ISOLATION AND FUNCTIONAL CHARACTERIZATION OF NOVEL ABUNDANTLY EXPRESSED SEED STORAGE PROTEIN GENES IN MATURING GRAINS OF COMMON BUCKWHEAT

ABSTRACT

By

CRESSIDA JAMIR

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

NORTH EASTERN HILL UNIVERSITY
SHILLONG-793022, INDIA
2009

Thesis

...U LIBRARY
...No. 103994
26-10-10

DS
583.57
JAM

Seed formation is a complex process that involves fertilization of the egg cell, morphogenesis of the embryo and a maturation phase which is characterized by accumulation of storage reserves and loss of water from the tissues (Harada, 1997). The maturation of embryos ends with a desiccation phase when the embryo enters into a quiescent state, thereby permitting its maintenance and survival under a range of environmental conditions. Development of the embryo occurs in two distinct phases which comprise (i) morphogenesis, during which the basic body plan of the plant is established and (ii) maturation, during which the processes related to the embyro's entry into metabolic quiescence and subsequent germination occur. In many plants, the zygote elongates and then divides asymmetrically to form daughter cells of different sizes and cytoplasmic densities. The apical daughter cell, after two rounds of longitudinal and one round of transverse divisions, gives rise to an eight-cell embryo proper. At the eight-cell stage, four regions with different developmental fates can be recognized: (1) the apical embryo domain, (2) the central embryo

domain, (3) the basal embryo domain or hypophysis and (4) the extra embryonic suspensor. Clonal analyses have confirmed that the contribution of each cell to the seedling body plan is highly predictable (Laux *et al.,* 2004). However, rare variations in the cell division pattern do occur where each cell differentiates according to its final position, establishing that developing plant cells are flexible and assume their fate corresponding to positional information (Poethig *et al.,* 1986; Saulsberry *et al.,* 2002). A dramatic transformation of embryo morphology occurs during the transition from the globular to the heart stage. The shift in embryo symmetry from radial at the globular stage to bilateral at the heart stage represents the initial delineation of the two major embryonic organ systems, the cotyledons and axis. Following their formation, the cotyledons and axis elongate rapidly as a result of cell division and cell expansion (Tykarska, 1979, 1980). While embryo morphogenesis occurs as a monophasic phenomenon in lower vascular plants with no distinct end to embryonic development or a definite beginning to postembryonic growth (Steeves and Sussex, 1989), morphogenesis of higher plant embryos is interrupted by a period of maturation during which storage reserve macromolecules, including storage proteins, lipids, and carbohydrates, accumulate in virtually all cells of the embryo (Walbot, 1978; Jenner, 1982; Slack and Browse, 1984; Casey *et al.,* 1986; Tykarska, 1987; Shotwell and Larkins, 1989).

Plant development is integrated by long-range signals, such as growth factors, which are transported along the shoot-root axis (Schiavone and Cooke, 1987; Lyndon, 1990; Schiavone and Racusen, 1990; Fischer and Neuhaus, 1996; Creelman and Mullet, 1997; Kende and Zeevaart, 1997). Abscisic acid (ABA) has been invoked as the key hormone which is required throughout the process of seed maturation (Nambara and Marion-Poll, 2003; Finch-Savage and Leubner-Metzger,

2006). ABA signaling has been shown to be intimately connected to the expression of four master regulator genes viz. ABSCISIC ACID INSENSITIVE 3 (*ABI3*), FUSCA3 (*FUS3*), LEAFY COTYLEDON 1 (*LEC1*) and *LEC2* (Finkelstein, *et al.* 2002). Cadman, *et al.* (2006) have suggested that the induction of dormancy during later late-maturation phase was associated with a high ratio of ABA to GA which was mainly regulated by the dynamic balance between ABA biosynthesis and degradation in the seeds. Although the trigger for ABA synthesis is yet to be identified, expression studies of genes involved in ABA biosynthesis (including ZEAXANTHIN EPOXIDASE [ZEP] and 9-cis-EPOXYCAROTENOID DIOXYGENASES [NCEDs]) and catabolism (such as CYP707As encoding ABSCISIC ACID 80-HYDROXYLASES) have highlighted their key roles in the spatial and temporal accumulation of ABA in seeds (Nambara and Marion-Poll, 2005). Studies involving mutant phenotypes as the test material have, however, underlined the role of auxins in cell division and vascular differentiation in the basal domain which comprises hypocotyl, radical and primary root meristem (Berleth and Chatfield, 2002, Jenik and Barton, 2005; Leyser, 2006; Krogan and Berleth, 2007, Jenik *et al.*, 2007). Spatial and temporal changes in metabolite profiles including metabolite concentration gradients have also been suggested to provide signals for the onset and regulation of the seed maturation. Changes in ratio of sucrose to hexose in embryos have been suggested to be related to the formation of transfer cells, establishing an epidermis-localized sucrose uptake system (Offler *et al.,* 2003).

In recent years, molecular approaches have been used extensively to investigate seed embryogenesis. Seed proteins comprise one of the major and important storage materials, which accumulate in high amounts during the

maturation phase of seed development. These proteins remain stable throughout the developmental arrest and are then specifically degraded to serve as a source of carbon and nitrogen at the initial stages of germination and seedling growth. Seed storage protein genes represent abundantly transcribed genes that are controlled in a precise developmental manner during various stages of seed development. While a large proportion of the proteins found at early developmental stages in soybean seeds have been shown to belong to the 2S albumin fraction those accumulating during the maturation phase have been shown to belong to the 7.5S and 11.8S fractions (Eldridge *et al.*, 1966; Catsimpoolas *et al.,* 1969; Catsimpoolas and Leuthner, 1969).

Consