair AP2-SS12R showing the conserved positions of exons and introns among different SSP gene sequences in buckwheat.

Fig. 5.11a : Electrophoresis profile of total RNA isolated from grains of common buckwheat harvested 10, 20, 30 and 40 days after flowering.

Fig. 5.11b : Dot blot of RNA isolated from grains of common buckwheat harvested 10, 20, 30 and 40 days after flowering with [α- 32P]-dATP- labelled 1127 bp buckwheat genome walking library product Bw*Eco*RV lib.GWP3

Fig. 5.12 : Deduced amino acid sequence of the coding region of 1127 bp amplicon Bw*Eco*RV lib.GWP3 generated with buckwheat *Eco*RV adapter library and primer pair AP1- SS11R as the template and oligonucleotide primer pair AP2-SS12R

Fig. 5.13 : Clustal W (1.81) multiple alignment of the deduced amino acid sequence of the 1127 bp amplicon Bw*Eco*RV lib.GWP3 with amino acid sequences of some SSPs of other plants.

Fig. 5.14 : Phylogenetic tree based on alignment matrix of the deduced amino acid sequence of the 1127 bp amplicon Bw*Eco*RV lib.GWP3 with the amino acid sequences of legumin-type proteins from various angiosperms and gymnosperms.

Fig. 5.15 : Nucleotide sequence of the amplicon Bw*Eco*RV lib.GWP4 generated as a result of nested PCR with *Eco*RV adapter

library and primer pair AP1-SS13R as the template and oligonucleotide primer pair AP2-SS13R

Fig. 5.16 : Nucleotide sequence of the amplicon Bw*Dra*I lib.GWP2 generated as a result of nested PCR with *Dra*I adapter library and primer pair AP1- SS13R as the template and oligonucleotide primer pair AP2-SS13R

Fig. 5.17a : Electrophoresis profile of the amplified products of PCR with buckwheat genomic DNA as the template and primer pair SS15F-SS13R and primer pair SS15F-SS7R

Fig. 5.17b : Restriction digestion profile of recombinant plasmids pJET1166 and pJET1028. digested with *Nco*I and *Not*I.

Fig. 5.17c : Electrophoresis profile of the amplified products of PCR with recombinant plasmid clone pJET1028 as the template and oligonucleotide primer pair T_3-T_7 (L1) and recombinant plasmid clone pJET1028 as the template and oligonucleotide primer pair T_3-T_7

Fig. 5.18 : Nucleotide sequence of 1028 bases for the insert DNA of the recombinant plasmid clone pJET1028.

Fig. 5.19 : Nucleotide sequence of 1166 bases for the insert DNA of the recombinant plasmid clone pJET1166.

Fig. 5.20 : BLASTp output of the predicted amino acid sequence of the ORF at position 774 in the nucleotide sequence of the 1028 bp insert DNA of the recombinant plasmid clone pJET1028.

Fig. 5.21 : The promoter region and the corresponding TSS predicted by Neural Network promoter Prediction (NNPP tool) in the nucleotide sequence of the 1028 bp insert DNA of the recombinant plasmid clone pJET1028

Fig.5.22 : Nucleotide sequence of 1028 bases for the insert DNA of the recombinant plasmid clone pJET1028 showing various regulatory elements identified in the sequence.

Fig. 5.23 : Nucleotide sequence of 1028 bases for the insert DNA of the recombinant plasmid clone pJET1028 showing the position of the

primers designed for generating PCR-based deletion fragments of the 1028 bp DNA.

Fig. 5.24a : Electrophoresis profile of the amplicons generated by PCR based deletions.

Fig. 5.24b : Southern hybridization of the deletion fragments generated by PCR based deletions with [α-^{32}P]-dATP labelled insert DNA from the recombinant plasmid clone pJET1028.

Fig. 5.25 : Photomicrographs of fluorescing *E.coli* cells transformed with constructs pGlowTOPO$_{NULL}$, pGlowTOPO$_{-731}$, pGlowTOPO$_{-648}$, pGlowTOPO$_{-529}$, pGlowTOPO$_{-487}$, pGlowTOPO$_{-347}$, pGlowTOPO$_{-228}$, pGlowTOPO$_{-146}$, pGlowTOPO$_{-25}$.

Fig 5.26 : Schematic representation of the positions of various regulatory elements on the deletion fragments.

Fig. 5.27a : Electrophoresis profile of the 3.0 kb DNA fragment amplified from genomic DNA of common buckwheat by PCR with primer pair SS15F- SS16R .

Fig. 5.27b : Autoradiograph of the 3.0 kb fragment hybridized with [α-32P]-dATP- labelled 1028 bp amplicon generated with buckwheat genomic DNA as template using primer pair SS15F-SS13R.

Fig. 5.27c : Autoradiograph of the 3.0 kb fragment hybridized with [α-32P]-dATP- labelled buckwheat legumin-like protein gene specific probe (DQ200889).

Fig. 5.28 : Electrophoresis profile of the 3.0 kb DNA fragment digested with *Eco*RI, *Nco*I, *Hind*III, *Eco*RV and *Stu*I.

LIST OF TABLES

Table 2.1 : The content of essential amino acids in seed storage protein of common buckwheat (*Fagopyrum esculentum* Moench) as compared to other cereals.

Table 3.1 : List of primers used for PCR amplification of legumin-type seed storage protein gene and its promoter.

Table 4.1 : List of primer combinations and PCR annealing temperatures used for the amplification of SSP genes from buckwheat genomic DNA

Table 5.1 : List of primer combinations used for PCR-based genome walking and the sizes of bands amplified.

Table 5.2 : *Cis*-elements registered in RegSite database which were identified in the 1028 bp putative promoter by TSSP tool.

Table 5.3 : *Cis*-elements registered in PLACE database which were identified in the 1028 bp putative promoter by TSSP tool.

Table 5.4 : List of gene promoters used for analysis with MEME tool.

Table 5.5 : Common motifs found among 23 SSP gene promoters as detected by MEME tool.

Table. 5.6 : Alignment of the context sequence around TATA, TSS and ATG-start codon of buckwheat seed storage protein gene with the corresponding regions of seed storage protein genes from dicot and monocot plants. (after Joshi, 1987).

CHAPTER: I
INTRODUCTION

Plant genetic resources, representing the sum total of diversity accumulated through years of cultivation under domestication and natural selection, are considered as one of the most important gifts of nature to mankind. Many of these plants are important sources of high nutritive value foods for human consumption. Out of the total genetic diversity available, grain crops constitute one of the major sources of food for human nutrition. However, of the total available genetic diversity, mankind has utilized only a few plants as major food sources.

Although cereal grains and legume seeds are the major sources of vegetarian dietary proteins for human consumption, 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 and Larkins, 1989). Over the years many attempts have been made 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.

One of the approaches in such direction has been to manipulate the regulation of amino acid biosynthesis to increase the abundance of a particular amino acid. Mutant selection and engineering genes encoding key enzymes of amino acid biosynthetic pathways have been used to increase amino acids in crop plants (Zeh *et al.*, 2001). However, an increase in the free essential amino acids may not necessarily lead to an increase in the content of fixed amino acids; the amino acids could be leached out from the plant tissue and lost during boiling and other processing (Falco *et al.*, 1995). Alternatives to the manipulation of regulation of amino acid biosynthesis have focused around either (i) the manipulation of the primary sequence of a seed protein gene by addition, substitution or deletion of nucleotides by site directed mutagenesis and expression of the altered gene in place of or in addition to the native gene (Lago *et al.*, 1990; Guerche *et al.*, 1990; Blechl and Anderson, 1996) or (ii) introduction of new heterologous genes, coding for proteins with a higher level of essential amino acids, under the seeds own promoter or a more efficient heterologous promoter (Nielsen *et al.*, 1995; Saalbach *et al.*, 1995). Efficient

manipulation of the amino acid composition of seed storage proteins also relies to a large extent, on the stability of the foreign proteins in a heterologous system. Stable expression of the altered/heterologous protein in the seed can be a difficult task owing to the complex biochemical processes associated with seed storage protein assembly and accumulation (Lambert and Yarwood, 1992; Lending *et al.*, 1992). Rational alteration of the native protein or the insertion of a heterologous protein may therefore also depend, to a large extent, on a detailed knowledge of the secondary structure of these proteins.

As a first step towards production of transgenic plants with improved amino acid composition, 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 are a pre-requisite. Such proteins and their genes have been isolated from soyabean (Hill and Briendenbach, 1974), pea (Higgins *et al.*, 1986; Hoffman *et al.*, 1988), *Lupinus albus* (Melo *et al.*, 1994), rice (Takaiwa *et al.*, 1987; Krishnan and 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; Kansal *et al.*, 1995; Mandaokar and Koundal, 1996, Saha and Koundal 1998); *Chenopodium* (Dey and Mandal, 1993); *Brassica* (Dasgupta and Mandal, 1991, Utsumi *et al.*, 1993; Dasgupta *et al.*, 1995); grain amaranth (Raina and Datta, 1992), buckwheat (Bharali, 2003) .

Considering the ever-increasing demand for food materials, it is not only necessary to use the available rich diversity and wide genetic resources and to improve the existing conventional cultivars but also to look for non-conventional lesser-known and underutilized food crops. The North Eastern region of India is

Fig. 1.1: (a) Plants of common buckwheat (*Fagopyrum esculentum* Moench) under cultivation in the experimental garden of the Depatment of Botany, North Eastern Hill University Shillong.

(b) Close up of the plant of common buckwheat (*Fagopyrum esculentum* Moench) in flowering stage.

extremely rich in floristic wealth and it is home to a variety of traditional crops that could form an important component of human diet in times to come These crops could also constitute an important genetic base to look for suitable heterologous proteins and their genes, which could be used as tools in crop improvement programmes. Amongst the existing known plant resources, the International Plant Genetic Resources Institute (IPGRI) and Consultative Group on International Agriculture (CGIAR) have identified common buckwheat, grain amaranth and *Chenopodium* as important but underutilized nutraceutical crops which could be used as genetic base for identification and isolation of suitable heterologous genes coding for biomolecules of potential economic importance. Due to the high protein content of their grains and nutritionally balanced amino acid composition of their grain proteins, such minor crops are considered to have great potential in contributing towards crop improvement programmes. Common buckwheat (*Fagopyrum esculentum* Moench) belongs to the family Polygonaceae. Due to short growth span, capability to grow at high altitudes and the high quality protein content of its grains it is an important crop in mountainous regions of India, China, Russia, Ukraine, Kazakhstan, parts of Eastern Europe, Canada, Japan, Korea and Nepal. Pictures of common buckwheat are shown in Fig. 1.1. The protein content is higher than that reported for any other cereals and the amino acid composition matches the WHO recommended values for a nutritionally rich protein with a balanced amino acid composition (Rout and Chrungoo, 1996). The nutritionally rich component of protein is a 26kDa basic subunit, which has more than 6% lysine and nearly 2% methionine (Rout and Chrungoo, 1996, 1997). Due to the balanced amino acid composition, high nutrient value and homology with seed storage proteins of leguminous group of

plants, this protein could be an important candidate for compensation of limiting amino acid in plants deficient in such amino acids.

The low level of expression of foreign genes in transgenic plants is a key limitation in the development of transgenic plants with improved quality traits. Since constitutive promoters, such as CaMV 35S promoter, do not drive the expression of their downstream genes in a tissue and developmental stage specific manner, they have limited utility in developing transgenics with enhanced levels of essential amino acids in their seed storage proteins. In contrast, the seed-specific promoter only expresses its downstream genes from mid to late stage of seed maturation, and there is no expression or much lower expression in other tissues. Successful introgression of a heterologous gene in plants for nutritional enhancement of seed storage proteins would therefore necessitate the identification and isolation of seed/endosperm specific promoters to drive seed/endosperm specific expression of the transgene in the heterologous system. Such promoters should have levels and patterns of expression which would synchronize the expression of transgenes with the physiological event during certain stage of plant development. The SSP genes are highly expressed and tightly regulated both spatially and temporally. In a number of reports it has been demonstrated that the promoters of SSP genes drive gene expression strictly in the seed tissues; during the course of seed maturation. Thus, promoters of many seed storage protein genes have been isolated and characterized. Examples of such promoters include the 5' regulatory regions from such genes as cruciferin (Sjödahl *et al.*, 1995), napin (Kridl *et al.*, 1991), phaseolin (Butos *et al.*, 1989), soybean beta-conglycinin (Lessard *et al.*, 2004; Chen *et al.*, 1986), maize zein (Matzke *et al.*, 1990; Thompson *et al.*, 1990; Brown *et al.*, 1986), rice glutelin

(Takaiwa *et al.*, 1991; Qu *et al.*, 2008), pea legumin (Lycett *et al.*, 1985), sunflower helianthenin (Nunberg *et al.*, 1995) etc. Unravelling the molecular basis of seed-specific gene expression has been mainly focused on the identification specific of *cis*-elements or elements within the control regions and the identification of transcription factors (TFs) associated with them. The cis-elements are relatively short sequences (signal/motif) in the region surrounding a gene; they vary in length, position, redundancy, orientation in DNA chain, and bases (Wasylyk, 1988; Johnson & McKnight, 1989; Mitchell & Tjian, 1989; McKnight & Yamamoto, 1992; Zawel and Reinberg, 1993; Fassler and Gussin, 1996). Identification and functional characterization these transcription factor binding site (TFBS) motifs or *cis*-elements in the upstream region of genes is therefore critical towards the understanding of the regulations of gene expression.

The study and increased understanding of gene promoters: their structure, function and mechanism of gene regulation will open up the possibility of modulation of gene expression in homologous as well as heterologous systems. The phenomenon of homology based gene silencing, frequently encountered in many transgenic systems (Jorgensen, 1995; Matzke and Matzke, 1995; Meyer, 1995), suggest that it is important to have an available selection of promoters, offering a range of alternative expression patterns. Lack of suitable gene promoters for driving expression of the heterologous genes in transgenes in the seed/endosperm is still a major limitation in obtaining the required level and pattern of expression. This emphasizes the need for concerted efforts to isolate genes and their promoters from indigenous crop plants so that transgenic development process is not hampered under the IPR regimes.

CHAPTER: II
REVIEW OF LITERATURE

Seed storage proteins are synthesized during the seed development and serves as the principal source of nitrogen for germination and seedling growth. Osborne (1924) pioneered the systematic study of plant proteins and categorised them on the basis of their solubility into (i) albumins, which are water soluble, (ii) globulins, which are salt soluble, (iii) prolamins, which are soluble in aqueous alcohol and (iv) glutelins which are soluble in dilute acid or alkaline solutions. In addition to this classification, seed storage proteins of crops have also been assigned specific names. Thus, *glycinin, glutelin, zein*, represent the specific SSP fractions from soybean, rice and maize respectively. While the albumins have sedimentation coefficient of 1.7 to 2S, the globulins have sedimentation coefficients ranging from 7-8S for vicilins and 11-13S for legumins. Albumins and globulins comprise the SSPs of dicots, whereas, prolamins and glutelins are major proteins in monocots.

The legumin subfamily of proteins has been originally described from seeds of Leguminosae (Derbyshire *et al.,* 1976). Seed storage proteins homologous to legumins are, however, widely distributed in both monocots and dicots and are selectively known as 11S globulins or legumin-like proteins (Margoliash *et al.,* 1970, Reichlin *et al.,* 1970, Dudman and Millerd 1975, Hagar *et al.,* 1995, Shutov *et al.,* 1998, Mishra and Green, 1990, Hakman *et al.,* 1990). These proteins are stored as aggregate bodies in vacuoles localized either in cotyledon or endosperm tissue of developing seeds (Casey, 1999). During germination vacuolar processing enzymes (VPE's) alter the proteins' conformation, opening them to unlimited proteolysis. Degradation of the storage globulins by these VPE's provides the developing plant with a supply of elements and amino acids essential to growth (Muntz *et al.*, 2002). The legumin-like proteins are present as hexameric complexes, with each subunit consisting of a heavy α-chain and a light β- chain; both the chains are processed as a single precursor- prolegumin, which is cleaved at the asparaginyl endopeptidase (Asn-Gly) recognition site; the two resultant polypeptides remain bound to each other by a disulphide bridge (Derbyshire *et al.,* 1976; Sun *et al.,* 1978; Moreira *et al.,* 1979; Dagalarrondo *et al.,* 1986; Melo *et al.,* 1994; Hara *et al.,* 1978; Dey and Mandal, 1993; Walburg and Larkins, 1983).

The globulin gene sequence encodes two cupin domains, and the gene product forms a radially symmetric homodimer. Each of these homodimers combines with two others to form a radially symmetric trimer (Hirano *et al.*, 1985). This is accomplished via non-covalent bonding between hydrophobic regions in a pocket formed by a helical region at either end of the bicupin structure. While the radially

(a)

(b)

symmetric trimer is the final quaternary product in case of the 7S globulin family, the trimers stack in pairs to form hexamers in the 11S legumins. The cupin domains found in the protein products of 7S and 11S genes share remarkable structural similarities at the tertiary level. Though the primary sequence structure seems to deviate significantly in several areas, the general cupin motif can be found in the peptide sequence of both domains. Each domain has a cupin motif followed by a helix and turn. The cupin motif is composed of two beta strands separated by an internal motif spacer; the three elements together forming the main cupin beta-barrel with the two beta-strands forming an anti-parallel jelly-roll, with the inter motif spacer forming a hairpin turn at its center point. The internal motif spacer shows considerable variations in its length between the 11S and 7S globulins, as well as across orthologous and paralogous copies of the globulins. The differences in the bonding pattern between cupin domains in the 11S and 7S globulin has been ascribed to the variations between the multimeric structure of these two classes of proteins and the need of cystein disulfide bridges between separate cupin domains to stabilize the overall structure (Rodin *et al.*, 1990; Adachi *et al.*, 2003). The high level of similarity in the tertiary structure of between cupin domains of various proteins has lead to the postulation that all cupin domain containing genes share a point of common origin (Dunwell *et al.*, 2000). The proposed models of domain duplication near the origin of the plant storage globulins also hint at a connection with some more ancient relationships with single cupin proteins (Dunwell *et al.*, 2001). The cupin fold, which describes the beta coil domain of the globulins, as well as the single cupin GLPs, is found in a wide array of proteins ranging from bacterial oxalate

decarboxylases, to fungal phosphatase isomerases. While some proteins with cupin domains have been reported to be associated with DNA binding and retrotransposition genes, functions of many others have not been elucidated as yet (Khuri *et al.*, 2001).

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 (Greenwood and Chrispeels, 1985; Sengupta-Gopalan *et al.*, 1985, Bray *et al.*,1987; Hall *et al.*, 1999; Baker *et al.*, 1988; Perez-Grau and Goldberg, 1989; Rahman *et al.*, 1983; Fujino *et al.*, 2001; Milisavljevic *et al.*, 2004; Jain, 2004). In cereals such as wheat, barley and maize, prolamin, the major storage protein is deposited only in the starchy endosperm, whereas albumins and globulins tend to be largely concentrated in the embryo and aleurone tissue (Payne, 1983). Gatehouse *et al.*, (1986) have demonstrated that in legumes the albumins and globulins were distributed mainly in the cotyledons and embryonic axis. Developmental regulated expression of SSP genes has been demonstrated in *Phaseolus* (Hall *et al.*, 1999; Sengupta-Gopalan *et al.*, 1985; Murray and Kennard, 1984), Brassica (Ellerstrom *et al.*, 1996; Stalberg *et al.*, 1993), Arabidopsis (Guerche *et al.*, 1990), Barley (Rahman *et al.*, 1983) and Sunflower (reviewed by Goldberg *et al.*, 1989), buckwheat (Fujino *et al.*, 2001). It has been suggested that the spatial and temporal control of SSP gene expression is exerted at the level of transcriptional (Bartels and Thompson, 1986; Sorensen *et al.*, 1989; Bustos *et al.*, 1991; Mandal and Mandal, 2000).

Table 2.1: The content of essential amino acids* in seed storage protein of common buckwheat (*Fagopyrum esculentum* Moench) as compared to other cereals.

Food Grain	**Lysine**	**Methionine**	**Tryptophan**	**Leucine**
Buckwheat	5.9	3.7	1.4	5.8
Amaranth	5.0	4.4	1.4	4.7
Wheat	2.6	3.5	1.2	6.3
Rice	3.8	3.0	1.0	8.2
Maize	1.9	3.2	0.6	13.0
FAO/WHO recommendation	5.5	3.5	1.0	7.0

*as percent of total protein

Seed storage proteins have attracted the attention of researchers mainly on account of their importance in human nutrition. Seed storage proteins, intended as a source of nitrogen during the initial stages of germination and seedling growth, constitute an important source of dietary proteins for human consumption. Although cereal grains and legume seeds are the major sources of vegetarian dietary proteins for human consumption, the nutritional quality of the proteins in both does not match the WHO standards for dietary proteins with a nutritionally balanced amino acid composition. While the major amino acid deficiency in legume seed proteins is their low content of sulphur containing amino acids, cereal proteins are deficient in lysine (Boulter, 1981; Shotwell and Larkins, 1989). Rout and Chrungoo (1996) have compared the amino acid composition of seed/grain storage proteins of some of the conventional crops with the WHO recommended values for a nutritionally balanced protein (Table 2.1). They have emphasized the nutraceutical importance of some underutilized crops like *Chenopodium*, grain amaranth and buckwheat.

Over the years, many attempts have been made to improve the level of essential amino acids in seed storage proteins of important crop plants through conventional breeding programmes (Larkins, 1983; Coulter and Bewely, 1990). However, in most cases the attempts have either led to a severe depletion in seed storage protein levels or abnormalities in seed development (Bliss *et al.*, 1972). A variety of barley Riso1508 produced in this way had very high lysine content but a severe depletion of the storage protein, hordein (Hermann and Larkins, 1991). The negative correlation between the seed protein content and the level of essential amino acids per unit protein has come as a major handicap in improving the amino acid

composition of seed proteins in crops. Because of inherent limitations in inter-specific hybridizations, molecular approaches have provided alternative strategies to conventional breeding programmes. One of the approaches in this direction has been to manipulate the regulation of amino acid biosynthesis to increase the abundance of a particular amino acid. Mutant selection and engineering of genes encoding key enzymes of amino acid biosynthetic pathways have been used to increase amino acids in crop plants (Zeh *et al.*, 2001; Gidamis *et al.*, 1995; Salbaach *et al.*,1995a). However, an increase in the level of free essential amino acids may not necessarily lead to an increase in the content of fixed amino acids, Falco *et al.* (1995) have suggested that the amino acids could be even be leached out from the plant tissues. Such free amino acids may even be lost during post harvest processing of the grains. Alternatives to the manipulation of regulation of amino acid biosynthesis have focused around either (i) the manipulation of the primary sequence of a gene by addition, substitution or deletion of nucleotides by site directed mutagenesis and expression of the altered gene in place of or in addition to the native gene (Lago *et al.*, 1990; Guerche *et al.*, 1990; Blechl and Anderson, 1996) or (ii) heterologous gene transfer across species barriers (Altenbach *et al.*, 1989, 1992; Saalbach *et al.*, 1995b; Muntz *et al.*, 1997; Molvig *et al.*, 1997; Townsend and Thomas, 1994.). Efficient manipulation of the amino acid composition of seed storage proteins also relies to a large extent, on the stability of the foreign proteins in a heterologous system. Stable expression of the altered/heterologous protein in the seed can be a difficult task owing to the complex biochemical processes associated with seed

storage protein assembly and accumulation (Lambert and Yarwood, 1992; Lending *et al.*, 1992).

As a first step towards production of transgenic plants with improved amino acid composition, 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 are a pre-requisite. Such proteins and their genes have been isolated from soyabean (Hill and Briendenbach, 1974), pea (Higgins *et al.*, 1986; Hoffman *et al.*, 1988), *Lupinus albus* (Melo *et al.*, 1994), rice (Takaiwa *et al.*, 1987; Krishnan and 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 (Koundal *et al.*, 1989, 1993; Khitha *et al.*, 1995; Kansal *et al.*, 1995; Mandaokar and Koundal, 1996, Saha and Koundal 1998); *Chenopodium* (Dey and Mandal, 1993); *Brassica* (Dasgupta and Mandal, 1991, Utsumi *et al.*, 1993; Dasgupta *et al.*, 1995; Ericson *et al.*, 1991); grain amaranth (Raina and Datta, 1992), buckwheat (Bharali, 2002). Most of the attempts aimed at nutritional enhancement of seed crops have, however, focussed on increasing the level of sulphur amino acids in legumes rather than the lysine content in cereals. The sulphur-rich SSPs identified as candidate for this approach are10 kDa and 21kDa zeins (Kirihara *et al.*, 1988; Chui and Falco,1995), 2S Brazil Nut Albumin (BNA) (Altenbach, 1987) and 2S albumin (SSA) from sunflower (Kortt *et al.*, 1991). Attempts have also been made to use *in vitro* mutagenesis to increase the methionine codons in genes that encode subunits of legume globulins (Nielsen *et al.*, 1989, 1995; Saalbach *et al.*, 1995; Hoffman *et al.*, 1988). Hoffman *et al.* (1988) inserted a 45bp high methionine coding DNA fragment

from *zein* into the 7S phaseolin gene of *Phaseolus vulgaris.* Even though transcribable mRNA was produced, there was no accumulation of mutant protein in the seeds. Since the 3D structure of phaseolin indicates that the additional methionine residues could cause misfolding of the modified phaseolin (Lawrence *et al.,* 1990), Hoffman *et al.,* (1988) concluded that the failure in accumulation of the mutant protein could be due to the problems associated with stability of the mutant protein in the heterologous system. Rao *et al.* (1994) used a rate limiting strategy to increase the lysine content of L-hordothionin in barley seed storage protein by mutation of appropriate residues of lysine. Based on three dimensional model of the protein, they identified surface residues amenable to substitution with lysine. This approach allowed the creation of a modified L-hordothionin protein that had about 27% lysine.

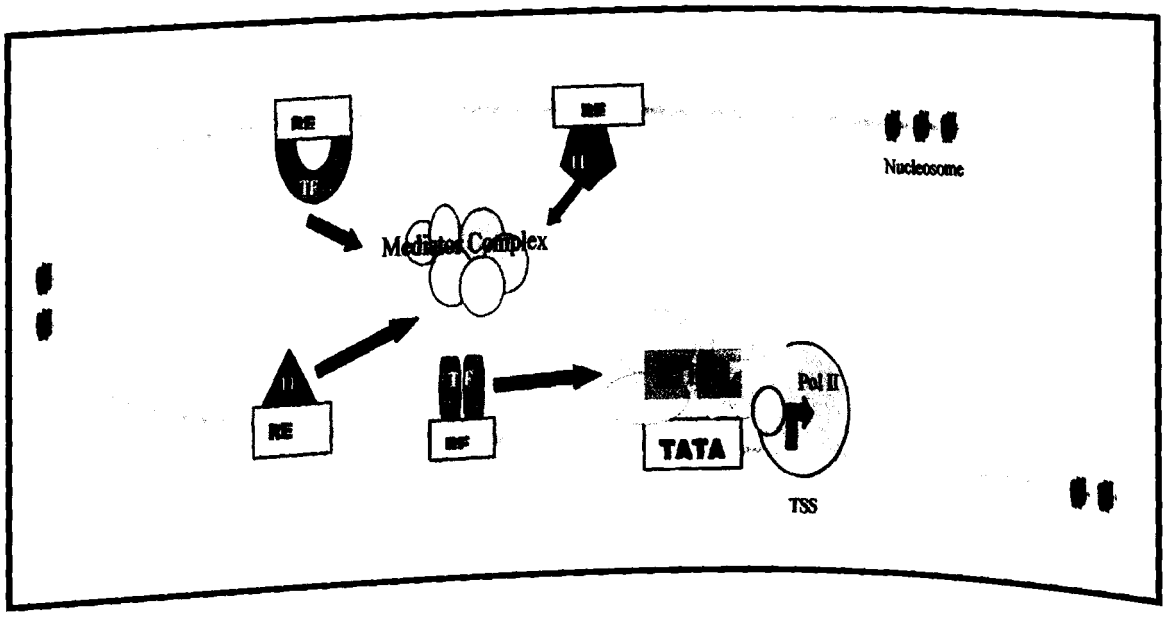
In 1987, Susan Altenbaach and her colleagues proposed that one way to increase the methionine content of legume seed protein and hence their nutritional quality was to introduce the sulphur-rich 2S albumin gene from Brazil nut for which they isolated a cDNA clone. The first successful experiment to increase the methionine content of legume seed proteins and hence their nutritional quality was reported by Salbaach *et al.*, 1995. They introduced the sulphur rich 2S albumin gene from Brazil nut into narbon bean and demonstrated a seed specific high level expression of the gene in narbon bean under the influence of LeB4 promoter. The foreign protein accumulated to approximately 3-8% of the total seed protein, resulting in upto 30% increase in the methionine content of seeds. Altenbach *et al.* (1992) transferred the 2S albumin gene from Brazil nut to canola and developed transgenic lines expressing the 2S albumin gene. They reported a 33% increase in the

content of salt extractable protein in transgenic canola. Salbaach *et al.* (1995) have developed transgenic lines of tobacco and narbon bean expressing the 2S Brazil nut albumin gene under the influence of (CaMV) 35S promoter. The presence of a constitutive promoter caused the expression of the transgene in all the tissues of the plant. Molvig *et al.* (1997) developed transgenic lines of lupin (*Lupinus agustifolius*) with the sunflower seed albumin (SSA) gene under the control of pea vicilin promoter. The transgenic lines accumulated SSA to approximately 5% of the salt extractable seed protein with a nearly two fold increase in the methionine content of the seeds. They also demonstrated a 15% increase in the biological value of proteins from the seeds of transgenic lines of lupin. The first report of significant increase in seed lysine content due to seed-specific expression of pea legumin heterologous gene in transgenic tobacco was reported by Keeler *et al.* (1997). Similar reports have been made in transgenic wheat by Casey *et al.* (2001), rice (Zhenweiz *et al.*, 1995). Zhenweiz *et al.* (1995) developed transgenic lines of rice expressing β-phaseolin gene under the influence of rice glutelin promoter. The transgenic plants accumulated β-phaseolin upto 4% of the total salt extractable endosperm proteins. These results have supported the possibility of producing novel plants with improved nutritional quality of their seed storage proteins.

Crop improvement by heterologous gene transfer essentially requires a promoter that would regulate the spatial and temporal expression of the trangenes. As a first step, potentially useful promoters need to be evaluated in view of their developmental stage-specificity, seed-specificity and expression levels. The study and increased understanding of gene promoters including their structure, function

and mechanism of gene regulation will open up the possibility of modulation of gene expression in homologous as well as heterologous systems. While many seed specific promoters have been identified and characterized not many promoters from genes of seed storage proteins have been characterized till date. Examples of such promoters include the 5' regulatory regions from such genes as cruciferin (Sjödahl *et al.*, 1995), napin (Kridl *et al.*, 1991), phaseolin (Bustos *et al.*, 1989), soybean beta-conglycinin (Chen *et al.*, 1986), maize zein (Matzke *et al.*, 1990; Thompson *et al.*, 1990; Brown *et al.*, 1986), rice glutelin (Takaiwa *et al.*, 1991; Qu *et al.*, 2008), pea legumin (Lycett *et al.*, 1985), sunflower helianthenin (Nunberg *et al.*, 1995) etc. The primary regulatory sequences are generally located within 1000bp upstream of the transcription start site (TSS) in plant genes, although there are cases where regulatory sequences are found further upstream (Zhang *et al.*,1996) or are downstream (3') (Dietrich *et al.*, 1992) of the coding sequences. The basic eukaryotic promoter sequence comprises of three regions: the core promoter, the proximal promoter, and the distal promoter. A typical core promoter encompasses DNA sequences between approximately -40 and +50 relative to transcription start site and consists of several sequence motifs which include TATA box, Initiator (Inr), TF IIB recognition element (BRE), and downstream core promoter element (DRE) (Struhl, 1987; Weis and Reinberg, 1992; Smale, 1994, 1997, 2001; Burke *et al.*, 1998). Transcription initiation of protein coding genes by RNA Polymerase II involves the stepwise assembly of general transcription factors (GTFs) to the core promoter to form a stable pre-initiation complex (Novina and Roy, 1996). The proximal promoter is about few hundred base pairs upstream of the TSS, containing

Fig. 2.1: Schematic representation of the assembly of transcription machinery on a typical eukaryotic promoter.


TF
Mediator Complex
Nucleosome
TATA
Pol II
TSS

several regulatory elements (*cis*-elements), while the distal promoter, which can range several thousands of base pairs upstream of the TSS, contains additional regulatory elements called enhancers, silencers and insulators (Roeder, 1996; Blackwood and Kadonaga, 1998; Carey and Smale, 2000; West *et al.*, 2002, Butler and Kadonaga, 2000). These elements contain recognition sites for a variety of sequence-specific DNA binding proteins (termed *trans*-acting factors) which act to regulate transcription by acting on the transcription complex. A diagrammatic representation of a typial promoter region is shown in Fig 2.1. Although gene expression is regulated at many levels, including chromatin packing (reviewed by Kingston *et al.*, 1996), transcription initiation, polyadenylation (reviewed by Wahle and Keller, 1996), splicing (reviewed by McKeown, 1992), mRNA stability (reviewed by Decker and Parker, 1994), translation initiation (reviewed by Kozak, 1992), it is generally thought that the single most important point of regulation is at initiation of transcription. Synthesis of seed storage proteins is primarily controlled at the transcriptional level and the seed-specific expression has been shown to be conferred upon the promoter regions of many storage protein genes (Devic *et al.*, 1996; Lee *et al.*, 2007; Moreno-Risueno *et al.*, 2008). Functional analysis of SSP gene promoters by expressing the complete genes or reporter gene expression analysis in transgenic systems have led to the conclusion that the promoters of SSP genes are directly involved in seed-specific expression of the SSP genes in the endosperm/embryo tissues of the seeds/grains (Sengupta-Gopalan *et al.*, 1985; Bäumlein *et al.*, 1987, 1991a; Murai *et al.*, 1983; Ellerstrom *et al.*, 1996). 5' serial deletion and nested internal deletion analyses of the upstream regions of SSP gene

promoters have revealed that a few hundred base pairs are sufficient for conferring tissue-specific regulation and temporal control on the expression of the SSP genes (Chen *et al.*, 1986, 1989, Radke et al., 1988; Colot *et al.*, 1987; Robert *et al.*, 1989; Shirsat *et al.*, 1989). However, additional upstream sequences have been shown to be important for enhanced activity (Goldberg *et al.*, 1989; Chen *et al.*, 1986; Shirsat *et al.*, 1989, 1989). The promoter elements of SSP genes studied most include coding for β-phaseolin, glycinin, conglycinin, helianthenin, nupin and legumin in dicots and zein, gliadin, hordein and glutelin in monocots have been well studied. Goldberg *et al.* (1989) have suggested that while the proximal elements upto -77 and -66 were essential for correct spatial and temporal expression of soybean lectin (*Le*1) and glycinin (*Gy*1) genes respectively, the more distal elements were required for their quantitative regulation. Bogue (1990), Thomas *et al.* (1991) and Nunberg *et al.* (1994) have suggested a bipartite organization of the regulatory elements in the 5' UTR of helianthinin gene with the proximal region (-116 to +24) involved in conferring seed specific expression of the gene and the more distal regions involved in refining and enhancement of the basic expression patterns conferred by the proximal region. This bipartite model applies to other dicot SSP genes as well. Similarly the proximal promoter region between -140 and +13 of β-conglycinin α'-subunit gene and between -125 and +1 of lectin (*dlec2*) gene have been reported to be sufficient for tissue specific expression of the genes (Chen *et al.*, 1986; Voelker *et al.*, 1987; Lessard *et al.*, 1993). On the basis of 5' deletion analyses of the 782bp upstream region of β-phaseolin gene fused to Gus reporter gene, Bustos *et al.* (1989, 1991) have identified three positive and two negative regulatory elements that are

necessary for spatial and temporal gene regulation of the β-phaseolin gene. Besides this Burow *et al.* (1992) have identified several *cis*-acting regulatory elements in the 782bp promoter region of the phaseolin gene. While temporal regulation of the gene expression has been reported to be manifested by regulatory elements in the -295/-107 region, spatial regulation of the expression is controlled by at least by three elements which include a major positive element between -295 and -228, a minimal promoter region between -64 and -14 and a negative element located -295/-107 of the TSS that represses phaseolin expression in stem and root (Burow *et al.*, 1992).

Nunberg *et al.* (1994) have identified at least three distinct sequences in the proximal promoter region (PPR) of *helianthinin* gene HaG3A which interact with nuclear proteins thereby regulating the expression of the gene. Of these, two motifs AGATGT ("A" motif) and TGATCT ("T" motif) occur twice in the helianthinin proximal promoter region. The "A" motif occurs at –111 (A2) and –58 (A1) and the "T" motif is situated at –83 (T2) and –41 (T1). Mutation of the A and T motifs in the PPR of the *helianthinin* promoter resulted in loss of GUS expression or ectopic GUS expression in non-embryonic tissues (Nunberg *et al.*, 1994). These results indicate an important negative regulatory component in the tissue specific expression of the gene. Presumably the A and T motifs are involved in such a negative regulatory loop. The third DNA binding motif identified in the helianthininin PPR is the "Y" box having the sequence CCAAAT. This box is similar to the C/EBP binding motif involved in communication between upstream enhancers and basal promoter elements (Rorth and Montell, 1992). These results suggest a bipartite structure for seed protein regulatory ensemble. The proximal promoter elements direct tissue

specific expression while the more distal elements enhance and modulate this basic pattern. Stalberg *et al.* (1993) have demonstrated that deletion of sequence between -1101 and -309 in the promoter region *nap*A gene coding for 2S SSP of *Brassica napus* resulted in increased reporter gene expression. The region between -309 to +44 was shown to be sufficient to direct high levels of correct tissue-specific expression of the gene. Stalberg *et al.* (1993) further demonstrated that while a 98bp deletion from position - 309 to position - 211 relative to the transcriptional start site decreased the expression, deletion upto -126 lead to complete inactivation of the promoter. Ellerström *et al.* (1996) have identified a region between -309 and -152 of TSS in the *nap*A promoter which was involved in regulating the quantitative expression of the gene. They have demonstrated that the *cis* element with similarity to ABRE overlapping with an E-Box was crucial for quantitative expression of the gene. On the other hand, deletion of the region containing $(CA)_n$ element increased promoter activity in both leaves and endosperm and a decreased its activity in the embryo indicating that this element is important for conferring seed specific expression by serving both as an activator as well as a repressor element (Ellerström *et al.*, 1996). Ezacurra *et al.* (1999) later confirmed that the seed-specific activity of the *nap*A promoter relied on the combinatorial interaction between the RY/G complex *viz.* the complex between −78 to −50 containing the two RY repeats and the G-box in *nap*A promoter and the B-box *viz.* ABA-responsive complex during seed development. Several investigations have been carried out to characterize the promoter regions of monocot SSP genes in in heterologous transgenic dicot systems as well as homologous monocot systems (Colot *et al.*, 1987; Robert *et al.*, 1989;

Zheng *et al.*, 1993; Wu *et al.*, 1998). A deletion series of the 5'UTR sequence of the LMW glutenin gene indicated that sequences present between 326bp and 160bp upstream of transcriptional start point were essential for endosperm-specific expression of the gene (Colot *et al.*, 1987). Similarly a 174bp element upstream from the transcription start site of maize *zein* gene has been shown to be essential for endosperm specificity and temporal control of the gene (Matzke *et al.*, 1990).

Unravelling the molecular basis of seed-specific gene expression has focused mainly on the identification specific of *cis*-elements and transcription factors (TFs) associated with them. Protein-protein interactions involving looping out of the intervening DNA have been suggested to be involved in regulation of gene expression in cases where the TF binding sites are located far from the promoter region (Adhya, 1989; Matthews, 1992). Recognition of these transcription factor binding site (TFBS) motifs or *cis*-elements in the upstream region of genes is therefore critical towards the understanding of the regulations of gene expression. On the basis of a comparison of the nucleotide sequences of 11S genes of broad bean, pea and soybean, a 28bp sequence located within 250bp upstream of the transcription start site (TSS) in the 5' flanking region of these genes has been found to be more conserved than most of the coding region (Baumlein *et al.*,1986). This region has been designated as "legumin box" and has the consensus sequence 5'TCCATAGCCATGCATGCTGAAGAATGTC3'. Legumin box has been implicated in the regulation of expression of legumin genes at the transcriptional level (Chamberland *et al.*, 1992). The core motif of the legumin box represents an alternating succession of purine and pyrimidine nucleotides known as "RY" motif

and consists of variants of the sequence CATGCATG. The RY motif is known to occur in multiple copies in the promoter region of genes coding for legumin type proteins in most legumes (Dickinson *et al.*, 1988) and cereals (Forde *et al.*, 1985). Consequently, the RY motif has become one of the most popular motifs for *cis* analysis of the promoter region of SSP genes. RY repeat motif has been identified as a key *cis* acting element for seed specific gene expression (Baumlein *et al.*, 1992; Chamberland *et al.*, 1992; Fujiwara and Beachy, 1994; Sakata *et al.*, 1997; Bobb *et al.*, 1997). While the 5' UTR of genes coding for 11-13S legumin type proteins are known to have a legumin box, that of genes coding for 7S vicilins have the "vicilin box" having the core sequence 5'GCCACCTCAT3' located within 150bp of the TSS (Gatehouse *et al.*, 1986). The vicilin box has been shown to be essential for embryo-specific expression of the 7S vicilin genes (Gatehouse *et al.*, 1986). Chandrasekharan *et al* (2003) have shown that out of the 4 RY elements present in *phas* promoter, mutation in the three distal RY elements lead to the expression of the reporter gene in the entire embryo including the radical. This could indicate that the three distal RY elements had a role in restricting the expression of the gene in the radical tissues of the seeds. While mutation in the proximal RY motif lead to reduced expression in embryo it abolished the expression of the gene in the radical. These results indicated the involvement of the three distal RY elements as a negative controller of legumin gene expression in the radicle. Similarly deletion of the RY motif in legumin and napin gene promoter abolished most of the seed-specific activity associated with the promoter (Baumlein *et al.*, 1992; Stalberg *et al.*, 1993; Ellerstrom *et al.*, 1996; Reidt *et al.*, 2000). On the basis of a comparison of the

promoter sequences of cereal prolamin genes a conserved region comprising of two conserved motifs *viz.* a 7bp 5'TGTAAAG3' element and a 9bp 5'(G/C)TGA(G/C)TCA(T/C)3', element located 300bp upstream of the transcriptional start site has been identified as an essential component of the promoter region of genes coding for prolamin type seed storage proteins (Kreis *et al.*, 1985; Forde *et al.*, 1985). This region has been designated as -300 element and is also known as the "endosperm box". While the 5'TGTAAAG3' element has been designated as the prolamin-box (P-box) or endosperm motif (E-motif), the 5'(G/C)TGA(G/C)TCA(T/C)3' element has been designated as the GCN4-like motif (Muller *et al.*, 1995). The bipartite prolamin box/ endosperm motif and the ACAA motif are important regulatory elements repeatedly found in the promoters of *SSP* genes (Takaiwa *et al.*, 1996; Albani *et al.,* 1997; Carbonero *et al.*, 2000; Diaz *et al.*, 2002). The prolamin box/ endosperm motif as well as the GCN4 and AACA motifs have been demonstrated to be essential for the regulation of expression of endosperm-specific genes (Zheng *et al.*, 1993; Takaiwa *et al.*, 1996; Yoshihara *et al.*, 1996; Mena *et al.*, 1998; Diaz *et al.*, 2005). The GCN4- like motif has been reported to form a palindromic structure and is the target of basic leucine zipper transcription factor (bZIP) proteins that belong to the Opaque2 subfamily (Albani *et al.*, 1997; On ˜ate *et al.*, 1999; Wu *et al.*, 2000; Onodera *et al.*, 2001). On the other hand, the prolamin box is recognized by the DOF class of zinc finger proteins (Mena *et al.*, 1998; Vicente-Carbajosa *et al.*, 1997) and the AACA element is recognized by MYB proteins (Suzuki *et al.*, 1998). The regulatory locus opaque2 (o2) has been shown to promote endosperm-specific expression of the 22 kDa and 19 kDa zein

storage protein genes in maize (Kodrzycki *et al.*, 1989). The o2 locus has been reported to encode a bZIP transcriptional activator (Lohmer *et al.*, 1991; Schmidt *et al.*, 1990; Ueda *et al.*, 1992; Yunes *et al.*, 1994) which binds to a hybrid G/A box (TCCACGTAGA) in the promoter of the 22 kDa zein gene (Schmidt *et al.*, 1992). Mutations in either the *o2* gene or in its binding sites within the endosperm box of zein promoters resulted in a dramatic reduction in the level of zein transcription thereby confirming its role in regulating the expression of *zein* genes (Schmidt *et al.*, 1992). The *cis*-elements function in concert to recruit trans- acting, DNA- binding proteins that will interact with RNA Pol II at the precise time and location needed for gene to become active. Most of the transcription factors involved in transcription of SSP genes have been identified by their mutant phenotypes. Thus, factors affecting seed storage protein accumulation and desiccation tolerance, such as LEAFY COTYLEDON I [LEC 1] (Kagaya *et al.*, 2005), LEAFY COTYLEDON 2 [LEC2] (Santos Mendoza *et al.*, 2005), FUSCA 3 [FUS3] (Bäumlein *et al.*, 1994; Keith *et al.*, 1994), ABSISSIC ACID INSENSITIVE 3 [ABI3] (Parcy *et al.*, 1997) have been shown to play crucial role during seed development. The loci for these mutants code for nuclear proteins that *trans*-activate seed storage protein synthesis. The ABI3 gene encodes a transcription factor, which is an ortholog of maize VP1 gene (Giraudat *et al.*, 1992). Further, ABI3 has been shown to *trans*-activate SSP transcription by binding to RY repeats in the promoter region of legumin genes (Reidt *et al.*, 2000). While the B3 domain of ABI3 has been shown to be essential for RY repeat mediated activation of the *napin* gene, the B2 domain of ABI3 is required for ABA-dependent activation through ABRE motif (Ezcurra *et al.*, 2000). Many genes

encoding for transcriptional factors that regulate the expression of storage protein genes have already been cloned (Hartings *et al.*, 1989; Schmidt *et al.*, 1992; Chern *et al.*, 1996).

Among the many environmental factors affecting SSP gene expression, plant hormone ABA stands out to be the most important one. It has already been established that ABA is a key regulator of gene expression during seed maturation (Marion-Poll, 1997; Phillips *et al.*, 1997). Accumulation of *napin* transcripts is known to be responsive to ABA (DeLisle and Crouch, 1989) and so is the reporter gene activity driven by a *napin* promoter (Jiang *et al.*, 1996). ABA has been shown to be required for modulation of helianthenin gene expression (Thomas *et al.*, 1991), accumulation of 12S cruciferin in *Brassica* embryos (Finkelstein *et al.* 1985) and continued synthesis of β-conglycinin in soybean cotyledons (Bray and Beachy, 1985). Promoter elements mediating ABA-responsive gene expression have been identified in seed storage protein genes (Hattori *et al.*, 1995; Shen and Ho, 1995; Vasil *et al.*, 1995; Ono *et al.*, 1996). Collectively these studies suggest that the expression of seed storage protein genes depends on a combinatorial array of distinct regulatory modules and a specific complement of *trans*- acting factors. Thus, each gene has a unique combination of *cis*-acting DNA sequences that function to direct its expression. Although the exact information to delineate a promoter and its regulatory elements requires experimental approaches like comparative sequence alignment, promoter deletions (Li *et al.*, 2001), substitutions (Harlow *et al.*, 1996) and linker scanning (Li and Shapiro, 1993), prior computational analysis of the sequence can serve as a guide to establish a platform for further promoter analysis. A

number of public databases and software tools are available for analysing the putative *cis*-acting motifs and regulatory elements in a promoter region (Wingender *et al.*, 1996; Fickett and Hatzigeorgiou, 1997; Rombauts *et al.*, 2003; Molina and Grotewold, 2005). Successful introgression of the target genes into crop plants requires the development of constructs carrying the target gene with an appropriate tissue specific promoter. Lack of suitable gene promoters for driving expression of the heterologous genes in transgenes in the endosperm is still a major limitation in obtaining the required level and pattern of expression. This emphasizes the need for concerted efforts to isolate genes and their promoters from indigenous crop plants so that transgenic development process is not hampered under the IPR regimes.

Amongst the existing known plant resources, the IPGRI and Consultative Group on International Agriculture have identified common buckwheat, grain amaranth and *Chenopodium* as important but underutilized nutraceutical crops which could be used as the genetic base for identification and isolation of suitable heterologous genes/promoters for use in crop improvement programmes aimed at improvement of nutritional quality of seed proteins. Common buckwheat (*Fagopyrum esculentum* Moench) grows extensively along the Himalayan foothills and is used by people living in these areas as a staple diet. The plant is a pseudocereal of high economic importance because of short growth span, capacity to grow on poor soils and the high protein content of its grains. The importance of the plant lies in the high protein content of its grains, short growth span and hardiness; besides the foliage is used as a green vegetable and is an important commercial source of the glucoside "rutin" which is an important drug used for treatment of hypertension and

coronary heart diseases. The main seed storage protein in the crop is a 13S globulin belonging to the legumin family of proteins having high level of essential amino acids such as lysine (Pomeranz *et al.*, 1975; Eggum *et al.*, 1981). The content of the protein is much higher than that reported for many other cereals and the amino acid composition matches the WHO recommended values for a nutritionally rich protein with a balanced amino acid composition (Rout land Chrungoo, 1996). Certain features that distinguish the basic subunit of buckwheat legumin from the basic subunits of other legumins of other plants include the high ratio of lysine to arginine and methionine to arginine. The ratio of lysine to arginine in buckwheat protein was found to be >1.0 as compared to that of soyabean, vicia and pea legumin where it is <1.0. The Met/Arg ratio of buckwheat protein is 2 fold higher than the basic subunit of soyabean glycinin and 4 fold higher than that of pea legumin. Due to the balanced amino acid composition and high nutrient value, the gene coding for this protein and its promoter could be an important candidate for compensation of limiting amino acids in plants deficient in such amino acids. The present investigation was therefore undertaken to isolate and characterize the endosperm/seed-specific promoter region of buckwheat legumin-like protein gene. Such endosperm/seed specific promoter would find application in:

1. Transgenic programmes aimed at improvement of the nutritional quality of conventional crops deficient in essential amino acids.
2. Elucidation of regulatory mechanisms controlling temporal as well as tissue-specific gene expression.

CHAPTER: III
MATERIALS AND METHODS

1. MATERIALS:

1.1 Plant Material:

Grains of common buckwheat (*Fagopyrum esculentum* Moench) [Accession No. IC18890] were procured from North Eastern Regional Station of National Bureau of Plant Genetic Resources, Shillong. The germplasm was multiplied in the experimental garden of the Department of Botany, North Eastern Hill University, Shillong.

1.2 Membranes and filters:

Positively charged nylon membrane (0.45μm pore size) for southern blotting, was purchased from Fluka. Membrane filters (cellulose nitrate, 0.2 μm) was purchased from Axiva.

1.3 Bacterial Strains:

Bacterial strains used for culture purposes included *E. coli* (DH5α) of the genotype *sup*E44 *lac*U 169 (φ 80 *lac*Z M15) *hsd R17 recA1 end*A1 gyr A9 thi-1 *rel*A1 and E. coil (TOP10) of geneotype F$^-$*mrc*A Δ(*mrr-hsd*RMS-*mcr*BC) Φ80*lacZ*ΔM15

Δ*lac*X74 *rec*A1 *ara*D139 Δ(*ara-leu*) 7697 *gal*U *gal*K *rps*L (Str^R) *end*A1 *nup*G. The bacterial strains were maintained as glycerol stocks with 15% glycerol (v/v) and stored at -80°C. For further use, the cells from the glycerol stocks were streaked on LB agar (amp^-) plate to generate single colonies.

2. PROTOCOLS:

2.1 Preparation of Competent Cells:

Cells of E.coli were made competent following the method described by Sambrook *et al.* (1989). A single colony of the host cells [*E.coli* (DH5α)] was cultured overnight in 50 ml of LB broth (amp^-) in an orbital shaker incubator at 37°C with continuous shaking at 200 rpm. 10 ml of the overnight grown culture was transferred to 50ml capacity polypropylene tube containing 30 ml of pre-warmed LB broth (amp^-). The cells were allowed to grow at 37°C with shaking at 200 rpm to A_{600} of 0.5 to 0.6. The optimally grown culture was chilled in ice for 2 hours and the cells were pelleted by centrifugation at 1000xg for 10 minutes. The pelleted cells were resuspended in 40 ml of ice cold 100mM calcium chloride and allowed to stand on ice for 45 minutes. The cells were pelleted again at 1000xg at 4°C for 10 minutes and the supernatant was discarded. The pelleted cells were resuspended in 5 ml of ice-cold 100mM $CaCl_2$; the cell suspension was distributed into aliquots of 170µl each in 1.0 ml microcentrifuge tubes followed by addition to 30 µl of sterile glycerol to make a 15% glycerol stock of cells. The glycerol stocks of competent cells were frozen in liquid nitrogen and stored at -80°C.

2.2 Transformation of *E. coli* cells with plasmid DNA:

Transformation of *E. coli* with plasmid DNA was carried out as per Sambrook *et al.* (1989). One vial of glycerol stock of competent cells (200 µl) was thawed on ice to

get the suspension of competent cells. 100-200 ng of plasmid DNA was added to the tubes containing competent cells of *E. coli* and the suspension incubated over ice for 30 minutes. The chilled mixture was subjected to heat shock at 42°C for 2 minutes and then chilled immediately over ice. 1ml of LB medium was added to the tubes containing the cell suspension and the cells were allowed to grow in the medium for 1 hour at 37°C. After incubation, the cells were concentrated by a brief spin of 30 seconds at 13,000 rpm, at 4°C. The supernatant was discarded and the cell pellet was resuspended in 200 µl of fresh LB medium. The cell suspension was plated on LB agar plates containing appropriate antibiotics and incubated in an incubator for 12 hour at 37°C. Appropriate positive and negative control plates were also incubated along with the experimental plates to check for contaminations, if any.

2.3 Isolation of Plasmid DNA:

Plasmid DNA was isolated on a mini scale from the overnight grown cultures of *E. coli* (DH5α) cells which had been transformed with the appropriate plasmid following the alkaline lysis method (Birnboim and Doly, 1979). Bacterial cultures were raised by inoculating a single colony of cells into 50 ml of LB broth containing appropriate antibiotics. The culture was allowed to grow overnight in an orbital shaker incubator at 37°C with orbital speed of 200 rpm. 1.5 µl of the overnight grown culture was centrifuged 4,000 rpm in a table top refrigerated centrifuge (Heraeus Biofuge Fresco) with 24x1.5ml rotor for 5 minutes at 4°C. The pelleted cells were resuspended in 100 µl of ice cold alkaline lysis buffer and incubated at room temperature for 5 minutes after which 200 µl of freshly prepared solution containing 0.2N NaOH and 1% SDS was added to the tubes. The suspension was mixed by inverting 4-5 times and the incubated

over ice for 5 minutes. Incubation over ice was followed by addition of 150 µl of ice cold 5M potassium acetate (pH 4.8) after which the tubes were again placed over ice for 5 minutes. At this step, when the pH of the suspension is lowered by the addition of potassium acetate solution and under high salt condition, the denatured bacterial genomic DNA precipitates out with the insoluble protein-SDS complex while the plasmid DNA is retained in the supernatant. The suspension in the tubes was centrifuged at 13,000 rpm for 5 minutes at 4°C in order to pellet the insoluble complexes and the clear supernatant containing the plasmid DNA was transferred to a fresh microcentrifuge tube. To remove the RNA, the plasmid DNA preparation was incubated with RNase A (added to a final concentration of 1µg ml^{-1}) at 37°C for 45 minutes. For further purification, equal volume of phenol: chloroform: isoamyl alcohol (25:24:1) was added to the RNase treated solution and the tubes were vortexed for 1 minute. The tubes were centrifuged at 13,000 rpm for 5 minutes at 4°C. The upper aqueous phase was transferred to a fresh tube and extracted ones with chloroform: isoamyl alcohol (24:1). Plasmid DNA was precipitated by addition of $1/10^{th}$ volume of 3M sodium acetate solution (pH 5.2) and two volumes of 100% ethanol, followed by incubation at -20°C overnight. Precipitated DNA was pelleted by centrifugation at 13,000 rpm in a microcentrifuge for 5 minutes, washed twice with 70% ethanol, vacuum dried and dissolved in nuclease-free ultra pure water.

2.4 Isolation of Genomic DNA:

Total genomic DNA was isolated from 14 days old etiolated seedlings by a modified CTAB extraction protocol (Murray and Thompson, 1980). Healthy grains of common buckwheat were surface sterilised by immersing in 0.01% $HgCl_2$ for 5 minutes followed by repeated rinsing with sterile distilled water. The washed grains were

germinated on sterile germination paper in a Plant growth chamber maintained at 27°C±2°C and 75± 10% relative humidity, under dark condition. The seedlings were harvested after 14 days of incubation in the plant growth chamber, washed with sterile distilled water, wrapped in aluminium foils, freeze dried in liquid nitrogen and stored at -80°C. The frozen tissue was crushed to fine powder in a sterile pestle and mortar under liquid nitrogen. 500 mg each of the powdered tissue was transferred to a 1.5 ml microcentrifuge tubes containing 500 µl of 2X CTAB buffer. The powder was mixed gently in the CTAB buffer to avoid formation of clumps and then incubated in a water bath at 65°C for 1 hour. The incubation was followed by addition of chloroform: isoamyl alcohol (24:1) in the ratio of 1:1, with gentle inversions to mix the cell lysate thoroughly with chloroform- isoamyl alcohol. The mixture was centrifuged for 5 minutes at 4°C in a Heraeus Biofuge (Fresco) table top refrigerated microcentrifuge to separate the aqueous phase, containing the nucleic acids, from the organic phase containing the proteins, carbohydrates and cell debris. The aqueous phase, containing the nucleic acids, was collected into a new tube and $1/10^{th}$ volume 3M sodium acetate (pH 5.2) and two volumes of 100% ethanol was added to it to precipitate the DNA. For complete precipitation, the mixture was allowed to stand at -20°C, overnight. The precipitated DNA was pelleted by centrifugation at 13,000 rpm for 5 minutes at 4°C. The pellet was washed twice with 70% ethanol, vacuum dried and dissolved in nuclease-free ultra pure water. The isolated DNA was electrophoressed on 0.8% agarose gel at 80V for 1 hour in 1X TBE buffer (pH 7.5). After the electrophoresis, the gel was stained with ethidium bromide solution (0.5µg ml^{-1}) for 15 minutes and destained in water till the background

fluorescence disappeared. DNA was visualised as fluorescent bands under UV light on a UV transilluminator.

Quality check of isolated DNA: For checking the quality of the isolated genomic DNA, 5 µl of the DNA solution was electrophoressed on 0.8% agarose gel along with *Eco*RI-*Hind*III digested Lamda DNA. A single band showing molecular mass of >21 kB on the agarose gel indicated a good quality preparation with no shearing of the isolated DNA. The purity of the DNA preparation was also checked by measurement of absorbance of the sample at 260 nm and 280 nm in Perkin Elmer Lamda35 UV/VIS Spectrophotometer and working out the A_{260}/A_{280} ratio of the sample. Samples with A_{260}/A_{280} of >1.8 were used for further experiments.

Quantity check of isolated of DNA: DNA was quantified by visual observation of ethidium bromide stained Agarose gel on a transilluminator and comparison of the intensity of fluorescence of DNA bands with fluorescence of known amounts of λ DNA electrophoressed along with the isolated DNA sample (Sambrook *et al.*, 1989). Quantification of the DNA sample was also carried out spectrophotometrically. 10 µl of DNA solution was mixed with 990 µl of ultrapure water and the absorbance of the solution recorded at 260 nm in Perkin Elmer Lamda35 UV/VIS Spectrophotometer. Concentration of DNA in the solution was calculated using the equation:

A_{260} x dilution factor x 50 = µg/ml DNA.

2.5 Polymerase Chain Reaction (PCR) Amplification:

PCR amplification of the target DNA was carried out with buckwheat genomic DNA as the template and primers derived from the conserved regions of nucleotide sequences of legumin-like protein genes available in the databases of gene banks. The

primers were designed using Primer3 online tool (Rozen and Skaletsky, 2000). The oligonucleotide primers used in the present investigation are listed in table 3.1. Amplification reactions carried out in 200 µl tubes in a Perkin Elmer thermal cycler with the reaction cycle comprising of one cycle of "hot start" (94°C, 5 min); 35 cycles of denaturation (94°C, 1 min); annealing (60-67°C, 1 min), and polymerization/primer extension (72°C, 1 min) and one cycle of chain elongation (72°C, 15 min). The PCR cycle is diagrammatically represented in Fig 3. 1. A typical 25 µl reaction volume (in 0.2 ml reaction tube) contained the following components:

Sterile water	: 15.2 µl
10X reaction buffer	: 2.5 µl
$MgCl_2$ (25mM Stock)	: 2.5 µl
10mM dNTP mix	: 0.5 µl
Primer 1 (5mM stock)	: 1.0 µl
Primer 1 (5mM stock)	: 1.0 µl
Genomic DNA template (~0.5µg)	: 2.0 µl
Taq DNA Polymerase (3 unit/µl)	: 0.3 µl
Total	**: 25.0 µl**

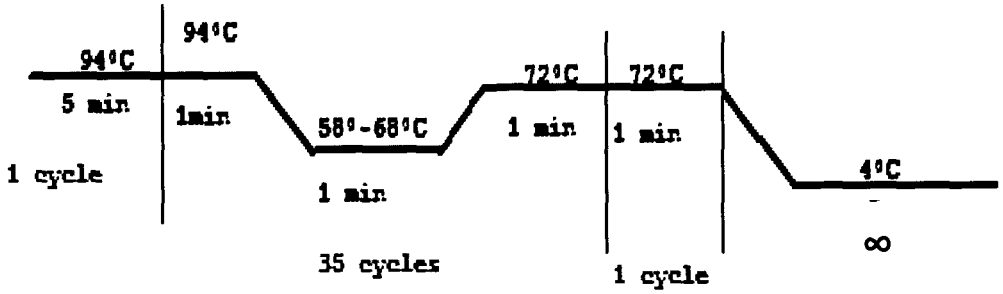
The set up included positive (without primers) as well as negative (without template DNA) controls. After the amplification reaction, the entire reaction mixture was electrophoressed on 1.2% Agarose gel at 80 V for 3 hours. After the electrophoresis was complete the gel was submerged for 15 minutes in ethidium bromide solution (0.5 µg ml^{-1}) followed by destaining in sterile distilled water. The DNA on the agarose gel was visualized under UV light on a transilluminator. Each amplification and subsequent agarose gel electrophoresis of the amplified DNA was repeated thrice to check reproducibility of the results. The region of agarose gel having the amplified DNA was excised from the gel with a sterile blade and the DNA was eluted from the agarose block

Table 3.1: List of primers used for PCR amplification of legumin-type seed storage protein gene and its promoter.

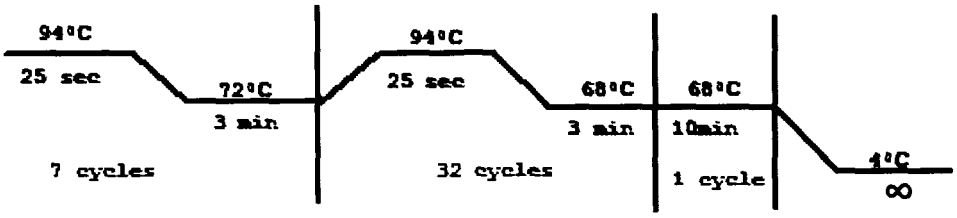
Primer	Sequence	Source
SS1R	5'GGAGAAGGTGGCCTTTCTCG 3'	BLAST analysis
SS2F	5'ATGCTTCATGGGGTGCTTCTATGC3'	BLAST analysis
SS3R	5'TTAAGACCTTCCTTCCGAAAGC 3'	BLAST analysis
SS4F	5'GGCGACACCAGCACACTGGAACTGAT3'	BLAST analysis
SS6R	5'GAAACGCTCCCTCTCCTTCTCATC3'	BLAST analysis
SS7R	5' GGTCCTACGGGCCTCTGTA3'	BLAST analysis
SS8F	5'CCAGGATGCCCGGAGACAT3'	BLAST analysis
SS10R	5'CGTCGAACACACTTCTTCCTTCATCTCC3'	BLAST analysis
SS11F	5' ATCGGAATTCGAGTACCC TCAGTCTCA3'	BLAST analysis
SS12F	5'AGCACCAGAAGATTTTCAGGATCAGAGA3'	BLAST analysis
SS13R	5'GAATTACGACACGAACGGCAACAAAT3'	BLAST analysis
SS14R	5'GCAGCTTAGTACCATAAGGCACAGTGA3'	BLAST analysis
SS15F	5'CTGCATGGATGTACCAGAGGGAAAATA3'	Present study
Ap1	5'GTAATACGACTCACTATAGGGC	Clontech
Ap2	5' ACTATAGGGCACGCGTGGT 3'	Clontech
SS18R	5' TAA(G/C)C(T/C)AGAAG(T/A)GGAAGAGG GA(G/C)TTG 3'	BLAST analysis
SS19R	5'CGAGAAATTCAC(C/T)CTTTTTATTGACA CT 3'	BLAST analysis

Fig 3.1: (a) Schematic representation of a typical three step cycle parameter for PCR amplification.

Fig 3.1: (b) Schematic representation two step cycle parameter used for PCR-based genome walking using Universal GenomeWalker Kit (Clontech).


94°C
5 min
1 cycle
94°C
1min
58°-68°C
1 min
35 cycles
72°C
1 min
72°C
1 min
1 cycle
4°C
∞

(a)


94°C
25 sec
72°C
3 min
7 cycles
94°C
25 sec
68°C
3 min
32 cycles
68°C
10min
1 cycle
4°C
∞

(b)

with Genei Spin Gel Extraction Kit (Bangalore Genei, India) as per the manufacturer's protocol. The eluted DNA was lyophilized and stored at -80°C till it was used.

2.6 Restriction Digestion:

For restriction digestion a 25 µl of the DNA solution was transferred to an eppendorf tube and made to 17 µl with sterile MiliQ pyrogen free water. 5 µl of appropriate buffer was added to the tubes followed gentle mixing of the solutions by repeatedly inverting the tubes 3-5 times. The reaction was started by adding 50 units of the restriction enzyme to the reaction mixture and incubating the tubes at 37°C for 4 hours in a circulatory water bath. A typical 50 µl restriction digestion mixture consisted of:

DNA (5 µg)	:	25.0 µl
10X Restriction enzyme buffer	:	5.0 µl
Restriction enzyme	:	3.0 µl
Ultra pure water	:	17.0 µl
Total	:	**50.0 µl**

For restriction digestion of plasmid DNA approximately 0.5 µg of plasmid DNA was incubated with 1.5U of the enzyme(s) with its respective buffer(s) at 37°C for 4 hours in a circulatory water bath. A typical 25µl digestion mixture consisted of:

Plasmid DNA (0.5µg)	:	5.0 µl
10X Restriction enzyme buffer	:	2.5 µl
Restriction enzyme (3U/µl)	:	0.5 µl
Ultra pure water	:	17.0 µl
Total	:	**25.0 µl**

The reaction was stopped by rapid heating of the reaction mixture at 65°C for 5 minutes in a dry heating bath. The digested DNA was immediately size fractionated by electrophoresis on 0.8% agarose gel at 50V for 6 hours. DNA was visualised by exposure of ethidium bromide stained gel to UV light under UV transilluminator.

2.7 Southern Blotting:

The size fractionated/ PCR amplified DNA was transferred from the gel to the to positively charged nylon membrane by alkaline transfer according to Southern (1975). After the electrophoresis was over, the gel was removed from the tray and soaked twice in 0.4M NaOH for 20 minutes each with gentle rocking at room temperature. The denatured DNA was subsequently transferred to positively charged nylon membrane in 0.4M NaOH solution under vacuum at 20 mbars for 90 minutes using Trans-Vac TE 80 vacuum blotter (Hoefer). Prior to transfer, the nylon membrane was cut to the size of the gel and moistened with distilled water for 5 minutes. A sheet of Whatman no. 3 blotting paper was saturated with 0.4 M NaOH, placed over the steel mesh on top of the lower chamber of the vacuum blotter. The nylon membrane was placed over the 3mm Whatman in a manner that it was positioned under the window of the silicon sheet mask placed over the blotting paper. The agarose gel was placed in the window of the silicon sheet directly over the membrane taking care that no air bubbles are trapped between the gel and the membrane. The blotter tank was filled with transfer solution (0.4M NaOH) to submerge the gel before applying the vacuum. Transfer was initiated by connecting the blotter to the vacuum pump to generate a negative pressure of 20 mbars. After the completion of the transfer, the gel was stained with ethidium bromide to check the extent of transfer. The nylon membrane was rinsed with 2X SSC buffer for 10 minutes and then air-dried. The DNA transferred to the membrane was covalently linked to the membrane by exposure to UV light at 254 nm in a UV-crosslinker.

2.8 Methylene blue staining:

Staining of the membrane for detection of molecular weight markers was proceeded by soaking it in 5% acetic acid solution for 15 minutes at room temperature. The membrane was subsequently transferred to staining solution comprised of 0.04% methylene blue prepared in 0.5M sodium acetate (pH 5.2). Staining was carried out at room temperature for 10 minutes. The stained membrane was destained by repeated rinsing with water till the background stain disappeared and the marker bands become visible. The membrane was air-dried and stored for further reference with autoradiograms.

2.9 Preparation of radio labelled probe:

The DNA probe used for hybridization with template DNA covalently linked to nylon membrane was radiolabelled with α-P^{32}[dATP] by random primed labelling using Random Primer Labelling Kit (Bangalore Genei) as per the manufacturer's protocol. The reaction mixture, in final volume of 25 µl, comprised of:

Denatured template DNA (25.0 ng)	: 10.0 µl
Labelling Buffer	: 2.5 µl
DTT	: 2.5 µl
Random Hexamer	: 1.0 µl
dCTP	: 1.0 µl
dGTP	: 1.0 µl
dTTP	: 1.0 µl
[α-P^{32}]dATP (~50µCi)	: 5.0 µl
Klenow Fragment	: 1.0 µl
Nuclease-free water	: 5.0 µl
Total	**: 25.0 µl**

For preparation of the labelling mixture, the DNA to be labelled was diluted to a final volume of 10 µl in a microfuge tube and denatured by heating in a water bath at 100°C for 10 minutes followed by quick chilling over ice. The eppendorf tube containing

the denatured DNA solution was centrifuged briefly to spin down the contents of the tube. 2.5 µl each of labelling buffer and DTT, 1.0 µl each of random hexamer, dTTP, dCTP, dGTP, 5.0 µl each of nuclease free water and [α-P^{32}]dATP (~50µCi) and 1.0 µl of klenow fragment were added to the tube and the mixture vortexed briefly to mix the contents. The mixture was incubated at 37°C for 1 hour in a dry-bath after which the radiolabelled DNA was separated from unincorporated dNTPs by spin chromatography on Sephadex-G25. The spin column was prepared by fitting a1.5 ml eppendorf tube to the tip of a 1 ml disposable syringe for collection of the eluent and packing the outer jacket of the syringe with Sephadex-G25 pre-soaked in TE Buffer. The packed column was briefly spun in a laboratory centrifuge to remove the excess buffer after which a fresh eppendorf tube fitted to its tip for collection of the labelled DNA after centrifugation. The column was loaded with the entire labelling mixture and subjected to centrifugation at 3,000 rpm for 3 minutes in a laboratory centrifuge. The eppendorf tube containing the eluted radiolabelled DNA was removed from the column for estimation of radioactivity associated with the DNA using a liquid scintillation analyser/counter (Perkin Elmer Tri-Carb 2810TR). Usually ^{32}P-labelled DNA with a specific activity of 5-6x10^{-7}cpu μg^{-1} was obtained.

2.10 Southern hybridization:

Prior to hybridization with the radiolabelled probe, the membrane carrying the DNA was incubated with 10 ml 5X SSPE buffer containing 100 mg ml^{-1} denatured Herring sperm DNA and 1% sarkosyl. Pre-hybridization was carried out in Pyrex hybridization bottles in a Hybridization oven at 55°C for 8 hours with continuous orbital rotation. Hybridization was initiated by addition of [α-P^{32}]dATP labelled DNA probe to

the tube containing the nylon membrane immersed in the pre-hybridization buffer. The membrane along with the buffer containing the radiolabelled DNA probe was incubated in the hybridization oven with continuous rotation for 16 hours at 55°C-68°C (depending on the required annealing temperature for DNA and the probe) to allow the probe to bind to the target DNA. The membrane was subsequently washed with SSC buffer under increasing stringency conditions in order to remove the unbound/ non-specifically bound probe from the membrane. The sequence of wash cycles was as under:

Wash buffer composition	**Duration (min.)**	**Temperature (°C)**
5X SCC containing 0.1% SDS	10	65
2X SSC containing 0.1% SDS	15	65
0.2X SSC containing 0.1% SDS	15	67
0.1X SSC	15	67

During each step of washing, radioactivity on the membrane was measured with a portable Gieger Counter. After washing, the membrane was wrapped in cling Saran Wrap and exposed to a Kodak X-OMAT X-ray film kept in between intensifying screens in a cassette for an appropriate duration in an ultrafreezer at -80°C. The film was subsequently developed in the developing solution for detection of regions where the radiolabelled probe had bound to the template DNA.

2.11 Construction of Genome walking library:

The genome walking library was constructed using Clontech Universal GenomeWalker Kit according to the manufacturer's protocol. The steps in library construction and subsequent amplification of included:

1. Restriction digestion of buckwheat genomic DNA with appropriate restriction endonucleases.
2. Ligation of digested genomic DNA to GenomeWalker Adaptors
3. PCR-based DNA Walking

Restriction digestion of genomic DNA: Genomic DNA isolated from etiolated seedlings of common buckwheat (*Fagopyrum esculentum* Moench) was digested separately with *DraI, StuI, Eco*RV, *and Pvu*II for construction of *DraI, StuI, Eco*RV, *and Pvu*II genomewalker libraries The digestion mixture comprised of:

Genomic DNA (0.05 µg/µl)	:	60.0 µl
Restriction enzyme buffer (10X)	:	10.0 µl
Restriction endonuclease (10 U/µl)	:	5.0 µl
Water	:	25.0 µl
Total volume	:	**100.0 µl**

The mixture was incubated in a circulatory water bath at 37°C for 2 hours to digest the DNA. The contents of the tube were vortexed at low speed for 5-10 seconds and the tubes were maintained in a circulatory water bath for 16-18 hours at 37°C to complete the digestion. The digested DNA was separated from other components of the digestion reaction reaction mixture by addition of TE buffer-saturated phenol (pH 8.0) in the ratio of 1:1. The mixture was vortexed at slow speed for 5-10 seconds to mix the components of the solution and then centrifuged at 13,000 rpm for 5 minutes in a Heraeus biofuge microcentrifuge with 22x1.5 rotor to separate the aqueous and the organic phases. The aqueous phase was transferred to a new tube and extracted with equal volume of chloroform: isoamyl alcohol (24:1) in the ratio of 1:1. The mixture was centrifuged again at 13,000 rpm for 5 minutes in a microcentrifuge to separate the aqueous and the organic phases. The aqueous phase was transferred to a new tube and DNA was precipitated from the solution by addition of 2 volumes of chilled 100% ethanol and $1/10^{th}$ volume of 3M sodium acetate solution (pH 5.2). The mixture was kept overnight at -20°C for complete precipitation of DNA. The precipitated DNA was pelleted by centrifugation at 13,000 rpm in a microcentrifuge for 10 minutes, washed twice with 70% ethanol to

remove traces of salts, vacuum dried and dissolved in 20 µl of nuclease-free ultra pure water.

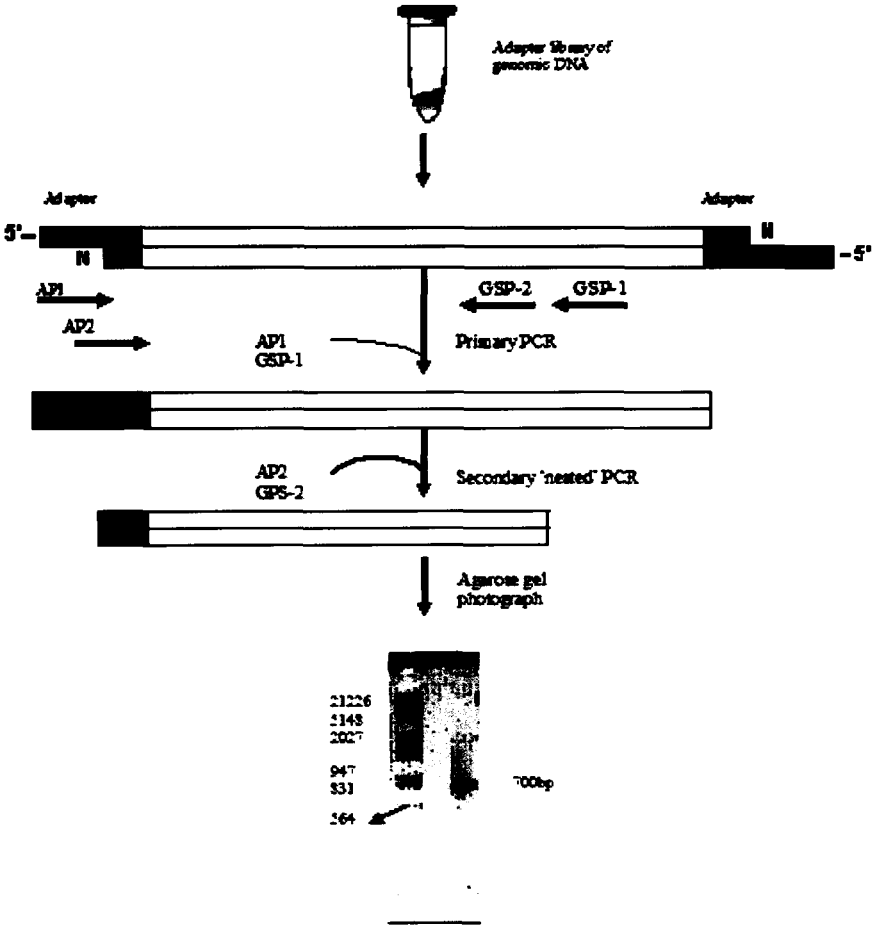
Ligation of digested genomic DNA to GenomeWalker Adaptors: Genome walker libraries for DNA digested with *DraI, StuI, Eco*RV, *and Pvu*II were constructed by ligation of the digested genomic DNA with genome walker adaptors (Clonetech) according to the manufacturer's protocol. The reaction mixture for generation of genome walker libraries consisted of :

Digested DNA (0.3 µg)	:	4.0 µl
GenomeWalker Adaptor (25 µM)	:	1.9 µl
10X Ligation Buffer	:	1.6 µl
T4 DNA Ligase (6U/µl)	:	0.5 µl

The mixture was incubated at 16°C for 16 hours in a circulatory water bath to ligate the adaptors to the 5' and 3' ends of the digested DNA. The reaction was stopped by heat denaturation at 70°C for 5 minutes in a dry heating bath. The mixture was made upto 80 µl with nuclease-free ultra pure water and then vortexed at low speed for 5-10 seconds. The vortexed mixture was centrifuged at 1000 rpm in a microcentrifuge to spin down the contents. The libraries were maintained at -20°C for further use.

PCR-based DNA Walking in GenomeWalker Libraries: PCR based genome walking on Genome walker libraries was carried out according to the manufacturers protocol which comprised of the initial step of primary amplification using the adaptor primer AP1 provided along with the kit and the gene specific primer 1 (GSP1) and the 2nd step of nested PCR with the amplification product of the primary amplification reaction as the template with adaptor primer AP2 and gene specific primer 2 (GSP2) as the oligonucleotide primers.

Fig 3.2: Strategy for PCR-based gene walking using Universal GenomeWalker Kit from Clontech


Adapter library of genomic DNA
Adapter
Adapter
5'
N
N
5'
AP1
AP2
GSP-2
GSP-1
AP1
GSP-1
Primary PCR
AP2
GPS-2
Secondary 'nested' PCR
Agarose gel photograph
21226
5148
2027
947
831
564
700bp

The amplification mixture having a total volume of 50 µl for the primary PCR of the adapter libraries comprised of:

Sterile water	:	36.5 µl
10X *Taq* Assay Buffer	:	5.0 µl
$MgCl_2$ (25mM Stock)	:	2.5 µl
dNTP mix (10mM)	:	0.5 µl
Primer AP 1 (10µM stock)	:	1.0 µl
Gene specific Primer 1 (5µM stock)	:	2.0 µl
DNA-Adapter Library	:	2.0 µl
Taq DNA Polymerase (3 unit/µl)	:	0.5 µl

PCR amplification was done using the two-step cycle parameters, which is diagrammatically represented in Fig 3.2. After the amplification reaction, an aliquot of the reaction mixture was electrophoressed on 0.8% Agarose gel at 80 V for 3 hours. After the electrophoresis was complete the gel was submerged for 15 minutes in ethidium bromide solution (0.5 µg ml^{-1}) followed by destaining in sterile distilled water. The DNA on the agarose gel was visualized under UV light on a transilluminator.

2 µl of the amplified mixture from the primary PCR reactions were diluted with water to a volume of 50 µl and 2µl of the diluted amplified mixture from the primary PCR was used as the template for the secondary (nested) PCR. The amplification mixture having a total volume of 50 µl for the secondary (nested) PCR comprised of:

Sterile water	:	37.5 µl
10X *Taq* Assay Buffer	:	5.0 µl
$MgCl_2$ (25mM Stock)	:	2.5 µl
dNTP mix (10mM)	:	0.5 µl
Primer AP 2 (10 µM stock)	:	1.0 µl
Gene specific Primer 2 (10mM stock)	:	1.0 µl
Diluted primary PCR product	:	2.0 µl
Taq DNA polymerase (3 unit/µl)	:	0.5 µl

The PCR cycle parameters were same as that of the primary PCR reaction. After the amplification reaction, the entire reaction mixture was electrophoressed on 1.5%

agarose gel at 80 V for 3 hours. After the electrophoresis was complete the gel was submerged for 15 minutes in ethidium bromide solution (0.5 μg ml^{-1}) followed by destaining in sterile distilled water. The DNA on the agarose gel was visualized under UV light on a transilluminator. Each amplification and subsequent agarose gel electrophoresis of the amplified DNA was repeated thrice to check reproducibility of the results.

The amplified DNA was eluted from the agarose gels Elution of DNA from the agarose gels with Genei Spin Gel Extraction Kit (Bangalore Genei, India) as per the manufacturer's protocol. The region of agarose gel having the amplified DNA was excised from the gel with a sterile blade; the excised gel pieces were weighed and immersed in 3(w/v) volumes of gel solubilization buffer containing a drop of 3M sodium acetate solution (pH 5.0). The mixture was incubated at 50°C till the agarose pieces dissolved completely in the solublization buffer. Isopropanol was added to the solution in the proportion of 100 μl for every 100 μg of agarose gel transferred to the tube. The solution was subjected to spin chromatography on a spin column provided along with the kit. The column was fitted to the collection tube provided with the kit for collection of the eluent, loaded with the solublization solution and centrifuged at 10,000 rpm for 1 minute at 20 °C. The flow through was discarded and the column was placed back in the same collection tube. The column was now loaded with 700 μl of the diluted wash buffer and centrifuged again for 1 minute at 10,000 rpm. The flow through was discarded and the column centrifuged again 2 minutes at 10,000 rpm to ensure complete removal of the wash buffer. The collection tube was removed and the spin column fitted with 1.5 ml microfuge tube at the tip for collection of the eluent. 50 μl of water was added to the

centre of the membrane in the spin column and the whole setup allowed to stand at room temperature for 5 minutes in order to solubilize the DNA in water. The set up was centrifuged for 2 minutes to recover the DNA solution. The eluted DNA was lyophilized and stored at -80°C till it was used.

2.12 Cloning of PCR amplified DNA:

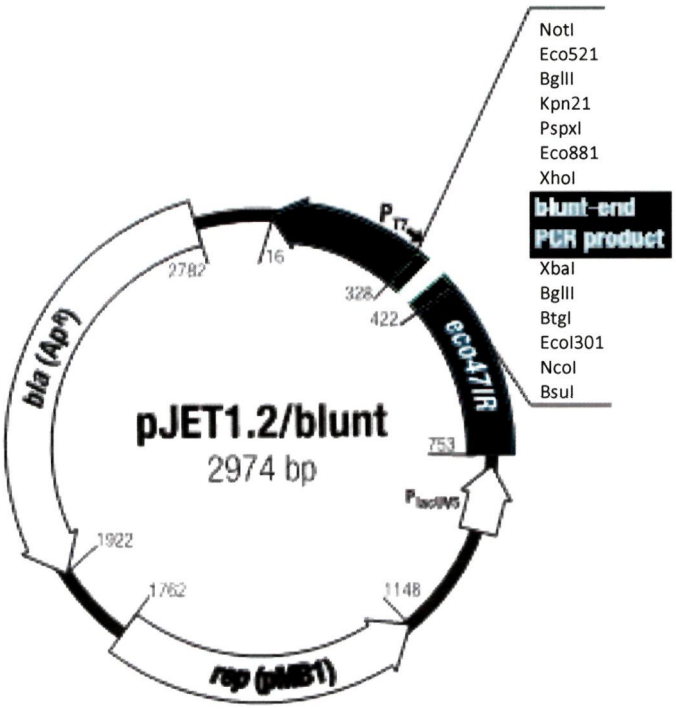
The eluted PCR products were cloned into the polyclonal clonal site of pJET1.2/blunt cloning vector (Fermentas) using CloneJET PCR clonng kit according to the manufacturer's protocol. Prior to cloning the purified PCR fragment was blunted with DNA blunting enzyme supplied along with the kit according to the manufacturer's protocol. The 18 µl blunting reaction mixture contained:

2X reaction buffer	:	10.0 µl
PCR product (50 ng)	:	2 .0 µl
DNA blunting enzyme	:	1 .0 µl
Nuclease-free water	:	5.0 µl

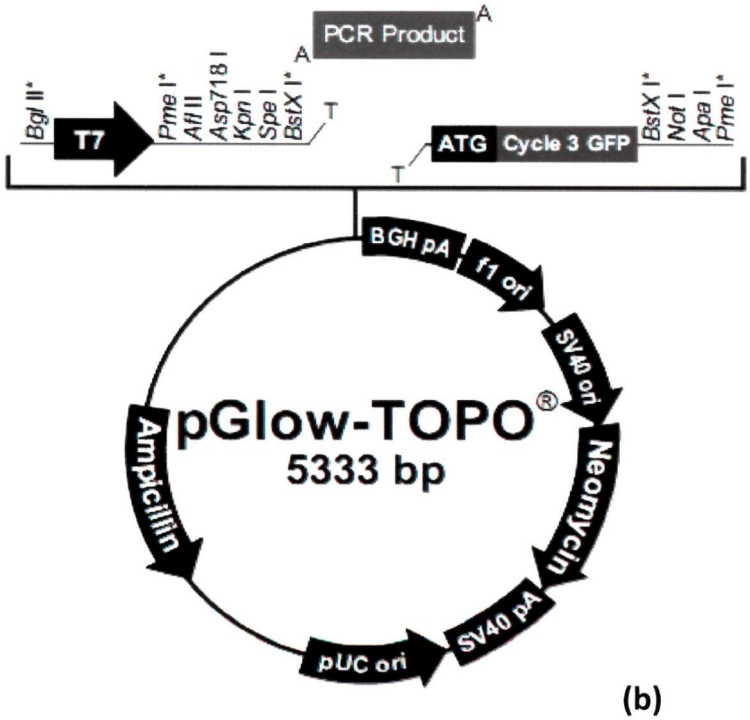
The mixture was vortexed briefly and incubated at 70°C for 5 minutes in a circulatory water bath and then chilled briefly on ice. The cloning reaction was started by addition of 1 µl of pJET1.2/blunt cloning vector and 1 µl of T4 DNA ligase (5U/µl) followed by a incubation for 5 minutes at room temperature. The ligation mixture was used directly for transformation of competent cells of *E. coli* (DH5α) with the recombinant vector as described under section 2.2. The transformation mixture was spread on LB-Agar (amp^+) plates and cultured overnight at 37°C in an incubator. For checking the recombinant clones for the presence of insert DNA, 50 ml aliquots of LB broth (amp^+) were inoculated with single colonies of the host cells and the cultures were incubated at 37°C overnight in an orbital shaker incubator with continuous orbital shaking at 200 rpm. Plasmid DNA was isolated from the overnight grown cultures

Fig 3.3: (a) Map and features of pJET1.2/blunt vector (Fermentas) used for cloning of PCR products.

Fig 3.3: (b) Map and features of pGlow-TOPO (Invitrogen), a promoterless reporter vector with GFP as the reporter gene.


NotI
Eco521
BglII
Kpn21
PspxI
Eco881
XhoI
blunt-end
PCR product
XbaI
BglII
BtgI
EcoI301
NcoI
BsuI
pJET1.2/blunt
2974 bp
bla (Ap^R)
eco47IR
rep (pMB1)
2782
16
328
422
753
1922
1762
1148

(a)


PCR Product
A
A
Bgl II*
T7
Pme I*
Afl II
Asp718 I
Kpn I
Spe I
BstX I*
T
T
ATG
Cycle 3 GFP
BstX I*
Not I
Apa I
Pme I*
BGH pA
f1 ori
SV40 ori
Neomycin
SV40 pA
pUC ori
Ampicillin
pGlow-TOPO®
5333 bp

(b)

following the protocol described under section 2.3. The isolated plasmid DNA was digested with *Not*I and *Nco*I following the protocol described under section 2.6 to excise the cloned insert DNA from the vector. The restriction digestion mixture having a total volume of 10 µl comprised:

Plasmid DNA (0.5 µg)	:	5.0 µl
10X reaction buffer	:	1.0 µl
*Not*1 enzyme (3U/µl)/ *Nco*I enzyme (3U/µl)	:	0.5 µl
Water	:	3.5 µl
Total	:	**10.0 µl**

This mixture was incubated at 37°C for 4 hours in a circulatory water bath. The reaction was stopped by heat inactivation at 65°C for 5 minutes. The restriction digestion was carried out in two steps, with two different enzymes. The linearized plasmid DNA from the Ist digestion step was extracted with TE buffer saturated phenol-chloroform-isoamyl alcohol (25:24:1) in the ratio of 1:1. The DNA was precipitated from the aqueous phase by addition of $1/10^{th}$ volume 3M sodium acetate (pH 5.2) and two volumes of 100% ethanol as described under section 2.3. The precipated DNA was washed twice with 70% ethanol, dried and sissolved in minimal amount of nuclease free water. The linearized plasmid DNA was digested with the 2^{nd} restriction endonuclease to excise the insert DNA following the same procedure as described above.

The digestion mixture from the 2^{nd} digestion step was directly electrophoressed on 1.2% agarose gel at 80V for 1 hour along with the undigested plasmid DNA which served as the control. DNA bands were visualized by exposure of ethidium bromide stained agarose gel to UV light in a UV transilluminator. The presence insert DNA was confirmed by the differences in the DNA profiles as well as the mobility of the plasmid DNA between the undigested and digested plasmid DNA. The presence of the insert

DNA in the vectors was also confirmed by PCR amplification of the insert from the recombinant clone using the pJET1.2 forward and reverse sequencing primers provided with the kit.

2.13 Nucleotide sequencing and sequence analysis:

Nucleotide sequencing of the amplified DNA was carried out by automated sequencing service rendered by Bangalore Genei and Axygen, India. For nucleotide sequencing, the amplified DNA was gel purified, lyophilized and sent for sequencing to the nucleotide sequencing facility at Bangalore Genei and Axygen, India. Nucleotide sequence of the coding as well as the non-coding strands of the double stranded DNA was carried out by using appropriate sequence specific primers. The nucleotide sequence was subjected to BLAST analysis to determine the extent of homology of the nucleotide sequence with the sequences in database.

2.14 5' Progressive Deletion and Cloning into Reporter Vector:

Unidirectional deletion fragments of the putative promoter region of the endosperm specific legumin were generated by PCR with forward primers (DF1, DF2, DF3, DF4, DF5, DF6 and DF7) and the reverse primer (DR1). The reverse primer, DR1 was designed in such a way so as to include a ribosomal binding site (-AGGA-) in the PCR product. The position of the primers on the nucleotide sequence of the putative promoter is shown in Fig 3.3.

The DNA fragments generated as a consequence of PCR with forward primers designed for different regions of the 5' upstream region of buckwheat legumin gene and the forward primer were purified using Qiagen gel extraction and purification kit (Qiagen) as per the manufacturer's protocol. Promoter driven Green Fluorescent Protein

(GFP) gene constructs was created by subcloning 1028 bp fragment containing the 5' upstream region of buckwheat legumin gene, being the putative promoter, as well as the unidirectional deletion fragments of the 1028 bp 5'uptream region of the buckwheat legumin into the TOPO cloning site of linearized promoterless expression vector, pGlow-TOPO (Invitrogen) as per the manufacturer's protocol. The reaction mixture with a total volume of 6.0 µl comprises:

PCR product	:	2.0 µl
Salt solution	:	1.0 µl
Sterile water	:	2.0 µl
pGlowTOPO vector	:	1.0 µl
Total	**:**	**6.0 µl**

The reaction mixture was gently mixed and incubated for 5 minutes at room temperature. 2 µl of ligation mixture was then used for transformation of TOP10 competent cells supplied along with the kit as described under section 2.2. The transformed cells were plated on LB-agar (amp^+) plates and incubated at 37°C for 48 hours in an incubator.

Fluorescence as a result of expression of GFP gene in each of the clones was observed in Carl Zeiss fluorescence microscope using filter no. 09 with excitation at 450 to 490 nm and emission at 515 nm. For this purpose cells from individual colonies were picked up with a sterile toothpick and dispersed in a drop of sterile water placed over a sterile glass slide. The cell suspension in the glass slide was covered with a cover slip and viewed under fluorescence microscope. Images were captured by CDD camera fitted to the microscope.

2.15 Isolation of Total RNA:

Total RNA was isolated from grains of common buckwheat harvested at early-, mid- and late- maturation stages as per the protocol described by Hosein (2001) with minor modifications. The harvested grains were rinsed thoroughly with water treated with DEPC (0.2% DEPC in water), freeze dried in liquid nitrogen and stored at -80° C till further use. All the glass-wares and plastic-wares, used for RNA isolation were given prior DEPC treatment. All aqueous solutions, except Tris-HCl, were treated with DEPC (final volume 0.2%) and autoclaved (Sambrook *et al*, 1989). About 1g of seed was crushed in liquid nitrogen to fine powder in a DEPC treated pestle and mortar. 500 mg of powdered tissue was transferred to a microcentrifuge tube and 700 µl of 100mM Tris-HCl (pH 9.0) was added to it. The mixture was vortexed for 30 seconds to disperse the clumps of powdered mass formed during the addition of Tris buffer. $1/10^{th}$ volume of 5% (W/V) SDS was added to the tubes and the mixture shaken on ice for 2 minutes. Incubation over ice was followed by addition of equal volume of phenol (pH 8.0) and periodic vortexing to mix the aqueous and organic phases. The tubes was allowed to stand over ice for 10 minutes to allow settling of phases and then centrifuged at 13,000 rpm for 2 minutes at 4°C in Heraeus Biofuge (Fresco) table top refrigerated microcentrifuge to separate the aqueous and the organic phases. The upper aqueous layer was transferred to new tube and re-extracted with phenol, several times, until no residue was observed at the interface between the aqueous and organic phases. Final extraction was done with equal volume of phenol: chloroform: isoamyl alcohol (25:24:1). The aqueous phase was transferred to a fresh microcentrifuge tube and LiCl solution to a final concentration of 2M added to it. The tubes were inverted several times to mix the contents and then kept for 1 hour in an ultrafreezer at 80°C to precipitate RNA. The

precipitated RNA was pelleted by centrifugation at 13,000 rpm for 10 min at 4°C, the supernatant was discarded and the pellet resuspended by vortexing in 500 µl of 2M LiCl. The tubes were centrifuged again at 13,000 rpm for 10 minutes and the pellet washed again 2M LiCl. The RNA pellet was finally resuspended in 400µl of DEPC-treated water and $1/10^{th}$ volume of 3M sodium acetate (pH 5.2) and 2 volumes of absolute ethanol were added to it. The solution was left overnight at -20°C for precipitation of RNA. The precipitated RNA was pelleted by centrifugation at 13,000 rpm in a microcentrifuge for 10 minutes, washed twice with 70% ethanol, vacuum dried and dissolved in DEPC-treated water. The isolated RNA was electophoressed on 0.8% agarose gel at 50V for 3 hours in 1X TBE buffer (pH 7.5). After the electrophoresis was over, the gel was stained with ethidium bromide (0.5µg ml^{-1}) for 15 minutes and destained in water till the background fluorescence disappeared. RNA was visualised by exposure to UV on a UV transilluminator.

RNA isolated from was dot blotted on positively charged nylon membrane as per Sambrook *et al.* (1989). RNA samples were first denatured in a reaction mixture comprising of:

RNA (1 µg)	:	10.0 µl
20X SSC	:	10.0 µl
37% Formaldehyde	:	7.0 µl
Formamide	:	20.0 µl

Denaturation was carried out at by heating the mixture at 65°C for 10 minutes in a dry heat bath followed by rapid cooling over an ice bath. The nylon membrane was saturated with DEPC-treated water and assembled into the blotter apparatus. Denatured RNA solution was loaded on the slots on the membrane using the dot blot manifold under

and negative pressure of 20 mbars. After loading the denatured RNA on the membrane the dot blot manifold was disassembled and the membrane with the denature RNA was rinsed with 2X SSC buffer. The rinsed membrane was air dried and DNA cross-linked to it by exposure to UV light at 254 nm in a UV-crosslinker. The membrane was subsequently used for hybridization of the RNA with appropriate radiolabelled DNA probes.

3. CHEMICALS AND SOLUTIONS

3.1 Chemicals:

Bacto-trypton, yeast extract, agar, Tris base, glycerol, EDTA, sodium chloride, calcium chloride, SDS, tri-sodium citrate, sodium hydroxide, sucrose, sarkosyl, ethidium bromide, potassium acetate, PVP, boric acid, bromophenol, ß-mercaptoethanol, sorbitol, lithium chloride, MOPS, formamide, formaldehyde, 8-hydroxy quinoline and ampicillin were purchased from Hi-media Laboratories Ltd.. Isopropanol, phenol, chloroform, glacial acetic acid, isoamyl alcohol, and other chemicals of routine use were purchased from Sisco Research Laboratories Pvt. Ltd. India. Agarose was purchased from Bioline and CTAB was procured from Bangalore Genei Pvt.Ltd. (India). [α-P^{32}]dATP was purchased from Bhabha Atomic Research Centre, Mumbai.

3.2 Molecular weight markers:

DNA molecular weight markers which included λ DNA *Eco*RI/*Hind*III double digest, 100bp ladder and pBR322/*Hinf* I digest and were purchased from Bangalore Genei Pvt. Ltd, India.

3.4 Enzymes and reagent kits:

Restriction endonucleases (with restriction buffers) and ribonuclease A were procured from Bangalore Genei Pvt Ltd. (India). Random Primer Labelling Kit was purchased from Bangalore Genei Pvt. Ltd. (India). Spin Gel Extaction Kit were purchased from Bangalore Genei Pvt. Ltd (India) and Qiagen Inc. (UK). PCR Cloning Kit was purchased from Fermentas Life Sciences (USA) and Universal GenomeWalker Kit was purchased from Clontech Laboratories Inc. (California). TOPO Reporter Kit was purchased from Invitrogen Corporation (USA).

3.5 Oligonucleotide primers:

The oligonucleotide primers were synthesized at M/S Bangalore Genei Pvt Ltd. (India) as lyophilized powders. Prior to use the oligonucleotide primers were dissolved in appropriate amount of nuclease-free water to make 100 µM stock solutions and aliquoted into 5 µM working solutions. The aliquots were stored at -20°C.

3.6 Deoxyribonucleotides (dNTP) solutions:

dATP, dTTP, dGTP and dCTP were procured from Bangalore Genei, India as 10mM stock solutions. The stock solutions were mixed together and diluted with appropriate amount of nuclease free ultra pure water to give a 5 mM dNTP mix working solution.

3.7 *Taq* DNA polymerase:

Taq DNA polymerase (3units/µl) was purchased from M/S Bangalore Genei, India.

3.8 Buffers and Solutions:

TE Buffer, pH 8.0:
10 mM Tris-Cl, pH 8.0
1 mM EDTA pH 8.0

5X Tris-borate (TBE) buffer (pH 7.5):
800 ml water
20 ml of 0.5M EDTA, pH8
54 .0 gm Tris base
27.5 gm boric acid

The components of the respective buffer were mixed together and made upto 1000 ml with distilled water. The buffers were stored in glass bottles at room temperature.

Gel loading buffer (5X):

25 mg of bromophenol blue was dissolved in 50 ml water and 50 ml of glycerol was added to it. The pH of the solution was adjusted to 7.0 and the final volume was made to 100 ml with water. The solution was filter sterilised and stored at -20°C.

2X CTAB buffer:

100mM Tris-HCl, pH 8.0
20mM EDTA
5M NaCl
2% CTAB
0.2% ß-mercaptoethanol
1% PVP
5mM glutamic acid
ß-mercaptoethanol, PVP and glutamic acid were added just before use.
Alkaline-lysis buffer:
50 mM Sucrose
25mM Tris-HCl, pH 8.0
10mM EDTA.

10X MOPS buffer:
0.2M morpholinopropane sulfonic acid,
0.5M sodium acetate
0.01M EDTA.

The pH was adjusted to 7.0 with 1NaOH.

20X SSC buffer, pH 7.0:
3 M NaCl
0.3 M tri-sodium citrate $2H_2O$

5X SSPE Buffer, pH 7.0:
20mM EDTA
200mM NaHPO4
3.6M NaCl

5M Potassium acetate solution, pH 4.8:
60 ml 5M potassium acetate solution
11.5 ml of glacial acetic acid
28.5 ml water (to make the volume to 100 ml)

(This solution is 3M with respect to potassium and 5M with respect to acetate having pH 4.8).

Luria Bertani broth (LB broth):
10 g Bactotryptone
5 g Bacto-yeast extract
5 g NaCl

The components were mixed in 800 ml water with constant stirring and then made up to 1000 ml with water. The pH of the broth was adjusted to 7.5 with NaOH solution.

Luria Bertani agar (LB agar):
10 g bactotryptone
5 g bacto-yeast extract
5 g NaCl
15 g agar

The components were mixed in 800 ml water with constant stirring and then made up to 1000 ml with water. The pH of the broth was adjusted to 7.5 with NaOH solution. The broth was heated to 60oC in a water bath and 15 gm of agar added to it with constant shaking till the agar dissolved completely. The solution autoclaved and poured into petriplates under sterile conditions.

Transfer solution:
0.4 M NaOH

Tris Equillibrated Phenol:

8-hydroxy quinoline
Liquefied phenol, redistilled
50 mM Tris base (adjusted to pH~10.5)
TE buffer, pH 8.0

Distilled phenol containing 0.1% (w/v) 8-hydroxyquinoline was mixed with equal volume of 50mM Tris and stirred continuously with repeated changes of 50mM Tris till the pH of the phenol was brought to 8.0. The aquous phase was removed from the solution and replaced by $1/10^{th}$ volume of TE buffer to form a thin layer of about 1cm on top of the phenol layer. The buffer saturated phenol was stored in dark bottles at 4°C.

CHAPTER: IV

AMPLIFICATION AND SEQUENCE ANALYSIS OF LEGUMIN GENE OF COMMON BUCKWHEAT

Experimental:

Grains of common buckwheat [*Fagopyrum esculentum* Moench (accession No. IC18890)] were procured from the North Eastern Regional Station of the National Bureau of Plant Genetic Resources, Shillong. Healthy grains of uniform size were screened and used for the present study. A portion of the grains were sowed in the experimental garden of Department of Botany, North Eastern Hill University, Shillong for multiplication.

The investigations carried out under the present study focused around amplification of SSP genes by polymerase chain reaction using oligonucleotide primers designed from the conserved regions of legumin-type SSP gene sequences available in the Genbank databases. For achieving the identified target, healthy grains of uniform size were germinated in dark in a seed germinator at 27°C and 85% R.H. The germinating seedlings were maintained in a plant growth chamber for 14 days in dark till the Ist leaves emerged fully. Genomic DNA was isolated from 14

days old etiolated seedlings of buckwheat following the CTAB buffer extraction protocol. PCR amplification was carried out with primers designed from conserved regions of legumin-type SSP gene sequences in the database. The primers are shown in table 3.1 and the annealing temperatures for PCR with different combinations of primers are shown in table 3.2.

Results:

The modified CTAB protocol used in the present investigation for isolation of genomic DNA yielded fairly good quality DNA from the etiolated seedlings of common buckwheat. Under UV light, the isolated DNA was detected on the agarose gel as a single fluorescent band having apparent molecular mass of 21 kb (Fig. 4.1). There was no fluorescence due to RNA in the electrophoressed DNA sample. The electrophoresis profile of the isolated DNA did not reveal any degradation or shearing of the isolated DNA. The concentration of DNA in the preparation was quantified by visual observation of the ethidium bromide stained DNA in the agarose gel on a UV-transilluminator and comparison of the fluorescence of the band of known amount of λ DNA electrophoressed alongside the isolated DNA using the KODAK 1D image analysis software. The concentration of DNA was also measured as a function of absorbance shown by the sample at 260nm. The yield of DNA ranged between 20-25μg DNA gm^{-1} tissue. Purity of the isolated DNA was assessed by measurement of absorbance at 260 and 280 nm and calculation of the ratio of absorbance at the two wavelengths. The ratio of absorbance at 260 and 280 nm (A_{260}/A_{280}) for the DNA samples isolated during the present investigation ranged between 1.7-1.9 indicating that the isolated DNA was fairly pure for use in the

Fig. 4.1: Total genomic DNA isolated from etiolated seedlings of buckwheat (*Fagopyrum esculentum* Moench.) L1-L4: Isolated genomic DNA, M: *Eco*RI/*Hind*III digested λ DNA.

Fig. 4.2: Restriction digestion profile of genomic DNA isolated from etiolated seedlings of buckwheat (*Fagopyrum esculentum* Moench).
(a) M: *Eco*RI/*Hind*III digested λ DNA, L3: genomic DNA digested with *Eco*RI
(b) M: *Eco*RI/*Hind*III digested λ DNA, L4: genomic DNA digested with *Hind*III
(c) M: *Eco*RI/*Hind*III digested λ DNA, L2: genomic DNA digested with *Nco*I

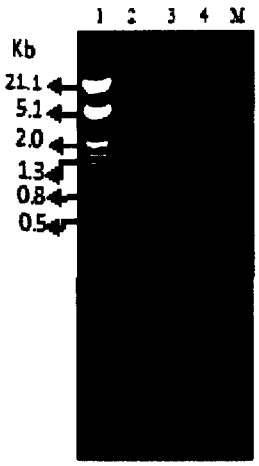


Fig: 4.1

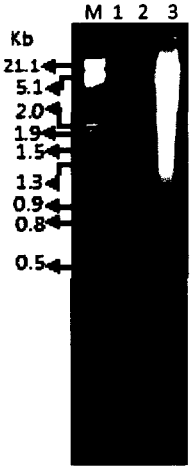


Fig: 4.2 (a)

Fig.4.2 (b)

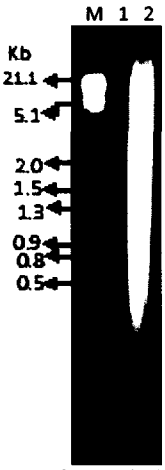


Fig. 4.2 (c)

investigation. The genomic DNA isolated from etiolated seedlings of common buckwheat was digested with *Eco*RI, *Hind*III and *Nco*I and size fractionated on 0.8% agarose gel. The electrophoretic profile of *Eco*RI digested DNA revealed a uniform smear ranging in size from 0.5 kb to 21 kb. The smear also showed four bands having molecular masses of 0.904 kb, 0.832 kb, 0.625 kb and 0.5 kb. (Fig. 4.2a). The agarose gel electrophoresis profile of *Hind*III and *Nco*I digested DNA showed uniform smears ranging from 0.3 kb to 20 kb. However, no distinct bands were visible in the electrophoretic profile of either *Hind*III or *Nco*I digested DNA (Fig. 4.2 b,c).

Amplification of SSP genes from buckwheat genomic DNA was attempted by polymerase chain reaction with combinations of oligonucleotide primers designed from conserved regions of SSP gene sequences available in the databases. The oligonucleotide primer pairs and the annealing temperature for PCR with each primer pair are presented in table 4.1. PCR amplification with oligonucleotide primer pair SS2F and SS3R with buckwheat genomic DNA as the template and amplification reaction comprising hot start at 94°C for 5 minutes, 35 cycles comprising of denaturation at 94°C for 1minute, annealing at 58°C for 1minute and chain extension at 72°C for 1minute followed by one reaction of chain elongation at 72°C for 10 minutes, amplified a DNA fragment having an apparent molecular mass of 850bp (Fig. 4.3). The nucleotide sequence for 835 bases of the amplified DNA is presented in Fig. 4.4. The sequence showed a maximum of 91% homology with mitochondrial NADH dehydrogenase subunit 1 of *Petunia* (acc. no. X60402) and *Brassica napus* (acc. no. AP006444). The sequence similarity with mitochondrial NADH

Table 4.1: List of primer combinations and PCR annealing temperatures used for the amplification of SSP genes from buckwheat genomic DNA

PRIMER PAIR	ANNEALING TEMPERATURE (°C)
SS2F-SS3R	58
SS4F-SS3R	67
SS8F-SS1R	67
SS8F-SS3R	67
SS2F-SS1R	67
SS15F-SS16R	68
SS15F-SS17R	68

Fig. 4.3: PCR amplification profile with buckwheat genomic DNA as template and primer pair SS2F-SS3R. M: *Eco*RI/*Hind*III digested λ DNA, L1: 0.835 kb amplicon generated with buckwheat genomic DNA as template and primer pair SS2F-SS3R, L2: positive control.

Fig. 4.4: Nucleotide sequence of 0.835 kb DNA fragment amplified with buckwheat genomic DNA as template and primer pair SS2F-SS3R

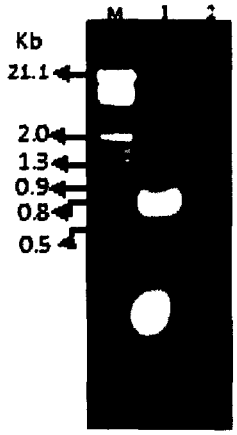


Fig: 4.3

```
1   TCTGTGCTCG TACTTGCCCT CAATGTTGTC AAAATGTTTG CTTACTTCAG
51  CCACTCTTCC GGGATCACCA ACCGTAATAA CGGTATTGGC CAACTCATCC
101 GGCCTCACAT TTAAATGATA AACTGCACCT CTGTCATTAA TGATTAATTC
151 ACTTTCAGCA ATTCTGAGCA TATTTATTTC AATTAGACCC TAAAATTCGG
201 GATTTCTTAT TTAGAATTTT AAAATTCCTT TCTTTCCTCT ATTTTTGCAT
251 CCCTAAATAG GGTCCTGTGG CCGAGTGGCT AGGCAGAGCT CTCCGGCCGA
301 AGACTAGTGG TAGGTGGTCC TGCGGAGCTT TCGGAGAAGG GTAGCCTAAT
351 GTGTAAGCAC AGCAATGAAC CGCGGCGAAC CCTCAGACGA CCTATCTAAG
401 ATTAGGAAGG AGATCCTCAG TAGTGGTGAC CCTTTCACTC TTCCACGGAC
451 TGATACATGT ACCGAATGCT CATACGGGAA AGTGGATTCC TGGGTCTGGC
501 AACCTCCCAG GTTGCTCCGA GAAATCCTTT CTTTCTCGTC CACTCAGGGG
551 GGGTGCGGAC ACCCCTGCGC GGGATTACAG GTGACGGGTA CAAGAATGGC
601 GGGGAAAGTG AACAGTACCC GTACGACATT CAGGGGATGG ATGTAGACCC
651 ATCGGCAGGG GATATCATTT CCGGGTCCTG GGGAGGTAGC CGCTTCCAGG
701 AACCCCAAAA GGCTGGAACT GGGGGGAACC CCTTGAGTCC TGAAACAACG
751 CAGGGCAATT CATTGGCTAT CAAAAAGGAG CAAAAACGGC TTGGCTCCCT
801 TTTTCCTAGA ACGAAAAGGT CGATTTGCCT ATCCC
```

Fig: 4.4

dehydrogenases of *Oenothera* (acc. no. M63034), *Arabidopsis* (acc. no. X98300), rice (acc. no. DQ167807), wheat (acc. no. X57966), *Anthoceros* (acc. no. AY354978), *Equisetum* (acc. no. AY354940), *Sphagnum* (acc. no. AY354932) and *Cycas* (acc. no. AP009381) ranged between 74-90%. The nucleotide sequence has been deposited to NCBI Genbank and assigned the accession number DQ852632. Fig. 4.5 represents the alignment of the nucleotide sequence with the NADH dehydrogenase gene sequences of other plants. The alignment reveals a high degree of sequence conservation across bryophytes, pteridophytes, gymnosperms and angiosperms. The most distinguishing feature of the sequence amplified in the present study is a 6 base "CCTCCC" insertion/substitution at position 203 in the partial nucleotide sequence of the buckwheat NADH dehydrogenase amplified in the present study. An important difference between the sequences of mitochondrial dehydrogenases of angiosperms and gymnosperms, observed in the alignment, is the insertion of 4 bases each *viz.* "CGCG" and "CTTT" respectively at positions 55 and 231 corresponding to buckwheat mitochondrial dehydrogenase (DQ852632) in the mitochondrial NADH dehydrogenase of angiosperms compared to that of the gymnosperms. On the other hand, the nucleotide sequences of NADH dehydrogenases of pteridophytes and bryophytes showed a deletion of 3 bases *viz.* "AGT" at position 122 corresponding to the buckwheat mitochondrial dehydrogenase.

PCR amplification with oligonucleotide primer pairs (i) SS8F and SS1R and (ii) SS8F/SS3R with buckwheat genomic DNA as the template and amplification reaction comprising hot start at 94°C for 5 minutes, 35 cycles comprising of

Fig. 4.5: Clustal W (1.81) multiple alignment of 835 bp amplicon generated with buckwheat genomic DNA as template and primer pair SS2F-SS3R with nucleotide sequences of some *nad*1 gene sequences available in the database ('*' mark represents conserved residues, invariant/similar residues are represented as ':').

CLUSTAL W (1.81) multiple sequence alignment

```
buckwheat_nad1                  AGACTAGTGGTAGGT-GGTCCTGCGGAGCTTTCGGAGAAGGGTAGCCTAA
Glycine_max_nad1_U09988.1       AGACTAGTGGTAGGT-GGTCCCGCGGAGCTTTCGGAGAAGGGTAGCCTAG
Zea_mays_nad1_U09986.1          AGACTAGTGATAGGT-GGTCCCGCGGAGCTTTCGGAGAAGGGTAGCCTAG
Tea_nad1_AY839914.1             AGACTAGTGGTAGGT-GGTCCCGCGGAGCTTTCGGAGAAGGGTAGCCTAG
Oenothera_nad1_M63034.1         AGACTAGTGGTAGGT-GGTCCCGCGGAGCTTTCGGAGAAGGGTAGCCTAG
AP009381.1_Cycas_taitungensis_  AGACTAGTGGTAGGT-GGTCCCGCGAAGCTTTCGGAGAAGGGTAGCCTAG
Equisetum_nad1_AY354940.1       -----------AGGTAGGTCCCGCGAAGCTTTCAGAGAAGGGTAGCCTAG
Polytrichum_nad1_gene_AY354966  ----------------GGTCCCGCGAAGCTTTCGGAGAAGGGTAGCCTGG
Anthoceros_nad1_AY354978.1      ----------------GGTCCCGCGAAGCTTTCGGAGAAGGGTAGCCTGG
                                               ***** *** ******* **************

buckwheat_nad1                  TGTGTAAGCACAGCAATGAACCGCGGCGAACCC-TCAGACGACCTATCTA
Glycine_max_nad1_U09988.1       TGTGTAAGCACAGCAATGAACCGCGGCGAACCC-TCAGACGACCTATCTA
Zea_mays_nad1_U09986.1          TGTGTAAGCACAGCAATGAACCGCGGCGAACCC-TCAGACGACCCTTCTA
Tea_nad1_AY839914.1             TGTGTAAGCACAGCAATGAACCGCGGCGAACCC-TCAGACGACCTATCTA
Oenothera_nad1_M63034.1         TGTGTAAGCACAGCAATGAACCGCGGCGAACCC-TCAGACGCCCTATCTA
AP009381.1_Cycas_taitungensis_  TGTGTAAGCACAGCAATGGAC----GCGAACCC-TCGGACGACCTATCTA
Equisetum_nad1_AY354940.1       TGTGTAAGCGCAGCAAAGAACC---GCGAACCCGTCAGACGACCTATCTA
Polytrichum_nad1_gene_AY354966  TGTGTAAGCATAGCAATGAAGC---GCGAACCA-TCGGACGACCTATCTA
Anthoceros_nad1_AY354978.1      TGTGTAAGCACAGCAACGAACC---GCGAACCCATTGGACGACCTATCTA
                                *********  ***** * *     *******  *  **** **  ****

buckwheat_nad1                  AGATTAGGAAGGAGATCCTCAGTAGTGGTGACCCTTTC----ACTCTTCC
Glycine_max_nad1_U09988.1       AGATTAGGGGGGAGATCCTCAGTAGTGGTGACCCTTTC----ACTCTTCC
Zea_mays_nad1_U09986.1          AGATAAGGGGGGAGACCCTCAGTAGTGGTGACCCTTTG----ACTCTTCC
Tea_nad1_AY839914.1             AGATTAGGGGGGAGATCCTCAGTAGTGGTGACCCTTTC----ACTCTTCC
Oenothera_nad1_M63034.1         AGATTAGGGGGG-GATCCTCAGTAGTGGTGACCCTTGCTTGAACTCTTCC
AP009381.1_Cycas_taitungensis_  GGATTAGGGGGGAGATCCTAAGTAGTGGTGACCTTGTA----ACTCTTCC
Equisetum_nad1_AY354940.1       AGATGGGGGGGGATATCCCAAGT---GGTTACCTCGTA----ACTCTTCC
Polytrichum_nad1_gene_AY354966  AGATTAGGGAGA-TATCTTAAGT---GGGTACCTCGTA----ACTTTTCC
Anthoceros_nad1_AY354978.1      AGATTAGGGGAT--ATCCT---------TTTCCTCGTA----ACTCCTTC
                                 ***  **      * *             **         ***  * *

buckwheat_nad1                  ACGGACTGATACATGTACCGAATGCTCATACGGGAAAGTGGATTCCTGGG
Glycine_max_nad1_U09988.1       ACGGACTGATACATGTACCGAATGCTCATACGGGAAAGTTGACTCCTGGG
Zea_mays_nad1_U09986.1          ACTGACTTATATATGTACCGAATGCTCATACGGGAAAGTGAACTCCTGGG
Tea_nad1_AY839914.1             ACGGACTGATACATGTACCGAATGCTCATACGGGAAAGTTGACTCCTGGG
Oenothera_nad1_M63034.1         TCGGACTGATACATGTACCGAATGCTCATACGGGAAAGTTGACTCCTGGG
AP009381.1_Cycas_taitungensis_  ACGGACCTATACATGTACCGAATGCTCATACGGGAAAGTGCACTCCTGGG
Equisetum_nad1_AY354940.1       CCAGACTTATACGTGTACCGAGTGCTCATACGGGAAAGTGCGCTCCTAGG
Polytrichum_nad1_gene_AY354966  CCAGACCTATACGTGTACCGAATGCTCATACGGGAAAGTGCGCTCCTAGG
Anthoceros_nad1_AY354978.1      CCAGGCCTATACGTGTACCGAATGCTCATACGGGAAAGTGCGCTCCTAGA
                                 * * *  ***  ******** *****************    **** *

buckwheat_nad1                  TCTGGCAACCTCCCAGGTTGCTCCGAGAAATCCTTTCTTTCTCGTCCACT
Glycine_max_nad1_U09988.1       TCTGGAAC------GAAGTCTCCGAAAAA-TCCTTTCTTTCTCGTCCACT
Zea_mays_nad1_U09986.1          TCTGGGGG------TTGCTCCGAGAAAAAATCCTTTCTTTCTCGTCCACT
Tea_nad1_AY839914.1             TCTGGAACC-TGGGGGGTTGCTCCGAGAAATCCTTTCTTTCTCGTCCACT
Oenothera_nad1_M63034.1         TCTGGAAC--TGGGGGGTTGCTCCGAGAAATCCTTTCTTTCTCGTCCACT
AP009381.1_Cycas_taitungensis_  TCTGGAACCAGGGAGGGTTGCTCCGAGAAATCCTTT----CTCGTCCACT
Equisetum_nad1_AY354940.1       TCTGGAACCAGGGAGGGTTGCTCTGAGAAATCGTCT--------CCAAAT
Polytrichum_nad1_gene_AY354966  TCTGGAACCAGGGAGGGTTGCTCCGAGAAAAAACTT----CTCGTCTCAT
Anthoceros_nad1_AY354978.1      TCTGGAACCAGGGAGGGTTGCTCCGAGAAAAAATTT----CTCGTCTTAT
                                *****             *      * **      *         *   *

buckwheat_nad1                  CAGGGGGGGTGCGGACACCCCTGCGCGGGATTACAGGTGACGGGNTACAA
Glycine_max_nad1_U09988.1       CAG-GGGGGTGCGGACACACCTGCGCGG-ATTACAGGTGACGG-TTACAA
Zea_mays_nad1_U09986.1          CCA-GGGGGTGCGGACACACCTGCGCGG-ATTACAGGTGACGG-TTACAA
Tea_nad1_AY839914.1             CAG-GGGGGTGCGGACACACCTGCGCGG-ATTACAGGTGACAG-TTACAA
Oenothera_nad1_M63034.1         CAG-GGGGGTGCGGACACACCTGCGCGG-ATTACAGGTGACGG-TTACAA
AP009381.1_Cycas_taitungensis_  CCA-GGGGGTGCGGACACACCTGCGCGG-ATTACAGGTGACGG-TTACAA
Equisetum_nad1_AY354940.1       TGT-TGGGGTGCGGACACACCCGCGCGG-ATTACAGGTGACGG-CTACGG
Polytrichum_nad1_gene_AY354966  CC--GGGGGTGCGGACACACTTGCGCGA-ATTACAGGTGACAG-CTACAA
Anthoceros_nad1_AY354978.1      CTT-GGGGGTGCGGACACACTTGCGCGG-ATTACAGGTGACGG-CTACAA
                                     ************* *  *****  ************ *  ***
```

Fig: 4.5

denaturation at 94°C for 1minute, annealing at 67°C for 1minute and chain extension at 72°C for 1minute followed by one reaction of chain elongation at 72°C for 10 minutes amplified DNA fragments having apparent molecular masses of 1.1 kb for the primer pair SS8F and SS1R and 1.6 kb for the primer pair SS8F-SS3R. The amplicons resolved as distinct bands corresponding to molecular masses of 1.1 kb and 1.6 kb respectively on 0.8% agarose gel (Fig. 4.6a). Similarly PCR amplification with oligonucleotide primer pairs (i) SS4F and SS3R and (ii) SS2F and SS1R with chain reaction comprising hot start at 94°C for 5 minutes, 35 cycles comprising of denaturation at 94°C for 1minute, annealing at 67°C for 1minute and chain extension at 72°C for 1minute followed by one reaction of chain elongation at 72°C for 10 minutes amplified DNA fragments having apparent molecular masses of 830bp for the primer pair SS4F and SS3R and 1.3 kb for the primer pair SS2F and SS1R (Fig. 4.6b,c). The amplicons resolved as distinct bands corresponding to molecular masses of 0.83 kb and 1.3 kb respectively on 1.2% agarose gel (Fig. 4.6b,c).

Southern hybridization of the 1.1 kb, 1.6 kb and 1.3 kb amplicons with [α-^{32}P]-dATP labelled 0.83 kb amplicon generated positive signals, corresponding to the position of 1.1 kb, 1.6 kb and 1.3 kb amplicons, on the exposed X-ray film (Fig. 4.6d,e). These results indicated that the 0.83 kb, 1.1 kb, 1.6 kb and 1.3 kb DNA fragments, which were amplification products of primer pairs SS8F-SS1R, SS8F-SS3R, SS4F-SS3R and SS2F-SS1R shared a high degree of sequence homology. Fig. 4.8 shows the nucleotide sequence of 1613 bases for the 1.6 kb DNA fragment amplified by primer pair SS8F-SS3R. In order to determine the identity of the nucleotide sequence and to determine its homology with other nucleotide sequences

Fig. 4.6: (a) PCR amplification profile with buckwheat genomic DNA as template and primer pairs SS8F-SS1R (L1) and SS8F-SS3R (L2). M: *Eco*RI/*Hind*III digested λ DNA

(b) PCR amplification profile with buckwheat genomic DNA as template and primer pair SS4F-SS3R (L1). M: *Eco*RI/*Hind*III digested λ DNA

(c) PCR amplification profile with buckwheat genomic DNA as template and primer pair SS2F-SS1R (L1). M: *Eco*RI/*Hind*III digested λ DNA

(d) Autoradiograph of the amplification products of primer pairs SS8F-SS1R (L1) and SS8F-SS3R (L2) hybridized with [α-^{32}P]-dATP labelled 0.835 kb DNA fragment amplified by primer pair SS4F-SS3R

(e) Autoradiograph of the amplification products of primer pair SS2F-SS1R hybridized with [α-^{32}P]-dATP labelled 800 bp DNA fragment amplified by primer pair SS4F-SS3R

Fig. 4.7: Diagramatic representation showing the relative positions of oligonucleotide primers SS2F, SS8F, SS4F, SS1R and SS3R on the target DNA.

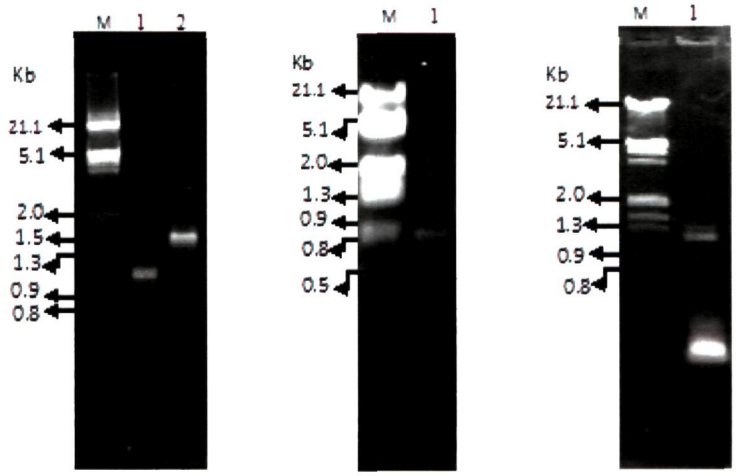
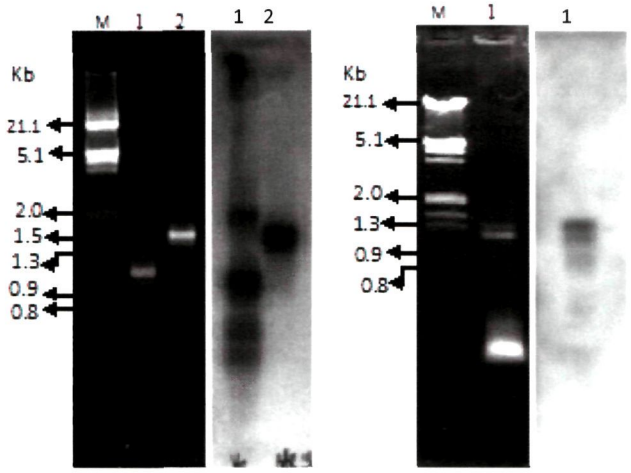


Fig:4.6 (a) (b) (c)



(d) (e)

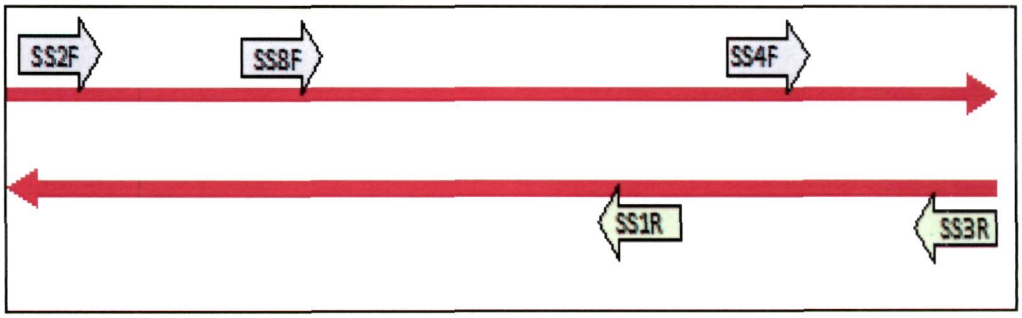


Fig 4.7

Fig. 4.8: Nucleotide sequence of 1613 bases for the 1.6 kb fragment amplified from buckwheat genomic DNA using the primer pair SS8F-SS3R.

```
1     CCCGGAGACA TACTCACTCT CGTGGAATAG GAAATGAGAG GTGATGAAGA
51    AGAAGAGGGA TTCGAGAGTG GTAGACGGAT GACCGACGCG CACCAGCCCA
101   CACGTCGTGT TCGCAAGGGC GACGTAGTCG CACTCCCTCA GGGAACCGTT
151   CACTGGTGCT TCAACGACGG CCAAGAGGAT TTGGTCGTTG TCGCCGTCCA
201   CAACCTCAAC ACCGACGCCA ATCAACTCGA CCAATCTCTC AAGGTATATA
251   TATTTTATAC CTGTTTATAT TCAAGCTCGG GTTGCTTCGG GCCTCGGGTC
301   GGTTCGGGCC AAAATTTACA AAGTATTAAG CAGTTTATCC GGAAATGGGA
351   CGGACCTGAA ATGGCTGCCG TGGACGTTGT TATCTTTTTT AAAAAAATAT
401   TTAAACTTTT TTTAATATCA TTTAAAAAAA ATTATACGTA AGATATTTTA
451   TATAAAATAT ATTATATTTA TACCTTATAT AATATGTACA ATATATTTTT
501   TTCAATGGGG CAAAATGGCG AACAGGGCCC GGGACAAAAT GGGGGCGGAA
551   CCCGCCCGAC GAATTGGCCG ACCGGGTTTT TTAGCCCGGC GGGGCTTTTT
601   TTTTTTTTTA CCCCCCCCTT TCTCCCGGGT ACGGGAAAGC CTATTATATA
651   CATATATCCA CAACTACGAT ATATATGTAT ATAATCAATT AATCAGGGCT
701   AATTAATGCA TGGTTGGGTT TTGGACTGGC AGACATTCTT CTTGGCAGGT
751   GGAGTGCAAG GAGGATCAAA AGAGAGGCAG AGCCAGAAGC TACCGCTACC
801   TTCAACAACA TACTAGAGCG CCTTCTGACG ACCAAGCTTC TAGCTGAGGC
851   ACTAGGGACT GAGGAGGAGA CCGTCAGGAA AATGCAGGAG AGTGACGAGC
901   GTGGACCTAT CGTCAAAGCT AGGAAGAATA TGCGACAAAT GGTCACCCCG
951   CCGAGATTCG GGAGGGAACA AGATGAGGAC GAGACAAACG GGTTAGAAGA
1001  ATCCTTCTGC AACATGAGGT TCCGTCACAA CCTCGGCCCT AGGACCGAGG
1051  CCGACATCGC CTCCAGACAG GCTGGAAGGA TCCACTCCGT CGACCAGAAC
1101  AAGCTTCCCA TCCTCGAGTT TATCGACATG AGCGCCGAGA AAGCCCCCTT
1151  CTCCCGGTAT ATATATACCT AAATGTTCAT TCACTGAAAT ATATGAGCTT
1201  GTAAACTTTA TGTTTTATTT TTTTTTTGTG CAGAACGCCA TGTTAGCGCC
1251  GGCATGGGCG GTGTCAGGCC ACAGAGTGTT CTACGTGTTG AGGGGTGAGG
1301  CTCAGATGCA GATAGTAGAC GACAACGGTC AGACGGTGTT AGATGACAGA
1351  GTCTCACAAG GATCCATGGT TGTGATCCCA CAGTTCTACA CCTCGACATG
1401  TAGAGCAGGA AGAGACGGAC TCGAGTACGT TTCATTCGAG ACCACAGCAA
1451  ACCCGATGAG TAGTCCATTG AACGGCCACG CTTCGGTTTT CAAGGGAATG
1501  CCAATTCCTG TGCTCAGCAA CTCGTACCAG ATCTCGCCCC GTGCCGCCTA
1551  TGAGCTGAAG CAGACAAGGA GCCACGAGCA CGGGCAGTTC TCGCCTTTCG
1601  GAGGAAGGTC GTA
```

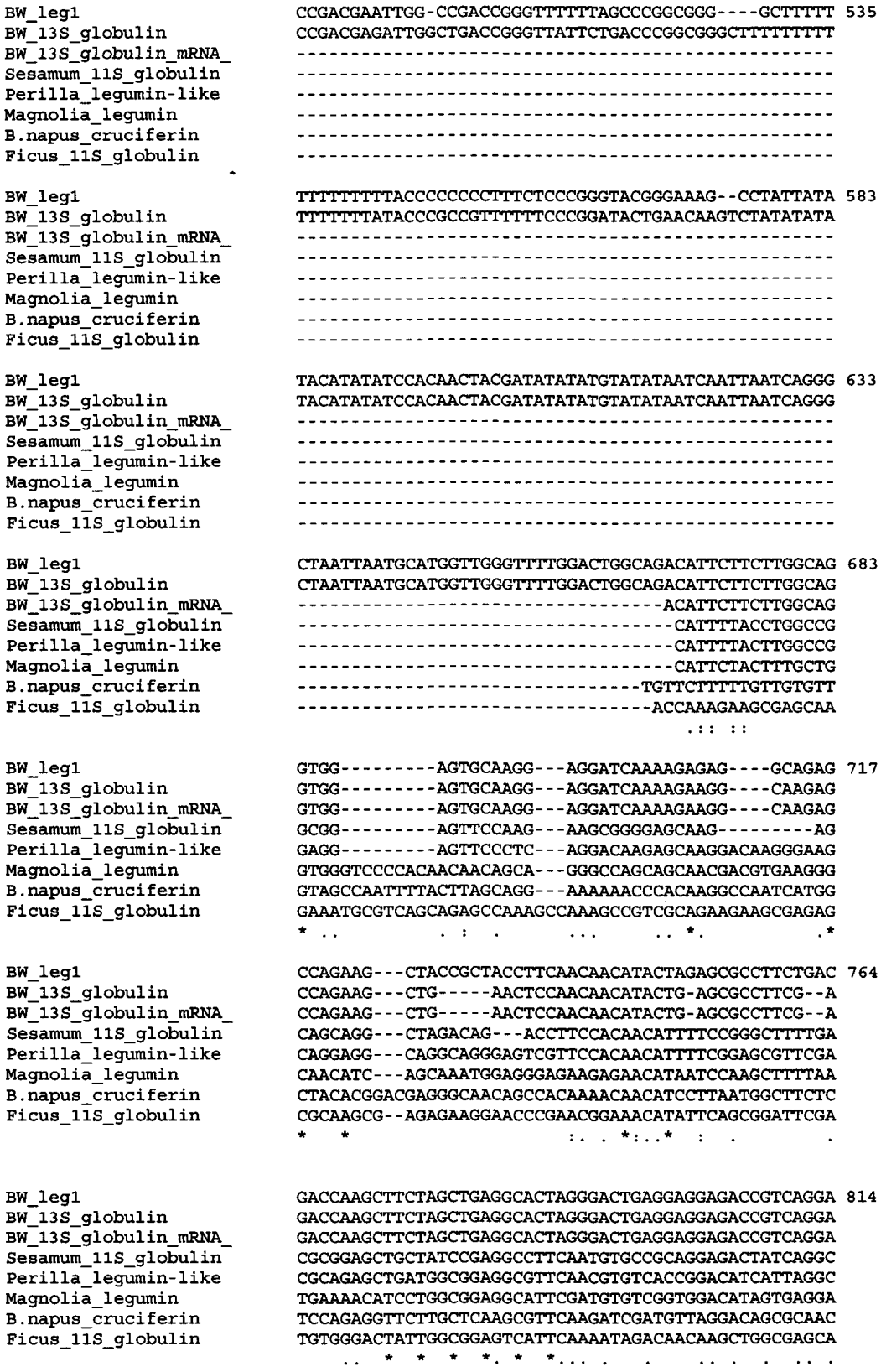
Fig: 4.8

Fig. 4.9: Clustal W (1.81) multiple alignment of the nucleate sequence of 1613 bases of the 1.6 kb amplicon generated with buckwheat genomic DNA as template and primer pair SS8F-SS3R with nucleotide sequences of SSP genes available in the database ('*' mark represents conserved residues, invariant/similar residues are represented as ':').

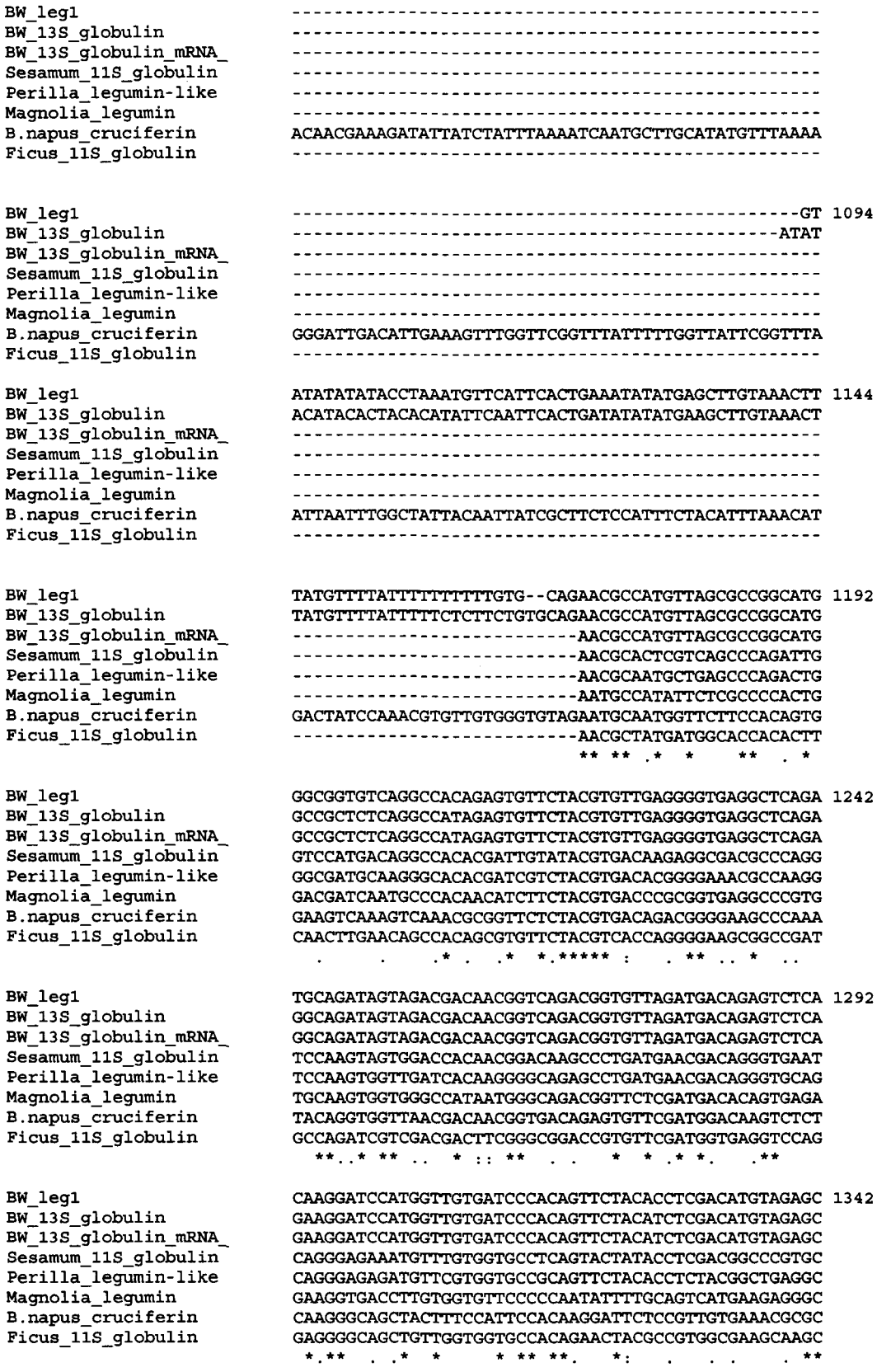
CLUSTAL W (1.81) multiple sequence alignment

```
BW_leg1                 ----------------------------------CCCGGAGACATACTC 15
BW_13S_globulin         --GAGAAGGTGTGTTTGGGTTGTCACTACCAGGATGCCCGGAGACATACC
BW_13S_globulin_mRNA_   --GAGAAGGTGTGTTTGGGTTGTCACTACCAGGATGCCCGGAGACATACC
Sesamum_11S_globulin    --GTCAGGGCTTGATAAGCATCATGGTCCCCGGCTGTGCTGAAACGTATC
Perilla_legumin-like    --GTGAGGGATTCATCGGCGTCATCTTCCCCGGCTGCGCCGAGACATACC
Magnolia_legumin        --GAAGAGGGCTGCTAGGAATAACGTACCCAGGGTGCGCAGAGACCTACC
B.napus_cruciferin      AGGACAAGGCCTTATGGGGAGAGTGGTCCCTGGATGCGCTGAGACTTTCC
Ficus_11S_globulin      --GCAGGGGATTTCACGGAGCTGTGATCCCTGGTTGTCCAGAAACTTTTG
                                                          ..... :

BW_leg1                 AC--------------------------------TCTCGTGGAAT-AGGA 32
BW_13S_globulin         AG--------------------------------TCTCGTGGAAT-GGAA
BW_13S_globulin_mRNA_   AG--------------------------------TCTCGTGGAAT-GGAA
Sesamum_11S_globulin    AG--------------------------------GTCCACAGGAGCCAGA
Perilla_legumin-like    AC--------------------------------GGCGGTGAGG---AGG
Magnolia_legumin        GC--------------------------------TCTAG-------AGGA
B.napus_cruciferin      AGGACTCATCAGTATTCCAACCAGGCAGTGGCAGCCCCTTCGGAGAAGGT
Ficus_11S_globulin      AAG---------------------------------------------
                        .
BW_leg1                 AATG--AGAGGTGATGAAGAAGAAGAGGGA-------------------- 60
BW_13S_globulin         ATGAGAGGTGATGAAGAAGAAGAAGAGGGA--------------------
BW_13S_globulin_mRNA_   ATGAGAGGTGATGAAGAAGAAGAAGAGGGA--------------------
Sesamum_11S_globulin    GAACCATGGAGCGCACAGAGGCATCGGAG----------------------
Perilla_legumin-like    CCTCGTTCGAGGGCAGAAGGGAAGGCGAG----------------------
Magnolia_legumin        CAGCCTCAGAGAACCGGAGGTGAGCAGCAG---------------------
B.napus_cruciferin      CAAGGCCAAGGTCAGCAGGGTCAGGGGCAAGGTCAGGGTCAGGGTCAAGG
Ficus_11S_globulin      ------------AATCAGGAT-----------------------------
                                  .  ....

BW_leg1                 ------------------------------------------------
BW_13S_globulin         ------------------------------------------------
BW_13S_globulin_mRNA_   ------------------------------------------------
Sesamum_11S_globulin    ------------------------------------------------
Perilla_legumin-like    ------------------------------------------------
Magnolia_legumin        ------------------------------------------------
B.napus_cruciferin      CAAGGGCCAACAGGGTCAAGGCAAGGGCCAACAGGGTCAATCCCAGGGCC
Ficus_11S_globulin      ------------------------------------------------

BW_leg1                 --TTCGAGAGTGGTAGACGGATGACCGACGCGCACCAGCCCACACGTCGT 108
BW_13S_globulin         --TTCGAGAGTGGTAGACGGATGACCGATGCGCACCAGCCCACACGTCGT
BW_13S_globulin_mRNA_   --TTCGAGAGTGGTAGACGGATGACCGATGCGCACCAGCCCACACGTCGT
Sesamum_11S_globulin    --CAGCAAGATAGAGGAAGCGTGAGGGATTTGCATCAGAAGGTGCACCGC
Perilla_legumin-like    --CAGAAGGAAGGAGGAAGAGCGAGAGACTCGCACCAGAAGATCCACCGC
Magnolia_legumin        --CAGCAACGTGGTGAGAGTATCAGTGACCAACACCAGAAGATCCATCGG
B.napus_cruciferin      AACAGGGTCAAGGTCAAGGTTTCCGTGATATGCACCAGAAAGTAGAGCAC
Ficus_11S_globulin      ---TCCAACAATCACAGAGGTCTCAAGAGCAGCACCAGAAGGTTCGCGAA
                           :  .  .:  : .. *    .  **   .** ***.. .   .  .

BW_leg1                 GTTCGCAAGGGCGACGTAGTCGCACTCCCTCAGGGAACCGTTCACTGGTG 158
BW_13S_globulin         GTTCGCAAGGGCGACGTAGTCGCACTCCCTCAGGGAACCGTTCACTGGTG
BW_13S_globulin_mRNA_   GTTCGCAAGGGCGACGTAGTCGCACTCCCTCAGGGAACCGTTCACTGGTG
Sesamum_11S_globulin    CTTCGCCAAGGAGACATTGTTGCTATTCCGTCTGGTGCTGCACATTGGTG
Perilla_legumin-like    ATTCGTCAAGGCGACATCGTGGTTCTGCCACCGGGCGCAGTCCACTGGTG
Magnolia_legumin        ATCCGACGTGGAGATATCGTAGCCCTCCCGGCTGGCGTGGCCCACTGGTG
B.napus_cruciferin      ATAAGGAGCGGCGACACCATTGCTACACATCCCGGTGTAGCTCAATGGTT
Ficus_11S_globulin      ATCTGCGAAGGCGACATCGTTGCCGCCCCCGCCGGAGTGGCTCAGTGGGT
                         *  *  . **.** .  .* *     *.  . ** .  *  ** ***

BW_leg1                 CTTCAACGACGGCCAAGAGGATTTGGTCGTTGTCGCCGTCCACAACCTCA 208
BW_13S_globulin         CTTCAACGACGGCCAGGAGGATTTGGTGGTTGTCGCCGTCCACAACCTCA
BW_13S_globulin_mRNA_   CTTCAACGACGGCCAGGAGGATTTGGTGGTTGTCGCCGTCCACAACCTCA
Sesamum_11S_globulin    CTACAACGACGGGAGCGAAGATTTAGTCGCCGTCTCCATCAATGACGTCA
Perilla_legumin-like    CCACAACGACGGCAGCGAGGACTTGGTCGCCGTCTCCATCAACGATCTCA
Magnolia_legumin        CTATAACGACGGCAACGAAGAGCTTGTGGCCCTCTCTATTACGGATTTCA
B.napus_cruciferin      CTACAACAATGGAAACCAACCTCTTGTCATCGTTGCCGTCATGGATTTAG
Ficus_11S_globulin      CTACAACAACGGCGACTCTCCGCTTGTTCTCGTCTCCTTCACCGACGTCG
                        * : ***.* **  .  .  .  * **     *  *  *  .  .*  *..
```

```
BW_leg1                 ACACCGACGCCAATCAACTCGACCAATCTCTCAAGGTATATATATTTTAT 258
BW_13S_globulin         ACACCGACGCCAATCAACTCGACCAATCTCTCAAGGTATATATATTTTAT
BW_13S_globulin_mRNA_   ACACCGACGCCAATCAACTCGACCAATCTCTCAAG--------------- 193
Sesamum_11S_globulin    ACCACCTCTCAAACCAGCTGGATCAAAAATTCAGAG---------------
Perilla_legumin-like    ACCATCAATCAAACCAGCTGGACCAGAAATTCAGAG---------------
Magnolia_legumin        ATAGCGAGTCCAACCAGCTCGACCAGAGGCCTAGGT---------------
B.napus_cruciferin      CTAGCCACCAGAACCAGCTTGACCGCAACCCAAGGTATATATATACCCCA
Ficus_11S_globulin      GCAATCAGGCTAACCAGCTTGACCTAAACACAAGAAGATTCCACATCGGA
                          .   :  . ** **.** ** *  :     *..

BW_leg1                 ACCTGTTTATATTCAAGCTCGGGTTGCTTCGGGCCTCGGGTCGGTTCGGG 243
BW_13S_globulin         ACCTGTTTATATTCAAGCTCGGGTTGCTTCGGGCCTCGGGTCGGTTCGGG
BW_13S_globulin_mRNA_   ---------------------------------------------------
Sesamum_11S_globulin    ---------------------------------------------------
Perilla_legumin-like    ---------------------------------------------------
Magnolia_legumin        ---------------------------------------------------
B.napus_cruciferin      AAACCTCAAAATAAATTATTAAAATTTTATTATAAAG---------------
Ficus_11S_globulin      GGTAACCCACTGAAGG-------------------------------------

BW_leg1                 CCAAAATTTACAAAGTATTAAGCAGTTTATCCGGAAATGGGACGGACCTG 293
BW_13S_globulin         CCAAAATTTACAAAATATTAAGCAGTTTATCCGGAAATGGGACGGACCTG
BW_13S_globulin_mRNA_   ---------------------------------------------------
Sesamum_11S_globulin    ---------------------------------------------------
Perilla_legumin-like    ---------------------------------------------------
Magnolia_legumin        ---------------------------------------------------
B.napus_cruciferin      ---------------------------------------------------
Ficus_11S_globulin      ---------------------------------------------------

BW_leg1                 AAATGGCTGCCGTGGACGTTGTTATCTTTTTTAAAAAAATATTTAAACTT 343
BW_13S_globulin         AAATGGCTGCCGTGGACGTTGTTATCTTTTTTAAAAAAATATTTAAACTT
BW_13S_globulin_mRNA_   ---------------------------------------------------
Sesamum_11S_globulin    ---------------------------------------------------
Perilla_legumin-like    ---------------------------------------------------
Magnolia_legumin        ---------------------------------------------------
B.napus_cruciferin      ---------------------------------------------------
Ficus_11S_globulin      ---------------------------------------------------

BW_leg1                 TTTTTAATATCATTTAAAAAAAATT-ATACGTAAGATATTTTATATAAAA 392
BW_13S_globulin         TTTTTAATATCATTTAAAAAAAAATTATACGTAAGATATTTTATATAAAA
BW_13S_globulin_mRNA_   ---------------------------------------------------
Sesamum_11S_globulin    ---------------------------------------------------
Perilla_legumin-like    ---------------------------------------------------
Magnolia_legumin        ---------------------------------------------------
B.napus_cruciferin      ---------------------------------------------------
Ficus_11S_globulin      ---------------------------------------------------

BW_leg1                 TATATTATATTTATACCTTATATAATATGTACAATATATTTTTTTCAATG 442
BW_13S_globulin         TATATTATATTTATACTTTATATAATATGTACAATATATTTTTTTCACTG
BW_13S_globulin_mRNA_   ---------------------------------------------------
Sesamum_11S_globulin    ---------------------------------------------------
Perilla_legumin-like    ---------------------------------------------------
Magnolia_legumin        ---------------------------------------------------
B.napus_cruciferin      ---------------------------------------------------
Ficus_11S_globulin      ---------------------------------------------------

BW_leg1                 GGGCAAAATGGCGAACAGGGCCCGGGACAAAATGGGGGCGG--AACCCGC 490
BW_13S_globulin         GGCAGACATGCTGAACAGGGCTCGGGACTAGAGTGGTGTCGGGAACCCGC
BW_13S_globulin_mRNA_   ---------------------------------------------------
Sesamum_11S_globulin    ---------------------------------------------------
Perilla_legumin-like    ---------------------------------------------------
Magnolia_legumin        ---------------------------------------------------
B.napus_cruciferin      ---------------------------------------------------
Ficus_11S_globulin      ---------------------------------------------------
```


BW_leg1 CCGACGAATTGG-CCGACCGGGTTTTTTAGCCCGGCGGG----GCTTTTT 535
BW_13S_globulin CCGACGAGATTGGCTGACCGGGTTATTCTGACCCGGCGGGCTTTTTTTTT
BW_13S_globulin_mRNA_
Sesamum_11S_globulin
Perilla_legumin-like
Magnolia_legumin
B.napus_cruciferin
Ficus_11S_globulin
BW_leg1 TTTTTTTTTACCCCCCCCTTTCTCCCGGGTACGGGAAAG--CCTATTATA 583
BW_13S_globulin TTTTTTTATACCCGCCGTTTTTTCCCGGATACTGAACAAGTCTATATATA
BW_13S_globulin_mRNA_
Sesamum_11S_globulin
Perilla_legumin-like
Magnolia_legumin
B.napus_cruciferin
Ficus_11S_globulin
BW_leg1 TACATATATCCACAACTACGATATATATGTATATAATCAATTAATCAGGG 633
BW_13S_globulin TACATATATCCACAACTACGATATATATGTATATAATCAATTAATCAGGG
BW_13S_globulin_mRNA_
Sesamum_11S_globulin
Perilla_legumin-like
Magnolia_legumin
B.napus_cruciferin
Ficus_11S_globulin
BW_leg1 CTAATTAATGCATGGTTGGGTTTTGGACTGGCAGACATTCTTCTTGGCAG 683
BW_13S_globulin CTAATTAATGCATGGTTGGGTTTTGGACTGGCAGACATTCTTCTTGGCAG
BW_13S_globulin_mRNA_ ----------ACATTCTTCTTGGCAG
Sesamum_11S_globulin ----------CATTTTACCTGGCCG
Perilla_legumin-like ----------CATTTTACTTGGCCG
Magnolia_legumin ----------CATTCTACTTTGCTG
B.napus_cruciferin ----------TGTTCTTTTTGTTGTGTT
Ficus_11S_globulin ----------ACCAAAGAAGCGAGCAA
BW_leg1 GTGG---------AGTGCAAGG---AGGATCAAAAGAGAG----GCAGAG 717
BW_13S_globulin GTGG---------AGTGCAAGG---AGGATCAAAAGAAGG----CAAGAG
BW_13S_globulin_mRNA_ GTGG---------AGTGCAAGG---AGGATCAAAAGAAGG----CAAGAG
Sesamum_11S_globulin GCGG---------AGTTCCAAG---AAGCGGGGAGCAAG---------AG
Perilla_legumin-like GAGG---------AGTTCCCTC---AGGACAAGAGCAAGGACAAGGGAAG
Magnolia_legumin GTGGGTCCCCACAACAACAGCA---GGGCCAGCAGCAACGACGTGAAGGG
B.napus_cruciferin GTAGCCAATTTTACTTAGCAGG---AAAAAACCCACAAGGCCAATCATGG
Ficus_11S_globulin GAAATGCGTCAGCAGAGCCAAAGCCAAAGCCGTCGCAGAAGAAGCGAGAG
BW_leg1 CCAGAAG---CTACCGCTACCTTCAACAACATACTAGAGCGCCTTCTGAC 764
BW_13S_globulin CCAGAAG---CTG-----AACTCCAACAACATACTG-AGCGCCTTCG--A
BW_13S_globulin_mRNA_ CCAGAAG---CTG-----AACTCCAACAACATACTG-AGCGCCTTCG--A
Sesamum_11S_globulin CAGCAGG---CTAGACAG---ACCTTCCACAACATTTTCCGGGCTTTTGA
Perilla_legumin-like CAGGAGG---CAGGCAGGGAGTCGTTCCACAACATTTTCGGAGCGTTCGA
Magnolia_legumin CAACATC---AGCAAATGGAGGGAGAAGAGAACATAATCCAAGCTTTTAA
B.napus_cruciferin CTACACGGACGAGGGCAACAGCCACAAAACAACATCCTTAATGGCTTCTC
Ficus_11S_globulin CGCAAGCG--AGAGAAGGAACCCGAACGGAAACATATTCAGCGGATTCGA
BW_leg1 GACCAAGCTTCTAGCTGAGGCACTAGGGACTGAGGAGGAGACCGTCAGGA 814
BW_13S_globulin GACCAAGCTTCTAGCTGAGGCACTAGGGACTGAGGAGGAGACCGTCAGGA
BW_13S_globulin_mRNA_ GACCAAGCTTCTAGCTGAGGCACTAGGGACTGAGGAGGAGACCGTCAGGA
Sesamum_11S_globulin CGCGGAGCTGCTATCCGAGGCCTTCAATGTGCCGCAGGAGACTATCAGGC
Perilla_legumin-like CGCAGAGCTGATGGCGGAGGCGTTCAACGTGTCACCGGACATCATTAGGC
Magnolia_legumin TGAAAACATCCTGGCGGAGGCATTCGATGTGTCGGTGGACATAGTGAGGA
B.napus_cruciferin TCCAGAGGTTCTTGCTCAAGCGTTCAAGATCGATGTTAGGACAGCGCAAC
Ficus_11S_globulin TGTGGGACTATTGGCGGAGTCATTCAAAATAGACAACAAGCTGGCGAGCA

```
BW_leg1                 AAATGCAGGAGAGTGACGAGCGTGGACCTATCGTCAAAG-CTAGGAAGAA 863
BW_13S_globulin         AAATGCAGGAGAGTGACGAGCGTGGACCTATCGTCAAAG-CTAGGAAGAA
BW_13S_globulin_mRNA_   AAATGCAGGAGAGTGACGAGCGTGGACCTATCGTCAAAG-CTAGGAAGAA
Sesamum_11S_globulin    GGATGCAATCGGAGGAGGAAGAGAGAGGGCTCATCGTCATGGCCAGAGAA
Perilla_legumin-like    GTATGCAGGCGTCGGAGGAGGAGAGAGGGCTAAGCGTGATGGCTCATGAG
Magnolia_legumin        AGATGCAGAGAAATGACGATAGGGGCTACATTGTCAAAGTCAAAAGAGGA
B.napus_cruciferin      AACTTCAGAACCAGCAAGATAACCGGGGAAACATTGTCCGTGTCCAAGGC
Ficus_11S_globulin      GGCTCCAGAACCAGAACGACAGGAGGGAGAGGATCGTCAGAGTGAGGGAA
                        . .* **.       * **  .  *    .  .  .:        . *.

BW_leg1                 TATGCGACAAATGGTCACCC---------------------------- 883
BW_13S_globulin         TATGCGACAAATGGTCACCC----------------------------
BW_13S_globulin_mRNA_   TATGCGACAAATGGTCACCC----------------------------
Sesamum_11S_globulin    CGCATGACATTTGTCCGGCC----------------------------
Perilla_legumin-like    AGCATGAGGTACATCCGCCC----------------------------
Magnolia_legumin        GAGATGAGCATGGTTCGGCC----------------------------
B.napus_cruciferin      CCCTTCGGTGTTATTAGGCCG---------------------------
Ficus_11S_globulin      GACCTTCATATAGTGAGCCCAGGCAGGATTCAAGAGGAAGAACGCCGACA
                              : .  .. **
BW_leg1                 --------------------------CGCCGAGATTCGGGAGGGAACAA 906
BW_13S_globulin         --------------------------CGCCGAGATTCGGGAGGGAACAA
BW_13S_globulin_mRNA_   --------------------------CGCCGAGATTCGGGAGGGAACAA
Sesamum_11S_globulin    --------------------------CGATGAAGAAGAAGGAGAGCAAG
Perilla_legumin-like    --------------------------CGAAGAAATG---AGGGAACATA
Magnolia_legumin        --------------------------TGACGAAGAGGCAGAAGATGAAG
B.napus_cruciferin      --------------------------CCATTGAAAAGCCAGAGACCACAG
Ficus_11S_globulin      CGAGTACAGAAGAGGATCGTACGACGAAGAAGAGTACAGAAGGAGATCGA
                                                .  **. :    ....   . .

BW_leg1                 GATGAGG--------ACGAGACAAACGGGTTAGAAGAATCCTTCTGCAAC 948
BW_13S_globulin         GATGAGG--------ACGAGACAAACGGGTTAGAAGAATCCTTCTGCAAC
BW_13S_globulin_mRNA_   GATGAGG--------ACGAGACAAACGGGTTAGAAGAATCCTTCTGCAAC
Sesamum_11S_globulin    AGCATAGGGGAAGACAATTGGACAACGGCCTGGAAGAAACTTTCTGCACC
Perilla_legumin-like    GCAGAAGATCATCATCAAATGAAAATGGATTGGAAGAATCTTTCTGCTCC
Magnolia_legumin        AGCAATATCAACAAGGTAGGAGAAACGGATTTGAAGAGGTCTACTGCAAT
B.napus_cruciferin      GAGACAG-----------AAGCTAACGGTCTAGAAGAGACCATATGCAGC
Ficus_11S_globulin      ACGGTGGTGAGTACATGATGACGAACGGCTTGGAAGAGACATTCTGCACT
                        .  .  .              .  ** **  * *****.   ::.***:

BW_leg1                 ATGAGGTTCCGTCACAACCTCGGCCCTAGGACCGAGGCCGACATCGCCTC 998
BW_13S_globulin         ATGAGGTTCCGTCACAACCTCGGCCCTAGGACCGAGGCCGACATCGCCTC
BW_13S_globulin_mRNA_   ATGAGGTTCCGTCACAACCTCGGCCCTAGGACCGAGGCCGACATCGCCTC
Sesamum_11S_globulin    ATGAAATTCCGCACCAACGTTGAGAGCCGGAGGGAAGCAGATATTTTCTC
Perilla_legumin-like    ATGAAAATCATGTCCAACCTCGACAACACAAGAGAAGCAGATGTGTATTC
Magnolia_legumin        ATGAGAGTAAACCATTACATGGACAACCCTAGAGAAGCCGACATCTACAG
B.napus_cruciferin      GCAAGGTGCACGGATAACCTCGATGACCCATCTAACGCGGATGTGTATAA
Ficus_11S_globulin      CTGAGGATGAGGCACAACATCGACCGCCCTTCTCAGGCCGACATCTTCAA
                          .*..   .   . :** * *.    .  :   * ** ** .*    :

BW_leg1                 CAGACAGGCTGGAAGGATCCACTCCGTCGACCAGAACAAGCTTCCCATCC 1048
BW_13S_globulin         CAGACAGGCTGGTAGGATCCACTCCGTCGACCAGAACAAGCTTCCCATCC
BW_13S_globulin_mRNA_   CAGACAGGCTGGTAGGATCCACTCCGTCGACCAGAACAAGCTTCCCATCC
Sesamum_11S_globulin    CCGACAGGCCGGAAGAGTGCACGTCGTCGACAGGAACAAGCTTCCCATCC
Perilla_legumin-like    ACGACAGGCCGGAAAACTCAACGTCGTCGACATGCATAAGCTTCCCATCC
Magnolia_legumin        TAGGCAGGCAGGCCGTCTCAACTCAGTCAACATGAACAAGCTTCCCATCC
B.napus_cruciferin      GCCACAGCTTGGTTACATCAGCATTCTTAACAGTTATGATCTACCCATCC
Ficus_11S_globulin      CCCCCGCGGCGGACGCGTCACCACGGTCAACAACTTCAACCTTCCCATCC
                          .  *.    **  .  * . *    * .**.   : .* **:*******

BW_leg1                 TCGAGTTTATCGACATGAGCGCCGAGAAAG-CCCCCTTCTCCCG------ 1092
BW_13S_globulin         TCGAGTTTATCGACATGAGCGCCGAGAAAGGCCACCTTCTCCCGGTAT--
BW_13S_globulin_mRNA_   TCGAGTTTATCGACATGAGCGCCGAGAAAGGCCACCTTCTCCCG------
Sesamum_11S_globulin    TCAAATACATGGATTTGAGTGCTGAAAAAGGCAATCTTTACTCG------
Perilla_legumin-like    TCAGGGCCGTCGACATGAGCGCAGAGAAAGGAACCCTTTTCCCC------
Magnolia_legumin        TTCGGATGCTGGGCATGAGCTCTGAAAAGGGCTACCTCTACCAG------
B.napus_cruciferin      TTCGCGTACTTCGCCTCTCAGCCCTCCGTGGATCAATCCGTCAAGTGAGT
Ficus_11S_globulin      TCCGTTTCCTCCGCCTCACCGCCGAGAGAGGTGTCCTCTACAAG------
                        *  .     *  .  * :   *  : .. *     .*     .
```


BW_leg1
BW_13S_globulin
BW_13S_globulin_mRNA_
Sesamum_11S_globulin
Perilla_legumin-like
Magnolia_legumin
B.napus_cruciferin
Ficus_11S_globulin
ACAACGAAAGATATTATCTATTTAAAATCAATGCTTGCATATGTTTAAAA
GT 1094
ATAT
GGGATTGACATTGAAAGTTTGGTTCGGTTTATTTTTGGTTATTCGGTTTA
ATATATATACCTAAATGTTCATTCACTGAAATATATGAGCTTGTAAACTT 1144
ACATACACTACACATATTCAATTCACTGATATATATGAAGCTTGTAAACT
ATTAATTTGGCTATTACAATTATCGCTTCTCCATTTCTACATTTAAACAT
TATGTTTTATTTTTTTTTTGTG--CAGAACGCCATGTTAGCGCCGGCATG 1192
TATGTTTTATTTTTCTCTTCTGTGCAGAACGCCATGTTAGCGCCGGCATG
AACGCCATGTTAGCGCCGGCATG
AACGCACTCGTCAGCCCAGATTG
AACGCAATGCTGAGCCCAGACTG
AATGCCATATTCTCGCCCCACTG
GACTATCCAAACGTGTTGTGGGTGTAGAATGCAATGGTTCTTCCACAGTG
AACGCTATGATGGCACCACACTT
GGCGGTGTCAGGCCACAGAGTGTTCTACGTGTTGAGGGGTGAGGCTCAGA 1242
GCCGCTCTCAGGCCATAGAGTGTTCTACGTGTTGAGGGGTGAGGCTCAGA
GCCGCTCTCAGGCCATAGAGTGTTCTACGTGTTGAGGGGTGAGGCTCAGA
GTCCATGACAGGCCACACGATTGTATACGTGACAAGAGGCGACGCCCAGG
GGCGATGCAAGGGCACACGATCGTCTACGTGACACGGGGAAACGCCAAGG
GACGATCAATGCCCACAACATCTTCTACGTGACCCGCGGTGAGGCCCGTG
GAAGTCAAAGTCAAACGCGGTTCTCTACGTGACAGACGGGGAAGCCCAAA
CAACTTGAACAGCCACAGCGTGTTCTACGTCACCAGGGGAAGCGGCCGAT
TGCAGATAGTAGACGACAACGGTCAGACGGTGTTAGATGACAGAGTCTCA 1292
GGCAGATAGTAGACGACAACGGTCAGACGGTGTTAGATGACAGAGTCTCA
GGCAGATAGTAGACGACAACGGTCAGACGGTGTTAGATGACAGAGTCTCA
TCCAAGTAGTGGACCACAACGGACAAGCCCTGATGAACGACAGGGTGAAT
TCCAAGTGGTTGATCACAAGGGGCAGAGCCTGATGAACGACAGGGTGCAG
TGCAAGTGGTGGGCCATAATGGGCAGACGGTTCTCGATGACACAGTGAGA
TACAGGTGGTTAACGACAACGGTGACAGAGTGTTCGATGGACAAGTCTCT
GCCAGATCGTCGACGACTTCGGGCGGACCGTGTTCGATGGTGAGGTCCAG
CAAGGATCCATGGTTGTGATCCCACAGTTCTACACCTCGACATGTAGAGC 1342
GAAGGATCCATGGTTGTGATCCCACAGTTCTACATCTCGACATGTAGAGC
GAAGGATCCATGGTTGTGATCCCACAGTTCTACATCTCGACATGTAGAGC
CAGGGAGAAATGTTTGTGGTGCCTCAGTACTATACCTCGACGGCCCGTGC
CAGGGAGAGATGTTCGTGGTGCCGCAGTTCTACACCTCTACGGCTGAGGC
GAAGGTGACCTTGTGGTGTTCCCCCAATATTTTGCAGTCATGAAGAGGGC
CAAGGGCAGCTACTTTCCATTCCACAAGGATTCTCCGTTGTGAAACGCGC
GAGGGGCAGCTGTTGGTGGTGCCACAGAACTACGCCGTGGCGAAGCAAGC

```
BW_leg1                 AGGAAGAGACGGACTCGAGTACGTTTCATTCGAGACCACAGCAAACCCGA 1392
BW_13S_globulin         AGGAAGAGACGGACTCGAGTACGTTTCATTCGAGACCACAGCAAACCCGA
BW_13S_globulin_mRNA_   AGGAAGAGACGGACTCGAGTACGTTTCATTCGAGACCACAGCAAACCCGA
Sesamum_11S_globulin    AGGGAACAACGGCTTTGAATGGGTTGCTTTCAAGACCACCGGGAGTCCGA
Perilla_legumin-like    TGGAAACGAGGGGTTCGAGTGGGTGGCGTTCAAGACCAGTGGCTTCCCCA
Magnolia_legumin        GGGCAACAACGGCTTTGAATGGGTGTCGTTCAAAACGTCGGCCTCTCCCA
B.napus_cruciferin      AACAAGCGATCAGTTCAGGTGGATAGAATTCAAGACAAACGCAAACGCCC
Ficus_11S_globulin      CAGCAACCGCGGATTCGAGTGGATCGCCATCAAGACCAACGACAACGCCA
                         .  *.. .  .  * ...*. .*  . :**.*.** :  *  :   * .

BW_leg1                 TGAGTAGTCCATTGAACGGCCACGCTTCGGTTTTCAAGGGAATGCCAATT 1442
BW_13S_globulin         TGAGTAGTCCATTGAACGGCCACGCTTCGGTTTTCAAGGGAATGCCAATT
BW_13S_globulin_mRNA_   TGAGTAGTCCATTGAACGGCCACGCTTCGGTTTTCAAGGGAATGCCAATT
Sesamum_11S_globulin    TGCGCAGCCCTCTGGCTGGTTACACATCGGTGATCAGAGCAATGCCTCTT
Perilla_legumin-like    TGCGCAACCAGGTGGCCGGATACACGTCGGCCTTGAGAGCAATGCCGCTC
Magnolia_legumin        TGAGGAGCCCACTGGCGGGGTCCACATCAACTATCAAGGGAATGCCGTTG
B.napus_cruciferin      AGATCAACACTCTTGCTGGACGTACCTCAGTCATGAGAGGTTTACCATTA
Ficus_11S_globulin      TGAGAAACCCGCTCGCCGGAAGGATCTCGGCCATCCGAGCACTGCCGGAG
                        :*.  *. ..  * .. **    .  **..  :* ...* : *.**  :

BW_leg1                 CCTGTGCTCAGCAACTCGTACCAGATCTCGCCCCGTGCCGCCTATGAGCT 1492
BW_13S_globulin         CCTGTGCTCAGCAACTCGTACCAGATCTCGCCCCGTGCCGCCTATGAGCT
BW_13S_globulin_mRNA_   CCTGTGCTCAGCAACTCGTACCAGATCTCGCCCCGTGCCGCCTATGAGCT
Sesamum_11S_globulin    CAGGTCATCACAAATTCGTATCAGATTTCGCCGAACCAGGCTCAGGCCCT
Perilla_legumin-like    CAAGTTCTGACCAACGCCTACCAGATGTCTCCCAACGAGGCTAGGGCGAT
Magnolia_legumin        GAGGTGCTGACAAATGCATATCAGGTGTCATACAGAGAGGCCCAAAACCT
B.napus_cruciferin      GAGGTCATAGCCAATGGGTACCAAATCTCACTTGAAGAAGCAAGAAGGGT
Ficus_11S_globulin      GACTTGCTATCTAACGCATTTCGCATCTCGAGAGAGCAAGCCAGGAACTT
                         .  * .*    **    *: *. .* **     .  . **  . .   *

BW_leg1                 GAAG-----CAGACAAGGAGCCACGAGCACGGGCAGTTC-TCGCCTTTCG 1486
BW_13S_globulin         GAAG-----CAGACAAGGAGCCACGAACACGGTCTGTTC-TCACCTTTCG
BW_13S_globulin_mRNA_   GAAG-----CAGACAAGGAGCCACGAACACGGTCTGTTC-TCACCTTTCG
Sesamum_11S_globulin    GAAG-----ATGAACAGGGGCAGCCAGAG---TTTCTTG-CTGTCTCCGG
Perilla_legumin-like    CAAG-----ACCAATCGAGGTAGCCAGAC---TTTCTTG-TTGTCACCCA
Magnolia_legumin        GAAA-----TTCAATAGGGAACACCAGCTGATGTTCTTC-CCCCCTCCTA
B.napus_cruciferin      TAAGTTCAACACAATAGAGACCACTTTGACCCACAGTAG-TGGCCCAGCG
Ficus_11S_globulin      GAAGAACAACAGGGAAGAGGTCACTGTTTTCAGCACCCAGTCTTCTCAAC
                         **.        .  .*... ..*          :         *

BW_leg1                 G----AGGAAGGTCGTA-------------------------------- 1499
BW_13S_globulin         G----AGGAAGGTCTTAA------------------------------
BW_13S_globulin_mRNA_   G----AGGAAGGTCTTAAGTGTATGAACTAGGGCTTCTTTTATG------
Sesamum_11S_globulin    G----AGGGCGGCGATCTTAGATGTCTCTGCTGCTAATTACTAG------
Perilla_legumin-like    C----CCG-------------------CGCCCATGGAAGAAAG------
Magnolia_legumin        A----TCGCTCCTCTTAGACCTGAATTTTGCTATGTTTTCTTAG------
B.napus_cruciferin      AGCTACGGAAGGCCAAGGAAGGCTGATGCTTGAGTGCTTAAATGGCTGCG
Ficus_11S_globulin      AAGGCCGGAAGTACTAATTAAGAAGCTCTTGCTAGCTATAAATG------
```

Fig: 4.9

in the data bases, the sequence (accession no. GQ358523) was subjected to BLASTn analysis using the BLAST tool of NCBI. BLAST analysis clearly identified the sequences as a nucleotide sequence for the legumin gene. Clustal W alignment of the sequence with corresponding regions of nucleotide sequences of legumin genes from other plants is presented in (Fig. 4.9). Using an alignment that permitted maximum homology, the 1613bp nucleotide sequence amplified in the present study bearing accession no. GQ358523 showed 96% and 94% homology with 13S globulin (AY256960) and buckwheat Feleg51 13S globulin gene (AY359286.3) respectively. The percentage homology varied between 68 % with *Magnolia* legumin gene (X82465) to 75% with *Brassica napus* cruciferin (X592951). Besides these, the alignment revealed 71% and 69% homology with sesame 11S globulin (AF091842) and *Perilla* legumin (AF180392) respectively. Some of the distinguishing features observed in the 1613 bp nucleotide sequence amplified in the present study included an 8 base deletion at position 914 and a 6 base "ACCGCT" insertion/ substitution at position 727. The 6 base insertion/substitution distinguished the 1613 bp nucleotide sequence amplified in the present study from nucleotide sequence of other buckwheat globulin genes available in the data bases. The sequence was subjected to analysis by GENSCAN 1.0 for determination of intron/exon architecture. GENESCAN1.0 identified three exons within the 1613bp nucleotide sequence bearing accession no. GQ358523 amplified in the present study. While the I^{st} exon had a length of 210 bases and was located at position 34 to243, the 2^{nd} exon had a length of 320 bases and was located at position 837 to 1156. The 3^{rd} exon had a length of 240 bases and was located at position 1404 to1607. GENESCAN1.0

Fig. 4.10: Diagramatic representation of the intron/exon architechture in the 1613 bp nucleotide sequence of 1.6 kb amplicon generated buckwheat genomic DNA as the template and primer pair SS8F-SS3R.

Fig. 4.11: Deduced amino acid sequence of the coding region of 1613 bp nucleotide sequence of the 1.6 kb amplicon generated buckwheat genomic DNA as the template and primer pair SS8F-SS3R

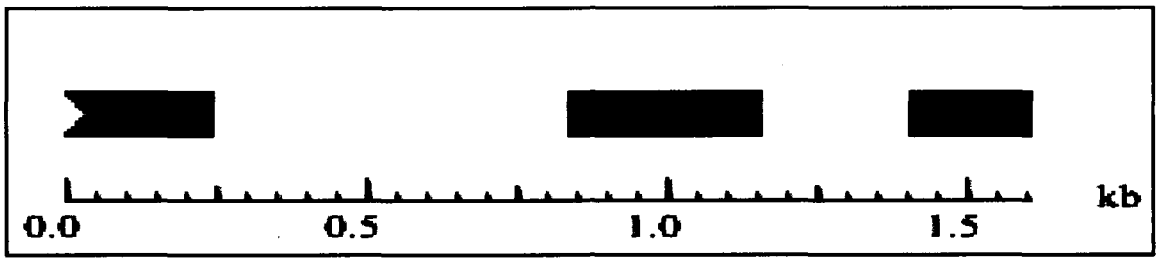


Fig: 4.10

```
10          20          30          40          50          60
MRGDEEEEGF ESGRRMTDAH QPTRRVRKGD VVALPQGTVH WCFNDGQEDL VVVAVHNLNT

         70          80          90         100         110         120
DANQLDQSLK LLAEALGTEE ETVRKMQESD ERGPIVKARK NMRQMVTPPR FGREQDEDET

        130         140         150         160         170         180
NGLEESFCNM RFRHNLGPRT EADIASRQAG RIHSVDQNKL PILEFIDMSA EKAPFSRAGR

        190         200         210         220         230         240
DGLEYVSFET TANPMSSPLN GHASVFKGMP IPVLSNSYQI SPRAAYELKQ TRSHEHGQFS

  245
PFGGX
```

Fig: 4.11

```
CLUSTAL W (1.81) multiple sequence alignment

BW_leg1                    --------MRGDEEEEGFESGRRMTDAHQPTRRVRKGDVVALPQGTVHWC
BW__13S_globulin           --------RGDEEEEEGFESGRRMTDAHQPTRRVRKGDVVALPQGTVHWC
sesame_globulin            --------TMERTEASEQQDRGSVRDLHQKVHRLRQGDIVAIPSGAAHWC
Rice_glutelin              --------SGQAQLTESQSQSQKFKDEHQKIHRFRQGDVIALPAGVAHWC
Zizania_glutelin           ------QQSMQAQLTGSQSQSQKFKDEHQKINRFRQGDVIALPAGVAHWC
Oat_12s_globulin           -----------PFDRAQGQSQSHLKDEHQRVHRFKQGDVIALPAGIVHWG
Amaranthus_11S_globulin    FQGGEDERIREQGSRKFGMRGDRFQDQHQKIRHLREGDIFAMPAGVSHWA
Ginkgo_11S_globulin        --------------TSRGGEGQQSQERSQKIRRVRRGDVVAIPAGVAYWL
Calocedrus_legumin         EEEEEERGRGREEECSRHERERARREESSQKIRRVRRGDVVAIYAGVAYWW
                                                 :  *  .:.:.**:.*:  *  :*

BW_leg1                    FNDGQEDLVVVAVHNLNTDANQLDQSLK----------------------
BW__13S_globulin           FNDGQEDLVVVAVHNLNTDANQLDQSLKTFFLAGGVQGGSK--------E
sesame_globulin            YNDGSEDLVAVSINDVNHLSNQLDQKFRAFYLAGGVPRSGE--------Q
Rice_glutelin              YNDGEVPVVAIYVTDLNNGANQLDP-RQRDFLLAGNKRNPQ-------AY
Zizania_glutelin           YNDGEVPVVAIYVIDINNAANQLDP-RQRDFLLAGNMRSPQ-------AY
Oat_12s_globulin           YNDGDAPVVAIYVFDVNNNANQLEP-RQKEFLLAGNNKEDQ-------QF
Amaranthus_11S_globulin    YNNGDQPLVAVILIDTANHANQLDKNFPTRFYLAGKPQQEHSGEHQFSRE
Ginkgo_11S_globulin        YNDGNRRLQIVAIADTSNHQNQLDQTYRPFYLAGSAPSG--------AQK
Calocedrus_legumin         YNDGDKPLRTVAIADASNHQNQLDKRYRPFFLAGSPATRER------SER
                           :*:*.  :  : : :     ***:

BW_leg1                    ------------------LLAEALGTEEETVRKMQ-ESDERGPIVKARKN
BW__13S_globulin           GKSQKLNSNNILSAFETKLLAEALGTEEETVRKMQ-ESDERGPIVKARKN
sesame_globulin            EQQARQTFHNIFRAFDAELLSEAFNVPQETIRRMQSEEEERGLIVMARER
Rice_glutelin              RREVEERSQNIFSGFSTELLSEALGVSGQVARQLQCQNDQRGEIVRVEHG
Zizania_glutelin           RREVENQSQNIFSGFSAELLSEALGISTGVARQLQCQNDQRGEIVRVEHG
Oat_12s_globulin           G-------QNIFSGFNIQLLSEALGISQQAAQRIQSQKEQRGEIIRVTQR
Amaranthus_11S_globulin    SRRGERNTGNIFRGFETRLLAESFGVSEEIAQKLQAEQDDRGNIVRVQEG
Ginkgo_11S_globulin        AAGATSIGDNILQGFDTDTLAEAMGISQDTARRIQQNQKKGLIVKVERGL
Calocedrus_legumin         AGEGEKYGGNVLAGFDANMLAEALGVRRQVVIDIQENNRESGLIVRVNDP
                                           *:*::.         :* :. .   :

BW_leg1                    MRQMVTPPRFGREQD------------------EDE--TNGLEESFCNMR
BW__13S_globulin           MRQMVTPPRFGREQD------------------EDE--TNGLEESFCNMR
sesame_globulin            MTFVRPDEEEGEQEH------------------RGRQLDNGLEETFCTMK
Rice_glutelin              LSLLQPYASLQEQEQGQVQSRERYQEGQYQQSQYGSGCSNGLDETFCTLR
Zizania_glutelin           LSLLQPYASLQEQEQKQEQPRERYQVTQHQQSQYGGGCSNGLDETFCAMR
Oat_12s_globulin           LQFLKPTMSQQDRS----------------------FNGLEENFCSLE
Amaranthus_11S_globulin    LHVIKPPSRAWEERE--------------QGSRGSRYLPNGVEETICSAR
Ginkgo_11S_globulin        RMPGP--------------------PSDDYERERREREGNNVEEFYCSMR
Calocedrus_legumin         RRAGPGGEGAPLFLN--------TVAEASEDMKSWGINPGGLHQFYCNMR
                                                          ..:.:  *  .

BW_leg1                    FRHNLGPRTEADIASRQAGRIHSVDQNKLPILEFIDMSAEKAPFS-----
BW__13S_globulin           FRHNLGPRTEADIASRQAGRIHSVDQNKLPILEFIDMSAEKGHLLPNAML
sesame_globulin            FRTNVESRREADIFSRQAGRVHVVDRNKLPILKYMDLSAEKGNLYSNALV
Rice_glutelin              VRQNIDNPNRADTYNPRAGRVTNLNTQNFPILSLVQMSAVKVNLYQNALL
Zizania_glutelin           IWQNIDNPNLADTYNPRAGRVTNLNSQKFPILNLIQMSAVKVNLYQNALL
Oat_12s_globulin           AKQNIENPKRADTYNPRAGRITRLHGQNFPILNLVQMSATRVNLYQNAIL
Amaranthus_11S_globulin    LAVNVDDPSKADVYTPEAGRLTTVNSFNLPILRHLRLSAAKGVLYRNAMM
Ginkgo_11S_globulin        LRHNADDSEDADVYVRNGGRLNTVNRLKLPALRSLRLGAERGILQPNAMF
Calocedrus_legumin         LRHNADRPDDADIFVRDGGRLNTVNRFKLHALSHLNLAAERGVLRPGAMF
                              *      **     .**:  :.  ::  *  : :.* :  :

BW_leg1                    --------------------------------------------------
BW__13S_globulin           APAWPLSGHRVFYVLRGEAQRQIVDDNGQTVLDDRVSEGSMVVIPQFYIS
sesame_globulin            SPDWSMTGHTIVYVTRGDAQVQVVDHNGQALMNDRVNQGEMFVVPQYYTS
Rice_glutelin              SPFWNINAHSVVYITQGRARVQVVNNNGKTVFNGELRRGQLLIIPQHYAV
Zizania_glutelin           SPFWNINSHSVVYVTQGCARVQVVNNNGKTVFNGELRRGQLLIIPQHYVV
Oat_12s_globulin           SPFWNINAHSVVYMIQGHARVQVVNNNGQTVFNDRLRQGQLLILPQHYVV
Amaranthus_11S_globulin    APHYNLNAHNIMYCVRGRGRIQIVNDQGQSVFDEELSRGQLVVVPQNFAI
Ginkgo_11S_globulin        APSWLN-AHAVMYVTRGQGRIQIVQNEGRRVFDGAVKEGQFLVIPQLHAI
Calocedrus_legumin         APSWVA-CHAILYATRGDARIEVVENRGRRVFDGRVQEGQFLVIPQFYAV
```

```
BW_leg1                    --RAGRDGLEYVSFETTANPMSSPLNGHASVFKGMPIPVLSNSYQISPRA
BW__13S_globulin           TCRAGRDGLEYVSFETTANPMSSPLNGHASVFKGMPIPVLSNSYQISPRA
sesame_globulin            TARAGNNGFEWVAFKTTGSPMRSPLAGYTSVIRAMPLQVITNSYQISPNQ
Rice_glutelin              VKKAQREGCAYIAFKTNPNSMVSHIAGKSSIFRALPNDVLANAYRISREE
Zizania_glutelin           VKKAQREGCAYIAFKTNPNSMVSHIVGKSSIFRALPTDVLANAYRISRED
Oat_12s_globulin           LKKTEREGCQYISFKTNPNSMVSHIAGKSSILRALPVNVLANAYRISRQE
Amaranthus_11S_globulin    VKQAFEDGFEWVSFKTSENAMFQSLAGRTSAIRSLPIDVVSNIYQISREE
Ginkgo_11S_globulin        AKQAGKDGLEWISFTTSDSPIRSTLTGRNSVLKAMPQEVVMNAYRINGKD
Calocedrus_legumin         MKRPGDQGFDWITFTTCHSPIRSSFTGRNSVLKGMPQEVVMNAYNISRRE
                             :.  :*  :::* *  ..: . : *  * ::.:*  *: * *.*. .

BW_leg1                    AYELKQTRSHEHGQFSPFGGX----------------------
BW__13S_globulin           AYELKQTRSHEHGLFSPFGGRS--------------------
sesame_globulin            AQALKMNRGSQSFLLSPGGRRS--------------------
Rice_glutelin              AQRLKHNRGDEFGAFTPIQYKSYQDVYNAAESS---------
Zizania_glutelin           AQRLKHNRGDELGAFTPLQYKSYQDVSSVAASS---------
Oat_12s_globulin           VRNLKNNRGQESGVFTPKFTQTSFQPYPEGEDESSLINKASE
Amaranthus_11S_globulin    AFGLKFNRPETTLFRSSGQGEYRRKISIA-------------
Ginkgo_11S_globulin        ARDLRRNREHETIILSPTPQHQQPRAIE--------------
Calocedrus_legumin         AHELRWNREHEFLILPPRGGQRESEQ----------------
                           .  *: .*       ..
```

Fig:4.12

identified the 210 bp exon as the initial exon and the other two exons as internal exons. The position the three exons on the 1613bp nucleotide sequence is shown in (Fig. 4.10).

The deduced amino acid sequence of the 1613 bp amplicon comprised of 245 amino acids with a theoretical molecular mass of 27 kDa. The sequence is presented in Fig. 4.11. Sequence homology analysis of the 245 residue amino acid sequence with BLASTp against non-redundant protein database clearly identified the protein as belonging to the legumin subfamily. Using an alignment that permitted maximum homology, the deduced amino acid sequence showed a maximum of 70% homology with 13S buckwheat globulin (AAP15457). The percentage of homology varied between 54-40% with deduced amino acid sequences of legumins from other crops including *Perilla frutescens* legumin-like protein (AAF19607) and legumin precursor of *Magnolia salicifolia* (CAA57848). Clustal W alignment of the sequence with amino acid sequences of legumin-type proteins available in the database is presented in Fig. 4.12. Motif search on the deduced amino acid sequence with MOTIF SCAN identified two cupin domains between 134 to 226 and 103 to 113 in the sequence.

The phylogenetic tree constructed from the alignment data of the deduced amino acid sequence with amino acid sequences of SSPs available in EMBL database described the relationship of the sequence with amino acid sequences of other seed storage proteins (Fig. 4.13). The tree showed a clear division into legumin type proteins from angiosperms and gymnosperms. The sequences of two legumin type seed storage proteins of buckwheat emerged as a separate group within the tree.

Fig. 4.13: Phylogenetic tree based on alignment matrix of the deduced amino acid sequence of the 1.6 kb DNA fragment amplified with buckwheat genomic DNA as the template and primer pair SS8F-SS3R with the amino acid sequences of legumin-type proteins from various angiosperms and gymnosperms. Accession numbers of the sequences selected for the analysis are indicated against each branch.

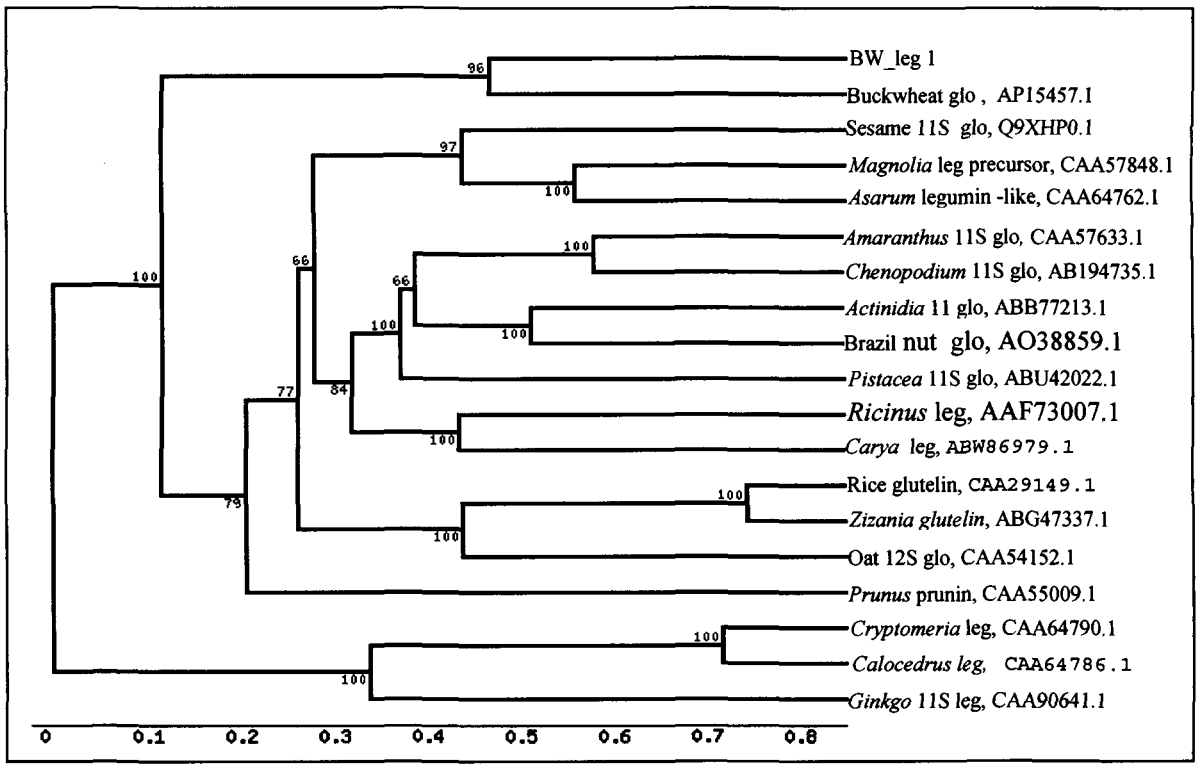

BW_leg 1
Buckwheat glo , AP15457.1
Sesame 11S glo, Q9XHP0.1
Magnolia leg precursor, CAA57848.1
Asarum legumin -like, CAA64762.1
Amaranthus 11S glo, CAA57633.1
Chenopodium 11S glo, AB194735.1
Actinidia 11 glo, ABB77213.1
Brazil nut glo, AO38859.1
Pistacea 11S glo, ABU42022.1
Ricinus leg, AAF73007.1
Carya leg, ABW86979.1
Rice glutelin, CAA29149.1
Zizania glutelin, ABG47337.1
Oat 12S glo, CAA54152.1
Prunus prunin, CAA55009.1
Cryptomeria leg, CAA64790.1
Calocedrus leg, CAA64786.1
Ginkgo 11S leg, CAA90641.1
0
0.1
0.2
0.3
0.4
0.5
0.6
0.7
0.8

Fig: 4.13

Fig. 4.14 (a) PCR amplification profile with buckwheat genomic DNA as template and primer SS8F-SS6R M: *Eco*RI/*Hind*III digested λ DNA, L1: Negative control (without template), L2: Positive control (without primers) L4 & L5: amplification product of the PCR..

(b) Autoradiograph of the 1.4 kb PCR product hybridized with [α-^{32}P]-dATP labelled buckwheat legumin gene specific probe (DQ200889).

Fig. 4.15 Partial nucleotide sequence of 564 bases for the 1.4 kb PCR product amplified from buckwheat genomic DNA with primer pair SS8F-SS6R

Fig: 4.14 (a) (b)

```
1   GCAATCCACG CGCCGGTTGT ATCATCACCG TAAACAGCAA CAAATCTCCC
51  AATCCTCGAA ATCATCAAGT ATAGCGCCCC AGCACGTCGT CCATATACAA
101 GGTAATATAA ACCCAAGTTG AAGACCTAAA ATCGACAAAT TTGAAACCAG
151 AACTCGAGTA GACTCAGATT GACGAGAATT CGTCGAACAT ATGCAGAATG
201 AGATCCTGAG ACCCAGATGT ATTTTGAAGA CGCAGAGGGC TCCGTAGCTG
251 ACGAGAGGAG AAGAAAGAAT CACAATTATG GTTGTTGTAG GAAGAAGTGT
301 GATCACTGCC AACGTGCAGC GAGGACAGAA GAGGGTGGCA CCACAGAGAA
351 TCCGGTCGTA TGCGAAGTCA GGAATTGATG TTCCCCATTA CCCCCAGAAT
401 GCCCGGGCTC CTGGCGCGGA AAATTATCTG GGAGCAGATA CAACTTCGGG
451 GGAGGCGGAG GAGCAAGAGG ATGAAGATGC CAACCAGACG ACCCCTCGAC
501 CTTCGTTCCC CGCGCCCCCG TGCCCCTCCC CCCCGTCCTC CCCGCCTGCC
551 CCAATCCAGG AGAT
```

Fig:4.15

PCR amplification with oligonucleotide primer pair SS8F and SS6R and chain reaction comprising hot start at 94°C for 5 minutes, 35 cycles comprising of denaturation at 94°C for 1minute, annealing at 68°C for 1minute and chain extension at 72°C for 1minute followed by one reaction of chain elongation at 72°C for 10 minutes amplified a DNA fragment having an apparent molecular mass of 1.4 kb. The amplicon resolved as a single band corresponding to a molecular mass of 1.4 kb on 0.8% agarose gel (Fig. 4.14a). Southern hybridization of the amplicon with [α-^{32}P]-dATP labelled buckwheat legumin cDNA (acc. no. DQ200889) at hybridization temperature of 65°C generated a positive signal on the exposed X-ray film corresponding to the position of the amplified DNA on the nylon membrane (Fig. 4.14 b). This indicated that the indicating that the DNA fragment amplified with oligonucleotide primer pair SS8F and SS6R shared sequence homology with buckwheat legumin-like protein gene bearing accession no. DQ200889. BLASTn analysis 564 bases partial nucleotide sequence of the 1.4 kb amplicon revealed 88% sequence homology with legumin-like protein gene bearing acc. No. D87980. The percentage homology with other buckwheat SSP gene sequences (AF216801, AY245536 and AF152003) available in the genebank databases varied from 78 to 86%.

Discussion:

Seed storage proteins constitute an important component of seed reserves and serve as a source of nitrogen during the initial stages of growth of the seedlings. Besides acting to provide soluble nitrogenous compounds during the initial stages of seedling growth, seed storage proteins also constitute an important source of dietary

proteins for human consumption. Cereal grains and legume seeds are the two major sources of vegetarian dietary proteins for human consumption. However, the nutritional quality of the proteins in both these crops does not match the WHO standards of a protein with nutritionally balanced amino acid composition. 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 and Larkins, 1989).

The main storage protein in grains of common buckwheat is a 13S globulin with molecular mass of 280 kDa. The protein comprises more than 50% of the total protein content of grains (Rout, 1996). The amino acid composition of the protein matches the WHO recommended values for a nutritionally rich protein with a balanced amino acid composition (Rout and Chrungoo, 1996). The nutritionally rich component of protein is a 26kDa basic subunit, which has more than 6% lysine and nearly 2% methionine (Rout and Chrungoo, 1996, 1997). Due to the balanced amino acid composition, high nutrient value and homology with seed storage proteins of leguminous group of plants, this protein could be an important candidate for compensation of limiting amino acid in plants deficient in such amino acids, through heterologous gene transfer across species. This requires isolation and cloning of full length genes and their promoters for use in crop improvement programmes aimed at development of transgenics with improved nutritional quality of seed storage proteins.

One of the biggest obstacles in isolation high quality DNA from starch/protein rich plant tissues is the interferences in purification procedures due to

presence of carbohydrates, proteins and even polyphenolic compounds. The modified CTAB protocol used in the present study yielded good quality genomic DNA. Analysis of the isolated genomic DNA on 0.8% agarose gel revealed that the sample was free of RNA contamination and has not undergone any shearing during isolation. The A260/280 ratio of >1.8 indicated the sample to be quite pure. The isolated genomic DNA was digested with to check the restriction digestion profile of the DNA isolated from shoot tissues of common buckwheat. The *Eco*RI digested resolved as a uniform streak ranging in size from 21.0 to 0.5 kb. The smear also showed four bands having molecular masses of 0.904 kb, 0.832 kb, 0.625 kb and 0.5 kb. The appearance of distinct bands in *Eco*RI digested DNA indicates the presence of *Eco*RI repeats of varying lengths in buckwheat genomic DNA. While the *Hind*III and *Nco*I digested DNA resolved as uniform smears ranging from in molecular mass from 0.3 kb to 20 kb, no distinct bands were visible in the electrophoretic profile of either *Hind*III or *Nco*I digested DNA. The absence of any distinct bands in *Hind*III and *Nco*I digested DNA indicates the absence of any repeats corresponding to either *Hind*III or *Nco*I in the buckwheat genomic DNA. Similar results have been observed by Bharali (2002) on the restriction digestion profile of buckwheat genomic DNA. Bharali (2002) has reported 6 bands representing *Eco*R1 repeats in genomic DNA of common buckwheat.

Even though isolation of genes from genomic DNA libraries by screening of the library with appropriate probes offers a reliable system for cloning full length genes, PCR amplification using oligonucleotide primers designed from the nucleotide sequences of conserved regions of 5'upstream and 3'downstream regions

of the genes offers a much quicker way of isolation of full length genes directly from genomic DNA isolated from the target plants. In the present study polymerase chain reaction with buckwheat genomic DNA as the template and combination of forward primer having the sequence 5' ATGCTTCATGGGGTGCTTCTATGC 3' and reverse primer having the sequence 5' TTAAGACCTTCCTCCGAAAG 3' amplified a 835 bp DNA fragment which showed sequence homology with mitochondrial NADH dehydrogenase gene of diverse plants ranging from bryophytes to angiosperms. Clustal W alignment of nucleotide sequence of the 835 bp amplicon with nucleotide sequences mitochondrial NADH dehydrogenase gene of bryophytes, pteridophytes, gymnosperms and angiosperms revealed a high degree of sequence conservation for the gene across these groups of plants. The amplification of NADH dehydrogenase gene by the primers designed for SSP genes could be due to the presence of short sequences homologous to the designed primers in the mitochondrial NADH dehydrogenase. The most distinguishing feature of the sequence amplified in the present study is a 6 base "CCTCCC" insertion/substitution at position 203 in the partial nucleotide sequence of the buckwheat NADH dehydrogenase amplified in the present study. An important difference between in the sequences of mitochondrial dehydrogenases of angiosperms and gymnosperms, observed in the alignment, is addition of 4 based each viz. "CGCG" and "CTTT" respectively at positions 55 and 231 corresponding to buckwheat mitochondrial dehydrogenase (DQ852632) in the mitochondrial NADH dehydrogenase of angiosperms compared to that of the gymnosperms. The nucleotide sequences of NADH dehydrogenases of pteridophytes

and bryophytes showed a deletion of 3 bases viz. "AGT" at position 122 corresponding to the buckwheat mitochondrial dehydrogenase.

Polymerase chain reaction with buckwheat genomic DNA as the template and primer combinations SS8F-SS3R, SS8F-SS1R, SS4F-SS3R and SS2F-SS3R respectively amplified DNA fragments showing apparent molecular mass of 1.6 kb, 1.1 kb, 0.85 kb and 1.3 kb. Southern hybridization of the of the 1.1 kb, 1.6 kb and 1.3 kb amplicons with [α-^{32}P]-dATP labelled 830bp amplicon clearly established a high degree of sequence homology between the amplicons. The nucleotide sequence of 1613 bases of the 1.6 kb DNA fragment, the longest fragment amplified by primer pair SS8F-SS3R, showed 94% and 96% sequence homology respectively with buckwheat Feleg51 13S globulin gene (AY359286.3) and 13S globulin mRNA sequences (AY256960). A statistical evaluation of the alignments revealed the sequence homologies were highly significant. Some of the distinguishing features observed in the nucleotide sequence of the 1613bp amplicon included an 8 base deletion at position 914 and a 6 base "ACCGCT" insertion/substitution at position 927. This distinguished the nucleotide sequence of the 1613bp DNA fragment amplified in the present study from nucleotide sequences of other globulin genes available in the data bases.

GENSCAN 1.0 detected three exons in the nucleotide sequence of the 1613bp DNA fragment amplified in the present study. While the 210 bp exon between positions 34-343 has been identified as the initial exon, the other two exons were identified as being internal to the sequence. Legumin genes, coding for the most widely distributed group of seed storage proteins show a fairly uniform intron-exon

organization having three introns strictly conserved in position (Shewry, 1995). However, the coexistence of two- and three- intron structure has led to the assumption that legumin genes have evolved by addition (Shotwell and larkins, 1989), or conversely, by loss of introns (Vonder *et al.*, 1988). The latter scenario gained support from the recently described exceptional five exon/four intron legumin genes from *Ginkgo biloba* which may be viewed as more ancient states of legumin gene structure (Hager *et al.*, 1995). The *Ginkgo biloba* legumin genes have introns I to III located at exactly the same conserved positions as is known for legumin genes of angiosperms along with an additional intron downstream from intron III which interrupts the sequence coding for the β-polypeptide C-terminal region. The legumin gene from *Welwitschia* (EMBL 250780) also has been reported to have an intron matching the *Ginkgo* legumin gene intron IV. This may indicate intron IV as being ancestral. The recognition of homology in intron/exon patterns is of importance in elucidation of structural similarities between legumin- and vicilin-type storage proteins. On the basis of their observations on intron-exon structures from most of the known legumin genes (Hager *et al.*, 1996) have suggested that the evolution of legumin genes has proceeded with loss of introns during the course of evolution. Our results on the intron-exon organization is in conformity with the observations of (Shewry, 1995).

The amino acid sequence deduced from the nucleotide sequence of GQ358523 comprised of 245 amino acids with a theoretical molecular mass of 27 kDa. Sequence homology analysis of the 245 residue amino acid sequence with BLASTp against non-redundant protein database clearly identified the protein as

belonging to the legumin subfamily. Using an alignment that permitted maximum homology, the deduced amino acid sequence showed a maximum of 70% homology with 13S buckwheat globulin (AAP15457). The percentage of homology varied between 54-40% with deduced amino acid sequences of legumins from other crops including *Perilla frutescens* legumin-like protein (AAF19607) and legumin precursor of *Magnolia salicifolia* (CAA57848). Further, identification of two cupin domains in the deduced amino acid sequence indicated that the protein belongs to the cupin superfamily. The cupins are a large superfamily of proteins, including seed storage proteins, which are thought to have originated by divergent evolution from a common ancestor. They share a common hexameric architecture, which has been described as 'double-stranded β-helix' with each helix having a conserved barrel domain.

The phylogenetic tree constructed from the alignment data of the deduced amino acid sequence with amino acid sequences of SSPs available in EMBL database showed a clear division into legumin type proteins from angiosperms and gymnosperms. The sequences of two legumin type seed storage proteins of buckwheat emerged as a separate group within the tree. Our results on phylogenetic relationships amongst legumin genes derived by neighbour joining distance method are in conformity with the observations of Hager *et al.* (1996).

CHAPTER: V
AMPLIFICATION OF 5' UPSTREAM FLANKING REGION BY PCR-BASED GENOME WALKING

Experimental:

Grains of common buckwheat [*Fagopyrum esculentum* Moench (accession No. IC18890)] were procured from the North Eastern Regional Station of the National Bureau of Plant Genetic Resources, Shillong. Healthy grains of uniform size were screened and used for the present study. The grains were also sown in experimental beds in the Botanical garden of Department of Botany, North Eastern Hill University, Shillong for multiplication.

The investigations carried out under the present study involved construction of genome walking adapter libraries from genomic DNA isolated from etiolated seedlings of common buckwheat and amplification of the SSP genes and their 5' upstream regions from the library using genome walking approach. The approach also involved amplification of SSP gene from genomic DNA of common buckwheat by polymerase chain reaction using oligonucleotide primers designed from sequence

data obtained during genome walking. For achieving the identified targets for both the approaches, healthy grains of uniform size were germinated in dark in a seed germinator at 27°C and 85% R.H. The germinating seedlings were maintained in a plant growth chamber for 14 days in dark till the Ist leaves emerged fully. Genomic DNA was isolated from 14 days old etiolated seedlings of buckwheat following the CTAB buffer extraction protocol. The isolated DNA was digested with appropriate restriction enzymes followed by ligation with adapter primers to construct a genome walking library. PCR amplification was carried out with primers designed from conserved regions of legumin-type SSP gene sequences in the database as well as the sequence data obtained during genome walking. The primers are shown in Table 3.1.

Results:

The 5' upstream region of buckwheat legumin-type SSP gene was isolated by PCR-based genome walking using Universal Genome Walker Kit from Clontech (USA). 0.5 µg of genomic DNA isolated from etiolated seedlings of common buckwheat was digested separately with *Dra*I, *Eco*RV, *Pvu*II and *Stu*I followed by size fractionation on 1% agarose gel at 80 volts for 3 hours. While the *Dra* I and *Eco*RV digested DNA resolved as a uniform streak ranging in molecular mass from 20.0 to 0.5kb, there was no streak of digested DNA visible on the agarose gel for *Pvu*II and *Stu*I treated DNA. The DNA in *Pvu*II and *Stu*I treated reaction mixtures was visible on the gel as an intact band corresponding to molecular mass of 21kb (Fig. 5.1). The *Dra*I and *Eco*RV digested DNA was separately ligated to adapter primers to generate the genome walker adapter libraries.

Fig.5.1: Electrophoresis profile of genomic DNA isolated from etiolated seedlings of common buckwheat and digested with D*ra*I (L1), *Eco*RV (L2), *Pvu*II (L3) and *Stu*I (L4).

Fig. 5.2: Electrophoresis profile of amplification products of primary PCR with L1: *Dra*I adapter library as the template and primer pair AP1-SS6R, L2: *Dra*I adapter library as the template and primer pair AP1- SS11F, L3: *Dra*I adapter library as the template and primer pair AP1- SS13R, L4: *Eco*RV adapter library as the template and primer pair AP1-SS6R, L5: *Eco*RV adapter library as the template and primer pair AP1-SS11F, L6: *Eco*RV adapter library as the template and primer pair AP1-SS13R, M: *Eco*RI/*Hind*III digested λ DNA.

Fig. 5.3: Electrophoresis profile of amplification products of nested PCR with products of primary PCR with *Dra*I adapter library and primer pair AP1- SS6R as the template and oligonucleotide primer pair AP2-SS10R (L2), L1; control, M: *Eco*RI/*Hind*III digested λ DNA.

Fig. 5.4: Electrophoresis profile of amplification products of nested PCR with products of primary PCR with L2: *Eco*RV adapter library and primer pair AP1- SS11R as the template and oligonucleotide primer pair AP2-SS12R, L3: *Eco*RV adapter library and primer pair AP1- SS13R as the template and oligonucleotide primer pair AP2-SS13R, L4: *Dra*I adapter library and primer pair AP1- SS13R as the template and oligonucleotide primer pair AP2-SS13R, L1: control, M: *Eco*RI/*Hind*III digested λ DNA.

Fig. 5.1

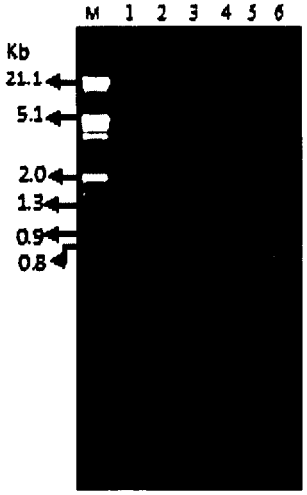


Fig. 5.2

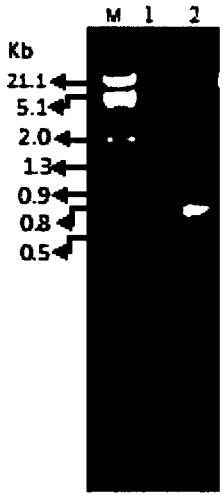


Fig. 5.3

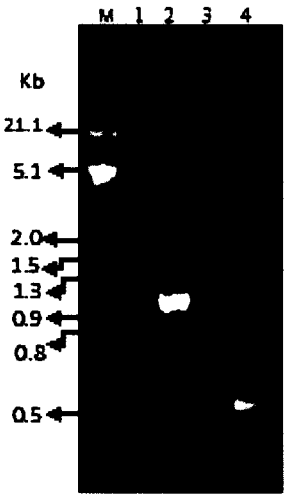


Fig. 5.4

The designing of gene specific primers is crucial for the success of the PCR-based genome walking. Buckwheat legumin gene sequences bearing accession Nos. AF152003, AY256960, AY245536 AF216800, AF216801, DQ200889, D87980, D87982 were retrieved from NCBI Genbank and analyzed by Webcutter to locate the *Dra*I *and Eco*RV restriction sites on the sequences. Based on the location of the restriction sites, gene specific primers were designed to anneal to the regions which were not in the immediate vicinity of the *Dra*I and *Eco*RV restriction sites. The oligonucleotide primer combinations used in the present study for PCR based genome walking and the size of amplified products obtained for each primer pair are shown in table 5.1. When the amplified products of primary polymerase chain reactions with *Dra*I and *Eco*RV adapter libraries as the templates and adapter primer AP1 and gene specific primers SS6R, SS11F and SS13R were electrophoressed on agarose gel, a uniform streak ranging in size from 0.2 kb to 2 kb could be detected under UV transilluminator (Fig. 5.2).

Nested PCR with the amplified DNA generated from PCR of *Dra*I adapter library and oligonucleotide primer pair AP1 and SS6R as the template and primer pair comprising adapter primer AP2 and gene specific primer SS10R amplified a DNA fragment showing apparent molecular mass of 0.7 kb. The amplicon was detected on agarose gel as a single band with molecular mass of 0.7 kb (Fig. 5.3). The amplicon has been designated as Buckwheat *Dra*I lib. genome walking product 1 (*BwDra*I lib.GWP1). Secondary (nested) PCR with the amplified DNA generated from PCR of *Dra*I adapter library and oligonucleotide primer pair AP1 and SS11F as the template and primer pair comprising adapter primer AP2 and gene specific

Table 5.1 List of primer combinations used for PCR-based genome walking and the sizes of bands amplified.

Template Adapter Library	**Primer pair for Primary PCR**	**Primer pair for Secondary PCR**	**Amplicon size (bp)**	**Named as**
Dra I	AP1-SS6R	AP2-SS10R	776	Bw*Dra*I lib. GWP1
	AP1-SS11F	AP2-SS12F	nil	-
	AP1-SS13R	AP2-SS14R	574	Bw*Dra*I lib. GWP2
*Eco*RV	AP1-SS6R	AP2-SS10R	nil	-
	AP1-SS11F	AP2-SS12F	1127	BwEcoRV lib. GWP3
	AP1-SS13R	AP2-SS14R	903	BwEcoRV lib. GWP4

primer SS12F did not generate any amplicon in the reaction mixture. However, nested PCR with the amplified DNA generated from PCR of *Dra*I adapter library and oligonucleotide primer pair AP1 and SS13R as the template and primer pair comprising of adapter primer AP2 and gene specific primer SS14R generated an amplicon having apparent molecular mass of 0.57 kb. The amplicon was detected on agarose gel as a single band showing molecular mass of 0.57kb (Fig. 5.4). The amplicon has been designated as Buckwheat *Dra*I lib. genome walking product 2 (*BwDra*I lib.GWP2).

While the nested PCR with the amplified DNA generated from primary PCR of *Eco*RV adapter library and oligonucleotide primer pair AP1 and SS6R as the template and primer pair comprising adapter primer AP2 and gene specific primer SS10R did generate any amplicons, PCR with the amplified DNA generated from primary PCR of *Eco*RV adapter library and and oligonucleotide primer pair AP1 and SS11F as the template and primer comprising adapter primer AP2 and gene specific primer SS12F generated an amplicon showing molecular mass of 1.1 kb on the agarose gel (Fig. 5.4). The amplicon has been designated as *BwEco*RV lib.GWP3. Nested PCR with the amplified DNA generated from PCR of *Eco*RV adapter library and oligonucleotide primer pair AP1 and SS13R as the template and primer pair comprising of adapter primer AP2 and gene specific primer SS14R caused amplification of a DNA fragment showing molecular mass of 0.9 kb on the agarose gel (Fig. 5.4). The amplicon has been designated as *BwEco*RV lib.GWP4.

Southern hybridization of the Buckwheat *Dra*I lib. genome walking product 1 (*BwDra*I lib.GWP1) with [α-^{32}P]-dATP labelled buckwheat legumin gene specific

DNA (acc. No. DQ200889) generated a strong positive signal on the exposed X-ray film; the position of the signal corresponding to the position of the amplified *BwDra*I lib.GWP1 on the nylon membrane (Fig. 5.5). The nucleotide sequence of 646 bases for the 0.7kb Buckwheat *Dra*I lib. genome walking product 1 *BwDra*I lib.GWP1 is presented in Fig. 5.6 BLAST analysis of the sequence revealed a maximum of 90% homology with cDNA clone for buckwheat legumin-like protein (acc. No. D87980). Southern hybridization of *BwDra*I lib.GWP2, *BwEco*RV lib.GWP3 and *BwEco*RV lib.GWP4 with [α-^{32}P]-dATP labelled 0.57 kb *BwDra*I lib.GWP2 generated positive hybridization signals on the X-ray film against the genome walking products *BwDra*I lib.GWP2, *BwEco*RV lib.GWP3 and *BwEco*RV lib.GWP4 (Fig. 5.7). The nucleotide sequences of the amplicons were determined by automated sequencing service offered by M/s Bangalore Genei India. The nucleotide sequence of 1127 bp for the 1.1 kb genome walking amplification product *BwEco*RV lib.GWP3 is presented in Fig. 5.8. BLASTn analysis of the 1127 bp sequence clearly identified it with the legumin gene family. Using an alignment that permitted maximum homology, the sequence showed >80% sequence homology with cDNA clones AY256960 and D87980 for buckwheat legumin-like protein. The sequence also showed high degree of similarity with legumin genes of *Pisum*, *Vicia faba*, soybean, *Pistacia, Carya, raphanus*, *Gingko biloba*, *Populus* and brazil nut. The sequence has been deposited in the genebank data base with accession no. GQ358524. CLUSTAL W alignment of the nucleotide sequence of 1.1 kb amplicon *BwEco*RV lib.GWP3 (acc. no. GQ358524) with the nucleotide sequences of some of the legumin genes available in the database is presented in Fig 5.9. Using an alignment that permitted maximum

Fig. 5.5: Electrophoresis profile of amplification products of nested PCR with products of primary PCR with *Dra*I adapter library and primer pair AP1-SS6R as the template and oligonucleotide primer pair AP2-SS10R along with corresponding autoradiograph showing the hybridization of the amplification product of the nested PCR with [α-^{32}P]-dATP labelled buckwheat legumin gene specific probe (DQ200889).

Fig. 5.6: Partial nucleotide sequence of 646 bases for the amplicon generated by nested PCR with products of primary PCR with *Dra*I adapter library and primer pair AP1- SS6R as the template and oligonucleotide primer pair AP2-SS10R.

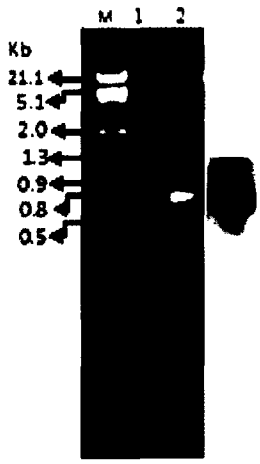


Fig. 5.5

```
1   ACCCAATTTT CTCTTTCTCT TCCAACCAGG TATCTCCACT CATGCAGATT
51  TGGGGAAAGC GAAGACACCA ATCCCGGGAG AGGGAAGATG AAGCCCTCTC
101 GGAGGCATAC ATCTCATGGA GTGGCGCCGG GTCTTCAAAG AATTTCAAAA
151 AGAGGAGCGG GGATCCTCAG CAAGCGCGGG GGGACACCGC CACGGGGATT
201 CACGTCCCGC TAGAGCCCCA AAAACATATC CCGAGTGCGG GTACAGAACT
251 TCCTCCGGAC AATGGTCCTG CGATAGCCGT AGTGCAGGCT CATCTTTCAT
301 TTCGAGATTT GTCTCTTCCT ATCGGCTCGG AGTTTACAGT CCCCAGCGCG
351 CGCGAAGATC CATCTCGGTC CGAGTCTCGC ATTACCCCGA AAATCGCCGA
401 CTGTTCACCA ATCTGAGAAT TCTTCCAGTT TAGTGTCCAG GGGAGTCGTC
451 CTAAGCAAGG TAATATAAAC CCCAGTTGAA GACCTAAAAT CGACAAATAT
501 GGAAAGTGAA GCTTTATTAA CTTCAGATTG ACGATAATTC GTGAATTTAT
551 GCAGAATGCG ATCCTCGGAC CGAGATGGAA CTGAACGCGC ACAGCGCACT
601 GTACGTGACG AGAGGAGAAG GAAGAGTCCA AGTTTCTAAA AAATAT
```

Fig. 5.6

Fig. 5.7: Electrophoresis profile of amplification products of nested PCR with products of primary PCR with L2: *Eco*RV adapter library and primer pair AP1- SS11R as the template and oligonucleotide primer pair AP2-SS12R, L3: *Eco*RV adapter library and primer pair AP1- SS13R as the template and oligonucleotide primer pair AP2-SS13R, L4: *Dra*I adapter library and primer pair AP1- SS13R as the template and oligonucleotide primer pair AP2-SS13R, L1: control, M: *Eco*RI/*Hind*III digested λ DNA and corresponding autoradiograph showing the hybridization of the amplification products of the nested PCRs with [α-^{32}P]-dATP labelled amplicon generated by nested PCR with *Dra*I adapter library and primer pair AP1- SS13R as the template and oligonucleotide primer pair AP2-SS13R.

Fig. 5.8: Nucleotide sequence of 1127 bases for the amplicon generated by nested PCR with *Eco*RV adapter library and primer pair AP1- SS11R as the template and oligonucleotide primer pair AP2-SS12R

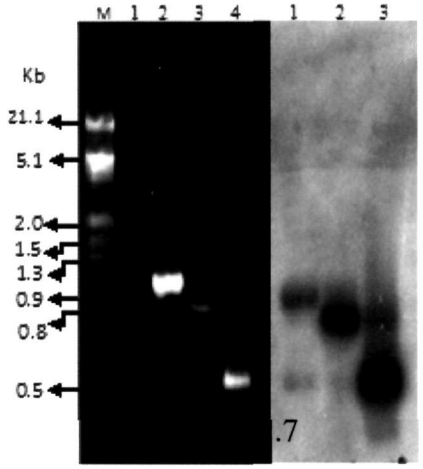


Fig. 5.7

```
1     TCTCCTGTTG TTCTGCGACG TCATCCCATC TCCCGCCGGT GTCGTGCAGT
51    GGACTCACAA CGACGGTGAC AACGATCTCA TCAGTATCAC TCTTTACGAT
101   GCCAACAGCT TCCAGAACCA GCTCGATGAG AACGTTAGGG TCAGTAATTA
151   AAACACTTGA TCAATCCCAT TTTCAATTCA AAACCATTTT ATAACTTACA
201   TAACCTGGAA TTTTGAAAAA AAACTCCTTC CTACCGGGCC AGAGCAGGCA
251   GAGCAGGGAG GACCGCCGCA GCCAGCGACG GACTAGGGAG GATGGCATGG
301   ACCGCCATTC CCTTGAGAGC GATGACGACA ACCACTCCTC GAAGCAAACA
350   TCTTGATTGG ATTCAAGGAC GAGATCCTCC ACCACCTCTT CGGAATTGGT
401   GACCAGGAAA CCTCCGCAGG CTCAAAGGGG AAACGACCGG AGGGGATTCT
451   CGCCCAGGCC CGGGAACCCA ACCCGGGGTC CCAAAGAATT TGAAAAAAAC
501   CCCCGAGGAA AAGAGGTGAC AGGAAAAGAG GTGGAAGCGG GAGGAGCAAT
551   GGATTGGAGC AAGCGTTCTG CAACCTGAAA TTCAAGCAAA ATGTTAACAG
601   GCCTTCTCGC GCCGAAGTCT TCAACCCACG CGCTGGTCGT ATCAACACCG
651   TTAACAGCAA CAATCTCCCA ATCCTTGAAT TCATCCAACT TAGCGCCCAG
701   CACGTCGTCA TATACAAGGT AATATAAACC CCAGGGAAGC CATAAAAGCG
751   ACAAATCTGA AATGAAGCTT GATTAACTTC AGATTGACGA TAATTCGTTG
801   AATTTATGCA GAATGCGATC CTCGGACCGA GATGGAACTT GAACGCGCAC
851   AGCGCACTGT ACGTGACGAG AGGAGAAGGA AGAGTCCAGG TTGTGGGAGA
901   TGAAGGAAGA AGTGTGTTTG ACGACAACGT GCAGAGAGGA CAGATCCTTG
951   TGGTCCCACA GGGATTCGCA GTGGTGTTGA AGGCAGGAAG AGAAGGACTG
1001  GAGTGGGTGG AGTTGAAGAA CGACGACAAC GCCATAACCA GTCCGATTGC
1051  CGGGAAGACT TCGGTGTTGA GGGCGATCCC TGTGGAGGTT CTGCCAACTC
1101  CTACGAACCA GCCCGGCCGC GCGGGCC
```

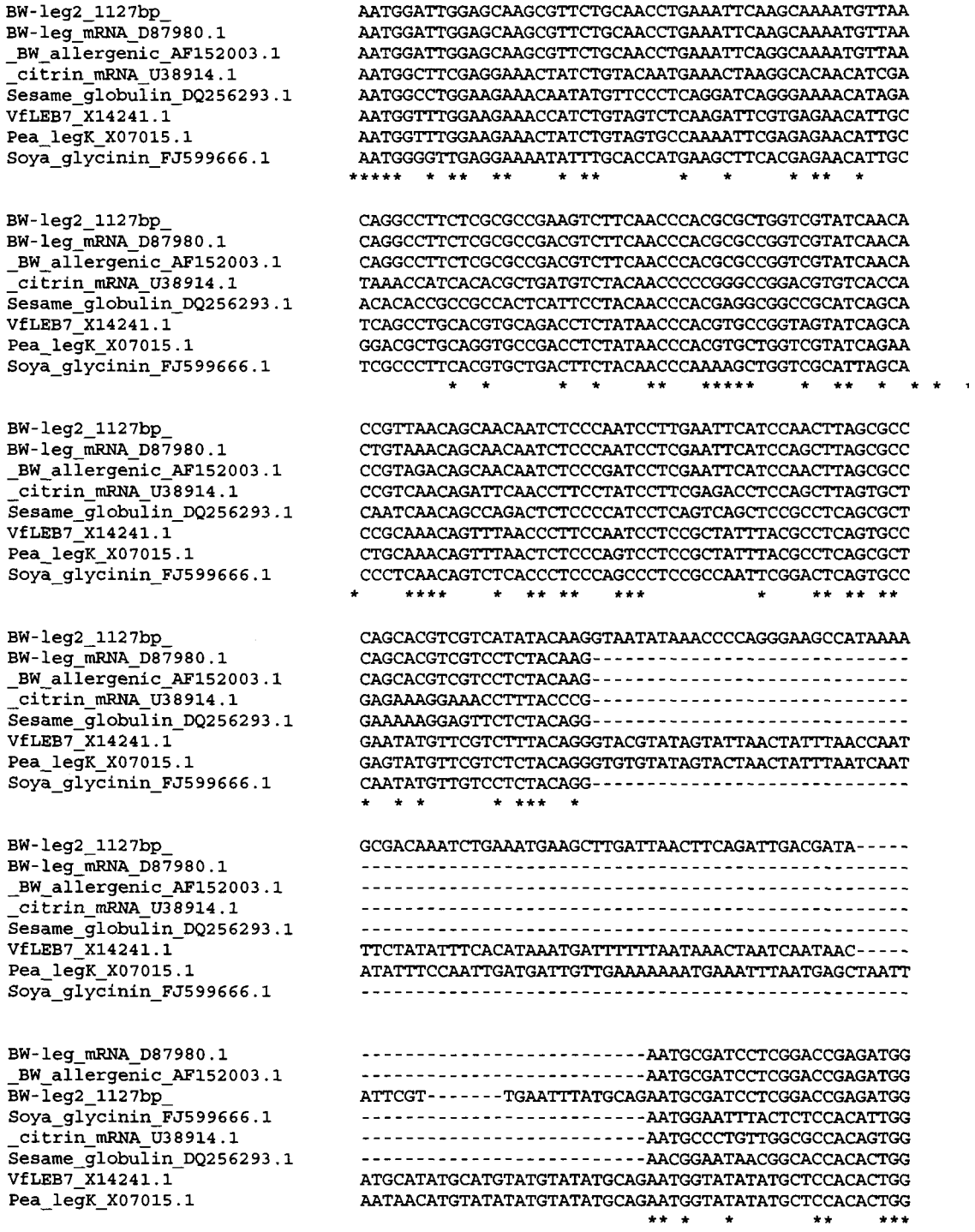
Fig. 5.8

Fig. 5.9 Clustal W (1.81) multiple alignment of 1127 bp nucleotide sequence of Bw*Eco*RV lib.GWP3 with nucleotide sequences of some SSP genes of other plants. '*'mark represents conserved residues, invariant/similar residues are presented as ':'

CLUSTAL W (1.81) multiple sequence alignment

```
BW-leg2_1127bp_                 ----------------TCTCCTGTTGTTCTGCGACGTCATCCCAT-CTCC
BW-leg_mRNA_D87980.1              GCACCAGGCAGAGTGAGAGCGAAGAATTCAGCCGTGGAGACCAACGCACC
BW_allergenic_AF152003.1        --------------------AAGAATCCAGCCGTGGAGACCAACGCTCC
citrin_mRNA_U38914.1            --------------------CCGGATGTGCTGAGACTTTCCAAGATTCACA
Sesame_globulin_DQ256293.1      ----------------GCGCGGAGACTTATGAATCGGAATCAGGCGTCGG
VfLEB7_X14241.1                 ----------------------------------GGAATTCCATATTGG
Pea_legK_X07015.1               ----------------------------------GAATTCCTTATTGG
Soya_glycinin_FJ599666.1        --------------------CCTATAACACTGGCGATGAACCAGTTGTTGC


BW-leg2_1127bp_                 CGCCGGTGTC-GTGC-AGTGGACTCA----CAACGACGGTGACAACGATC
BW-leg_mRNA_D87980.1            AGGCAGAGTGAGAGC-GAAGAATTCAGCCGTGGAGACCAGCACCAGAAGA
_BW_allergenic_AF152003.1       AGGCAGAGTGAGAGC-GAAGAATTCAGCCGTGGAGACCAGCACCAGAAGA
_citrin_mRNA_U38914.1           AGCAAGCAGTCGTTC-CAGGGCAGTAAATCCCAAGAACAACACCAAAAGG
Sesame_globulin_DQ256293.1      TTCTACTGGAGAAGAAGAAGGGCGCCAGAGAACAGATCGCCACCAGAAGC
VfLEB7_X14241.1                 ACATATAACAATGGC-GATGAACCGCTTGTTGCCATTAGTCTTCTTGACA
Pea_legK_X07015.1               ACATATAACCATGGC-GATGAACCTCTTGTTGCCATTAGCCTTCTTGACA
Soya_glycinin_FJ599666.1        CATCAGTCTT--CTT-GACACCTCCAACTTCAACAATCAGCTTGATCAAA
                                                                                *


BW-leg2_1127bp_                 TCATCAGTATCACTCTTTACGATGCCAACAGCTTCCAGAACCAGCTCGAT
BW-leg_mRNA_D87980.1            TTTTCAGGATCAGAGACGGTGACGTCATCCCATCTCCCGCCGGTGTCGTG
_BW_allergenic_AF152003.1       TTTTCAGGATCAGAGACGGCGACGTCATCCCATCTCCCGCCGGTGTCGTG
_citrin_mRNA_U38914.1           TCAGACAACTACGTGAGGGTGATGTCATTGCATTGCCTGCTGGAGCAGCT
Sesame_globulin_DQ256293.1      TACGGCGGTTCCGTCGTGGTGATGTCCTCGCATTGAGGGAGGGTGTCACC
VfLEB7_X14241.1                 CTTCCAACATTGCAAAC--CAGCTTGATTCAACCCCAAGAGTAAGTAATA
Pea_legK_X07015.1               CTTCCAACATTGCAAAC--CAGCTCGATTCAACCCCAAGAGTAAGTGATA
Soya_glycinin_FJ599666.1        ACCCCAGAGTATTTTACCTTGCTGGGAACCCAGATATAGAGCACCCAGAG
                                       *


BW-leg2_1127bp_                 -GAGAACGTTAGGGTCAGTAATTAAAAC-ACTTGATCAATCCCA---TTT
BW-leg_mRNA_D87980.1            -CAGTGGACTCACAATGACGGTGACAACGATCTCATCAGTATTACTCTTT
_BW_allergenic_AF152003.1       -CAGTGGACCCACAACAACGGTGACAACGATCTCATCAGTATCACTCTTT
_citrin_mRNA_U38914.1           -CACTGGATTTACAACAATGGCCGGGACCAGCTTGTTTTGGTCGCCCTCG
Sesame_globulin_DQ256293.1      -CACTGGGCTTATAACGACGGAGACACCCCGATAATCAGCGTCTCGATCC
VfLEB7_X14241.1                 ---GTCTGTTATAAATTGTT-CTAGAACTTTAAACTCAATCAAATCGCAT
Pea_legK_X07015.1               ---GTGTATCCATTCATACAGTATGCTCTTTCGATTATAACTTAAAAGTT
Soya_glycinin_FJ599666.1        ACCATGCAACAACAGCAGCAGCAGAAGAGTCATGGTGGACGCAAGCAGGG
                                                             *


BW-leg2_1127bp_                 TC-AATTCAAAACCATTTTATAACTTACA--TAACCT----GGAATTTTG
BW-leg_mRNA_D87980.1            AC-GATGCCAACAGCTTCCAGAACCAGCT--CGATG-----GGAACGTT-
_BW_allergenic_AF152003.1       AC-GATGCCAACAGCTTCCAGAACCAGCT--CGATG-----AGAACGTT-
_citrin_mRNA_U38914.1           TTGAACGTTGGCAATTCTCAAAACCAGCT--TGATC-----AGTACTTC-
Sesame_globulin_DQ256293.1      GC-GACGTTGCAAACGAGGCCAACCAGCT--CGATTT-----GAAATTC-
VfLEB7_X14241.1                 GTAAATATGTGTTTTTCAGGTATTTTACCT-TGGTG-----GGAACCCGG
Pea_legK_X07015.1               TCTAATGTAAATATGTGTATGCAGGTATTT-TACCTTGGTGGAAACCCAG
Soya_glycinin_FJ599666.1        GC-AACACCAGCAGCCGGAGGAAGAAGGTGGCAGTGTGCTCAGTGGCTTC
                                   *


BW-leg2_1127bp_                  AAAAAAAACTCCTT--------CCTACCGGGCCAGA-GCAGGCAGAGCAG
BW-leg_mRNA_D87980.1             ---AGGAACTTCTT--------CCTAGCTGGTCAGA-GCAAGCAGAGCAG
_BW_allergenic_AF152003.1        ---AGGAACTTCTT--------CCTAGCTGGTCAGA-GCAAGCAGAGCAG
_citrin_mRNA_U38914.1            ---AGGAAATTCTA--------CCTCGGTGGCAACCCACAACCACAACTC
Sesame_globulin_DQ256293.1       ---AGAAAATTTTT--------CCTGGCTGGAAACCCTCAAACAGCGCAA
VfLEB7_X14241.1                  AGGTAGAGTTCCCTGA---AACACAGGAGGAGCAGCAAGAAAGACATCAA
Pea_legK_X07015.1                AAACAGAGTTCCCCGA---AACACAGGAGGAACAACAAGGAAGGCATCGG
Soya_glycinin_FJ599666.1         AGCAAACATTTCTTAGCACAATCCTTCAACACCAAC-GAGGACACAGCTG
                                        *            *        *            *
```

```
BW-leg2_1127bp_                 GGAGGACCGCCGCAGCCAG-CGACGGACTAGGGAGGATGGCATGGACCGC
BW-leg_mRNA_D87980.1            GGAGGACCGCCGCAGCCAG-CGACAGACTAGGGAGGAAGGCAGTGACCGC
_BW_allergenic_AF152003.1       GGAGGACCGCCGCAGCCAG-CGACAGACTAGGGAGGAAGGCAGTGACCGC
_citrin_mRNA_U38914.1           CAAG----GTTTCAGTCAA----------AGTCAAGGTGGCAGAAGTCAG
Sesame_globulin_DQ256293.1      -------TTCCAAGGCCAA-CAGGAGAGAGAACAACAACCCAGAGGCGAG
VfLEB7_X14241.1                 CAAAAACATAGTTTACCAG-TTGGACGTCGGGGTGGACAGCACCAACAAG
Pea_legK_X07015.1               CAAAAGCATAGTTACCCTG-TTGGACGTAGGAGTGGACATCACCAACAAG
Soya_glycinin_FJ599666.1        AGAA-ACTTCGGTCTCCAGATGACGAAAGGAAGCAGATCGTGACAGTGGA
                                                *

BW-leg2_1127bp_                 CATTCCCTTGAGAGCGATGACGAC-AACCACTCCTCGAAGCAAACATCTT
BW-leg_mRNA_D87980.1            CAATCCCGTGAGAGCGATGACGACGAAGCACTTCTCGAAGCAAACATCTT
_BW_allergenic_AF152003.1       CAATCCCGTGAGAGCCAAGACGACGAAGCACTTCTCGAAGCAAACATCTT
_citrin_mRNA_U38914.1           GGAAGCCAAGGCAGCGACGACGGGAGAG--------GTGGCAATC-TCTT
Sesame_globulin_DQ256293.1      GGCAGGAGAGGTCAAGAAGAAGGGCAAGGA------ACAAGCAACATATT
VfLEB7_X14241.1                 AAGAGGAATCTGAAGAACAAAAGGACGGTA----------ACAGCGTTTT
Pea_legK_X07015.1               AAGAGGAATCCGAAGAACAAAACGAAGGTA----------ACAGCGTGCT
Soya_glycinin_FJ599666.1        GGGAGGCCTCAGCGTTATCAGCCCCAAGTGG--CAAGAACAAGAAGACGA
                                          *  *

BW-leg2_1127bp_                 GATTGGATTCAAGGACGA-GATCCTCCACCACCTCTTCGGAATTGGTGAC
BW-leg_mRNA_D87980.1            GACTGGATTCCAGGACGA-GATCCTCCAAGAAATCTTCCGAAATGTTGAC
_BW_allergenic_AF152003.1       GAGTGGATTCGAGGACGA-GATCCTCCAAGAAATCTTCCGAAATGTTGAC
_citrin_mRNA_U38914.1           TAGAGGCTTTGACGA--GCGGTTGTTGGCTGAGGCCTTC--AACGTGAAC
Sesame_globulin_DQ256293.1      CAATGGATTTAACGAGGA--ATTCTTGGCCGAGAGCTTC--AACACTGAT
VfLEB7_X14241.1                 GAGTGGCTTCAGTTCAGA--GTTTTTAGCACAGACATTC--AACACCGAA
Pea_legK_X07015.1               GAGTGGCGTCAGCTCAGA--GTTTTTAGCACAAACGTTC--AACACTGAA
Soya_glycinin_FJ599666.1        AGACGAAGATGAAGACGAAGAATATGAACAAACTCCCTCTTATCCTCCAC
                                   *                     *              *      *

BW-leg2_1127bp_                   CAGGA--AACC-TCCGCAGGCT-CAAAGGGGA-AACGACCGGAGGGGATT
BW-leg_mRNA_D87980.1              CAGGA--GACCATCAGCAAGCT-CAGAGGTGACAACGACCAGAGAGGATT
_BW_allergenic_AF152003.1         CAGGA--GACCATCAGCAAGCT-CAGAGGTGAGAACGACCAGAGAGGATT
_citrin_mRNA_U38914.1             CCAGA--TCTAATCAGGAGACTGCAGAGGCCACAGATACAGAGGGGCATT
Sesame_globulin_DQ256293.1        CCCCA--ACTAATAAGGAAATTGCAGTCAAGGGAGGACAACAGGGGCATC
VfLEB7_X14241.1                   GAGGA--TACAGCTAAGAGACTTCGTTCTCCACGCGACAAAAGGAATCAA
Pea_legK_X07015.1                 GAGGA--TACAGCGAAGAGACTTCGATCTCCACGAGACGAAAGGAGTCAA
Soya_glycinin_FJ599666.1          GACGACCAAGCCATGGAAAGCA-TGAAGATGACGAGGACGAGGACGAAGA
                                     *             *

BW-leg2_1127bp_                   C-TCGCCCAGGCCCGG--GAACCCAA-CCCGGGGTCCCAAAGAATT-TGA
BW-leg_mRNA_D87980.1              CATCGTCCAGGCTCGG--GACCTCAAACTCCGGGTCCCAGAGGAGTATGA
_BW_allergenic_AF152003.1         CATCGTCCAGGCTCGG--GACCTCAAACTCCGGGTCCCAGAGGAGTATGA
_citrin_mRNA_U38914.1             ATCATCAGAGTCGAGGAAGAGCTGAGAGTACTGTCTCCTCAAA-----GA
Sesame_globulin_DQ256293.1        ATCGTCCGAGCCGAAAGGCCGCTCAGATTGGTCTTGCCTGAATACG--GA
VfLEB7_X14241.1                   ATCGTGAGAGTTGAGGGCGGTCTCCGCATTATCAACCCTGAGGGGCAGCA
Pea_legK_X07015.1                 ATTGTGCGAGTTGAGGGAGGTCTCCGCATTATCAACCCCAAGGGGAAGGA
Soya_glycinin_FJ599666.1          AGGAGATCAACCTCG-----TCCTGATCACC---CTCCACAG-----CGA
                                     *             *             **  *        *
BW-leg_mRNA_D87980.1              AGAAGAACTCCAGAGGGA-----------------------------AA
_BW_allergenic_AF152003.1         AGAAGAACTCCAGAGGGA-----------------------------AA
BW-leg2_1127bp_                   AAAA-AACCCCCGAGGAA-----------------------------AA
_citrin_mRNA_U38914.1             GACAGAGAACAAGAACAG-----------------------------GA
Sesame_globulin_DQ256293.1        CGAGAAGAGCAGCAGCGA-----------------------------CA
VfLEB7_X14241.1                   GGAAGAAGAAGAACAAGA-----------------------------GG
Pea_legK_X07015.1                 AGAAGAAGAAGAAAAAGAACAGAGTCATTCTCACTCTCACAGAGAGGAGG
Soya_glycinin_FJ599666.1          CCAAGCAGGCCCGAACAA-----------------------------CA

BW-leg2_1127bp_                   GAGGTGACAGGAAAAGAGGTGGAAGC---------------GGGAGGAGC
BW-leg_mRNA_D87980.1              GAGGTGACAGGAAAAGAGGTGGAAGC---------------GGGAGGAGC
_BW_allergenic_AF152003.1         GAGGTGACAGGAAAAGAGGTGGAAGC---------------GGGAGGAGC
_citrin_mRNA_U38914.1             AGAATGCGAAGAAACTCCGTCATAT----------------GAAAGGGAC
Sesame_globulin_DQ256293.1        GCGGGAGCAAGGAAGAGGAGGAGGA------------------TACATG
VfLEB7_X14241.1                   AGGAGGAGAAGCAGAGAAGTGAGCA------------------GGGAAGG
Pea_legK_X07015.1                 AGGAAGAAGAAGAAGAAGATGAGGAGAAACAAAGAAGTGAGGAAAGAAAG
Soya_glycinin_FJ599666.1          AGAACCACGTGGAAGAGGATGTCAG------------------ACTAGA
                                        *
```


BW-leg2_1127bp_
BW-leg_mRNA_D87980.1
_BW_allergenic_AF152003.1
_citrin_mRNA_U38914.1
Sesame_globulin_DQ256293.1
VfLEB7_X14241.1
Pea_legK_X07015.1
Soya_glycinin_FJ599666.1
AATGGATTGGAGCAAGCGTTCTGCAACCTGAAATTCAAGCAAAATGTTAA
AATGGATTGGAGCAAGCGTTCTGCAACCTGAAATTCAAGCAAAATGTTAA
AATGGATTGGAGCAAGCGTTCTGCAACCTGAAATTCAGGCAAAATGTTAA
AATGGCTTCGAGGAAACTATCTGTACAATGAAACTAAGGCACAACATCGA
AATGGCCTGGAAGAAACAATATGTTCCCTCAGGATCAGGGAAAACATAGA
AATGGTTTGGAAGAAACCATCTGTAGTCTCAAGATTCGTGAGAACATTGC
AATGGTTTGGAAGAAACTATCTGTAGTGCCAAAATTCGAGAGAACATTGC
AATGGGGTTGAGGAAAATATTTGCACCATGAAGCTTCACGAGAACATTGC
CAGGCCTTCTCGCGCCGAAGTCTTCAACCCACGCGCTGGTCGTATCAACA
CAGGCCTTCTCGCGCCGACGTCTTCAACCCACGCGCCGGTCGTATCAACA
CAGGCCTTCTCGCGCCGACGTCTTCAACCCACGCGCCGGTCGTATCAACA
TAAACCATCACACGCTGATGTCTACAACCCCCGGGCCGGACGTGTCACCA
ACACACCGCCGCCACTCATTCCTACAACCCACGAGGCGGCCGCATCAGCA
TCAGCCTGCACGTGCAGACCTCTATAACCCACGTGCCGGTAGTATCAGCA
GGACGCTGCAGGTGCCGACCTCTATAACCCACGTGCTGGTCGTATCAGAA
TCGCCCTTCACGTGCTGACTTCTACAACCCAAAAGCTGGTCGCATTAGCA
CCGTTAACAGCAACAATCTCCCAATCCTTGAATTCATCCAACTTAGCGCC
CTGTAAACAGCAACAATCTCCCAATCCTCGAATTCATCCAGCTTAGCGCC
CCGTAGACAGCAACAATCTCCCGATCCTCGAATTCATCCAACTTAGCGCC
CCGTCAACAGATTCAACCTTCCTATCCTTCGAGACCTCCAGCTTAGTGCT
CAATCAACAGCCAGACTCTCCCCATCCTCAGTCAGCTCCGCCTCAGCGCT
CCGCAAACAGTTTAACCCTTCCAATCCTCCGCTATTTACGCCTCAGTGCC
CTGCAAACAGTTTAACTCTCCCAGTCCTCCGCTATTTACGCCTCAGCGCT
CCCTCAACAGTCTCACCCTCCCAGCCCTCCGCCAATTCGGACTCAGTGCC
CAGCACGTCGTCATATACAAGGTAATATAAACCCCAGGGAAGCCATAAAA
CAGCACGTCGTCCTCTACAAG----------
CAGCACGTCGTCCTCTACAAG----------
GAGAAAGGAAACCTTTACCCG----------
GAAAAAGGAGTTCTCTACAGG----------
GAATATGTTCGTCTTTACAGGGTACGTATAGTATTAACTATTTAACCAAT
GAGTATGTTCGTCTCTACAGGGTGTGTATAGTACTAACTATTTAATCAAT
CAATATGTTGTCCTCTACAGG----------
GCGACAAATCTGAAATGAAGCTTGATTAACTTCAGATTGACGATA-----

TTCTATATTTCACATAAATGATTTTTTAATAAACTAATCAATAAC-----
ATATTTCCAATTGATGATTGTTGAAAAAAATGAAATTTAATGAGCTAATT

----------AATGCGATCCTCGGACCGAGATGG
----------AATGCGATCCTCGGACCGAGATGG
ATTCGT-------TGAATTTATGCAGAATGCGATCCTCGGACCGAGATGG
----------AATGGAATTTACTCTCCACATTGG
----------AATGCCCTGTTGGCGCCACAGTGG
----------AACGGAATAACGGCACCACACTGG
ATGCATATGCATGTATGTATATGCAGAATGGTATATATGCTCCACACTGG
AATAACATGTATATATGTATATGCAGAATGGTATATATGCTCCACACTGG

```
BW-leg2_1127bp_                 AACTTGAACGCGCACAGCGCACTGTACGTGACGAGAGGAGAAGGAAGAGT
BW-leg_mRNA_D87980.1            AACTTGAACGCGCACAGCGCACTGTACGTGACGAGAGGAGAAGGAAGAGT
_BW_allergenic_AF152003.1       AACTTGAACGCGCACAGCGCACTGTACGTGACGAGAGGAGAAGGAAGAGT
_citrin_mRNA_U38914.1           AACTTGAATGCCCACAGCATAGTCTACGTAACAAGGGGCAACGGCAGGAT
Sesame_globulin_DQ256293.1      AGCACGAACTCCCACTCCGCATTGTACGTGACGAGAGGAAGCGCAAGAAT
VfLEB7_X14241.1                 AACATAAACGCAAACAGTCTGCTGTACGTGATAAGAGGAGAAGGAAGAGT
Pea_legK_X07015.1               AACATAAACGCCAACAGTCTGCTGTACGTGATTAGAGGAGAAGGAAGAGT
Soya_glycinin_FJ599666.1        AACTTGAACGCGAACAGTGTGATCTATGTGACTCGAGGGAAAGGAAGAGT
                                * *   **  *  **       * ** ** *   * **    *  **  *

BW-leg2_1127bp_                 CCAGGTTGTGGGAGATGAAGGAAGAAGTGTGTTTGACGACAACGTGCAGA
BW-leg_mRNA_D87980.1            CCAAGTTGTGGGAGATGAAGGAAGAAGTGTGTTCGACGACAACGTGCAGC
_BW_allergenic_AF152003.1       CCAGGTTGTCGGAGATGAAGGAAGAAGTGTGTTCGACGACAACGTGCAGC
_citrin_mRNA_U38914.1           GCAAATTGTAGCGGAGAACGGGGAGAATGTGTTCGACGGTCAAATCCGGG
Sesame_globulin_DQ256293.1      CCAGGTGGTGGGGCACAAGGGAAGATCGGTGTTGAACGAGGAGGTGAACG
VfLEB7_X14241.1                 TAGGATTGTGAACTCCCAAGGAAACGCAGTGTTCGACAACAAGGTGAGAA
Pea_legK_X07015.1               TAGGATTGTGAACTTCCAAGGAGACGCAGTGTTCGACAACAAGGTCAGAA
Soya_glycinin_FJ599666.1        TAGAGTGGTGAACTGCCAAGGGAATGCAGTGTTCGACGGTGAGCTAAGGA
                                     * **        * **       *****  **    *  *

BW-leg2_1127bp_                 GAGGACAGATCCTTGTGGTCCCACAGGGATTCGCAGTGGTGTTGAAGGCA
BW-leg_mRNA_D87980.1            GAGGACAGATCCTTGTAGTCCCACAGGGATTCGCAGTGGTGTTGAAGGCA
_BW_allergenic_AF152003.1       GAGGACAGATCCTTGTGGTCCCACAGGGATTCGCAGTGGTGTTGAAGGCA
Soya_glycinin_FJ599666.1        GGGGACAATTGCTAGTGGTGCCGCAGAACTTTGTGGTGGCTGAGCAAGGG
_citrin_mRNA_U38914.1           AGGGTCAGCTGATCGTTGTTCCGCAGGGCTTCGCCGTCGTGAAGAGGGCA
Sesame_globulin_DQ256293.1      AGGGGCAGCTTGTGGTGGTACCACAGAACTTTGCGCTGGCAATAAGAGCC
VfLEB7_X14241.1                 AGGGACAGTTGGTGGTGGTACCACAAAACTTTGTGGTGGCGGAACAAGCT
Pea_legK_X07015.1               AGGGACAGTTGGTGGTGGTACCACAAAACTTTGTGGTGGCGGAACAAGCT
                                  ** **  *  * ** ** ** **    ** *   * *        *

BW-leg2_1127bp_                 GGAAG---AGAAGGACTGGAGTGGGTGGAGTTGAAGAACGACGACAACGC
BW-leg_mRNA_D87980.1            GGAAG---AGAAGGACTGGAGTGGGTGGAGTTGAAGAACGACGACAACGC
_BW_allergenic_AF152003.1       GGAAG---AGAAGGACTGGAGTGGGTGGAGTTGAAGAACGACGACAACGC
_citrin_mRNA_U38914.1           GGTAA---CCGTGGACTGGAGTGGATATCATTCAAGACAAACGACGTCGC
Sesame_globulin_DQ256293.1      GGCGA---GCAGGGATTCGAGTATGTGACTTTCAGGACCAACGACAACGC
VfLEB7_X14241.1                 GGAGAGGAAGAAGGATTAGAGTATCTTGTGTTCAAGACAAATGACAGAGC
Pea_legK_X07015.1               GGGGAGGAAGAAGGATTAGAGTATGTGGTGTTCAAGACAAATGACAGAGC
Soya_glycinin_FJ599666.1        GGAGA---ACAAGGATTGGAATACGTAGTGTTCAAGACACACCACAACGC
                                **          *** * ** *   *    ** * **   *  **   **

BW-leg2_1127bp_                 CATAACCAGTCCGATTGCCGGGAAGACTTCGGTGTTGAGGGCGATCCCTG
BW-leg_mRNA_D87980.1            CATAACCAGTCCGATTGCCGGGAAGACTTCGGTGTTGAGGGCGATCCCTG
_BW_allergenic_AF152003.1       CATAACCAGCCCGATTGCCGGGAAGACTTCGGTGTTGAGGGCGATCCCTG
_citrin_mRNA_U38914.1           CATGACAAGCCAGCTGGCCGGAAGGGCTTCGGTGTTAAGAGGGCTTCCGT
Sesame_globulin_DQ256293.1      CATGAAAAGCGAGCTGGCAGGGCGGCTCTCCGCCATCCGGGCGATGCCGG
VfLEB7_X14241.1                 TGCTGTTAGCCACGT-------ACAACA--GGTGTTTAGGGCCACTCCTG
Pea_legK_X07015.1               TGCGGTTAGCCACGT-------ACAACA--GGTGCTTAGGGCCACTCCTG
Soya_glycinin_FJ599666.1        CGTGAGCAGCTACATT------AAGGAT---GTGTTTAGGGCAATCCCTT
                                      **     *                *   *  * *     **

BW-leg2_1127bp_                 TGGAGGTTCT-GCCAACTCCTACGAACCAGCCCGGCCGCGCGGGCC----
BW-leg_mRNA_D87980.1            TGGAGGTTCTTGCCAACTCATACGA--TATCTCGACGAAGGAAGCTTTCA
_BW_allergenic_AF152003.1       TGGAGGTTCTTGCCAACTCCTACGA--TATCTCGACGAAGGAAGCGTTCA
_citrin_mRNA_U38914.1           TGGACGTTATCCAGAACTCGTTCCA--AGTGTCGAGGGATGAAGCTCAGA
Sesame_globulin_DQ256293.1      ACGAAGTTGTGATGAACGCGTTTGG--TGTTTCCAGGGAGGACGCCAGGA
VfLEB7_X14241.1                 CAGATGTTCTTGCAAATGCTTTTGGTCTTCGTCAACGCC--AAGTCACCG
Pea_legK_X07015.1               CAGAGGTTCTTGCAAATGCTTTTGGTCTTCGTCAACGCC--AAGTCACGG
Soya_glycinin_FJ599666.1        CGGAGGTTCTTTCCAATTCTTACAACCTTGGCCAGAGTCAAGTGCGTC--
                                  ** *** *    **  * *           *          *
```

Fig. 5.9

homology, the nucleotide sequence of *BwEco*RV lib.GWP3 showed a 4 base "GCCG" deletion at position 57 and a 3 base "CTC" deletion at position 670. Further, the 1127 bp sequence showed an insertion of 5 bases "CTCCT" at position 324. GENSCAN 1.0 identified three exons and two introns in the sequence GQ358524. The Ist exon was detected as the initial exon and the other two as the internal exons of the gene. While the Ist intron detected in the sequence had a length of 82 bases and was positioned between bases 140 and 221, the 2nd intron had a length of 91 bases and was positioned between bases719 and 809. Multiple alignment of the sequences with CLUSTAL W also revealed the presence of two introns in the 1127 bp nucleotide sequence of *BwEco*RV lib.GWP3. The alignment also revealed the conserved positon of introns in seed storage protein genes. The position of the exons on the 1127 bp sequence *BwEco*RV lib.GWP3 is shown in Fig. 5.10. The software could not detect the 5' untranslated region or a putative polyadenylation site in the sequence.

The genome walking product *BwEco*RV lib.GWP3 was used as a probe to assess changes in the transcript levels corresponding to the gene in grains of common buckwheat at different stages of maturation. Hybridization of total RNA isolated from grains of common buckwheat at different stages of development with [α-^{32}P]-dATP labelled genome walking product buckwheat *BwEco*RV lib.GWP3 did not reveal any signals against RNA harvested at 10 days and 20 days after flowering (Fig. 5.11). However strong signals were detected against RNA harvested at 30 and 40 days after flowering; the intensity of signal being stronger against RNA isolated 40 days after flowering. (Fig. 5.11).

Fig. 5.10: (a) Diagramatic representation of the intron/exon architechture in the 1127 bp nucleotide sequence of the amplicon generated with buckwheat *Eco*RV adapter library and primer pair AP1- SS11R as the template and oligonucleotide primer pair AP2-SS12R

(b) Screen shot of the BLAST of the 1127 bp sequence of the amplicon Bw*Eco*RV lib.GWP3 generated with buckwheat *Eco*RV adapter library and primer pair AP1- SS11R as the template and oligonucleotide primer pair AP2-SS12R showing the conserved positions of exons and introns among different SSP gene sequences in buckwheat.

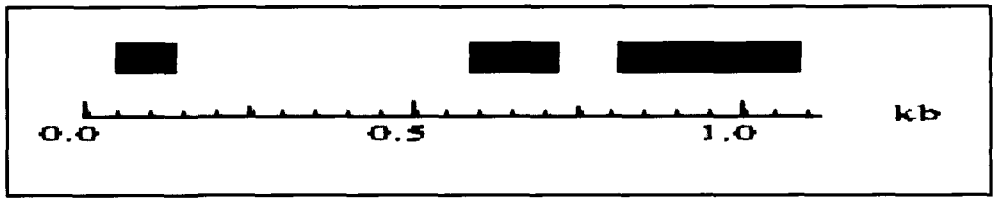


Fig. 5.10a

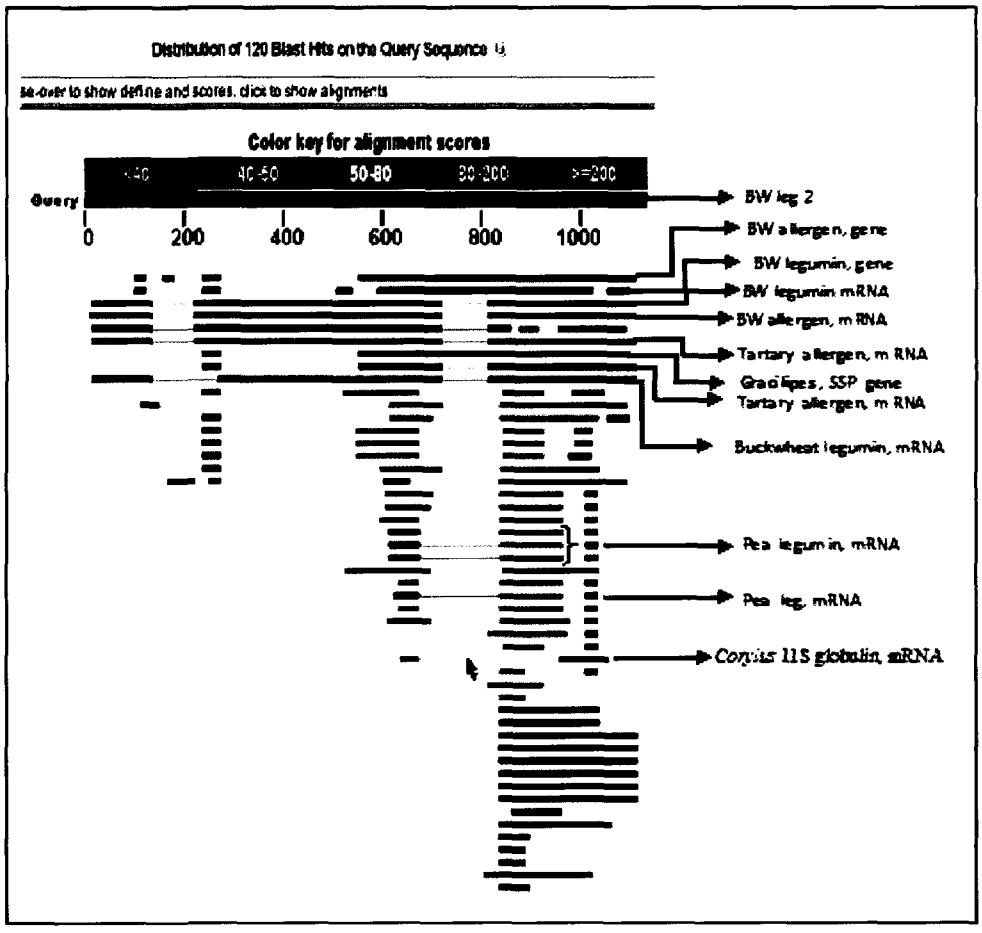


Fig: 5.10b

Fig. 5.11a : Electrophoresis profile of total RNA isolated from grains of common buckwheat harvested 10, 20, 30 and 40 days after flowering.

Fig. 5.11b : Dot blot of RNA isolated from grains of common buckwheat harvested 10, 20, 30 and 40 days after flowering with [α- 32P]-dATP- labelled 1127 bp buckwheat genome walking library product Bw*Eco*RV lib.GWP3

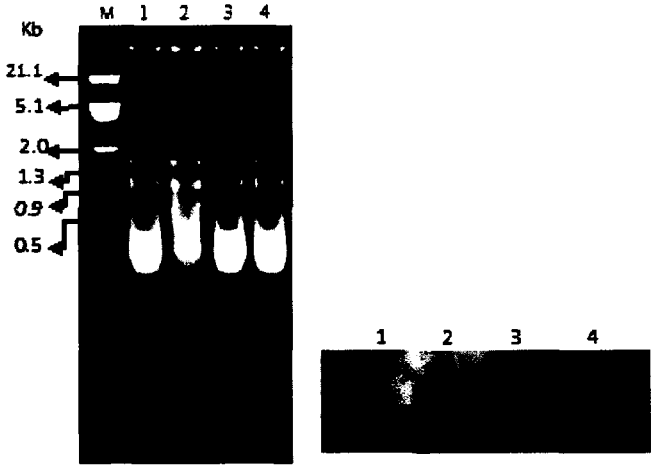


Fig. 5.11 (a) (b)

The deduced amino acid sequence of 167 residues for the 1127 bp genome walking product *BwEco*RV lib.GWP3 is presented in Fig. 5.12. BLASTp analysis of the deduced amino acid sequence clearly identified the sequence with the legumin subfamily of proteins. Using the alignment that permitted maximum homology, the deduced amino acid sequence showed 78% with buckwheat legumin-like protein (acc. no. BAA21758). The percentage of homology varied between 40 to 69% with deduced amino acid sequences of legumin genes from other crops including *Magnolia salicifolia* legumin precursor (acc. no. CAA57846) and *Coffea arabica* 11S globulin (acc. no. AAC61881). CLUSTAL W alignment of the deduced amino acid sequence with the amino acid sequences of SSPs from other plants is presented in Fig. 5.13. The phylogenetic tree constructed from the alignment data of the deduced amino acid sequence with amino acid sequences of SSPs available in EMBL database revealed clear grouping into monocots, dicots and gymnosperms (Fig. 5.14). The sequences from dicots clustered into one group along with the sequences of buckwheat SSPs.

The nucleotide sequences of 869 bp for the 0.9 kb genome walking product *BwEco*RV lib.GWP4 and of 438 bp for the 0.57 kb genome walking product *BwDra*I lib.GWP2 are presented in Figs. 5.15 and 5.16. Pairwise alignment of the 869 bp and 438 bp nucleotide sequences with CLUSTAL W clearly identified the 438 bp sequence as a subset of the 869 bp sequence. BLASTn analysis of the nucleotide sequence of genome walking product *BwEco*RV lib.GWP4 revealed that the 3' end of this sequence aligned with the 5' ends of the buckwheat legumin cDNA clones D87980, D87982 and AY152003. In order to exclude the adapter primer sequences

Fig. 5.12: Deduced amino acid sequence of the coding region of 1127 bp amplicon Bw*Eco*RV lib.GWP3 generated with buckwheat *Eco*RV adapter library and primer pair AP1- SS11R as the template and oligonucleotide primer pair AP2-SS12R

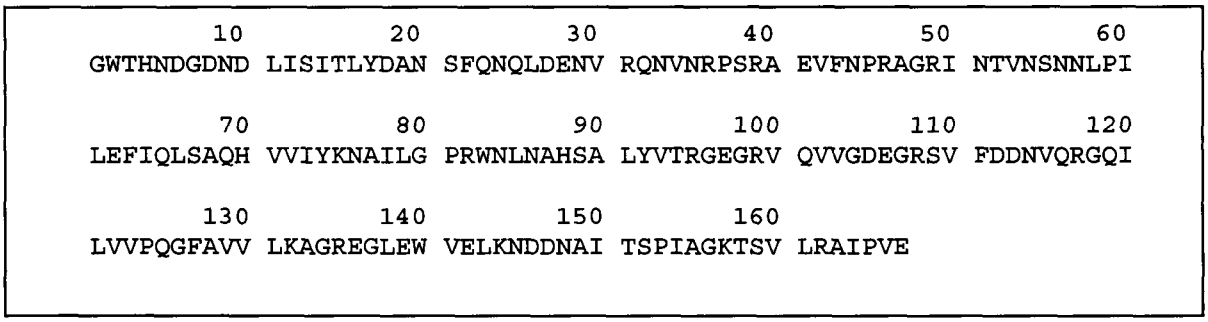

10 20 30 40 50 60
GWTHNDGDND LISITLYDAN SFQNQLDENV RQNVNRPSRA EVFNPRAGRI NTVNSNNLPI
70 80 90 100 110 120
LEFIQLSAQH VVIYKNAILG PRWNLNAHSA LYVTRGEGRV QVVGDEGRSV FDDNVQRGQI
130 140 150 160
LVVPQGFAVV LKAGREGLEW VELKNDDNAI TSPIAGKTSV LRAIPVE

Fig. 5.12

Fig. 5.13: Clustal W (1.81) multiple alignment of the deduced amino acid sequence of the 1127 bp amplicon Bw*Eco*RV lib.GWP3 with amino acid sequences of some SSPs of other plants. '*'mark represents conserved residues, invariant/similar residues are presented as ':'

CLUSTAL W (1.81) multiple sequence alignment

```
BW_leg2                          ---------------------------------------QNVNRPSRAEV 11
BW_globulin_O23878.1             RS----------------------NGLEQAFCNLKFKQNVNRPSRADV
BW_allergenic_Q9XFM4.1           RS----------------------NGLEQAFCNLKFRQNVNRPSRADV
Oat_globulin_CAA52764.1          QVGKSTPYQGGQSSQYQAGQSWDQSFNGLEENFCSLEARKNIENPQHADT
Zizania_glutelin_ABG47337.1      QYG------GGCS-----------NGLDETFCAMRIWQNIDNPNLADT
Magnolia_legumin_CAA57846.1      -------------------------NGLEEIQCSSKLTYNIADPTRADV
Coffea_globulin_AAC61881.1       GRG-------------------WRSNGLEETLCTVKLSENIGLPQEADV
Picea_legumin_CAA44874.1         SSG--------------------DENGVEELVCPLRVKHNADNPEDADV
Pine_globulin_CAA77569.1         RED--------------------SENGVEELVCPMRVKHNADNPEDADL
                                                                 *   *  *:

BW_leg2                          FNPRAGRINTVNSNNLPILEFIQLSAQHVVIYKNAILGPRWNLNAHSALY 61
BW_globulin_O23878.1             FNPRAGRINTVNSNNLPILEFIQLSAQHVVLYKNAILGPRWNLNAHSALY
BW_allergenic_Q9XFM4.1           FNPRAGRINTVDSNNLPILEFIQLSAQHVVLYKNAILGPRWNLNAHSALY
Oat_globulin_CAA52764.1          YNPRAGRITRLNSKNFPILNIVQMSATRVNLYQNAILSPFWNINAHSVIY
Zizania_glutelin_ABG47337.1      YNPRAGRVTNLNSQKFPILNLIQMSAVKVNLYQNALLSPFWNINSHSVVY
Magnolia_legumin_CAA57846.1      YNPQAGRITSLNSQKLPILNVLQLSAERGVLYRNALLAPQWNVNAHSLVY
Coffea_globulin_AAC61881.1       FNPRAGRITTVNSQKIPILSSLQLSAERGFLYSNAIFAPHWNINAHSALY
Picea_legumin_CAA44874.1         YVRDGGRLNRVNRFKLPVLKYLRLGAERVVLHPRASCVPSWRMNAHGIMY
Pine_globulin_CAA77569.1         YVRDGGRMNIVNRYKLPALKYLGLGAERVILRPRASFVPSWRMNAHAIMY
                                 :   .**:. ::  ::* *. : :.* :  :  .*   * *.:*:*. :*

BW_leg2                          VTRGEGRVQVVGDEGRSVFDDNVQRGQILVVPQGFAVVLKAGREGLEWVE 111
BW_globulin_O23878.1             VTRGEGRVQVVGDEGRSVFDDNVQRGQILVVPQGFAVVLKAGREGLEWVE
BW_allergenic_Q9XFM4.1           VTRGEGRVQVVGDEGRSVFDDNVQRGQILVVPQGFAVVLKAGREGLEWVE
Oat_globulin_CAA52764.1          MIQGHARVQVVNNNGQTVFNDILRRGQLLIVPQHFVVLKKAEREGCQYIS
Zizania_glutelin_ABG47337.1      VTQGCARVQVVNNNGKTVFNGELRRGQLLIIPQHYVVVKKAQREGCAYIA
Magnolia_legumin_CAA57846.1      ATRGNGRVQIVGEQGRPVFDGELREGQLVVVPQSFAVVKKAGNEGFEYVA
Coffea_globulin_AAC61881.1       VIRGNARIQVVDHKGNKVFDDEVKQGQLIIVPQYFAVIKKAGNEGFEYVA
Picea_legumin_CAA44874.1         VTRGEGRIEVVGDEGRSVFDGRVREGQFIVIPQFYAVIKQAGDEGFEWIT
Pine_globulin_CAA77569.1         VTRGEGRIEVVGDEGRSVFDGRVKEGQFIVIPQFYAVVKQAGEDGLEYIR
                                   :* .*:::*..:*. **:. ::.**::::** :.*: :*  :*  ::

BW_leg2                          LKNDDNAITSPIAGKTSVLRAIPVE------------------------ 136
BW_globulin_O23878.1             LKNDDNAITSPIAGKTSVLRAIPVEVLANSYDISTKEAFRLKNGR-QEVE
BW_allergenic_Q9XFM4.1           LKNDDNAITSPIAGKTSVLRAIPVEVLANSYDISTKEAFRLKNGR-QEVE
Oat_globulin_CAA52764.1          FKTNPNSMVSHIAGKSSILRALPIDVLANAYRISRQEARNLKNNRGEEFG
Zizania_glutelin_ABG47337.1      FKTNPNSMVSHIVGKSSIFRALPTDVLANAYRISREDAQRLKHNRGDELG
Magnolia_legumin_CAA57846.1      FKTNDNAMNSPLVGKTSVIRAMPEDVLINSYRISREEARRLKYNR-EEIA
Coffea_globulin_AAC61881.1       FKTNDNAMINPLVGRLSALRAIPEEVLRSSFQISSEEAEELKYGR-QEAL
Picea_legumin_CAA44874.1         FTTSDISFQSFLAGRQSVLKAMPEEVLSAAYRMDRTEVRQIMRNRERDTL
Pine_globulin_CAA77569.1         FTTSDNSYRSTLAGRQSVLKACRGSVACGLQNRPKRSP-SVMRNREHDTL
                                 :...  :  . :.*: * ::*   .
```

Fig. 5.13

Fig. 5.14: Phylogenetic tree based on alignment matrix of the deduced amino acid sequence of the 1127 bp amplicon Bw*Eco*RV lib.GWP3 with the amino acid sequences of legumin-type proteins from various angiosperms and gymnosperms.

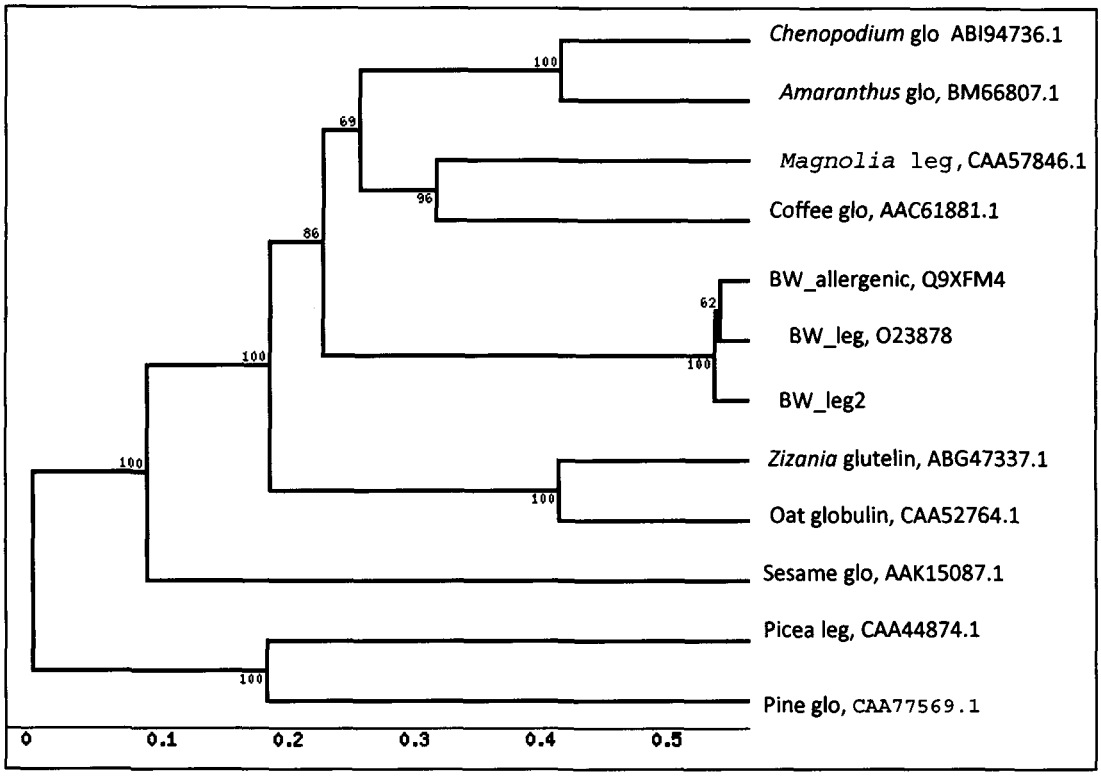

Chenopodium glo ABI94736.1
Amaranthus glo, BM66807.1
Magnolia leg, CAA57846.1
Coffee glo, AAC61881.1
BW_allergenic, Q9XFM4
BW_leg, O23878
BW_leg2
Zizania glutelin, ABG47337.1
Oat globulin, CAA52764.1
Sesame glo, AAK15087.1
Picea leg, CAA44874.1
Pine glo, CAA77569.1
100
69
96
86
62
100
100
100
100
100
0
0.1
0.2
0.3
0.4
0.5

Fig. 5.14

Fig. 5.15: Nucleotide sequence of the amplicon Bw*Eco*RV lib.GWP4 generated as a result of nested PCR with *Eco*RV adapter library and primer pair AP1-SS13R as the template and oligonucleotide primer pair AP2-SS13R

Fig. 5.16: Nucleotide sequence of the amplicon Bw*Dra*I lib.GWP2 generated as a result of nested PCR with *Dra*I adapter library and primer pair AP1- SS13R as the template and oligonucleotide primer pair AP2-SS13R

```
1   AAGTATCCGC GCGGCCGGGC TGGTTCCAGT TTGCCCCTGC ATGGATGTAC
51  CAGAGGGAAA ATATCACACT TTTAACTCGG AAAGAATCTA CAGTGGAGCT
101 TCCAAACAAT GGTGAGTGTC AGAAAAACTA TTTTAAGCTG TAATCAACCA
151 AAGGTTAGCA CAAGGTAACT AATTAAAAAA AAATTGGAGG TTCTCAGTAC
201 TAACGAGTAA CTTTATCATT TGATTACATG AGTTTCTCAT CAGGCCTTTA
251 CAAAAACAAA GAACAAACAA TATCAAATGC CTTGTACCTG TATAGATCCT
301 ATCATGACAA CTCTAGCCCA TATCATCCAT GCAGGTTTAT CAATATAAAT
351 ATGTCAGTAA ATCACAAAGC AGGCTAGGCA GCAGGTCAAT TATCACATTA
401 CTGGAACCTA AATTTGCAGC TTTGAAATTC ACTTAGAGAG TAAAAAATTT
451 GGCGCTCAAT ACATAATCAG AGTGTGGGAT GAATCAACCC CCTTTGATAC
501 TTCAATTATA AAAAATAAGT AGACGAACCT AGATCATTCA GAGGACCAAA
551 TTCATTCGAC CGAAAGCCTC CAAACCGAAA CATACGAGGT ATGTATGAGA
601 AGTCACCTGA AATCAAAATC AATTAAACTA CTACAAGTCG CCACATACAC
651 GACAGGTGTA AAGCCTCAGC ATACAGCAGG CGAATTGTCA TCCATGCAGA
701 TGATACAAGC CAAAGTCACC ATGCAACCAC ACTCAATACT TACTCTTCCA
751 TCCTCAGTAT CACAACTTCA AATCTTCCAC TATGTCAACT AAACTCATAC
801 TCTCCTTCTC ACTGTGCCTT ATGGTACTAA GCTGCTCTGC GCAGCTATTG
851 CCATGGCGGA AGGGACAAC
```

Fig. 5.15

```
1   CTTAGAGAGT AAAAAATTTG GCGCTCAATA CATAATCAGA GTGTGGGATG
51  AATCAACCCC CTTTGATACT TCAATTATAA AAAATAAGTA GACGAACCTA
101 GATCATTCAG AGGACCAAAT TCATTCGACC GAAAGCCTCC AAaCCGAAAC
151 ATACGAGGTA TGTATGAGAA GTCACCTGAA ATCAAAATCA ATTAAACTAC
201 TACAAGTCGC CACATACACG ACAGGTGTAA AGCCTCAGCA TACAGCAGGC
251 GAATTGTCAT CCATGCAGAT GATACAAGCC AAAGTCACCA TGCAACCACA
301 CTCAATACTT ACTCTTCCAT CCTCAGTATC ACAACTTCAA ATCTTCCACT
351 ATGTCAACTA AACTCATACT CTCCTTCTCA CTGTGCCTTA TGGTACTAAG
401 CTGCTCTGCG CAGCTATTGC CATGGCGGAA GGGACAAC
```

Fig. 5.16

from *BwEco*RV lib.GWP4 sequence and to include the coding region of the gene along with the promoter, a forward primer "SS15F" was designed from the 5' end of nucleotide sequence of *BwEco*RV lib.GWP4 and was used in combination with two reverse primers *viz.* SS13R and SS7R to amplify the target DNA from genomic DNA isolated from etiolated shoots of common buckwheat. While the polymerase chain reaction with primer pair SS15F and SS13R amplified a fragment with apparent molecular mass of 1.0 kb, that with primer pair SS15F and SS7R amplified a DNA fragment having apparent molecular mass 1.2 kb (Fig. 5.17). The two amplicons were gel purified, cloned in plasmid vector pJET1.2 (Fermentas) and sequenced, availing the nucleotide sequencing facility offered by Axygen Pvt. Ltd (India). The nucleotide sequences of 1028 bases for the 1.0 kb amplicon generated with primer pair SS15F-SS13R and of 1166 bases for the 1.2 kb amplicon generated with primer pair SS15F-SS7R are presented in Figs. 5.18 and 5.19. Pairwise alignment of the 1028 bp and 1166 bp nucleotide sequences using CLUSTAL W revealed 100% homology for the aligned region between the two sequences. The nucleotide sequence has been deposited in the gene bank data base with accession no. EU595873. BLASTn analysis with the 1028 base nucleotide sequences as the query revealed strong homology between the 3' region of the query sequence and the 5' region of buckwheat legumin cDNA sequences D87980, D87982 in the databank. The tool also identified the sequences as the 5' upstream regulatory region of SSP genes. While the clone with the 1166 bp amplicon was designated as pJETBw1166 that with the 1028 bp amplicon has been designated as pJETBw1028.

Fig. 5.17: (a) Electrophoresis profile of the amplified products of PCR with buckwheat genomic DNA as the template and primer pair SS15F-SS13R (L1) and primer pair SS15F-SS7R (L2). M: *Eco*RI/*Hind*III digested λ DNA

(b) restriction digestion profile of recombinant plasmids pJET1166 and pJET1028. L1: undigested plasmid DNA for the clone pJET1166, L2: undigested plasmid DNA for the clone pJET1028, L3: *Nco*I-*Not*I digestion profile of the recombinant plasmid pJET1166 showing the released 1.166 kb insert DNA, L4: *Nco*I-*Not*I digestion profile of the recombinant plasmid pJET10286 showing the released 1.028 kb insert. M: *Eco*RI/*Hind*III digested λ DNA

(c) Electrophoresis profile of the amplified products of PCR with recombinant plasmid clone pJET1028 as the template and oligonucleotide primer pair T_3-T_7 (L1) and recombinant plasmid clone pJET1028 as the template and oligonucleotide primer pair T_3-T_7 (L2). M: *Eco*RI/*Hind*III digested λ DNA

Fig. 5.18: Nucleotide sequence of 1028 bases for the insert DNA of the recombinant plasmid clone pJET1028.

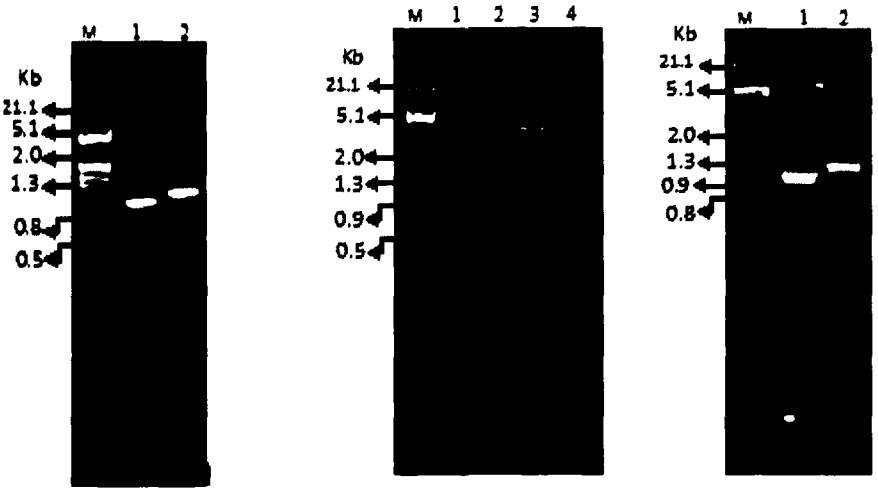


Fig: 5.17 (a) (b) (c)

```
1    CATACGATCT TACTCGGAAG ATCTAAGATG GAGCTTCCAA ACAATGGTGA
51   GTGTCAGAAA AACTATTTTA AGCTGTAATC AACCAAAGGT TAGCACAAGG
101  TAACTAATTA AAAAAAAATT GGAGGTTCTC AGTACTAACG AGTAACTTTA
151  TCATTTGATT ACATGAGTTT CTCATCAGGC CTTTACAAAA ACAAAGAACA
201  AACAATATCA AATGCCTTGT ACCTGTATAG ATCCTATCAT GACAACTCTA
251  GCCCATATCA TCCATGCAGG TTTATCAATA TAAATATGTC AGTAAATCAC
301  AAAGCAGGCT AGGCAGCAGG TCAATTATCA CATTACTGGA ACCTAAATTT
351  GCAGCTTTGA AATTCACTTA GAGAGTAAAA AATTTGGCGC TCATATACAT
401  AGCAGGTATT TAAATCAGAG TGTGGGATGA ATCAACCCCC TTTGATACTT
451  CAATTATAAA AAAGTAGACG AACCTAGATC ATTCAGAGGA CCAAATTCAT
501  TCGACCGAAA GCCTCCAAAC CGAAACATAC GAGGTATGTA TGAGAAGTCA
551  CCTGAAATCA AAATCAATTA AACTACTACA AGTCGCCACA TACACGACAG
601  GTGTAAAGCC TCAGCATACA GCAGGCGAAT TGTCATCCAT GCAGATGATA
651  CAAGCCAAAC TCACCATGCA ACCACACTCA ATACTTACTC TTCCATCCTC
701  AGTATAAAAT CCAACCCACG CCCGCTTCTT TCAATCACCA CCCCTCGATC
751  AACACAACTT CAAATCTTCC ACCATGTCAA CTAAACTCAT ACTCTCCTTC
801  TCACTGTGCC TTATGGTACT AAGCTGCTCT GGCGCAGCTT TTGGCCATGG
851  CAGAACGGAC AACGCAGCCG CCCCCCACCA TGGAACACCA GCATTCAGCA
901  TCAGTGTTGA TATCAGAGGC TCACCGCCTC TGAGCCCTCT CGTAAAGTCG
951  ATCGAGCTGA GTACTGGGAC TCGGACATGA CACCCTGAGT CGATGGCGAT
1001 TTGCGTATTT TGGAAATAAC AGAGAGAG
```

Fig. 5.18

Fig. 5.19: Nucleotide sequence of 1166 bases for the insert DNA of the recombinant plasmid clone pJET1166.

```
1    AAACTCTAGC TCAGGTAAGA TACTAAGAGT GTTCCTCCCA ATGATGGAGT
51   GGGGGTAAAA AAATTTATTT TAGGCGTAAA CCAACAAAGG GTAAAAAGAG
101  GGAACTAATT TAAAAAAATT TGGAGGTTCT CAGCACTAAC GAGTAACTTT
151  ATCATTTGAT TACATGATTT TTTCAATCGG ACCTTAACAA AAAAAAAAAA
201  AAAAAAAAAT CAATTGCTTT GACCAGGTTT AAATCCAATA TGGGCAATTC
251  AGGCTAATAT AACCAAGGAG GGTTATTAAT AAAAAATTTG TCAGTAAATA
301  AAAAGGAAGG TAAGGAAGAA GGTAATTAAT AAGGTAACTA AAAACTAATT
351  TTGCAGTTTT GAATTTAACA TAAAAGGTAA AAAATTTGGG CGCTAATAAA
401  AAAAATGTTT TATAAAAATA GGGGGGGGGG AATAAACACC CCCCGGGGGT
451  TCTTTTATTT AAAAATAGTA GACCACCCTT TTTTTTGAGG AGGAACTTCT
501  TCTTTCGACA GAAAGCCTCT TTGTATGAAA CAGACGGGTG TATGTATGAG
551  AAGTCCGCGG AAAAAAAATT TATTATACTA CAAGGGGTCC CATACATACA
601  CGAGGGGTGG GCCGCTTCAT CCATACAGGG GGATGATTTG TCCCTCCCTG
651  AAGATATAAC AGCAAAACTC CAGGATGCCC ACCCCCCCTC AATATTTACT
701  CCCTCCCCCC CTCAATATAA TAAACCCACC CCCGCGCTTC TTTTTATAAA
751  TCACCCACCC TCGATCAACA AAATCTAAAT ATTCTCCCAC GGGGTCAAAA
801  AAAACTCCCT CCTCCTCCTT CAAAGGTGGT GCCATAAGGA TAATAAAGCT
851  GCTGCGAAGA TATATAGTGG GGCAAAAGAG AACAACACAC CCCCCCCACC
901  ACGGGGGACG GCGAGCTTCT CCCGCATGGG TGTGATATAA AGATGTCTCC
951  ACCGCCTCAT GCCTCTCTGA TAAATTCCTG TACGGCGCGC AGAATAACAG
1001 AATGTGGGAA ACGATGCACC CCAGTTTTTT GCCATGCGGC AGAATTGTCT
1051 CCCCGTGCGG TGGTATATTT TATCTGGGAG GCTTTGTATC TTATTTCGAA
1101 CTCGAGGCGG GGGAGACTAA ATTTTTTTTG GGGGGGGAGA TTATGGGGAC
1151 CCACCCTCTT TTTTGA
```

Fig. 5.19

Fig. 5.20: BLASTp output of the predicted amino acid sequence of the ORF at position 774 in the nucleotide sequence of the 1028 bp insert DNA of the recombinant plasmid clone pJET1028.

```
>  sp|O23878.1|13S1_FAGES  RecName: Full=13S globulin seed storage
protein 1; AltName: Full=Legumin-like
protein 1; Contains: RecName: Full=13S globulin
seed storage protein 1 acidic chain; Contains: RecName:
Full=13S globulin seed storage protein 1 basic chain; Flags:
Precursor
 dbj|BAA21758.1|  legumin-like protein [Fagopyrum esculentum]
Length=565

 Score = 47.8 bits (112),  Expect = 4e-04, Method: Composition-based
stats.
 Identities = 26/48 (54%), Positives = 29/48 (60%), Gaps = 1/48 (2%)

Query  1   MSTKLILSFSLCLMVLSCSGAAFGHGRTDNAAAPHHGTPAFSISVDIR  48
           MSTKLILSFSLCLMVLSCS A     R    + PH G   F    D++
Sbjct  1   MSTKLILSFSLCLMVLSCS-AQLLPWRKGQRSRPHRGHQQFHHQCDVQ  47


>  sp|O23880.1|13S2_FAGES  RecName: Full=13S globulin seed storage
protein 2; AltName: Full=Legumin-like
protein 2; Contains: RecName: Full=13S globulin
seed storage protein 2 acidic chain; Contains: RecName:
Full=13S globulin seed storage protein 2 basic chain; Flags:
Precursor
 dbj|BAA21760.1|  legumin-like protein [Fagopyrum esculentum]
Length=504

 Score = 47.4 bits (111),  Expect = 4e-04, Method: Composition-based
stats.
 Identities = 24/37 (64%), Positives = 26/37 (70%), Gaps = 1/37 (2%)

Query  1   MSTKLILSFSLCLMVLSCSGAAFGHGRTDNAAAPHHG  37
           MSTKLILSFSLCLMVLSCS A     +    + PHHG
Sbjct  1   MSTKLILSFSLCLMVLSCS-AQLWPWQKGQGSRPHHG  36


>  gb|ACC62039.1|  legumin-like protein [Fagopyrum esculentum]
Length=18

 Score = 36.6 bits (83),  Expect = 0.76, Method: Compositional matrix
adjust.
 Identities = 18/18 (100%), Positives = 18/18 (100%), Gaps = 0/18 (0%)

Query  1   MSTKLILSFSLCLMVLSC  18
           MSTKLILSFSLCLMVLSC
Sbjct  1   MSTKLILSFSLCLMVLSC  18
```

Fig. 5.20

Fig. 5.21: The promoter region and the corresponding TSS predicted by Neural Network promoter Prediction (NNPP tool) in the nucleotide sequence of the 1028 bp insert DNA of the recombinant plasmid clone pJET1028

Promoter predictions for EU595873:

Start	End	Score	Promoter Sequence
364	414	0.96	TCACTTAGAGAGTAAAAAATTTGGCGCTCATATACATAGC**A**GGTATTTAA
445	495	0.95	ATACTTCAATTATAAAAAAGTAGACGAACCTAGATCATTC**A**GAGGACCAA
693	743	0.90	CCATCCTCAGTATAAAATCCAACCCACGCCCGCTTCTTTC**A**ATCACCACC

Fig. 5.21

Analysis of the GC content in the 1028 bp 5' upstream regulatory region of buckwheat legumin-type SSP gene (acc. no. EU595873) revealed a GC percentage of 41.83 in the sequence. This classifies the sequence as being AT rich. In order to identify the ATG start codon in the 1028bp sequence, the nucleotide sequence was subjected to ORF finder tool (NCBI). The tool identified 8 putative open reading frames *viz.* 774-980 (length 207 bases), 540-707 (length 168 bases), 749-901 (length 153 bases), 163-312 (length 150 bp), 921-1027 (length 108 bases) and 24-371 (length 108 bases). The predicted amino acid sequences for each of the 8 putative ORFs were subjected to BLASTp (NCBI). The predicted amino acid sequence of the ORF starting at position 774 showed alignment to the buckwheat 13S globulins SSPs bearing accession nos. O23878, O23880 and ACC26039 (Fig. 5.20). The first amino acid (methionine) of the predicted ORF coincided with the first amino acid (methionine) of all the three subject sequences. The ATG of the ORF at position 774-980 was thus assumed to be the initiating codon of the gene under study.

GENSCAN 1.0 identified the nucleotide sequence between position 698 to 737 of the 1028bp fragment as the 5' upstream regulatory region. Analysis of the 1028 bp sequence by promoter prediction tool (Neural Network Promoter Prediction) identified three probable promoter regions at positions 364-414, 445- 495 and 693-743 in the sequence (Fig. 5.21). The TSS for the three predicted promoters are 'A' at position 404 for the promoter at position 364-414, 'A' at position 485 for promoter at position 445-495 and the 'A' at position 773 for the promoter at position 693-743. Of the three predicted transcription start sites, the TSS at position 773 is located closest to the predicted ATG start codon. Thus, the adenine at position 733 could be the

transcription start site (TSS) of the promoter under study. Considering 'A' at position 773 as the predicted TSS and ATG at position 774 (+41) as the initiating codon, the TATA at position 703 (-30) would be the TATA box of the promoter under investigation. promHG also identified the adenine at position 733 as the prospective TSS.

Apart from TATA box, TSSP tool identified several *cis*-elements that are involved in the regulation of eukaryotic gene in general and seed-specific expression in particular in the nucleotide sequence of the putative promoter EU595873. These included motifs similar to 'TGTAAAG' (RegSite acc. no. RSP00028), 'CACA' (RegSite acc. no. RSP00131) and CATGCA (RegSite acc. no. RSP00327). The motifs identified by TSSP in 1028 bp amplicon and their characteristics are presented in table 5.2. PLACE Web-Signal-Scan software also identified several motifs, of which the ones corresponding to SSP genes are TGTAAAG, AACAAAC, CAAT, CANNTG, CATGCA in the nucleotide sequence of the putative promoter EU595873 (Table 5.3). The position of various elements /motifs in the 1028 bp putative promoter is shown in Fig. 5.22.

Attempts were also made to identify *cis*-elements by analysing the promoter sequence of other SSP genes to find out the conserved/overrepresented short sequences. 5' upstream regions of SSP gene sequences of different plants were downloaded from NCBI GenBank database. The name and accession numbers of the downloaded sequences are given in table 5.4. The downloaded sequences were trimmed to include nucleotides only upto 1000 bp upstream and 100 bp downstream of ATG start codon. The downloaded promoter sequences along with the 1028 bp

Table 5.2: *Cis*-elements registered in RegSite database which were identified in the 1028 bp putative promoter by TSSP tool.

Signal sequence	**Position**	**RegSite acc. no.**	**Characters**
TATA	-30.-454	RSP01090	TATA box
TGTAAAG	-131	RSP00028	22kD *zein* P-box, target of TF Opaque-2
CATGCA	68, -95, -470	RSP00327	*Brassica napus* napA RY element, TF: AB13
CACA	-60,-146, -435 & -639	RSP00131	B-phaseolin CACA box. TF: unknown nuclear factor.

Table 5.3: *Cis*-elements registered in PLACE database which were identified in the 1028 bp putative promoter by TSSP tool.

Signal sequence	**Position**	**PLACE acc. no.**	**Character.**
TGTAAAG	-131	S000001	-300 motif core element of α *zein* of maize
AACAAAC	-543	S000353	Core of AACA motifs Rice glutelin genes
CAAT	-54, -168, -282, -411, -457,	S000028	CAAT box *leg*A gene of pea.
CANNTG	-91, -135, -184, -524, -581	S000144	E-box of napA storage-protein gene of Brassica napus
CATGCA	-68, -95, -470	S000264	RY repeat" found in RY/G boxof napA gene in *Brassica napus*
TATAAT	-30, -554	S000203 S000109	TATA box found in the 5'upstream region of pea legA and beta-phaseolin promoter

Fig.5.22: Nucleotide sequence of 1028 bases for the insert DNA of the recombinant plasmid clone pJET1028. Various regulatory elements identified in the sequence are highlighted and their positions w.r.t TSS is indicated against each. The TSS is marked by an arrow, ATG start codon is encircled and the motif CANNTG is highlighted in red.

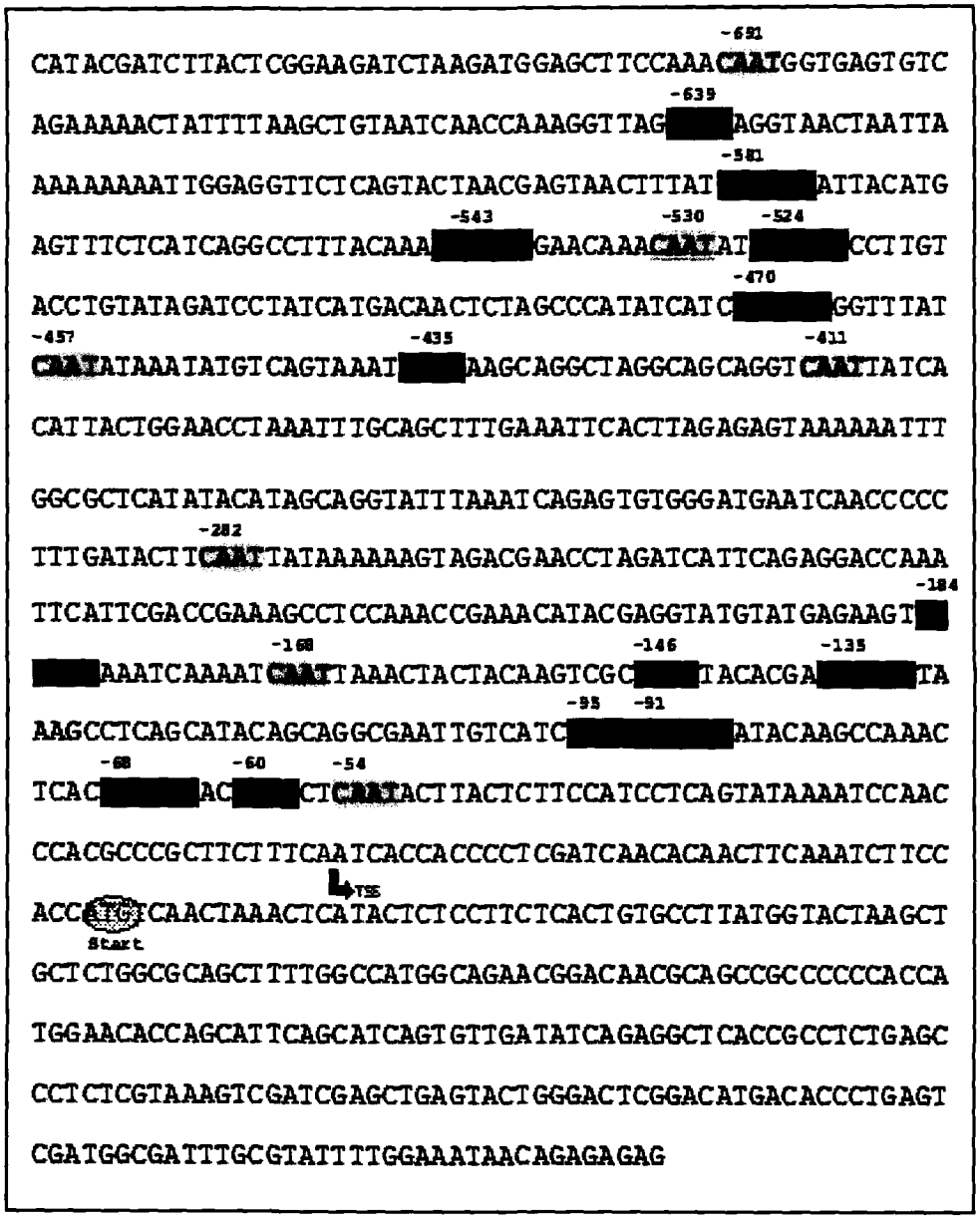

-691
CATACGATCTTACTCGGAAGATCTAAGATGGAGCTTCCAAACAATGGTGAGTGTC
-639
AGAAAAACTATTTTAAGCTGTAATCAACCAAAGGTTAG AGGTAACTAATTA
-581
AAAAAAAATTGGAGGTTCTCAGTACTAACGAGTAACTTTAT ATTACATG
-543 -530 -524
AGTTTCTCATCAGGCCTTTACAAA GAACAAACAATAT CCTTGT
-470
ACCTGTATAGATCCTATCATGACAACTCTAGCCCATATCATC GGTTTAT
-457 -435 -411
CAATATAAATATGTCAGTAAAT AAGCAGGCTAGGCAGCAGGTCAATTATCA
CATTACTGGAACCTAAATTTGCAGCTTTGAAATTCACTTAGAGAGTAAAAAATTT
GGCGCTCATATACATAGCAGGTATTTAAATCAGAGTGTGGGATGAATCAACCCCC
-282
TTTGATACTTCAATTATAAAAAAGTAGACGAACCTAGATCATTCAGAGGACCAAA
-184
TTCATTCGACCGAAAGCCTCCAAACCGAAACATACGAGGTATGTATGAGAAGT
-168 -146 -135
AAATCAAAATCAATTAAACTACTACAAGTCGC TACACGA TA
-95 -91
AAGCCTCAGCATACAGCAGGCGAATTGTCATC ATACAAGCCAAAC
-68 -60 -54
TCAC AC CTCAATACTTACTCTTCCATCCTCAGTATAAAATCCAAC
CCACGCCCGCTTCTTTCAATCACCACCCCTCGATCAACACAACTTCAAATCTTCC
TSS
ACCATGTCAACTAAACTCATACTCTCCTTCTCACTGTGCCTTATGGTACTAAGCT
Start
GCTCTGGCGCAGCTTTTGGCCATGGCAGAACGGACAACGCAGCCGCCCCCCACCA
TGGAACACCAGCATTCAGCATCAGTGTTGATATCAGAGGCTCACCGCCTCTGAGC
CCTCTCGTAAAGTCGATCGAGCTGAGTACTGGGACTCGGACATGACACCCTGAGT
CGATGGCGATTTGCGTATTTTGGAAATAACAGAGAGAG

Fig. 5.22

putative promoter of buckwheat legumin gene amplified in the present study were subjected to analysis by MEME 4.1.0. The software detected the motifs "TATAAA", "CATGCA", "GGTGT" in all the 23 promoters sequences, "CTCTCTTC" in 21 sequences, motif "ACGTGTC" in 12 sequences and the motif "GTAAG" in 8 sequences investigated in the present study. Consensus sequences/motifs identified by MEME 4.1.0 in the set of promoter sequences considered for the present investigation are shown in table 5.5.

PCR based unidirectional 5' progressive deletions of the putative promoter region in the construct pJETBw1028 generated deletion fragments of varying lengths for assessment of the role of various regulatory elements in regulating the expression of reporter genes. Seven forward primers *viz.* DF1, DF2, DF3, DF4, DF5, DF6 and DF7 and one reverse primer *viz.* DR1 were designed to amplify smaller fragments of various lengths from the 1.028 kb cloned DNA fragment. The positions of the primers on the 1.028 kb cloned DNA fragment are represented in Fig. 5.23. The primer DF7 was designed to eliminate the TATA box of the 1.028 kb putative promoter region of buckwheat SSP gene. PCR-based deletion of the putative promoter fragment generated a ladder of deleted fragments with the primer pair SS15F- DF1 generating a DNA fragment of highest molecular mass of 800 bp and the primer pair DF7- DR1 generating a DNA fragment with the lowest molecular mass of 74 bp (Fig. 5.24a). Southern hybridization of the deletion fragments with [α-^{32}P]-dATP labelled 1028 bp putative promoter region of buckwheat SSP gene generated positive signals, corresponding to the positions of each of the deletion fragment of the putative promoter region, on the X-ray film (Fig. 5.24b).

Table 5.4: List of gene promoters used for analysis with MEME tool.

Sl. No.	Plant source	Promoter of gene:	Acc. No.
1	buckwheat	Legumin (present study)	EU595873
2	Coffee	Csp1Gene	Y16975
3	Chick pea	Leg gene	Y13166
4	Pea	Lectin gene	Y00440
5	Pea	VicJ gene	X67428
6	*B.napus*	napB gene	X58142
7	soybean	A(1A)B(1B)and A(2)B(1A)boundary DNA	X53404
8	*Vicia faba*	pseudogene.	X14238
9	*Pisum*	leg J gene	X07014
10	*Vicia faba*	legB4 gene	X03677
11	Pea	Leg C gene	X02984P
12	Pea	leg B gene	X02983
13	Pea	Leg A Gene	X02982
14	*B.campestris*	Clone LambdaGCN 1-2 Nap A gene.	M64632
15	chenopods	Chenopodium 11S globulin gene	DQ917483
16	Coffee	11S globulin gene	DQ198078
17	Chenopodium	11S globulin gene	DQ917483
18	*Cajanus*	pPUMU2 seed-specific promoter-like sequence	AY771784
19	Soybean	conglycinin gene	AY605542
20	Buckwheat	13S globulin gene	AY359286
21	Vitis Vinifera	2S albumin precursor gene	AY267256
22	Phaseolus	Iap1 gene,promoter	AY166712
23	*B. napus*	Lea3, gene promoter.	AJ535107

Table 5.5: Common motifs found among 23 SSP gene promoters as detected by MEME tool.

Motif sequence logo	Position and sequence on promoterof present study	Motifs found in	Motif function
	279, 703 TATAAA	23/23	Canonical *cis*-element found in many eukaryotic gene promoters. Essential for transcriptional initiation.
	635, 260, TCCATCA 662 ACCATGCA	23/23	RY element found in many seed promoters, responsible for tissue specific gene expression.
	600 GGTGTAAA	23/23	unknown
	nil	12/23	Core motif of abscisic acid responsive element (ABRE), quantitative role, endosperm specific expression, ABA regulation.
	nil	21/23	Y-patch
	599 GGTGTAAAGC	8/23	Similar to enhancer- like element controlling the expression of soybean conglycinin gene (Chen et al., 1988).

Fig. 5.23: Nucleotide sequence of 1028 bases for the insert DNA of the recombinant plasmid clone pJET1028 showing the position of the primers designed for generating PCR-based deletion fragments of the 1028 bp DNA.

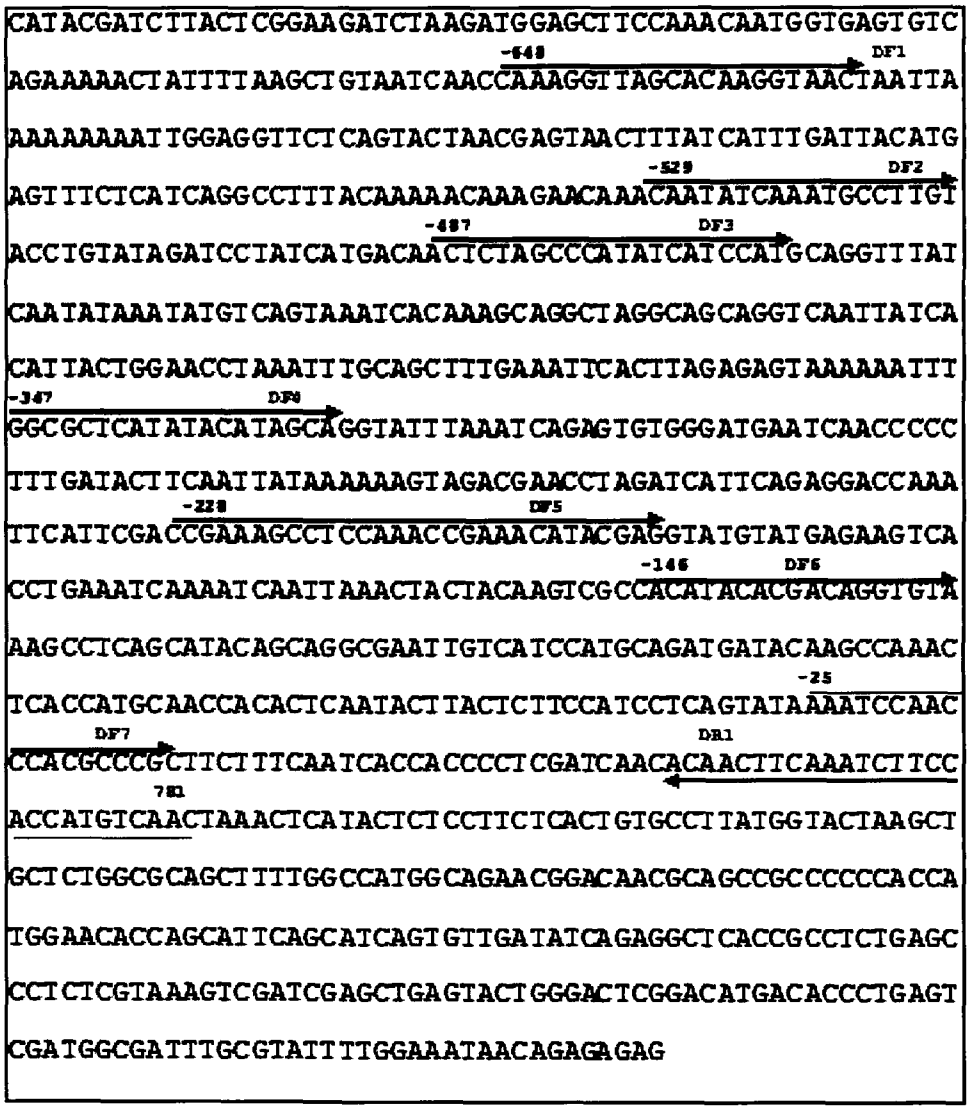

CATACGATCTTACTCGGAAGATCTAAGATGGAGCTTCCAAACAATGGTGAGTGTC
-648
DF1
AGAAAAACTATTTTAAGCTGTAATCAACCAAAGGTTAGCACAAGGTAACTAATTA
AAAAAAAATTGGAGGTTCTCAGTACTAACGAGTAACTTTATCATTTGATTACATG
-529
DF2
AGTTTCTCATCAGGCCTTTACAAAAACAAAGAACAAACAATATCAAATGCCTTGT
-487
DF3
ACCTGTATAGATCCTATCATGACAACTCTAGCCCATATCATCCATGCAGGTTTAT
CAATATAAATATGTCAGTAAATCACAAAGCAGGCTAGGCAGCAGGTCAATTATCA
CATTACTGGAACCTAAATTTGCAGCTTTGAAATTCACTTAGAGAGTAAAAAATTT
-347
DF4
GGCGCTCATATACATAGCAGGTATTTAAATCAGAGTGTGGGATGAATCAACCCCC
TTTGATACTTCAATTATAAAAAAGTAGACGAACCTAGATCATTCAGAGGACCAAA
-228
DF5
TTCATTCGACCGAAAGCCTCCAAACCGAAACATACGAGGTATGTATGAGAAGTCA
-146
DF6
CCTGAAATCAAAATCAATTAAACTACTACAAGTCGCCACATACACGACAGGTGTA
AAGCCTCAGCATACAGCAGGCGAATTGTCATCCATGCAGATGATACAAGCCAAAC
-25
TCACCATGCAACCACACTCAATACTTACTCTTCCATCCTCAGTATAAAATCCAAC
DF7
DR1
CCACGCCCGCTTCTTTCAATCACCACCCCTCGATCAACACAACTTCAAATCTTCC
781
ACCATGTCAACTAAACTCATACTCTCCTTCTCACTGTGCCTTATGGTACTAAGCT
GCTCTGGCGCAGCTTTTGGCCATGGCAGAACGGACAACGCAGCCGCCCCCCACCA
TGGAACACCAGCATTCAGCATCAGTGTTGATATCAGAGGCTCACCGCCTCTGAGC
CCTCTCGTAAAGTCGATCGAGCTGAGTACTGGGACTCGGACATGACACCCTGAGT
CGATGGCGATTTGCGTATTTTGGAAATAACAGAGAGAG

Fig. 5.23

Fig. 5.24: (a) Electrophoresis profile of the amplicons generated by PCR based deletions with plasmid DNA isolated from the recombinant clone recombinant plasmid pJET1028 as the template and primer pairs SS15F-DR1 (L1), DF1-DR1 (L2), DF2-DR1 (L3), DF3-DR1 (L4), DF4-DR1 (L5), DF5-DR1 (L6), DF6-DR1 (L7), DF7-DR1 (L8). M1: 100 bp Ladder, M2: *Hinf*II digested pBR322 DNA.

(b) Southern hybridization of the deletion fragments generated by PCR based deletions with plasmid DNA isolated from the recombinant clone recombinant plasmid pJET1028 as the template and primer pairs SS15F-DR1, DF1-DR1, DF2-DR1, DF3-DR1, DF4-DR1, DF5-DR1, DF6-DR1 and DF7-DR1 with with [α-^{32}P]-dATP labelled insert DNA from the recombinant plasmid clone pJET1028.

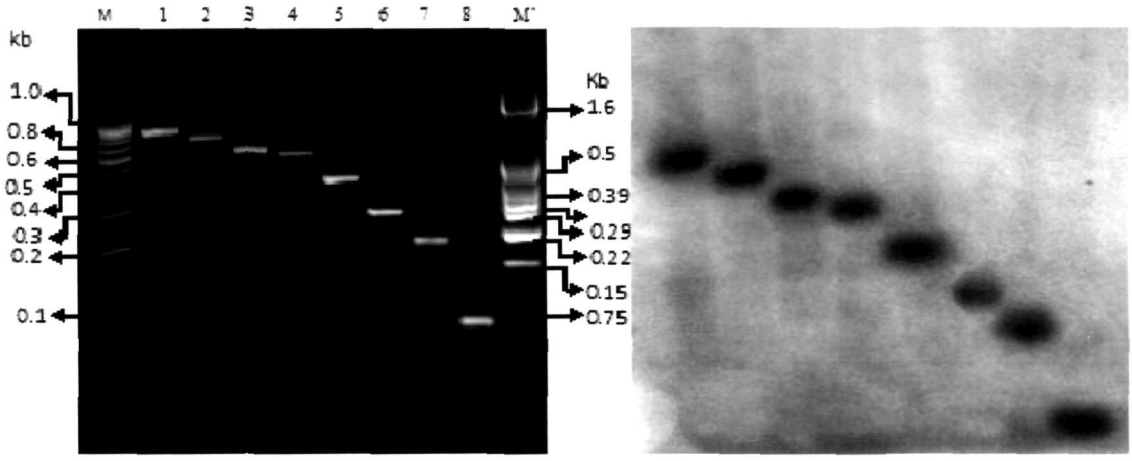


Fig. 5.24 (a) (b)

Each deletion fragment was gel purified and sub-cloned in promoterless vector pGlowTOPO vector (Invitrogen) carrying Green Fluorescent Protein (GFP) as the reporter gene. Constructs developed with each of the sub-cloned fragments were transformed into *E. coli* (TOP10) competent cells supplied along with the kit. A DNA fragment not forming a part of 5' upstream of any gene was also cloned simultaneously to act as the control. The control construct has been designated as $pGlowTOPO_{NUL}$ and the deletion constructs designated $pGlowTOPO_{-731}$, $pGlowTOPO_{-648}$, $pGlowTOPO_{-529}$, $pGlowTOPO_{-487}$, $pGlowTOPO_{-347}$, $pGlowTOPO_{-228}$, $pGlowTOPO_{-146}$ and $pGlowTOPO_{-25}$. The transformed *E. coli* cells were plated on LB(amp^+) agar plates and the plates incubated at 37°C for 48 hours. Numerous colonies became visible after 16 hours of incubation. However, incubation was continued for 48 hours to allow the transformed cell to express and accumulate GFP. Cells from the colonies on the culture plates were picked and viewed with Carl Zeiss fluorescence microscope with filter no. 9 which allowed excitation at 450-490 nm and fluorescence emission at 515 nm. The cells transformed with the $pGlowTOPO_{NUL}$ construct did not show any fluorescence when excited with UV light. Compared to the control, *E. coli* cells transformed with different deletion constructs showed varying intensity of GFP expression. Constructs with PCR generated deletion fragments -731, -648 and -529 did not show any differences in the expression of GFP in *E.coli* cells. Thus, deletion of -691 CAAT motif, -639 CACA motif, -581 CANNTG motif and -549 AACA motif did not appear to affect the activity of the promoter. However, construct with PCR generated deletion fragment -487, -347 and -228 showed significantly reduced activity of the promoter compared

to -731, -648 and -529 constructs. Thus deletion of -470 CATGCA, -530 CAAT, -457 CAAT, -411CAAT and -524 CANNTG and -435 CACA motifs caused a marked reduction in the expression of GFP in the *E.Coli* cells. While the constructs with deletion fragments -146 showed very negligible level of reporter gene expression that with deletion fragment -24 did not show any promoter activity. Fig 5.25 shows the photomicrographs of fluorescing *E.coli* cells transformed with different constructs of pGlowTOPO. The relative intensity of fluorescence emitted by *E. coli* cells transformed with different deletion constructs is represented in Fig. 5.26.

Amplification of the full length buckwheat legumin SSP gene along with its 5' upstream flanking region with buckwheat genomic DNA as the template and oligonucleotide primer pairs SS15F-SS16R amplified a fragment showing apparent molecular mass of 3.0 kb (Fig. 5.27a). The 3.0 kb amplicon was transferred to a nylon membrane and hybridized with [α-^{32}P]-dATP labelled 1028 bp putative promoter region of buckwheat SSP gene and buckwheat legumin like protein gene fragment corresponding to the coding region of the gene. Both the probes generated positive hybridization signals on the autoradiograph, corresponding to the position of the 3.0kb DNA fragment on the membrane (Fig. 5.27b,c). These results indicated that the 3.0 kb fragment has both the coding region as well as the promoter region of the buckwheat legumin-like protein gene. The 3.0 kb fragment was separately digested with *Eco*RI, *Nco*I, *Eco*RV, *Hind*III and *Stu*I. The restriction digestion profile of the 3.0 kb fragment for *Eco*RI, *Nco*I, *Eco*RV, *Hind*III and *Stu*I is presented in Fig 5.28. Digestion of the 3.0 kb amplicon with *Eco*RI restriction enzyme generated two bands having molecular mass of 2.2 kb and 0.5 kb. A similar profile

Fig. 5.25: Photomicrographs of fluorescing *E.coli* cells transformed with constructs $pGlowTOPO_{NULL}$, $pGlowTOPO_{-731}$, $pGlowTOPO_{-648}$, $pGlowTOPO_{-529}$, $pGlowTOPO_{-487}$, $pGlowTOPO_{-347}$, $pGlowTOPO_{-228}$, $pGlowTOPO_{-146}$, $pGlowTOPO_{-25}$.

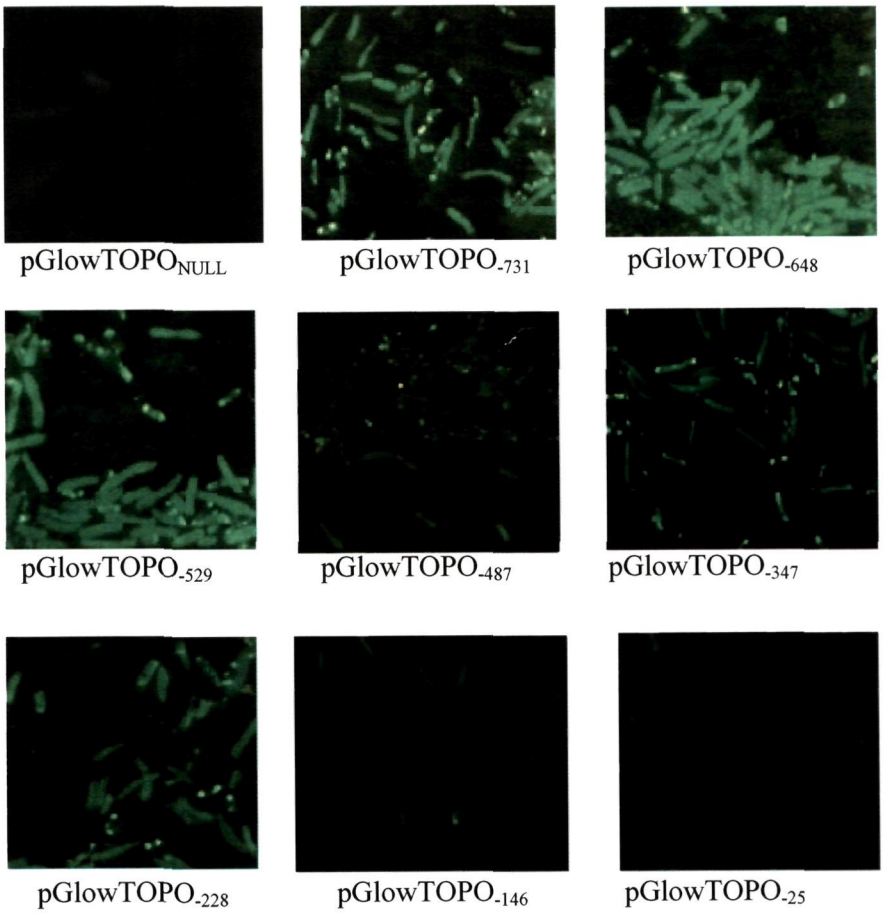

pGlowTOPO$_{NULL}$
pGlowTOPO$_{-731}$
pGlowTOPO$_{-648}$
pGlowTOPO$_{-529}$
pGlowTOPO$_{-487}$
pGlowTOPO$_{-347}$
pGlowTOPO$_{-228}$
pGlowTOPO$_{-146}$
pGlowTOPO$_{-25}$

Fig. 5.25

Fig 5.26: Schematic representation of the positions of various regulatory elements on the deletion fragments generated by PCR based deletions with plasmid DNA isolated from the recombinant clone recombinant plasmid pJET1028 as the template and primer pairs SS15F-DR1, DF1-DR1, DF2-DR1, DF3-DR1, DF4-DR1, DF5-DR1, DF6-DR1 and DF7-DR1. The relative level of expression of GFP by each deletion fragment in transformed *E.Coli* cells is shown along with. (+++ > ++ > + > -).

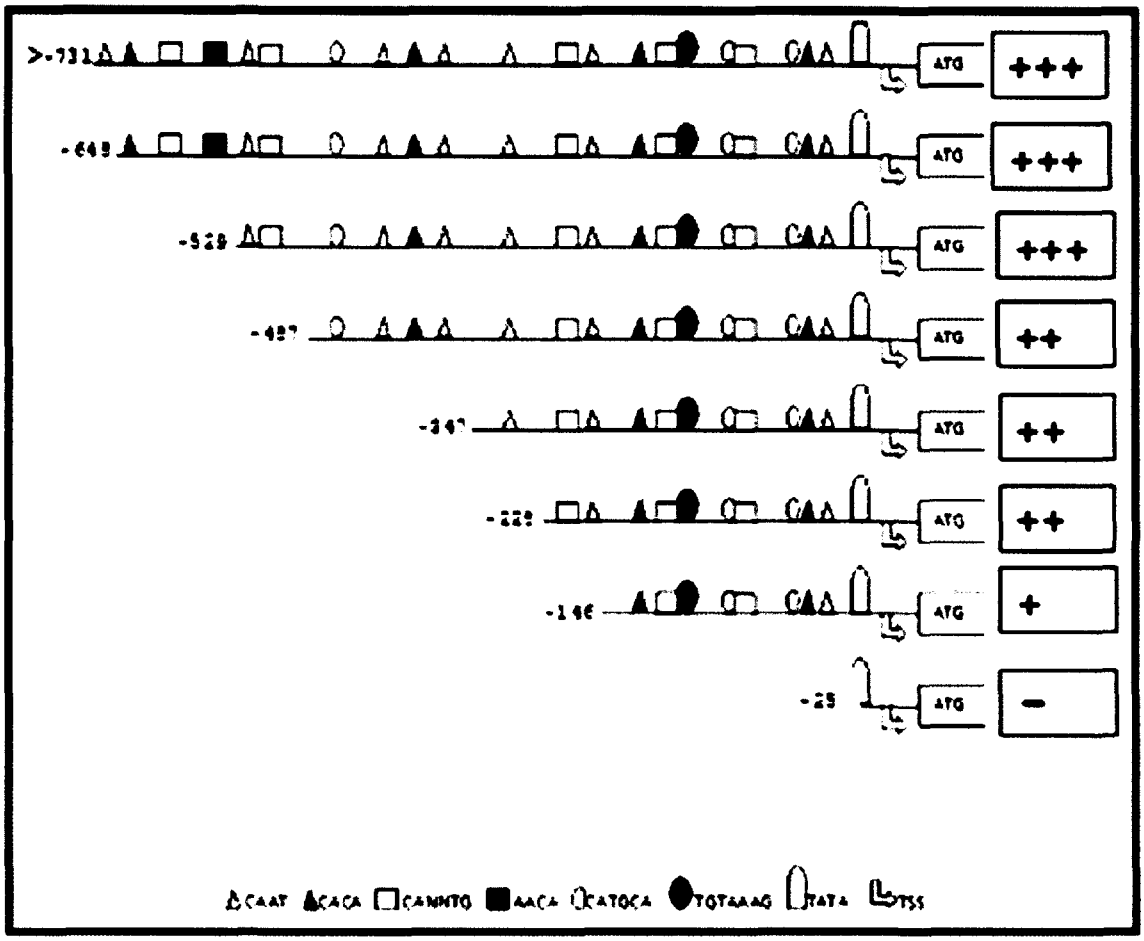


Fig. 5.26

Fig. 5.27: (a) Electrophoresis profile of the 3.0 kb DNA fragment amplified from genomic DNA of common buckwheat by PCR with primer pair SS15F-SS16R (L1-L3). M: *Eco*RI/*Hind*III digested λ DNA.

(b) Autoradiograph of the 3.0 kb fragment hybridized with [α-32P]-dATP- labelled 1028 bp amplicon generated with buckwheat genomic DNA as template using primer pair SS15F-SS13R.

(c) Autoradiograph of the 3.0 kb fragment hybridized with [α-32P]-dATP- labelled buckwheat legumin-like protein gene specific probe (DQ200889).

Fig. 5.28: Electrophoresis profile of the 3.0 kb DNA fragment digested with *Eco*RI (L1), *Nco*I (L2), *Hind*III (L3), *Eco*RV (L4), *Stu*I (L5). L6: undigested DNA, M: *Eco*RI/*Hind*III digested λ DNA

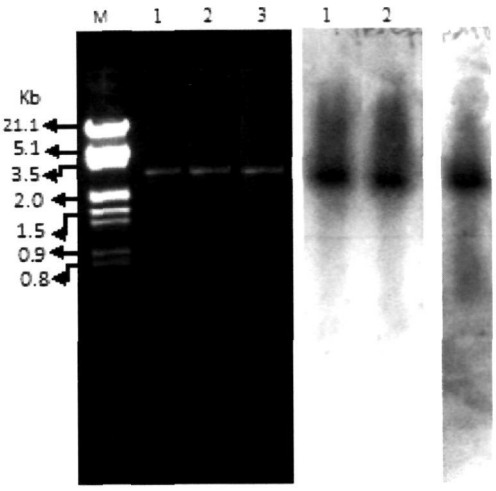


Fig. 5.27 (a) (b) (c)

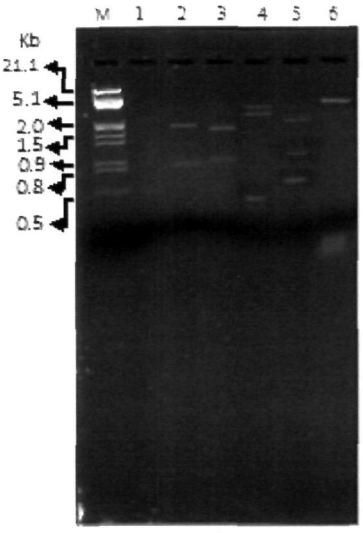


Fig. 5.28

was obtained when the 3.0 kb fragment was digested with *Nco*I and *Eco*RV. While restriction digestion with *Nco*I released two bands showing molecular masses of 2.1kb and 0.9 kb, that with *Eco*RV generated two bands having molecular mass of 1.9 kb and 1.1 kb. Restriction digestion with *Stu*I and *Hind*III generated two bands having molecular mass of 2.4 kb and 0.6 kb for the amplicon DNA digested with *Stu*I and two bands having molecular mass of 2.6 kb and 0.4 kb for the amplicon digested with *Hind*III.

Discussion:

Crop improvement by heterologous gene transfer essentially requires a promoter that would regulate the spatial and temporal expression of the trangenes. While many promoters have been identified and characterized not many seed storage protein gene promoters have been characterized till date. The primary regulatory sequences are generally located within 1000 bp upstream of the transcription start site (TSS) in plant genes, although there are cases where regulatory sequences are found further upstream (Zhang *et al.*, 1996) or downstream (3') (Dietrich *et al.*, 1992) of the coding sequences.

The gene specific primers used with adapter primers for genome walking were designed from the nucleotide sequence of the 5' upstream as well as the 3'downstream region of the target gene so as to walk along the library in 5' as well as 3' direction. The gene specific primers (GSPs) designed for genome walking were 26-30 bases in length with GC content ranging from 40-60%. To increase the stringency of the PCR reaction, the annealing temperature for all the genome walking PCR reactions carried out in the present investigation was kept at 68°C. In the

present investigation the "Touchdown" PCR protocol (Don *et al.*, 1991; Roux, 1995; Hecker and Roux, 1996) in which the annealing/extension temperature is several degrees higher than the T_m of the primers during the initial PCR cycles, thereby permitting efficient, exponential amplification of the gene-specific product, has been adopted. The higher temperature also enhances the suppression PCR effect with AP1, allowing a critical amount of gene-specific product to accumulate. Siebert *et al* (1995) have considered suppression PCR to be most critical for the success of genome walking.

BLASTn analysis of the nucleotide sequence of amplified genome walking products *BwDra*I lib.GWP1 and *BwEcoR*V lib.GWP3 revealed significant similarity with the nucleotide sequence of the cDNA clones D87890 and D87982. The results of BLAST did not reveal any homology with 5' upstream regions of any SSP gene. BLASTn of the nucleotide sequence of amplified genome walking products *Bw EcoR*V lib.GWP3 clearly identified it with the legumin gene family. Alignment of the 1127 bp nucleotide sequence of the 1.1 kb amplicon *BwEcoR*V lib.GWP3 revealed >85% sequence homology with buckwheat legumin-like cDNAs AY256960 and D87980. The sequence also showed high degree of similarity with legumin genes of other plants including *Pisum, Vicia faba,* soybean, *Pistacia, Carya, raphanus, Gingko biloba, Populus* and brazil nut. These results clearly establish a high degree of sequence conservation amongst legumin type SSP genes of various groups of plants. In comparison to other buckwheat legumin gene sequences, the nucleotide sequence of the 1.1 kb DNA amplified in the present study showed deletion of 4 bases "GCCG" at position 57 and 3 bases "CTC" at position 670. The distinguishing

feature of buckwheat legumin gene from that of legumin genes in other plants was an insertion of 5 bases "CTCCT" at position 324.

GENSCAN 1.0 detected three internal exons in the 1127 bp nucleotide sequence of the amplification product *BwEcoR*Vlib.GWP3. The graphic representation of results of BLASTn analysis of the 1127 bp nucleotide sequence as the query clearly established the conserved nature of intron-exon architecture in buckwheat legumin SSP genes. The amino acid sequence deduced from the 1127 bp nucleotide sequence of the 1.1 kb amplicon showed 78% homology with buckwheat legumin-like protein (BAA21758). The percentage of homology varied between 40-69% with deduced amino acid sequences of legumin genes from other crops including *Magnolia salicifolia* legumin precursor (CAA57846) and *Coffea arabica* 11S globulin (AAC61881). The phylogenetic tree constructed from the alignment data of the deduced amino acid sequence with amino acid sequences of SSPs available in EMBL database revealed clear clustering into three groups, one each for sequences belonging to monocots, dicots and gymnosperms. The sequences from dicots clustered into one group along with the sequences of buckwheat SSPs. These results are in conformity with the observations of Hager *et al.* (1995, 1996). Hager *et al.* (1996) have suggested that the legumin gene evolution has progressed towards loss of introns. They have also suggested a common origin of legumin and vicilin genes.

Pairwise alignment of the nucleotide sequence of 869 bases of the genome walking product *BwEco*RV lib.GWP4 and of 438 bases of the genome walking product *BwDra*1 lib.GWP2 clearly identified the 438 bp sequence as a subset of the

869 bp sequence. Even though the two amplicons belong to two different libraries, the primers used for amplification were same in both the cases. Further, the restriction site for *Dra*I is closer to the position of the primers than that of *Eco*RV, thus amplifying a shorter fragment from *Dra*I library than the *Eco*RV library.

In order to exclude the adapter primer sequences from *BwEco*RV lib.GWP4 sequence and to include the coding region of the gene along with the promoter, a forward primer "SS15F" was designed from the 5' end of nucleotide sequence of *BwEco*RV lib.GWP4 and was used in combination with two reverse primers *viz.* SS13R and SS7R to amplify the target DNA from buckwheat genomic DNA. Amplified products of molecular mass 1.0 kb and 1.2 kb were obtained by PCR of buckwheat genomic DNA with primer pairs SS15F-SS13R and SS15F-SS7R. While the amplicon generated with primer pair SS15F-SS13R gave a nucleotide sequence of 1028 bases, that generated with primer pair SS15F-SS7R gave a sequence of 1166 bases. Pairwise alignment of the 1028 bp and 1166 bp nucleotide sequences using CLUSTALW revealed 100% homology for the aligned region between the two sequences. The nucleotide sequence has been deposited in the gene bank data base with accession no. EU595873. BLASTn analysis of the sequence EU595873 showed considerable homology between the 3' region of the query sequences and the 5' region of buckwheat legumin gene sequences D87980, D87982 in the databank. The tool also identified the sequences as the 5' upstream regulatory region of legumin SSP genes. GENSCAN 1.0 identified the 1028 bp sequence (acc. no. EU595873) as a SSP gene promoter.

The ORF Finder tool of NCBI identified 8 openreading frames in the 1028 bp promoter sequence. The predicted amino acid sequence of the ORF starting at position 774 on BLASTp showed alignment to the 13S globulin SSPs bearing accession nos. O23878, O23880 and ACC26039. Further, the initiating methionine of the predicted ORF coincided with the first methionine of all the three subject sequences. These results indicate that the ATG_{774} of the ORF defined by the ORF Finder tool (NCBI) could be the initiating codon of the gene under study.

It is known that the efficiency of ATG codon recognition is modulated by the context sequence of the codon. The context sequence of ATG_{774} (TCCACCATGTCA) matches the optimal context sequence CCACCATG(G) derived by Kozak (1986, 1984). The YR rule, i.e., pyrimidine (C/T) at position -1 and purine (A/G) at position +1 of the TSS has been suggested by Yamamoto *et al.* (2007) and was suggested to be conserved in *Arabidopsis* and rice genes. The TSS at position 773 also follows the YR rule, having pyrimidine 'C' at -1 and purine 'A' at +1 position ($C^{-1}A^{+1}$).

Promoter prediction tool (Neural Network Promoter Prediction:NNPP) identified three probable promoter regions between position 364-414, 445- 495 and 693-743 in the nucleotide sequence of the amplified putative promoter EU595873. The TSS for the three predicted promoters are 'A_{404}' for the promoter at position 364-414, 'A_{485}' for promoter at position 445-495 and the 'A_{773}' for the promoter at position 693-743; the TSS at position 773 being located closest to the predicted ATG start codon. Yamamoto *et al.* (2007) have highlighted the significance of "YR" rule which defined the presence of a pyrimidine (C/T) at position -1 and a purine (A/G) at

Table. 5.6: Alignment of the context sequence around TATA, TSS and ATG-start codon of buckwheat seed storage protein gene with the corresponding regions of seed storage protein genes from dicot and monocot plants. (after Joshi, 1987).

A	B	C	D	E	F	G	H	
1.	Kidney bean	Phaseolin	CTCT[TATA]ATA	43	TTC[TCA]TCA	77	TCTACTATGATG	Dicots
2.	Pea	legumin(leg A)	TCTC[TATA]TTA	33	CGC[TCA]TCA	33	CTCTTCATGGCT	
3.	Pea	legumin(leg B)	TCTC[TATA]TTA	33	CGC[TCA]TCA	33	CTCTTCATGGCT	
4.	Pea	legumin(leg C)	TCTC[TATA]TTA	33	CGC[TCA]TCA	33	CTCTTCATGGCT	
5.	Vicia faba	legumin (LeB4)	TTCC[TCA]TCA	32	TTC[CCA]CCA	56	GTCACAATGTCC	
6.	Soybean	lectin (Le1)	CTAG[TAG]TAG	27	TGC[TAC]TAC	30	AAAGCAATGGCT	
7.	Kidney bean	lectin (pPVL134)	GTTG[TAG]TAG	29	AAT[CAT]CAT	10	GAATGCATGATC	
8.	Castor bean	lectin (ricin)	TCTG[TTT]TTT	48	GAC[GCC]GCC	34	TCAAGGATGAAA	
9.	Kidney bean	lectin (Lec1)	GTTG[TAG]TAG	31	TGC[TGA]TGA	12	GCATACATGGCT	
10.	Kidney bean	lectin (Lec2)	TCTC[TAG]TAG	33	TGC[TGA]TGA	12	GCATACATGGCA	
11.	Buckwheat	legumin-like	TCAG[ATC]ATC	30	TTC[ATC]ATC	41	TCCACCATGTCA	
12.	Wheat	gliadin (pW8233)	CAAC[TAG]TAG	27	ATG[TCC]TCC	77	TCCACCATGAAG	Monocots
13.	Wheat	gliadin (∞β)	GAGC[AAG]AAG	37	CTC[CCC]CCC	67	TCCACCATGAAG	
14.	Wheat	gliadin (γ)	TAGC[AAG]AAG	29	TCC[TAC]TAC	76	TCCACCATGAAG	
15.	Wheat	gliadin (yam-2)	CAAC[TAG]TAG	27	ATC[TCC]TCC	77	TCCACCATGAAG	
16.	Wheat	glutenin (pHSB-26)	TCCC[AGC]AGC	30	CTC[TCA]TCA	61	ATCGAAATGGCT	
17.	Wheat	glutenin (λc11)	TCCC[AGC]AGC	30	TTC[TCA]TCA	62	ACCGAGATGGCT	
18.	Maize	glutenin	AAGC[TAA]TAA	34	TCC[TCA]TCA	70	GACACCATGAGG	
19.	Barley	hordein (B1)	CTAC[TAG]TAG	27	ATC[TCA]TCA	52	TCCACCATGAAG	
20.	Maize	Zein (Z4)	TGTG[TAT]TAT	29	AAT[TAT]TAT	59	CCAATAATGGCA	
21.	Maize	Zein (19kD)	TGTG[TAT]TAT	28	CTA[TAT]TAT	61	CCAATAATGGCA	
22.	Maize	Zein (ZA1)	CCTA[TAG]TAG	31	TCC[TCA]TCA	68	ACAACAATGGCT	
23.	Maize	Zein (pML1)	AAAA[TGA]TGA	63	ATC[CCT]CCT	65	ACAACAATGGCT	

A. Serial Numbers
B. Plant
C. Gene
D. Flanking region of TATA sequence (Highlighted), 4 bases on the left and 5 bases on the right.
E. Distance between first 'T' of TATA sequence and nucleotide prior to TSS
F. TSS (highlighted)flank by 3 bases on each side
G. Distance between TSS and nucleotide prior to 'A' of ATG
H. ATG, proposed or observed translation start site (highlighted) flanked by 6 bases on left and 3 bases on the right.

position +1 of the TSS in the prediction of TSS in nucleotide sequences of genes. The TSS at position 773 identified in the present study has the $C^{-1}A^{+1}$ organization, thereby following the "YR" rule. Considering 'A_{773}' as the predicted TSS and ATG_{774}(+41w.r.t TSS) as the initiating codon, the TATA at position 703 (-30) would be the TATA box of the promoter under investigation. The TATA box in majority of plant promoters lies 25-30 bases upstream of TSS (Joshi, 1987).

Apart from TATA box, the 1028 bp sequence has several other *cis*-elements, that are involved in the regulation of eukaryotic gene in general and seed-specific expression in particular including 'TGTAAAG' (-131), CATGCA (-470, -95, -68) AACAAAC (-543), CAAT (-691, -530, -457, -411, -282, -168, -54), CACA (-639, -435, -146, -60) and CANNTG (-581, -524, -184, -135, -91). The CATGCA element with the core sequence "CATG" is conserved in seed-specific promoters of both dicots and monocots. CATG, popularly known as RY element, is considered to be the key element in regulating seed specific expression of genes (Dickinson *et al.*, 1988; Chamberland *et al.*, 1992; Baumlein *et al.*, 1992; Fujiwara *et al.*, 1994; Ellerstrom *et al.*, 1996; Stalberg *et al.*, 1993). Deletion of RY element in legumin gene of *Vicia faba* abolished most of the seed-specific promoter activity and resulted in low level expression in leaves (Baumlein *et al.*, 1992). Ezcurra *et al.* (1999) and Reidt *et al.* (2000) have shown that modification of the two RY elements present in the promoter of napin gene caused a drastic reduction in the seed specific expression of napin gene. Analysis of several other seed-specific promoters has confirmed the importance of the RY elements for quantitative expression of seed-specific genes as well as the potential of this motif in repression of expression in non-seed tissues

(Monke *et al.*, 2004). Analysis of the nucleotide sequence of the 1028 bp promoter EU595873 revealed the presence of seven CAAT motifs in the sequence. CAAT box has also been found in legumin genes from *Pisum sativum* (Shirsat *et. al.*, 1989). While CAAT box is a general *cis*-element found in many eukaryotic gene promoters the sequence "TGTAAAG" (P-box) is a core -300 element found in SSP genes of cereals and several other dicot seed protein genes (reviewed by Mortin and Quiggin, 1995; Vickers, 2004). However, the "P-box" was detected as a -131 element in the 1028 bp promoter identified in the present investigation. The P-box has been suggested to be an enhancer element involved in quantitave regulation of gene expression (Pettersson and Schaffner, 1987; Schirm et al., 1987; Wu *et al.*, 2000). Analysis of the 1028 bp promoter isolated in the present study revealed the presence of 4 CACA motifs in the sequence. CACA motif have also been reported in SSP gene promoters of many dicots and monocots (Chandrasekharan *et al.*, 2003; Doyle et al., 1986; Entwistle *et al.*, 1991; Ericson *et al.*, 1991, Goldberg *et al.*, 1986; Harada *et al.*, 1989; Reeves and Okita., 1987; Yoshino *et al.*, 2001). This motif has been shown to be important for tissue specificity of *Brassica napus napA* SSP gene promoter. While mutation in β-phaseolin promoter on one hand reduced the overall promoter activity it also induced expression of the reporter gene in the radicle (Chandrasekharan *et al.*, 2003). The CANNTG motif or MYC consensus box is found in the promoter of *Arabidopsis* dehydration responsive gene *rd*22 and is the recognition site for AtMYC2 transcription factor, which is involved in ABA signalling (Abe *et al.*, 2003). Deletion of this element together with an overlapping

ABRE abolishes expression of *Brassica napus* napin SSP gene promoter (Stalberg *et al.*, 1996).

Sequences important in gene expression are likely to be conserved among group of genes having the same pattern of expression (Davidson *et al.*, 1983). Therefore, seed storage protein gene promoters of diverse plants were aligned and scanned for the presence of short conserved sequences. MEME identified six conserved elements among the 23 promoters aligned, of which only four was found in the promoter of the present investigation. One of the universal motifs identified in all the sequences is GGTGT. However, there has been no report of this motif as a regulatory element so far. Another motif referred to as the "Y-patch" was detected in 21 of 23 promoters sequences considered for analysis. Molina and Grotewold (2005) reported the presence of Y-patch in the TSS region of *Arabidopsis* core promoters. The GC-skew in the vicinity of transcription start sites has been suggested to be the reflection of the presence of Y-patch (Tatarinova *et al.*, 2005; Fujimori *et al.*, 2005; Yamamoto *et al.*, 2007). Based on the distribution pattern of Y-patch and direction sensitivity of Y-patch, Yoshiharu *et al.* (2007) suggested that Y-patch has a potential to determine the direction of transcription. The GGTGT motif and the Y-patch were, however, not detected in the 1028 bp SSP promoter isolated in the present study.

PCR based unidirectional 5' progressive deletions of the 1028 bp putative promoter in the construct pJET1028 generated a ladder of deleted fragments with the primer pair SS15F- DF1 generating a DNA fragment of highest molecular mass of 800 bp and the primer pair DF7- DR1 generating a DNA fragment with the lowest molecular mass of 74 bp. Each deletion fragment was gel purified and sub-cloned in

promoterless vector pGlowTOPO vector (Invitrogen) bearing Green Fluorescent Protein (GFP) as the reporter gene. Constructs developed with each of the sub-cloned fragments were transformed into *E. coli* cells which were subsequently cultured on LB agar for 48 hours to allow expression of GFP gene in the cells. Fluorescence due to the accumulation of GFP in the cells as a consequence of reporter gene expression under the influence of the putative promoter was observed under fluorescence microscope with filter no. 9 which allowed excitation at 450-490 nm and fluorescence emission at 515 nm. Constructs with deletion fragments -731, -648 and -529 did not show any differences in the expression of GFP in *E.coli* cells. Thus deletion of -691 CAAT motif, -639 CACA motif, -581 CANNTG motif and -549 AACA motif did not appear to affect the activity of the promoter. However, construct with PCR generated deletion fragments -487, -347 and -228 showed significantly reduced activity of the promoter compared to -731, -648 and -529 constructs. Thus deletion of -530, -457, -411and -282 CAAT, and -524 CANNTG, -470 CATGCA and -435 CACA motifs caused a marked reduction in the expression of GFP in the *E.Coli* cells.

Baumlein *et al.* (1992) have suggested that in contrast to the region downstream of bp -566 the region distal of -701 in LeB4 promoter was of , is of little importance for high promoter activity. They have suggested that the proximal AT rich region up to bp -407 could have an important role in regulating the expression of genes coding for seed and other plant proteins. These sequences preferentially interact with high mobility group (HMG) proteins which seemingly recognize certain structural features instead of specific primary sequences (reviewed

by Weising and Kahl, 1991). The proximal region upto bp -529 of the 1028 bp promoter isolated in the present study has 60% AT content. Goldberg *et al.* (1989) have suggested that while the proximal elements upto -77 and -66 were essential for correct spatial and temporal expression of soybean lectin (*Le*1) and glycinin (*Gy*1) genes respectively, the more distal elements were required for their quantitative regulation. Bogue (1990), Thomas *et al.* (1991) and Nunberg *et al.* (1994) have suggested a bipartite organization of the regulatory elements in the 5' UTR of helianthinin gene with the proximal region (-116 to +24) involved in conferring seed specific expression of the gene and the more distal regions involved in refining and enhancement of the basic expression patterns conferred by the proximal region.

In the present study deletion of -470 CATGC element from the 1028 bp promoter caused a significant reduction in reporter gene expression. CATGC, the core motif of legumin box has been implicated as a key *cis* acting element for seed specific gene expression (Baumlein *et al.*, 1992; Chamberland *et al.*, 1992; Fujiwara and Beachy, 1994; Sakata *et al.*, 1997; Bobb *et al.*, 1997). Deletion of the CATGC motif within the 2.4 kb *Le*B4 upstream sequence has been reported to lead to a drastic reduction in reporter gene expression (Baumlein *et al.*, 1992). However, on the basis of their results on progressive deletions leaving the CATGCATG motif intact in the *Le*B4 promoter they have conclude that the CATGC motif was necessary but not absolutely essential for SSP for gene expression. Baumlein *et al.* (1992) have suggested that although originally described as an element specific for legume seed protein genes, the CATGCATG motif acted as a functional module in a wider range of plant promoters.

While the constructs with deletion fragments -146 showed very negligible level of reporter gene expression that with deletion fragment -24 did not show any promoter activity. This confirms the role of proximal elements in regulating the expression of SSP genes. The construct pGlowTOPO$_{-25}$ lacks the TGTAAG motif which is also known as P-box as well as the TATA box. The prolamin box/ endosperm motif as well as the GCN4 and AACA motifs have been demonstrated to be essential for the regulation of expression of endosperm-specific genes (Zheng *et al.*, 1993; Takaiwa *et al.*, 1996; Yoshihara *et al.*, 1996; Mena *et al.*, 1998; Diaz *et al.*, 2005).

CHAPTER VI
GENERAL SUMMARY & DISCUSSION

INTRODUCTION:

Plant genetic resources, representing the sum total of diversity accumulated through years of cultivation under domestication and natural selection, are considered as one of the most important gifts of nature to mankind. Many of these plants are important sources of high nutritive value foods for human consumption. Out of the total genetic diversity available, grain crops constitute one of the major sources of food for human nutrition. However, of the total available genetic diversity, mankind has utilized only a few plants as major food sources.

Seed storage proteins have attracted the attention of researchers mainly on account of their importance in human nutrition. Seed storage proteins, intended as a source of nitrogen during the initial stages of germination and seedling growth, constitute an important source of dietary proteins for human consumption. Although

cereal grains and legume seeds are the major sources of vegetarian dietary proteins for human consumption, the nutritional quality of the proteins in both does not match the WHO standards for dietary proteins with a nutritionally balanced amino acid composition. While the major amino acid deficiency in legume seed proteins is their low content of sulphur containing amino acids, cereal proteins are deficient in lysine (Boulter, 1981; Shotwell and Larkins, 1989). Rout and Chrungoo (1996) have compared the amino acid composition of seed/grain storage proteins of some of the conventional crops with the WHO recommended values for a nutritionally balanced protein. They have emphasized the nutraceutical importance of some underutilized crops like *Chenopodium*, grain amaranth and buckwheat.

Over the years, many attempts have been made to improve the level of essential amino acids in seed storage proteins of important crop plants through conventional breeding programmes (Larkins, 1983; Coulter and Bewely, 1990). However, in most cases the attempts have either led to a severe depletion in seed storage protein levels or abnormalities in seed development (Bliss *et al.*, 1972). A variety of barley Riso1508 produced in this way had very high lysine content but a severe depletion of the storage protein, hordein (Hermann and Larkins, 1991). The negative correlation between the seed protein content and the level of essential amino acids per unit protein has come as a major handicap in improving the amino acid composition of seed proteins in crops.

Because of inherent limitations in inter-specific hybridizations, molecular approaches have provided alternative strategies to conventional breeding programmes. One of the approaches in this direction has been to manipulate the

regulation of amino acid biosynthesis to increase the abundance of a particular amino acid. Mutant selection and engineering of genes encoding key enzymes of amino acid biosynthetic pathways have been used to increase amino acids in crop plants (Zeh *et al.*, 2001; Gidamis *et al.*, 1995; Salbaach *et al.*, 1995a;). However, an increase in the free essential amino acids may not necessarily lead to an increase in the content of fixed amino acids; the amino acids could be leached out from the plant tissue and lost during boiling and other processing (Falco *et al.*, 1995). Alternatives to the manipulation of regulation of amino acid biosynthesis have focused around either

(i) the manipulation of the primary sequence of a gene by addition, substitution or deletion of nucleotides by site directed mutagenesis and expression of the altered gene in place of or in addition to the native gene (Lago *et al.*, 1990; Guerche *et al.*, 1990; Blechl and Anderson, 1996) or

(ii) heterologous gene transfer across species barriers (Altenbach *et al.*, 1989, 1992; Saalbach *et al.*, 1995b; Muntz *et al.*, 1997; Molvig *et al.*, 1997; Townsend and Thomas, 1994.). Efficient manipulation of the amino acid composition of seed storage proteins also relies to a large extent, on the stability of the foreign proteins in a heterologous system. Stable expression of the altered/heterologous protein in the seed can be a difficult task owing to the complex biochemical processes associated with seed storage protein assembly and accumulation (Lambert and Yarwood, 1992; Lending *et al.*, 1992).

Crop improvement by heterologous gene transfer essentially requires a promoter that would regulate the temporal and spatial expression of the trangenes. As a first step, potentially useful promoters need to be evaluated in view of their

developmental stage-specificity, seed-specificity and expression levels. The study and increased understanding of gene promoters including their structure, function and mechanism of gene regulation will open up the possibility of modulation of gene expression in homologous as well as heterologous systems. While many seed specific promoters have been identified and characterized not many promoters from genes of seed storage proteins have been characterized till date. Examples of such promoters include the 5' regulatory regions from such genes as cruciferin (Sjödahl *et al.*, 1995), napin (Kridl *et al.*, 1991), phaseolin (Butos *et al.*, 1989), soybean beta-conglycinin (Lessard *et al.*, 2004; Chen *et al.*, 1986), buckwheat globulin (Millsavljevič *et al.*, 2004), maize zein (Matzke *et al.*, 1990; Thompson *et al.*, 1990; Brown *et al.*, 1986), rice glutelin (Takaiwa *et al.*, 1991; Qu *et al.*, 2008), pea legumin (Lycett *et al.*, 1985), sunflower helianthenin (Nunberg *et al.*, 1995) etc.

Successful introgression of the target genes into crop plants requires the development of constructs carrying the target gene with an appropriate tissue specific promoter. Lack of suitable gene promoters for driving expression of the heterologous genes in transgenes in the endosperm is still a major limitation in obtaining the required level and pattern of expression. This emphasizes the need for concerted efforts to isolate genes and their promoters from indigenous crop plants so that transgenic development process is not hampered under the IPR regimes. The present investigation was therefore undertaken to identify and characterize the endosperm/seed-specific promoter region of buckwheat legumin-like protein gene. Such endosperm/seed specific promoter would find application in:

1. Transgenic programmes aimed at improvement of the nutritional quality of conventional crops deficient in essential amino acids.
2. Elucidation of regulatory mechanisms controlling temporal as well as tissue-specific gene expression.

RESULTS:

Seed storage proteins constitute an important component of seed reserves and serve as a source of nitrogen during the initial stages of growth of the seedlings. Besides acting to provide soluble nitrogenous compounds during the initial stages of seedling growth, seed storage proteins also constitute an important source of dietary proteins for human consumption. Cereal grains and legume seeds are the two major sources of vegetarian dietary proteins for human consumption. However, the nutritional quality of the proteins in both these crops does not match the WHO standards of a protein with nutritionally balanced amino acid composition. 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 and Larkins, 1989).

The main storage protein in grains of common buckwheat is a 13S globulin with molecular mass of 280 kDa. The protein comprises more than 50% of the total protein content of grains (Rout, 1996). The amino acid composition of the protein matches the WHO recommended values for a nutritionally rich protein with a balanced amino acid composition (Rout and Chrungoo, 1996). The nutritionally rich component of protein is a 26kDa basic subunit, which has more than 6% lysine and nearly 2% methionine (Rout and Chrungoo, 1996, Rout *et al.,* 1997). Due to the balanced amino acid composition, high nutrient value and homology with seed storage proteins of leguminous group of plants, this protein could be an important

candidate for compensation of limiting amino acid in plants deficient in such amino acids, through heterologous gene transfer across species. This requires isolation and cloning of full length genes and their promoters for use in crop improvement programmes aimed at development of transgenics with improved nutritional quality of seed storage proteins.

One of the biggest obstacles in isolation high quality DNA from starch/protein rich plant tissues is the interferences in purification procedures due to presence of carbohydrates, proteins and even polyphenolic compounds. The modified CTAB protocol used in the present study yielded good quality genomic DNA. Analysis of the isolated genomic DNA on 0.8% agarose gel revealed that the sample was free of RNA contamination and has not undergone any shearing during isolation. The A260/280 ratio of >1.8 indicated the sample to be quite pure.

The isolated genomic DNA was digested with *Eco*RI, *Hind*III and *Nco*I to check the extent of digestion of the isolated DNA. The extent of digestion would be an indicator of the quality of isolated DNA. The *Eco*RI digested resolved as a uniform streak ranging in size from 21.0 to 0.5 kb. The smear also showed four bands having molecular masses of 0.904 kb, 0.832 kb, 0.625 kb and 0.5 kb. The appearance of distinct bands in *Eco*RI digested DNA indicates the presence of *Eco*RI repeats of varying lengths in buckwheat genomic DNA. While the *Hind*III and *Nco*I digested DNA resolved as uniform smears ranging from in molecular mass from 0.3 kb to 20 kb, no distinct bands were visible in the electrophoretic profile of either *Hind*III or *Nco*I digested DNA. The absence of any distinct bands in *Hind*III and *Nco*I digested DNA indicates the absence of any repeats corresponding to either

*Hind*III or *Nco*I in the buckwheat genomic DNA. Similar results have been observed by Bharali (2002) on the restriction digestion profile of buckwheat genomic DNA. Bharali (2002) has reported 6 bands representing *Eco*R1 repeats in genomic DNA of common buckwheat.

Even though isolation of genes from genomic DNA libraries by screening of the library with appropriate probes offers a reliable system for cloning full length genes, PCR amplification using oligonucleotide primers designed from the nucleotide sequences of conserved regions of 5'upstream and 3'downstream regions of the genes offers a much quicker way of isolation of full length genes directly from genomic DNA isolated from the target plants. In the present study polymerase chain reaction with buckwheat genomic DNA as the template and combination of forward primer having the sequence 5' ATGCTTCATGGGGTGCTTCTATGC 3' and reverse primer having the sequence 5' TTAAGACCTTCCTCCGAAAG 3' amplified a 835 bp DNA fragment which showed sequence homology with mitochondrial NADH dehydrogenase gene of diverse groups of plants ranging from bryophytes to angiosperms. CLUSTAL W alignment of nucleotide sequence of the 835 bp amplicon with nucleotide sequences mitochondrial NADH dehydrogenase gene of bryophytes, pteridophytes, gymnosperms and angiosperms revealed a high degree of sequence conservation for the gene across these groups of plants. The amplification of NADH dehydrogenase gene by the primers designed for SSP genes could be due to the presence of short sequences homologous to the designed primers in the mitochondrial NADH dehydrogenase. The most distinguishing feature of the sequence amplified in the present study was a 6 base "CCTCCC"

insertion/substitution at position 203. An important difference between in the sequences of mitochondrial dehydrogenases of angiosperms and gymnosperms, observed in the alignment, was the addition of 4 bases each *viz.* "CGCG" and "CTTT" respectively at positions 55 and 231 corresponding to buckwheat mitochondrial dehydrogenase (DQ852632) in the mitochondrial NADH dehydrogenase of angiosperms compared to that of the gymnosperms. The nucleotide sequences of NADH dehydrogenases of pteridophytes and bryophytes showed a deletion of 3 bases *viz.* "AGT" at position 122 corresponding to the buckwheat mitochondrial NADH dehydrogenase.

Polymerase chain reaction with buckwheat genomic DNA as the template and primer combinations SS8F-SS3R, SS8F-SS1R, SS4F-SS3R and SS2F-SS3R amplified DNA fragments showing apparent molecular mass of 1.6 kb, 1.1 kb, 0.85 kb and 1.3 kb respectively. Southern hybridization of the of the 1.1 kb, 1.6 kb and 1.3 kb amplicons with [α-^{32}P]-dATP labelled 830 bp amplicon clearly established a high degree of sequence homology between the amplicons. The nucleotide sequence of 1613 (acc. no. GQ358523) bases for the 1.6 kb DNA fragment, the longest fragment amplified by primer pair SS8F-SS3R, showed 94% and 96% sequence homology respectively with buckwheat Feleg51 13S globulin gene (acc. no. AY359286) and 13S globulin mRNA sequences (acc. no. AY256960). A statistical evaluation of the alignments revealed the sequence homologies were highly significant. Some of the distinguishing features observed in the nucleotide sequence of the 1613bp amplicon included an 8 base deletion at position 914 and a 6 base "ACCGCT" insertion/substitution at position 927. This distinguished the nucleotide

sequence of the 1613bp DNA fragment amplified in the present study from nucleotide sequences of other globulin genes available in the data bases.

GENSCAN 1.0 detected three exons in the nucleotide sequence of the 1613bp DNA fragment amplified in the present study. While the 210 bp exon at position between 34-343 was identified as the initial exon, the other two exons were identified as being internal to the sequence. Legumin genes, coding for the most widely distributed group of seed storage proteins show a fairly uniform intron-exon organization having three introns which have been reported to be strictly conserved in position (Shewry, 1995). However, the coexistence of two- and three- intron structure has led to the assumption that legumin genes have evolved by addition (Shotwell and Larkins, 1989), or conversely, by loss of introns (Vonder *et al.*, 1988). The latter scenario gained support from the recently described exceptional five exon/four intron legumin genes from *Ginkgo biloba* which may be viewed as more ancient states of legumin gene structure (Hager *et al.*, 1995). The *Ginkgo biloba* legumin genes have introns I to III located at exactly the same positions as is known for legumin genes of angiosperms and an additional intron downstream from intron III which interrupts the sequence coding for the β-polypeptide C-terminal region. The legumin gene from *Welwitschia* (EMBL 250780) also has been reported to have an intron matching the *Ginkgo* legumin gene intron IV. This may indicate intron IV as being ancestral. The recognition of homology in intron/exon patterns is of importance in elucidation of structural similarities between legumin- and vicilin-type storage proteins. On the basis of their observations on intron-exon architecture from most of the known legumin genes, (Hager *et al.*, 1996) have suggested that the

evolution of legumin genes has proceeded with loss of introns during the course of evolution. Our result on the intron-exon organization is in conformity with the observations of (Shewry, 1995).

The deduced amino acid sequence of the 1613 bp amplicon (acc. no. GQ358523) comprised of 245 amino acids with a theoretical molecular mass of 27 kDa. Sequence homology analysis of the 245 residue amino acid sequence by BLASTp against non-redundant protein database clearly identified the protein as belonging to the legumin subfamily. Using an alignment that permitted maximum homology, the deduced amino acid sequence showed a maximum of 70% homology with 13S buckwheat globulin (acc. no. AAP15457). The percentage of homology varied between 54-40% with deduced amino acid sequences of legumins from other crops including *Perilla frutescens* (acc. no. AAF19607) and *Magnolia salicifolia* (acc. no. CAA57848). Further, identification of two cupin domains in the deduced amino acid sequence indicated that the protein belongs to the cupin superfamily. The cupins are a large superfamily of proteins, including seed storage proteins, which are thought to have originated by divergent evolution from a common ancestor. They share a common hexameric architecture, which has been described as `double-stranded β-helix' with each helix having a conserved barrel domain.

The phylogenetic tree constructed from the alignment data of the deduced amino acid sequence with amino acid sequences of SSPs available in EMBL database showed a clear division into legumin type proteins from angiosperms and gymnosperms. The sequences of two legumin type seed storage proteins of buckwheat emerged as a separate group within the tree. Our results on phylogenetic

relationships amongst legumin genes derived by neighbour joining distance method are in conformity with the observations of Hager *et al.* (1996).

Crop improvement by heterologous gene transfer essentially requires a promoter that would regulate the spatial and temporal expression of the trangenes. While many promoters have been identified and characterized not many seed storage protein gene promoters have been characterized till date. The primary regulatory sequences of SSP genes are generally located within 1000bp upstream of the transcription start site (TSS), although there are cases where regulatory sequences are found further upstream (Zhang *et al.*, 1996) or downstream (3') (Dietrich *et al.*, 1992) of the coding sequences.

The gene specific primers used with adapter primers for genome walking were designed from the nucleotide sequence of the 5' upstream as well as the 3'downstream region of the target gene so as to walk along the library in 5' as well as 3' direction. The gene specific primers (GSPs) designed for genome walking were 26-30 bases in length with GC content ranging from 40-60%. To increase the stringency of the PCR reaction, the annealing temperature for all the genome walking PCR reactions carried out in the present investigation was 68°C. In the present investigation, the "Touchdown" PCR protocol (Don *et al*, 1991; Roux, 1995; Hecker and Roux, 1996) in which the annealing/extension temperature is several degrees higher than the T_m of the primers during the initial PCR cycles, thereby permitting efficient, exponential amplification of the gene-specific product, has been adopted. The higher temperature also enhances the suppression PCR effect with AP1, allowing a critical amount of gene-specific product to accumulate. Siebert *et al*

(1995) have considered suppression PCR to be most critical for the success of genome walking.

BLASTn analysis of the nucleotide sequence of amplified genome walking products *BwDra*I lib.GWP1 and *BwEcoR*V lib.GWP3 revealed significant similarity with the nucleotide sequence of the cDNA clones D87890 and D87982. The results of BLAST did not reveal any homology with 5' upstream regions of any SSP gene. BLASTn of the nucleotide sequence of amplified genome walking products *Bw EcoR*V lib.GWP3 clearly identified it with the legumin gene family. Alignment of the 1127 bp nucleotide sequence of the 1.1 kb amplicon *BwEcoR*V lib.GWP3 revealed >85% sequence homology with buckwheat legumin-like cDNAs AY256960 and D87980. The sequence also showed high degree of similarity with legumin genes of other plants like *Pisum*, *Vicia faba*, soybean, *Pistacia, Carya*, *Raphanus*, *Gingko biloba*, *Populus* and brazil nut. These results clearly establish a high degree of sequence conservation amongst legumin type SSP genes of various groups of plants. In comparison to other buckwheat legumin gene sequences, the nucleotide sequence of the 1.1 kb DNA amplified in the present study deletion of 4 bases "GCCG" at position 57 and 3 bases "CTC" at position 670. The distinguishing feature of buckwheat legumin gene from that of legumin genes in other plants is an insertion of 5 bases "CTCCT" at position 324.

GENSCAN 1.0 detected three internal exons in the 1127 bp nucleotide sequence of the amplification product *BwEcoR*Vlib.GWP3. The graphic representation of results of BLASTn clearly established the conserved nature of intron-exon architecture in buckwheat legumion SSP genes. The deduced amino acid

sequences of the 1.1 kb amplicon showed 78% homology with buckwheat legumin-like protein (acc. no. BAA21758). The percentage of homology varied between 40-69% with deduced amino acid sequences of legumin genes from other crops including *Magnolia salicifolia* legumin precursor (acc. no. CAA57846) and *Coffea arabica* 11S globulin (acc. no. AAC61881). The phylogenetic tree constructed from the alignment data of the deduced amino acid sequence with amino acid sequences of SSPs available in EMBL database revealed clear clustering into three groups, one each for sequences belonging to monocots, dicots and gymnosperms. The sequences from dicots clustered into one group along with the sequences of buckwheat SSPs. These results are in conformity with the observations of Hager *et al.* (1995, 1996). Hager *et al.* (1996) have suggested that the legumin gene evolution has progressed towards loss of introns. They have also suggested a common origin of legumin and vicilin genes.

Pairwise alignment of the nucleotide sequence for 869 bases of the genome walking product *BwEco*RV lib.GWP4 and for 438 bases of the genome walking product *BwDra*I lib.GWP2 clearly identified the 438 bp sequence as a subset of the 869 bp sequence. Even though the two amplicons belong to two different libraries the primers used for amplification were same in both the cases. Further the restriction site for *Dra*I is closer to the position of the primers than that of *Eco*RV thus amplifying a shorter fragment from *Dra*I library than the *Eco*RV library.

In order to exclude the adapter primer sequences from *BwEco*RV lib.GWP4 sequence and to include the coding region of the gene along with the promoter, a forward primer "SS15F" was designed from the 5' end of nucleotide sequence of

*BwEco*RV lib.GWP4 and was used in combination with two reverse primers *viz.* SS13R and SS7R to amplify the target DNA from buckwheat genomic DNA. Amplified products of molecular mass 1.0 kb and 1.2 kb were obtained by PCR of buckwheat genomic DNA with primer pairs SS15F-SS13R and SS15F-SS7R. While the amplicon generated with primer pair SS15F-SS13R gave a nucleotide sequence of 1028 bases that generated with primer pair SS15F-SS7R gave a sequence of 1166 bases. Pairwise alignment of the 1028 bp and 1166 bp nucleotide sequences using CLUSTALW revealed 100% homology for the aligned region between the two sequences. The nucleotide sequence has been deposited in the gene bank data base with accession no. EU595873. BLASTn analysis of the sequence EU595873 showed considerable homology between the 3' region of the query sequences and the 5' region of buckwheat legumin gene sequences in the databank. The tool also identified the sequences as the 5' upstream regulatory region of legumin SSP genes. GENSCAN 1.0 identified the 1028 bp sequence (acc. no. EU595873) the promoter region.

The ORF Finder tool of NCBI identified 8 open reading frames in the 1028 bp promoter sequence. The predicted amino acid sequence of the ORF starting at position 774 on BLASTp showed alignment to the 13S globulin SSP 1(acc. no. O23878), globulin SSP 2 (acc. no. O23880), and 13S globulin SSP 3 (acc. no. Q9XFM4) of buckwheat. Further, the initiating methionine of the predicted ORF coincided with the first methionine of all the three subject sequences. These results indicate that the ATG_{774} of the ORF defined by the ORF Finder tool (NCBI) could be the initiating codon of the gene under study. The context sequence of ATG_{774}

(TCCACCATGTCA) matches the optimal context sequence CCACCATG(G) derived by Kozak (1986, 1984). The YR rule, i.e., pyrimidine (C/T) at position -1 and purine (A/G) at position +1 of the TSS has been suggested by Yamamoto *et al.* (2007) and was suggested to be conserved in *Arabidopsis* and rice genes. The TSS at position 773 also follows the YR rule, having pyrimidine 'C' at -1 and purine 'A' at +1 position ($C^{-1}A^{+1}$).

Promoter prediction tool (Neural Network Promoter Prediction:NNPP) identified three probable promoter regions between position 364- 414, 445- 495 and 693-743 in the nucleotide sequence of the amplified putative promoter EU595873. The TSS for the three predicted promoters are 'A_{404}' for the promoter at position 364-414, 'A_{485}' for promoter at position 445-495 and the 'A_{773}' for the promoter at position 693-743; the TSS at position 773 being located closest to the predicted ATG start codon. Yamamoto *et al.* (2007) have highlighted the significance of "YR" rule which defined the presence of a pyrimidine (C/T) at position -1 and a purine (A/G) at position +1 of the TSS in the prediction of TSS in nucleotide sequences of genes. The TSS at position 773 identified in the present study has the $C^{-1}A^{+1}$ organization, thereby following the "YR" rule. Considering 'A_{773}' as the predicted TSS and ATG_{774} (+41) as the initiating codon, the TATA at position 703 (-30) would be the TATA box of the promoter under investigation. The TATA box in majority of the plant promoters lies 25-30 bases upstream of TSS (Joshi, 1987).

Apart from TATA box, 1028 bp sequence has several other *cis*-elements, that are involved in the regulation of eukaryotic gene in general and seed-specific expression in particular including 'TGTAAAG' (-131), CATGCA (-470, -95, -68)

AACAAAC (-543), CAAT (-691, -530, -457, -411, -282, -168, -54), CACA (-639, -435, -146, -60) and CANNTG (-581, -524, -184, -135, -91). The CATGCA element with the core sequence "CATG" is conserved in seed-specific promoters of both dicots and monocots. CATG, popularly known as RY element, is considered to be the key element in regulating seed specific expression of genes (Dickinson *et al.*, 1988; Chamberland *et al.*, 1992; Baumlein *et al.*, 1992; Fujiwara *et al.*, 1994; Ellerstrom *et al.*, 1996; Stalberg *et al.*, 1993). Deletion of RY element in legumin gene of *Vicia faba* abolished most of the seed-specific promoter activity and resulted in low level expression in leaves (Baumlein *et al.*, 1992). Ezcurra *et al.* (1999) and Reidt *et al.* (2000) have shown that modification of the two RY elements present in the promoter of napin gene caused a drastic reduction in the seed specific expression of napin gene. Analysis of several other seed-specific promoters has confirmed the importance of the RY elements for quantitative expression of seed-specific genes as well as the potential of this motif in repression of expression in non-seed tissues (Monke *et al.*, 2004). Analysis of the nucleotide sequence of the 1028 bp promoter EU595873 revealed the presence of seven CAAT motifs in the sequence. CAAT box has also been found in legumin genes from *Pisum sativum* (Shirsat *et. al.*, 1989). While CAAT box is a general *cis*-element found in many eukaryotic gene promoters the sequence "TGTAAAG" (P-box) is a core -300 element found in SSP genes of cereals and several other dicot seed protein genes (reviewed by Mortin and Quiggin, 1995; Vickers, 2004). However, in the present investigation the "P-box" was detected as a -131 element in the 1028 bp putative promoter. The P-box has been suggested to be an enhancer element involved in quantitative regulation of gene

expression (Pettersson and Schaffner, 1987; Schirm et al., 1987; Wu *et al.*, 2000). Analysis of the 1028 bp promoter isolated in the present study revealed the presence of 4 CACA motifs in the sequence. CACA motif have also been reported in SSP gene promoters of many dicots and monocots (Chandrasekharan *et al.*, 2003; Doyle et al., 1986; Entwistle *et al.*, 1991; Ericson *et al.*, 1991, Goldberg *et al.*, 1986; Harada *et al.*, 1989; Reeves and Okita., 1987; Yoshino *et al.*, 2001). This motif has been shown to be important for tissue specificity of *Brassica napus napA* SSP gene promoter. While mutation in β-phaseolin promoter on one hand reduced the overall promoter activity it also induced expression of the reporter gene in the radicle (Chandrasekharan *et al.*, 2003). The CANNTG motif or MYC consensus box is found in the promoter of *Arabidopsis* dehydration responsive gene *rd22* and is the recognition site for AtMYC2 transcription factor, which is involved in ABA signalling (Abe *et al.*, 2003). Deletion of this element together with an overlapping ABRE abolishes expression of *Brassica napus* napin SSP gene promoter (Stalberg *et al.*, 1996).

Sequences important in gene expression are likely to be conserved among group of genes having the same pattern of expression (Davidson *et al.*, 1983). Therefore, seed storage protein gene promoters of diverse plants were aligned and scanned for the presence of short conserved sequences. MEME identified six conserved elements among the 23 promoters aligned, of which only four were found in the promoter amplified in the present investigation. One of the universal motifs identified in all the sequences is GGTGT. However, there has been no report of this motif as a regulatory element so far. Another motif referred to as the "Y-patch" was

detected in 21 of 23 promoters sequences considered for analysis. Molina and Grotewold (2005) reported the presence of Y-patch in the TSS region of *Arabidopsis* core promoters. The GC-skew in the vicinity of transcription start sites has been suggested to be the reflection of the presence of Y-patch (Tatarinova *et al.*, 2005; Fujimori *et al.*, 2005; Yoshiharu *et al.*, 2007). Based on the distribution pattern of Y-patch and direction sensitivity of Y-patch, Yamamoto *et al.* (2007) suggested that Y-patch has a potential to determine the direction of transcription. The GGTGT motif and the Y-patch were, however, not detected in the 1028 bp SSP promoter isolated in the present study.

PCR based unidirectional 5' progressive deletions of the 1028 bp putative promoter in the construct pJET1028 generated a ladder of deleted fragments with the primer pair SS15F- DF1 generating a DNA fragment of highest molecular mass of 800 bp and the primer pair DF7- DR1 generating a DNA fragment with the lowest molecular mass of 74 bp. Each deletion fragment was gel purified and sub-cloned in promoterless vector pGlowTOPO vector (Invitrogen) bearing Green Fluorescent Protein (GFP) as the reporter gene. Constructs developed with each of the sub-cloned fragments were transformed into *E. coli* cells which were subsequently cultured on LB agar for 48 hours to allow expression of GFP gene in the cells. Fluorescence due to the accumulation of GFP in the cells, as a consequence of reporter gene expression under the influence of the putative promoter, was observed under a fluorescence microscope with filter no. 9 which allowed excitation at 450-490 nm and fluorescence emission at 515 nm. Constructs with deletion fragments -731, -648 and -529 did not show any differences in the expression of GFP in *E.coli*

cells. Thus, deletion of -691 CAAT motif, -639 CACA motif, -581 CANNTG motif and -549 AACA motif did not appear to affect the activity of the promoter. However, construct with PCR generated deletion fragments -487, -347 and -228 showed significantly reduced activity of the promoter compared to -731, -648 and -529 constructs. Thus deletion of -530, -457, -411and -282 CAAT, and -524 CANNTG, -470 CATGCA and -435 CACA motifs caused a marked reduction in the expression of GFP in the *E.Coli* cells.

Baumlein *et al.* (1992) have suggested that in contrast to the region downstream of bp -566 the region distal of -701 in LeB4 promoter was of little importance for high promoter activity. They have suggested that the proximal AT rich region up to bp -407 could have an important role in regulating the expression of genes coding for seed and other plant proteins. These sequences preferentially interact with high mobility group (HMG) proteins which seemingly recognize certain structural features instead of specific primary sequences (reviewed by Weising and Kahl, 1991). The proximal region upto bp -529 of the 1028 bp promoter isolated in the present study has 60% AT content. Goldberg *et al.* (1989) have suggested that while the proximal elements upto -77 and -66 were essential for correct spatial and temporal expression of soybean lectin (*Le*1) and glycinin (*Gy*1) genes respectively, the more distal elements being required only for their quantitative regulation. Bogue (1990), Thomas *et al.* (1991) and Nunberg *et al.* (1994) have suggested a bipartite organization of the regulatory elements in the 5' upstream of helianthinin gene with the proximal region (-116 to +24) involved in conferring seed specific expression of

the gene and the more distal regions involved in refining and enhancement of the basic expression patterns conferred by the proximal region.

In the present study deletion of -470 CATGC element from the 1028 bp promoter caused a significant reduction in reporter gene expression. CATGC, the core motif of legumin box has been implicated as a key *cis* acting element for seed specific gene expression (Baumlein *et al.*, 1992; Chamberland *et al.*, 1992; Fujiwara and Beachy, 1994; Sakata *et al.*, 1997; Bobb *et al.*, 1997). Deletion of the CATGC motif within the 2.4 kb *Le*B4 upstream sequence has been reported to lead to a drastic reduction in reporter gene expression (Baumlein *et al.*, 1992). However, on the basis of their results on progressive deletions leaving the CATGCATG motif intact in the *Le*B4 promoter they have conclude that the CATGC motif was necessary but not absolutely essential for SSP for gene expression. Baumlein *et al.* (1992) have suggested that although originally described as an element specific for legume seed protein genes, the CATGCATG motif acted as a functional module in a wider range of plant promoters.

While the constructs with deletion fragments -146 showed very negligible level of reporter gene expression that with deletion fragment -24 did not show any promoter activity. This confirms the role of proximal elements in regulating the expression of SSP genes. The construct pGlowTOPO$_{.25}$ lacks the TGTAAG motif which is also known as P-box as well as the TATA box. While the TATA box is essential for promoter activity, the prolamin box/ endosperm motif as well as the GCN4 and AACA motifs have been demonstrated to be essential for the regulation

of expression of endosperm-specific genes (Zheng *et al.*, 1993; Takaiwa *et al.*, 1996; Yoshihara *et al.*, 1996; Mena *et al.*, 1998; Diaz *et al.*, 2005).

A 3 kb full length buckwheat legumin gene along with its promoter was amplified by using primer SS15F as the forward primer and SS16R as the reverse. The primer SS16R has been designed from conserved 3' region of buckwheat legumin gene sequences available in the database. The 3 kb amplified fragment was hybridized with two different DNA fragments as probes. One of the probes corresponds to the promoter region, and the other probe corresponds to the coding region of buckwheat legumin gene. Positive signals were detected on the exposed X-ray films for both the probes. Thus, the 3 kb fragment amplified by the primer pair SS15F-SS16R includes both the promoter region as well as the coding region of the buckwheat legumin gene. The 3.0 kb fragment was separately digested with *Eco*RI, *Nco*I, *Eco*RV, *Hind*III and *Stu*I. The restriction digestion profile of the 3.0 kb fragment for *Eco*RI, *Nco*I, *Eco*RV, *Hind*III and *Stu*I is presented in Fig 5.27. Digestion of the 3.0 kb amplicon with *Eco*RI restriction enzyme generated two bands having molecular mass of 2.2 kb and 0.5 kb. A similar profile was obtained when the 3.0 kb fragment was digested with *Nco*I and *Eco*RV. While restriction digestion with *Nco*I released two bands showing molecular masses of 2.1kb and 0.9 kb, that with *Eco*RV generated two bands having molecular mass of 1.9 kb and 1.1 kb. Restriction digestion with *Stu*I and *Hind*III generated two bands having molecular mass of 2.4 kb and 0.6 kb for the amplicon DNA digested with *Stu*I and two bands having molecular mass of 2.6 kb and 0.4 kb for the amplicon digested with *Hind*III.

The restriction digestion profile of the 3.0 kb amplicon matched the restriction profile generated by webcutter.

CHAPTER: VII
REFERENCES

Abe, H., Urao, T., Ito, T., Seki, M., Shinozaki, K. and Yamaguchi-Shinozaki, K.(2003) Arabidopsis AtMYC2 (bHLH) and AtMYB2(MYB) function as transcriptional activators in abscisic acid signaling. Plant Cell., 15: 63-78.

Adachi, M., Kanamori, J., Masuda, T., Yagasaki, K., Kitamura, K., Mikami, B. and Utsumi, S. (2003) Crystal structure of soybean 11 S globulin: glycinin A3B4 homohexamer. Proc. Natl. Acad. Sci. U.S.A., 100: 7395–7400.

Adhya, S. (1989) Multipartite genetic control elements: communication by DNA loop; Annu. Rev. Genet. 23: 227–250.

Albani, D., Hammond-Kosack, M.C.U., Smith, C., Conlan, S., Colot, V., Holdsworth, M. and Bevan, M.W., (1997) The wheat transcriptional activator SPA: a seed-specific bZIP protein that recognizes the GCN4-like motif in the bifactorial endosperm box of prolamin genes. The Plant Cell., 9: 171–184.

Altenbach, S. B., Pearson, K. W., Meeker, G., Staraci, L. C. and Sun, S. S. M. (1989) Enhancement of the methionine content of seed proteins by the expression of a chimeric gene encoding a methionine-rich protein in transgenic plants. Pl. Mol. Biol., 13: 513-522.

Altenbach, S. B., Kuo, C-C., Staraci, L. C., Pearson, K. W., Wainwright, C., Georgescu, A. and Townsend, J. (1992) Accumulation of a Brazil nut albumin in seeds of transgenic canola results in enhanced levels of seed protein methionine. Pl. Mol. Biol., 18: 235-245.

Altenbach, S.B., Pearson, K.W., Leung, F.W. and Sun, S.S.M. (1987) Cloning and sequence analysis of a DNA encoding a Brazil nut protein exceptionally rich in methionine. Pl. Mol. Biol., 8: 239-250.

Baker, J.C., Steele, C. and Dure, L. III (1988) Sequence and characterization of 6 Lea proteins and their genes from cotton. Pl. Mol. Biol.. 11: 277-291.

Bartels, D. and Thompson, R.D. (1986) Synthesis of messenger-RNAs coding for abundant endosperm proteins during wheat grain development. Pl. Sci., 46:117-125

Bäumlein, H., Boerjan, W., Nagy, I., Panitz, R., Inzé, D. and Wobus, U. (1991a) Upstream sequences relating legumin gene expression in heterologous transgenic plants. Mol. Gen. Genet., 225: 121–128.

Baümlein, H., Miséra ,S., Luerssen, H., Kölle, K., Horstmann, C., Wobus, U. and Müller, A.J.(1994) The FUS3 gene of *Arabidopsis thaliana* is a regulator of gene expression during late embryogenesis. Plant J., 6: 379–387.

Bäumlein, H., Müller, A.J., Schiemann, J., Helbing, D., Manteuffel, R. and Wobus, U. (1987) A legumin B gene of *Vicia faba* is expressed in developing seeds of transgenic tobacco. Biologisches Zentralblatt .,106: 569–575.

Baumlein, H., Wobus, U. , Pustell, J. and Kalatos, 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.

Baumlein, H., Nagy, I., Villarroel, R., Inze,D. and 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: 233-239.

Bharali, S. (2002) Isolation, cloning and molecular analysis of the legumin gene of common buckwheat (*Fagopyrum esculentum* Moench) .Doctoral thesis, North Eastern Hill University, Shillong, India.

Birnboim HC, Doly J. (1979) A rapid alkaline extraction procedure for screening recombinant plasmid DNA. Nucleic Acids Res., 7(6):1513–1523.

Blackwood, E.M. and Kadonaga, J.T. (1998) Going the distance: a current view of enhancer action. Science, 281: 60-63.

Blechl,C. and Anderson, O.D. (1996) Expression of novel high molecular weight glutelin subunit gene in transgenic wheat. Nature Biotechnol, 14: 875-879.

Bliss, F.A., Mcleester, R.C. and Hall TC. (1972) Electrophoretic analysis of protein changes during the development of French bean fruit. Phytochemistry, 11: 647.

Bobb, A.J., Chern, M.S. and Bustos, M.M. (1997) Conserved RY-repeats mediate transactivation of seed-specific promoters by the developmental regulator PvALF. Nucleic Acids Res., 25: 641–647.

Bogue, M.A., Vonder Haar, R.A., Nuccio, M.L., Griffing, L.R., and Thomas, T.L. (1990) Developmentally regulated expression of a sunflower 11S seed protein gene in transgenic tobacco. Mol. Gen. Genet., 222: 49–57.

Boulter, D. (1981) Biochemistry of storage protein synthesis and deposition in the developing legume seed. In: Advances in botanical research vol.9.Woolhouse H.W. (Ed.). Academic Press, London.

Bray, E.A. and Beachy, R.N. (1985) Regulation by ABA of/3-conglycinin expression in cultured developing soybean cotyledons. Plant Physiol., 79: 746-750.

Bray, E.A., Naito, S., Pan, N-S., Anderson, E., DubeÂ, P. and Beachy RN (1987) Expression of the b-subunit of β-conglycinin in seeds of transgenic plants. Planta 172:364-370

Brown, J. W. S., Waneldt, C., Feix, G., Neuhaus, G. and Schweiger, H. G. (1986) The upstream regions of *Zein* genes: sequence analysis and expression in the unicellular green alga. Eur. J. Cell Biol., 42 (1): 161–170.

Burke, T.W., Willy, P.J., Kutach, A.K., Butler, J.E. and Kadonaga, J.T. (1998) The DPE, a conserved downstream core promoter element that is functionally analogous to the TATA box. Cold Spring Harb. Symp. Quant. Biol., 63: 75-82.

Burow, M.D., Sen, P., Chlan, C.A. and 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: 537-548.

Bustos, M.M., Guiltinan, M.J., Jordano, J., Begum, D., Kalkan ,F.A. and Hall, T.C. (1989) Regulation of GUS expression in transgenic tobacco plants by an A/T rich cis-acting sequence found upstream of a French bean 6-phaseolin gene. Plant Cell, 1: 839-853.

Bustos, M.M., Begum, D., Kalkan, F.A., Battraw, M.J. and Hall, T.C. (1991) Positive and negative *cis*-acting DNA domains are required for spatial and temporal regulation of gene expression by a seed storage protein promoter. EMBO J., 10: 1469-1479.

Butler, J.E. and Kadonaga, J.T.(2001) Enhancer-promoter specificity mediated by DPE or TATA core promoter motifs. Genes Dev., 15: 2515–2519.

Carbonero, P., Vicente-Carbajosa, J., Mena, M., Oñate, L., Lara, P. and Diaz, I. (2000) bZIP and DOF transcription factors in the regulation of gene expression in barley endosperm. In: Seed Biology (Black, M., Bradford, K.J. and Vazquez-Ramos, J. Wallingford, UK: CABI Publishing, pp. 27–41.

Carey, M. and Smale, S. (2000) Transcriptional regulation in eukaryotes: concepts, strategies, and techniques, Cold Spring Harbor, NY, Cold Spring Harbor Laboratory Press.

Casey, R. (1999) Distribution and some properties of seed globulins. In: Shewry,P.R., Casey, R., eds., Seed Proteins. Kluwer Academic Publishers, Dordrecht, the Netherlands: 159–169.

Casey, R., Christou, P., Domoney, C., Hedley, C., Hitchin, E., Parker, M., Stoger, E., Wang, T. and Zasiura, C. (2001) Expression of legumin and vicilin genes in pea mutants and the production of legumin in transgenic plants. Nahrung., 45(6):385-7.

Chamberland, S., Daigle, N. and Bemier, 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. Plant Mol. Biol., 19: 937-949.

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

Chen, Z.L., Schuler, M.A. and Beachy, R.N. (1986) Functional analysis of regulatory elements in a plant embryo-specific gene. Proc Natl Acad Sci USA., 83: 8560–8564.

Chen, Z.G., Stauffacher, C., Li, Y., Schmidt, T., Bomu, W., Kamer, G., Shanks, M., Lomonossoff G. and Johnson, J.E.(1989) Protein-RNA interactions in an icosahedral virus at 3.0 A resolution. Science. 245(4914): 154-9.

Chern, M.S., Eiben, H.S. and Bustos, M.M. (1996) The developmentally regulated bZIP factor ROM1 modulates transcription from lectin and storage protein genes in bean embryos. The Plant J., 10: 135-148.

Chui, C.F., and Falco, S.C. (1995) A new methionine-rich seed storage protein from maize. Plant Physiol., 107:291.

Colot, V., Robert, L.S., Kavanagh, T.A., Bevan, M.W., and Thompson, R.D., (1987) Localization of sequences in wheat endosperm protein genes which confer tissue-specific expression in tobacco. EMBO J., 6: 3559-3564.

Coulter, K.M. and Bewley, J.D. (1990) Characterization of a small Sulphur-Rich Storage albumin in Seeds of Alfalfa (*Medicago sativa*.L). J. Exptl. Bot., 41: 1541-1547.

Coupe, S.A., Taylor, J.E., Isaac, P.G. and 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. Biol., 23: 1223-1232.

Dagalarrondo, M., Raymond, J. and Azanza, J.L. (1986) Subunit composition of the globulin fraction of rapeseed (*brassica napus*) Plant Sci., 43: 115-124.

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

Dasgupta, S. and 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.

Davidson. E.H., Jacobs, H.T. and Britten, R.J. (1983) Eukaryotic gene expression: Very short repeats and coordinate induction of genes. Nature, 301: 468-470.

Decker, C.J., Parker, R. (1994) Mechanisms of mRNA degradation in eukaryotes. Trends Biochem Sci., 19(8):336–340.

DeLisle, A.J. and Crouch, M.L. (1989) Seed storage protein transcription and mRNA levels in *Brassica napus* during development and in response to exogenous abscisic acid. Plant Physiol ., 91: 617-623.

Derbyshire, E., Wright, D.J. and Boulter, D. (1976) Legumin and vicilin storage proteins of legume seeds. Phytochemistry., 15: 3-24.

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.

Dey, N. and Mandal, R.K. (1993) Characterization of 2S albumin with nutritionally balanced amino acid 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: 149-157.

Diaz, I., Vicente-Carbajosa, J., Abraham, Z., Martinez, M., Isabel-La-Moneda, I. and Carbonero, P. (2002) The GAMYB protein from barley interacts with the DOF transcription factor BPBF and activates endosperm-specific genes during seed development. Plant J., 29(4): 453-64.

Diaz, I., Martinez, M., Isabel-La-Moneda, I., Rubio-Somoza, I. and Carbonero, P. (2005) The DOF protein, SAD, interacts with GAMYB in plant nuclei and activates transcription of endosperm-specific genes during barley seed development. Plant J., 42: 652-62.

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. Nucliec Acids Res., 16: 371.

Dietrich, R.A., Radke, S.E., and Harada, J.J. (1992) Downstream DNA sequences are required to activate a gene expressed in the root cortex of embryos and seedlings. Plant cell, 4: 1371-1382.

Don, R.H., Cox, P.T., Wainwright, B.J., Baker, K. and Mattick, J.S. (1991) 'Touchdown' PCR to circumvent spurious priming during gene amplification. Nucleic Acids Res., 19: 4008.

Doyle, J.J., Schuler, M.A., Godette, W.D., Zenger, V., Beachy, R.N. and Slightom, J.L. (1986) The glycosylated seed storage proteins of *Glycine max* and *Phaseolus vulgaris*: structural homologies of genes and proteins. J. Biol.Chem., 261: 9228-9238.

Dudman , W.F. and Millerd, A. (1975) Immunological behaviour of legumin and vicilin from *Vicia faba*: a survey of related proteins in leguminosae subfamily faboideae. Biochem. Syst. Ecol., 3: 25-29.

Dunwell, J.M., Culham, A., Carter, C.E., Sosa-Aguirre, C.R. and Goodenough, P.W.(2001) Evolution of functional diversity in the cupin superfamily. Trends in Biochemical Science, 26: 740-746.

Dunwell, J. M., Khuri, S. and Gane, P. J. (2000) Microbial relatives of the seed storage proteins of higher plants: conservation of structure and diversification of function during evolution of the cupin superfamily. Microbiol. Mol. Biol. Rev., 64: 153–179.

Ellerstrom, M., Stalberg, 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. Pl. Mol. Biol., 32: 1019–1027.

Entwistle, J., Knudsen, S., Muller, M. and Cameron-Mills V. (1991) Amber codon suppression: The *in vivo* and *in vitro* analysis of two C-hordein genes from barley. Pl. Mol Biol., 17: 1217-1231.

Ericson, M. L., MurCn, E., Gustavsson, H.O., Josefsson, L.-G. and Rask, L. (1991) Analysis of the promoter region of napin genes from *Brassica napus* demonstrates binding of nuclear protein in vitro to a conserved sequence motif, Eur. J. Biochem., 197: 741 -746.

Ezcurra, I., Ellerstrom, M., Wycliffe, P., Stalberg, 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. Pl. Mol. Biol., 40: 699 -709.

Ezcurra, I., Wycliffe,P., Nehlin, L., Ellerström, M. and Rask, L. (2000) Transactivation of the *Brassica napus* napin promoter by ABI3 requires interaction of the conserved B2 and B3 domains of ABI3 with different *cis*-elements: B2 mediates activation through an ABRE, whereas B3 interacts with an RY/G-box. Plant J., 24: 57-66.

Falco, S.C., Guida,T., Locke, M., Mauvais, J., Sanders, C., Ward, R.T. and Webber, P. (1995) Transgenic canola and soybean seeds with increased lysine. Biotechnology., 13(6): 577–582.

Fassler, J.S. and Gussin, G.N. (1996) Promoters and basal transcription machinery in eubacteria and eukaryotes: concepts, definitions and analogies. Methods Enzymol., 273: 3–29.

Fickett, J.W. and Hatzigeorgiou, A.C. (1997) Eukaryotic promoter recognition. Genome Res., 7: 861–878.

Finkelstein, R.R., Tenbarge, K.M., Shumway, J.E. and Crouch, M.L. (1985) Role of ABA in maturation of rapeseed embryos. Plant Physiol .,78: 630-636.

Forde, B.G., Heyworth, A., Pywell, J. and Kreis, M. (1985) Nucleotide sequence of a Bi 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.

Fujimori, S., Washio, T. and Tomita, M (2005) GC compositional strand bias around transcription start sites in plants and fungi. BMC genomics, 6, 26

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. J. Agric. Food Chem., 49: 1825-1829.

Fujiwara, T. and Beachy, R.N. (1994) Tissue-specific and temporal regulation of a beta-conglycinin gene: roles of the RY repeat and other cis-acting elements. Plant Mol Biol., 24: 261–272.

Gatehouse, J. A., Evans, I. M., Croy, R.R.D. and Boulter, D. (1986) Differential expression of genes during legume seed development. Philos. Trans. R. Soc. Lond., 314: 367-384.

Gidamis, A.B., Wright,P., Haque, Z.U., Katsube, T., Kito, M. and Utsumi, S. (1995) Modification tolerability of soybean proglycinin. Biosci. Biotech. Biochem., 59: 1593–1595.

Giraudat, J., Hauge, B.M., Valon, C., Smalle, J., Parcy, F. and Goodman, H.M. (1992) Isolation of the Arabidopsis ABI3 gene by positional cloning. Plant Cell., 4: 1251–1261.

Goldberg, R.B. (1986) Regulation of plant gene expression. Philos. Trans. R. Soc. B314, 343–353.

Goldberg, R.B., Baker, S.J. and Perez-Grau, L. (1989) Regulation of gene expression during plant embryogenesis. Cell., 56: 149-157.

Goossens, A., Geremia, R., Bauw, G., Montague, M.V. and Angenson, G. (1994) Isolation and characterization of arcelin-5 proteins and cDNAs. Eur. J. Biochem., 225: 787-795.

Greenwood, J. and Chrispeels, M.J. (1985) Immunocytochemical localization of phaseolin and phytohemagglutinin in the endoplasmic reticulum and Golgi complex of developing bean cotyledons. Planta, 164: 295-302.

Guerche, P., De Almeida, E., Schwartztein, M.A., Gander, E., Krebbers, E. and Pelletier, G. (1990) Expression of 2S albumin from *Bertholletia exelsa* in *Brassica nupus*. Mol. Gen. Genet., 221: 306-314.

Hagar, K.P., Braun, H., Czihal, A., Muller, B. and Baumlein, H. (1995) Evolution of seed storage protein genes: legumin gene of *Ginkgo biloba*. J. Mol. Evol., 41: 457-466.

Hagar, K.P., Muller, B., Wind, C., Erbach. S. and Fischer, H. (1996) Evolution of legumin genes: loss of an ancestral intron at the beginning of angiosperm diversification. FEBS Letters., 387: 94-98.

Hakman, I., Stabel, P., Engstrom, P. and Erikson, I. (1990) Storage protein accumulation during zygotic and somatic embryo development in *Picea abies* (Norway spruce). Physiol. Plant., 80: 441-445.

Hall, T.C., Chandrasekharan, M.B. and Li, G. (1999) Phaseolin: its past, properties regulation and future. In: Shewry, P.R., Casey, R., eds., Seed proteins. Dordrecht, the Netherlands: Kluwer Academic Publishers: 209-240.

Hara, I., Ohmiya, M. and Matsubara, H. (1978) Pumpkin (*Cucurbita* sp.) seed globulin III. Comparison of subunit structures among seed globulins of various *Cucurbita* species and characterization of peptide components. Plant Cell Physiol., 19: 237-143.

Harada, J.J., Barker, S.J. and Goldberg, R.B. (1989) Soybean β-conglycinin genes are clustered in several DNA regions and are regulated by transcriptional and posttranslational processes. Plant Cell., 1: 415-425.

Harlow, P. P., Hobson, G. M. and Benfield, P. A. (1996). Constructions of linker-scanning mutations using PCR. Methods Mol. Biol., 57, 287-295.

Hartings, H., Maddaloni, M., Lazzaroni, N., Di Fonzo, N., Motto, M., Salamini, F. and Thompson R.D. (1989) The O2 gene which regulates *zein* deposition in maize encodes a protein with structural homologies to transcriptional activators. EMBO J., 8: 2795–2801.

Hattori, T., Terada, T. and Hamasuna, S. (1995) Regulation of the Osem gene by abscisic acid and the transcriptional activator VP1: analysis of *cis*-acting promoter elements required for regulation by abscisic acid and VP1. Plant J., 7: 913–925.

Hecker, K.H. and Roux, K.H. (1996) High and low annealing temperatures increase both specificity and yield in touchdown and stepdown PCR. BioTechniques, 20:478-485.

Heim, U., Baumlein, H. and Wobus, U. (1994) The legumin gene family: A reconstructed *Vicia faba* legumin gene encoding a high molecular weight subunit is related to typeB genes. Pl. Mol. Biol., 25: 131-135.

Hermann, R.G. and Larkins, B. (1991) In: Plant Molecular Biology, Plenum Press, New York.

Higgins, T.J.V., Chandler, P.M., Randall, P.J., Spencer, D., Beasch, L.R., Balagrone, R.J., Kroot, A.A. and Eingleis, A.S. (1986) Gene structure Protein structure, and Regulation of the Synthesis of a Sulphur rich Protein in Pea Seeds. J. Biol. Chem., 261: 11124-11130.

Hill, J.E. and Breidenbach, R.W. (1974) Proteins of soybean seeds. I. Isolation and characterization of the major components, Plant Physiology., 53: 742-46.

Hirano, H., Fukazawa, C. and Harada, L. (1985) The primary structures of the A4 and A5 subunits are highly homologous to that of the A3 subunit is the glycinin seed storage protein of soybean. FEBS 181:124-128

Hoffman, I. L., Donaldson, D.D. and Herman, E.M. (1988) A modified zein is synthesized, processed and degraded in the seeds of transgenic plants. Plant Mol. Biol., 11: 717-729.

Hosein, F. (2001) Isolation of high quality RNA from seeds and tubers of the Mexican Yam Bean (*Pachyrhizus erosus*). Pl. Mol. Biol. Rep., 19: 65a-65e.

Jain, M., Tyagi, S. B., Thakur, J. K., Tyagi, A. K. and Khurana, J. P., (2004) Molecular characterization of a light-responsive gene, breast basic conserved protein 1(*OsiBBC1*), encoding nuclear-localized protein homologous to

ribosomal protein L13 from *Oryza sativa indica*. Biochimica et Biophysica Acta., 1676: 182– 192.

Jiang, L., Abrams, S.R. and Kermode, A.R. (1996) Vicilin and napin storage-protein gene promoters are responsive to abscisic acid in developing tobacco seed but lose sensitivity following premature desiccation. Pl. Physiol., 110: 1135–1144.

Johnson, P.F. and McKnight, S.L.(1989) Eukaryotic transcriptional regulatory proteins. Annu. Rev. Biochem., 58: 799-839.

Jorgensen, R.A. (1995) Co-suppression, flower color patterns, and metastable gene expression states. Science 268:686-691.

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: 6643–6652.

Kagaya, Y., Toyoshima, R., Okuda, R., Usui, H., Yamamoto, A. and Hattori, T. (2005) Leafy Cotyledon1 controls seed storage protein genes through its regulation of Fusca3 and Abscisic Acid Insensitive3. Plant Cell Physiol., 46: 399–406.

Kansal, R., Ramkumar, T.P. and Koundal, K.R. (1995) Construction of chickpea cDNA library in λgt11. In: Genetic Research and Education: Current Trends and the Next Fifty Years. Ind. Soc. Genetics & Pl. Breed., pp.1727-1732.

Keeler S.J., Maloney C.L., Webber P.Y., Patterson , C., Hirata, L.T., Falco, S.C. and Rice, J.A. (1997) Expression of *de novo* high-lysine alpha-helical coiled-oil proteins may significantly increase the accumulation levels of lysine in mature seeds of transgenic tobacco plants. Plant Mol. Biol., 34: 15–29.

Keith, K., Kraml, M., Dengler, N.G. and McCourt, P. (1994) FUSCA3: a heterochronic mutation affecting late embryo development in Arabidopsis. Plant Cell, 6: 589–600.

Khitha, S., Kansal, R. and Koundal, K.R. (1995) Isolation and characterization of genomic DNA of chickpea. Pl. Physiol. & Biochem., 22: 1-5.

Khuri, S., Bakker, F.T. and Dunwell, J.M. (2001) Phylogeny, function, and evolution of the cupins, a structurally conserved, functionally diverse superfamily of proteins. Mol. Biol. and Evol., 18: 593-605.

Kingston, R.E., Bunker, C.A. and Imbalzano, A.N. (1996) Repression and activation by multiprotein complexes that alter chromatin structure. Genes Dev., 10(8): 905-920.

Kirihara, J.A., Hunsperger, J.P., Mahoney, W.C. and Messing, J.W. (1988) Differential expression of a gene for a methionine-rich storage protein in maize. Mol. Gen. Genet., 211(3): 477-484.

Kodrzycki, R., Boston, R.S. and Larkins, B.A. (1989) The opaque-2 mutation of maize differentially reduces *zein* gene transcription. Plant Cell, 1: 105-114.

Kortt, A.A., Caldwell, J.B. and Higgins, T.J.V.(1991) Amino acid and cDNA sequences of a methionine-rich 2S protein from sunflower seed (*Helianthus annuus* L.). Eur. J. Biochem., 195: 329-334.

Koundal, K.R., Mehta, S.L. and Sharma R.P. (1989) Construction of chickpea genomic library. Indian J. Exp. Biol., 27: 858-860.

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

Kozak, M. A. (1992) Consideration of alternative models for the initiation of translation in eukaryotes. Crit Rev Biochem Mol Biol., 27(4-5): 385-402.

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

Kreis, M., Shewry, P.R., Forde, B.G., Forde, J. and Miflin, B.J. (1985) Structure and evolution of seed storage proteins and their genes, with particular reference to those of wheat, barley and rye. In: Miflin, B. J. (ed.), Oxford Surveys of Plant Cell. Mol. Biol., 2: 253-317.

Kridl, J.C., McCarter, D.W., Rose, R.E., Scherer, D.E., Knutzon, D.S. and Radke, S.E. and Knauf, V.C. (1991) Isolation and characterization of an expressed napin gene from *Brassica rapa.* Seed Sci. Res. 1, 209-219.

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

Lago, W. J. P., Scott, M.P. and Neilsen, N.C. (1990) Assembly properties of modified subunits in the glycinin subunit family. In: Plant Molecular Biology. Herman, R. G., and Larkins, B. A. (eds.)., pp. 635-640.

Lambert, N. and Yarwood, J.N. (1992) In: Plant Protein Engineering. Shewry, P. R., and Gutteridge , S. Cambridge Univ. Press, Cambridge, U.K. (eds.). pp. 167-187.

Larkins, B.A. (1983) Genetic engineering of plants: An agricultural perspective. Plenum Press, New York pp. 92-118.

Lawrence, M. C., Suzuki, E., Varghese, J. N., Davis, P. C., Van Donkelaar, A.; Tulloch, P. A. and Colman, P. M. (1990) The three dimensional structure of the seed storage protein phaseolin at 3 resolution. EMBO J., 9: 9-15.

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., Thomas, T.R., Okita, W. and Jeon, J.S. (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: 531-546.

Lending, C.R., Wallace, J.C. and Larkins, B.A. (1992) In: Plant Protein Engineering Shewery,P.R. and Gutteridge, S.(eds.) Cambridge Univ. Press, Cambridge, U.K. pp. 209-218.

Lessard, P.A., Allen, R.D., Fujiwara, T. and Beachy, R.N. (1993) Upstream regulatory sequences from two 3-conglycinin genes. Plant Mol. Biol., 22: 873-885.

Li, X.M. and Shapiro, L. J. (1993). Three step PCR mutagenesis for linker scanning. Nucleic Acids Res. 21, 3745-3748.

Li, Z., Jayasankar, S. and Gray, D.J. (2001). Expression of a bifunctional green fluorescent protein (GFP) fusion marker under the control of three constitutive promoters and enhanced derivatives in transgenic grape (*Vitis vinifera*). Plant Sci. 160: 877-887

Lohmer, S., Maddaloni, M., Motto, M., Di Fonzo, N. and Hartings, H. (1991) The maize regulatory locus Opaque-2 encodes a DNA-binding protein which activates the transcription of the b-32 gene. EMBO J., 3: 617-624.

Lycett, G.W., Croy, R.R.D., Shirsat, A.H., Richards, DM and Boulter, D. (1985) The 5'-flanking regions of three pea legumin genes: comparison of DNA sequences. Nucleic Acid Res., 13(18):1633-1742.

Mandal, S. and Mandal, R.K. (2000) Seed storage proteins and approaches for improvement of their nutritional quality by genetic engineering. Curr. Sci., 79 (5): 576-589.

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

Margoliash, E, Nisonoff, A., and Reichlin, M. (1970) Immunological activity of cytochrome c. I. Precipitating antibodies to monomeric vertebrate cytochrome c. J. Biol. Chem., 245: 931-939.

Marion-Poll, A. (1997) ABA and seed development. Trends Plant Sci., 2:447–448

Matthews, K.S. (1992) DNA looping. Microbiol. Rev., 56: 123-136.

Matzke, A.J.M., Stoger, E.M., Schernthaner, J.P. and Matzke, M.A. (1990) Deletion analysis of a *zein* gene promoter in transgenic tobacco plants. Plant Mol. Biol., 14: 323-332.

Matzke, M.A., Matzke A.J.M. (1995) How and Why Do Plants Inactivate Homologous (Trans) genes? Plant Physiol., 107(3):679-685.

McKeown, M. (1992) Alternative mRNA splicing. Annu. Rev. Cell Biol., 8: 133-55.

McKnight, S.L. and Yamamoto, K. R. (1992) Transcriptional Regulation, "Combinatorial Regulation at a Mammalian Composite Response Element." Cold Spring Harbor Lab. Press, New York, 1169-1192

Melo, T.S., Ferreira, R.B. and Teixeira, A.N. (1994) The seed storage proteins from *Lapinus albus* Phytochemistry, 37:641-648.

Mena,M. Vicente- Carbajosa, J., Schmidt, R. and Carbonero, P. (1998) An endosperm-specific DOF protein from barley, highly conserved in wheat, binds to and activates transcription from the prolamin-box of a native B-hordein promoter in barley endosperm. Plant J., 16: 53-62.

Meyer, P. (1995) Understanding and controlling transgene expression. Trends Biotechnlogy, 13: 332-337.

Milisavljevic, M. D., Timotijevic, G. S., Radovic, S. R. , Brkljacˇ ic′, J.M. Konstantinovic,′ M. M., and Maksimovic, V. R. ′(2004) Vicilin-like storage globulin from buckwheat (*Fagopyrum esculentum* Moench) seeds. J. Agric. Food Chem., 52, 5258-5262.

Misra, S. and Green, M.J. (1994) Legumin-like storage polypeptide of conifer seeds and its antigenic cross-reactivity with 11S globulin of angiosperms. J. Expt. Bot., 45: 269-274.

Mitchell, P.J. and Tjian, R. (1989) Transcriptional regulation in mammalian cells by sequence-specific DNA binding proteins. Science, (4916): 371-378.

Molina, C. and Grotewold, E. (2005) Genome wide analysis of Arabidopsis core promoters. BMC Genomics., 6(1): 25.

Molvig, L., Tabe, L. M., Eggum, B. O., Moore, A. E., Craig, S., Spencer, D. and Higgins, T. J. V. (1997) Enhanced methionine levels and increased nutritive value of seeds of transgenic lupins (*Lupinus angustifolius* L.) expressing a sunflower seed albumin gene. *Proc. Natl. Acad. Sci. USA* 94: 8393-8398.

Monke, G., Altschmied, L., Tewes, A., Reidt, W., Mock, H. P., Baumlein, H. and Conrad, U. (2004) Seed-specific transcription factors AB13 and FUS3 : molecular interaction with DNA. Planta, 719: 158-166.

Moreira, M.A., Hermondson, M.A., Larkins, B.A. and Nielsen, N.C. (1979) Partial characterization of acidic and basic polypeptides of glycinin. J. Biol. Chem., 254: 9921-9929.

Moreno-Risueno, M.A., González, 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.

Morton, R.L., Quiggin, D., Higgins, T.J.V. (1995) Regulation of seed storage protein gene expression. *In* J Kigel, G Galili, eds, Seed Development and Germination. Marcel Dekker, NY, pp 103–138.

Müller, M., Muth, J.R., Gallusci, P., Knudsen, S., Maddaloni, M., Motto, M., Schmitz, D., Sørensen, M.B., Salamini, F., von Wettstein, D. and Thompson, R.D. (1995) Regulation of storage protein synthesis in cereal seeds: Developmental and nutritional aspects. J. Plant Physiol. 145:606–613.

Muntz, K., Blattner, F.R. and Shutov, A.D. (2002) Legumains: A family of asparagine-specific cysteine endopeptidases involved in propolypeptide processing and protein breakdown in plants. J. Plant Physiol., 160: 1281–1293.

Muntz, K., Cristov, V., Jung, R., Saalbach, G., Saalbach, I., Waddal, D., Pickardt, T. and Schleider, O. (1997) In: Sulphur Metabolism in Higher Plants: molecular, Ecophysiological and nutritional Aspects. W.J. Cram, L.J. De Kok, I. Stulen, C. Brunold and H. Rennenberg (Eds.), Backhuys, Leiden, 1997, pp. 71-86.

Murai, N., Sutton, D.W., Murray, M.G., Slightom, J .L,. Merlo D.J., Reichert, N.A ., Sengupta-Gopalan, C ., Stock, C.A., Barker, R .F., Kemp, J.D. and Hall, T.C. (1983) Phaseolin gene from bean is expressed after transfer to sunflower via tumor-inducing plasmid vectors . Science, 222: 476-482.

Murray, M., and Kennard, W. (1984) Altered Chromatin Conformation of the higher plant gene phaseolin. Biochemistry, 23:4225-4232.

Murray, M.G. and Thompson, W.F.(1980) Rapid isolation of high molecular weight DNA. Nucleic Acids Res. 8:4321-4325.

Neilsen, N. C., Jung, R., Nam,Y.W., Beaman,T.W., Oliveira, L.O., and Bassuner, R., (1995) Synthesis and assembly of 11S globulins. J. Plant. Physiol., 145: 641-647.

Neilsen, N.C., Dickenson, C.D., Cho, T.J., Thanh, V.H., Scallon, B. J., Fisher, R. L., Sims, T.L., Drews., G.N. and Goldberg, R.B. (1989) Characterization of the glycinin gene family in soybean. Plant Cell, 1: 313-328.

Novina, C.D. and A.L. Roy. (1996) Core promoters and transcription factor binding sites. J. Mol. Biol., 249: 923–932.

Nunberg, A.N., Li, Z., Bogue, M.A., Vivekananda, J., Reddy, A.S. and Thomas, T.L.(1994) Developmental and hormonal regulation of sunflower helianthinin genes: proximal promoter sequences confer regionalized seed expression. Plant Cell, 6: 473-486.

Nunberg, A. N., Li, Z., Chung, H-J., Reddy, A. S. and Thomas, T. L. (1995) Proximal promoter sequences of sunflower helianthinin genes confer regionalized seed-specific expression. J. Pl. Physiol., 45: 600-605

Oñate, L., Vicente-Carbajosa, J., Lara, P., Diaz, I. and Carbonero, P. (1999) Barley BLZ2: a seed- specific bZIP protein that interacts with BLZ1 in vivo and activates transcription from the GCN4-like motif of B-hordein promoters in barley endosperm. J. Biol. Chem., 274: 9175-9182.

Ono, A., Izawa, T., Chua, N.H. and Shimamoto, K. (1996). The rab16B promoter of rice contains two distinct abscisic acidresponsive elements. Plant Physiol., 112: 483-491.

Onodera, Y., Suzuki, A., Wu, C., Washida, H. and Takaiwa, F. (2001) A rice functional transcriptional activator, RISBZI, responsible for endosperm-specific expression of storage protein genes through GCN4 motif. J. Biol. Chem., 276: 14139-14152.

Osborne, T.B. (1924) The vegetable proteins, Monograph in Biochemistry, Longmans green and Co. London.

Parcy, F., Valon, C., Kohara ,A., Misera , S. and Giraudat , J. (1997) The absicic acid-Iinsensitive3, Fusca3, and Leafy Cotyledon1 loci act in concert to control multiple aspects of Arabidopsis seed development. Plant Cell, 9: 1265-1277.

Payne, P.I. (1983) Breeding for protein quantity and protein quality in seed crops. In: Seed Proteins, J. Daussant, J. Mosse and J Vaughan, eds. Academic Press, London. pp. 223

Perez-Grau, L. and Goldberg, R.B. (1989) Soybean seed protein genes are regulated spatially during embryogenesis. Plant Cell, 1(11): 1095-1109.

Petterson, M. and Schaffner, W. (1987) A purine-rich DNA sequence motif present in SV40 and lymphotropic papovavirus binds a lymphoid-specific factor and contributes to enhancer activity in lymphoid cells. Genes. Dev. 1(9):962–972.

Phillips, J., Artsaenko, O., Fiedler, U., Horstmann, C., Mock, H.P., Muntz, K. and Conrad U. (1997) Seed-specific immunomodulation of abscisic acid activity induces a developmental switch. The EMBO Journal 16: 4489–96.

Qu, S., Desai, A., Wing, R. and Sundaresan, V. (2008) A versatile transposon-based activation tag vector system for functional genomics in cereals and other monocot plants. Pl. Physiol., 146: 189–199.

Radke, S.E., Andrews, B.M., Moloney, M.M., Crouch , M.L., Kridl, J.C., Knauf, V.C.

(1988). Transformation of*Brassica napus* L. using Agrobacterium tumefaciens: developmentally regulated expression of a reintroduced napin gene. Theor. Appl. Genet., 75: 685-694

Rahman, S., Shewry, P. R., Forde, B. G., Kreis, M. and Miflin, B. J. (1983) Nutritional control of storage- protein synthesis in developing grain of barley (*Hordeum vulgare* L.). Planta, 159: 366-372.

Raina, A. and 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. USA, 89 : 11774-11778.

Rao, A.G., Hassan, M. and Hempel, J.C. (1994) Structure -function validation of high Lysine analogs of L-hordothions designed by protein modeling, Protein Engineering., 7: 1484-1493.

Reichlin, M., Nisonoff, A. and Margoliash, E. (1970) Immunological activity of cytochrome *c*. 3. Enhancement of antibody detection and immune response initiation by cytochrome *c* polymers. J. Biol. Chem., 245: 947–954.

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. Plant J., 21: 401–408.

Reeves, C.D. and Okita,T.W.(1987) Analysis of α/β-type gliadin genes from diploid and hexaploid wheats. Gene., 52: 257-266.

Robert, L.S., Thompson, R.D. and Flavell, R.B. (1989) Tissue-specific expression of a wheat high molecular weight glutenin gene in transgenic tobacco. Plant Cell., 1: 569-578.

Rodin, J., Ericson, M.L., Josefsson, L-G. and Rask, L. (1990) Characterization of a cDNA clone encoding a *Brassica napus* 12S protein (Cruciferin) subunit: Disulfide bonding between subunits. J. Biol. Chem. 5:2720-2723

Roeder, R.G. (1996) The role of general initiation factors in transcription by RNA polymerase II. Trends Biochem. Sci., 21(9): 327–335.

Rombauts, S., Florquin, K., Lescot, M., Marchal, K., Rouzé, P. and Van de Peer, Y. (2003) Computational approaches to identify promoters and *cis*-regulatory elements in plant genomes. Plant Physiol., 132: 1162–1176.

Rorth, P. and Montell, D.J.(1992) Drosophila C/EBP: a tissue-specific DNA-binding protein required for embryonic development. Genes Dev., 6: 2299–2311.

Rout, M.K. (1996) Studies on seed storage protein metabolism in common buckwheat (*Fagopyrum esculentum* Moench), PhD Thesis, North Eastern Hill University, Shillong, India.

Rout, M.K. and Chrungoo, N.K. (1996). Partial characterization of the lysine rich 280kD globulin from common buckwheat (*Fagopyrum esculentum* Moench): Its antigenic homology with seed proteins of some other crops. Biochem. & Mol. Biol. Intl., 40: 587-595.

Rout, M.K., Chrungoo, N.K. and Rao, K.S. (1997) Amino acid sequence of the basic subunit of 13S globulin of buckwheat. Phytochemistry, 45(5): 865-867.

Roux, K. H. (1995) Optimization and troubleshooting in PCR. PCR Methods Appl., 4: 5185-5194.

Rozen, S. and Skaletsky, H.J. (2000) Primer3 on the WWW for general users and for biologist programmers. In: Krawetz S, Misener S (eds) Bioinformatics methods and protocols: methods in molecular biology. Humana Press, Totowa, New Jersey, pp 365–386

Saalbach , I., Waddell, D., Pickardt, T., Schieder, O. and Muntz, K. (1995a) Stable expression of the sulphur rich 2S albumin gene in transgenic *Vicia narbonensis* increases the Methionine content of seeds. J. Plant. Physiol., 145: 674-681.

Saalbach, T., Pickardt, D.R., Waddell, S., Hillmer, O., Schieder and K. Müntz, K. (1995b) The sulphur-rich Brazil nut 2S albumin is specifically formed in transgenic seeds of the grain legume *Vicia narbonensis*. Euphytica 84:181-192.

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

Sakata, Y., Chiba, Y., Fukushima, H., Matsubara, N., Habu, Y., Naito, S. and 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. Biol., 34: 191-197.

Sambrook, J., E. F. Fritsch, and T. Maniatis. (1989) Molecular cloning: a laboratory manual. Second edition. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, New York, USA.

Santos Mendoza, M., Dubreucq, B., Miquel, M., Caboche, M. and Lepiniec, L. (2005) Leafy Cotyledon 2 activation is sufficient to trigger the accumulation of oil and seed specific mRNAs in Arabidopsis leaves. FEBS Lett., 579: 4666–4670.

Schirm, S., Jiricny, J., 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.

Schmidt, R.J., Ketudat, M., Aukerman ,M.J. and Hoschek, G. (1992) Opaque-2 is a transcriptional activator that recognizes a specific target site in 22-kD *zein* genes. Plant Cell, 4: 689–700.

Schmidt, R.J., Burr, F.A., Aukerman, M.J. and Burr, B. (1990) Maize regulatory gene opaque-2 encodes a protein with a leucine-zipper motif that binds to *zein* DNA. Proc. Natl. Acad. Sci. USA, 87: 46–50.

Sengupta-Gopalan, C., Reichert, N.A., Barker, R.F., Hall, T.L. and Kemp, J.O. (1985) Developmentally regulated expression of the bean beta-phaseolin gene in tobacco seed Proc. Natl. Acad. Sci. USA, 82: 3320–3324.

Shen, Q. and Ho, T.H.(1995) Functional dissection of an abscisic acid (ABA)-inducible gene reveals two independent ABA-responsive complexes each containing a G-box and a novel cis-acting element. Plant Cell., 7(3): 295–307.

Shewry, P.R. (1995) Plant storage proteins. Biol. Rev., Camb Philos. Soc., 70:375-426.

Shirsat, A., Wilford, N., Croy, R. and Boulter, D. (1989) Sequences responsible for the tissue specific promoter activity of a pea legumin gene in tobacco. Mol. Gen. Genet., 215: 326–331.

Shotwell, M. and Larkins, B.A. (1989) The biochemistry and molecular biology of seed storage proteins. In: Marcus E., (ed). The biochemistry of plants. Academic Press, San Diego.pp. 296-345.

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

Shutov, A., Braum, H., Chesnokov, Y.U.V. and Baumlein, H. (1998) A vicilin-like gene is specifically expressed in fern spores. Evolutionary pathway of seed storage globulins. Euro. J. Biochem., 252: 79-89.

Siebert, P.D., Chenchik, A., Kellogg, D.E., Lukyanov, K. A. and Lukyanov, S. A. (1995) An improved method for walking in uncloned genomic DNA. Nucleic Acids Res., 23: 1087-1088.

Sjödahl S., Gustavsson H.O., Rödin J. and Rask L. (1995) Deletion analysis of the B*rassica napus* cruciferin gene cru 1 promoter in transformed tobacco: promoter activity during early and late stages of embryogenesis is influenced by cis-acting elements in partially separate regions. Planta, 197: 264–271.

Smale, S.T. (1994) Core promoter architecture for eukaryotic protein coding genes. In Transcription: Mechanisms and Regulation. (Conaway RC, Conaway JW, eds), Raven Press, New York, NY: 63-81.

Smale, S.T. (1997) Transcription initiation from TATA-less promoters within eukaryotic protein-coding genes. Biochem. Biophys. Acta., 1351: 73-88.

Smale, S.T. (2001) Core promoters: active contributors to combinatorial gene regulation. Genes. Dev., 15:2503-8.

Sorensen, M.B., Cameron-Mills, V. and Brandt, A. (1989) Transcriptional and post-transcriptional regulation of gene expression in developing barley endosperm. Mol. Gen. Genet. 217:195-201

Southern, E.M. (1975) Detection of specific sequences among DNA fragments separated by gel electrophoresis. J. Mol.Biol., 98: 503-517.

Stalberg, K., Elierstrom, M., Josefsson, L.G., and Rask, L. (1993) Deletion analysis of a 2S seed storage protein promoter of *Brassica napus* in transgenic tobacco. Pl. Mol. Biol., 23: 671–683.

Stalberg, K., Ellerstom, M., Ezcurra, I., Ablov, S. and Rask, L. (1996) Disruption of an overlapping E-box/ABRE motif abolished high transcription of the *napA* storage-protein promoter in transgenic *Brassica napus* seeds. Planta, 199: 515-519.

Struhl, K. (1987) Promoters, activator proteins, and the mechanism of transcriptional initiation in yeast. Cell, 49: 295-7.

Sun, S.M., Mutschler, M.A., Bliss, F.A. and Hall, T.C. (1978) Protein synthesis and accumulation in bean cotyledons during growth. Pl. physiol., 61: 918-929.

Suzuki, A., Wu, C.Y., Washida, H. and Takaiwa , F. (1998) Rice MYB protein OSMYB5 specifically binds to the AACA motif conserved among promoters of storage protein glutelin. Pl. Cell Physiol., 39: 555-559.

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

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

Takaiwa, F., Oono, K. and Kato, A.(1991) Analysis of the 5'flanking region responsible for the endosperm-specific expression of a rice glutelin chimeric gene in transgenic tobacco. Pl. Mol.Biol., 16: 49-58.

Tatarinova, T, Brover, V., Troukhan, M. and Alexandrov, N (2003) Skew in CG content near the transcription start site in Arabidopsis thaliana. Bioinformatics 19(1): i313-i314.

Thomas, T.L., Vivekananda, J. and Bogue, M.A. (1991) ABA regulation of gene expression in embryos and mature plants. In: Abscisic Acid: Psychology and Biochemistry., (Ed. by Davis and Jones) pp.125-136.

Thompson, G.A., Boston, R.S., Lyznik, L.A., Hodges, T.K. and Larkins, B.A. (1990) Analysis of promoter activity from an a-zein gene 5' flanking sequence in transient expression assays. Pl.Mol. Biol., 15: 755-764.

Townsend, J.A. and Thomas, L.A. (1994) Factors which influence the *Agrobacterium*-mediated transformation of soybean. J. Cell. Biochem. Suppl., 18A: 78.

Ueda, T., Waverczak, W., Ward K., Sher, N., Ketudat, M., Schmidt, R.J. and Messing, J. (1992) Mutations of the 22– and 27-kD zein promoters affect transactivation by the Opaque-2 protein. The Plant Cell, 4:701–709.

Utsumi, S., Kitagawa, S., Katsube, T., Kang, I.J., Gidamis, A.B., Takaiwa, F. and Kito, M. (1993) Synthesis processing and accumulation of modified glycinins of

soyabean in the seeds, leaves and stems of transgenic tobacco. Pl. Sci., 92: 191-202.

Vasil,V., Marcotte, W.R., Jr, Rosenkrans, L., Cocciolone, S.M., Vasil, I.K., Quatrano, R.S. and McCarty ,D.R.(1995) Overlap of Viviparous1 (VP1) and abscisic acid response elements in the Em promoter: G-box elements are sufficient but not necessary for VP1 transactivation. Plant Cell, 7: 1511–1518.

Vicente-Carbajosa, J., Moose, S.P., Parsons, R.L. and Schmidt, R.J. (1997) A maize zinc-finger protein binds the prolamine box in zein genes promoters and interacts with the basic leucine zipper transcriptional activator Opaque2. Proc. Natl. Acad. Sci. USA., 94: 7685–7690.

Voelker, T., Sturm, A. and Chrispeels, M.J. (1987) Differences in expression between two seed lectin alleles obtained from normal and lectin-deficient beans are maintained in transgenic tobacco. EMBO J., 6: 3571-3577.

Vonder Haar, R.A., Allen, R.D., Cohen, E. A., Nessler, C.L. and Thomas, T.L. (1988) Organization of the sunflower 11S storage protein gene family. Gene, 74: 433-443.

Wahle, E. and Keller, W. (1996) The biochemistry of polyadenylation. Trends in Biochem. Sci.., 21(7): 247-50.

Walburg ,G. and Larkins, B.A. (1983) Oat seed globulin: Subunit characterization and demonstration of its synthesis as a precursor. Pl. Physiol., 72: 161-165.

Wasylyk, B., Imler, J. L., Chatton, B., Shatz, C. and Wasylyk, C. (1988) Negative and positive factors determine the activity of the polioma virus enhancer alpha domain in undifferentiated and differentiated cell types. Proc. Natl. Acad. Sci. USA., 85: 7952-7956.

Weis, L. and Reinberg, D. (1992) Transcription by RNA polymerase II: initiator-directed formation of transcription-competent complexes. FASEB J., 6: 3300-9.

Weising, K. and Kahl, G. (1991) Towards an understanding of plant gene regulation - the action of nuclear factors. Zeitschrift f. Naturforschg. 46c, 1-11.

West, A.G., Gaszner, M. and Felsenfeld, G. (2002) Insulators: many functions, many mechanisms. Genes Dev., 16: 271–288.

Wingender,E., Dietze,P., Karas,H. and Knüppel,R. (1996) TRANSFAC: A database on transcription factors and their DNA binding sites. Nucleic Acids Res., 24: 238-241

Wu, C.Y., Suzuki, A., Washida, H. and 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: 673-683.

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. Plant J., 23: 415–421.

Yamamoto YY, 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, 8:67.

Yoshihara, T. and Takaiwa, F.(1996) *cis*-regulatory elements responsible for quantitative regulation of the rice seed storage protein glutelin *GluA-3* gene. Pl. and Cell Physiol., 37: 107–111.

Yoshino, M., Kanazawa, A., Tsutsumi, K. I., Nakamura, I., Shimamoto, Y. (2001) Structure and characterization of the gene encoding alpha subunit of soybean beta-conglycinin. Genes. Genet. Syst., 76(2):99-105.

Yunes, J.A., Cord Neto, G., da Silva, M.J., Leite, A., Ottoboni, L.M. and Arruda, P. (1994) The transcriptional activator opaque-2 recognizes two different target sequences in the 22 kDlike a prolamin genes. Plant Cell, 6: 237-249.

Zawel, L. and Reinberg, D. (1993) Initiation of transcription by RNA polymerase II, a multi-step process.In: Prog. Nucleic Acids Res. Mol. Biol., 44, Eds.Cohn, W.E. and Moldave, K.Academic Press, San Dieago, pp. 67–108.

Zeh, M., Paola Casazza, A., Kreft, O., Roessner, U., Bieberich, K., Willmitzer, L., Hoefgen, R. and Hesse, H. (2001) Antisense Inhibition of Threonine Synthase leads to high methionine content in Transgenic Potato plants. Pl.Physiol., 127 (3): 792-802.

Zhang B, Foley RC, Singh KB (1993) Isolation and characterization of two related *Arabidopsis* ocs-element bZIP binding proteins. Plant J., 4:711-716.

Zhang, J.Z., Santes, C.M., Engel, M.L., Gasser, C.S. and Harada, J.J. (1996) DNA sequences that activate isocitrate lyase gene expression during late embryogenesis andduring postgerminative growth. Pl.Physiol. 110:1069-1079

Zheng, Z., Kawagoe, Y., Xiao, S., Li, Z., Okita, T.W., Hau, T.L., Lin, A. and Murai, N.(1993) 5' Distal and proximal cis-acting regulation elements are required for developmental control of a rice seed storage protein glutelin gene. Plant J., 4: 357-366.

Zhenweiz, Z., Sumi, K., Tanaka, K. and Murai, N. (1995) The bean seed storage protein β-phasolin is synthesized, processed, and accumulated in the vacuolar type-II protein bodies of transgenic rice endosperm Pl. Physiol., 109: 777-786.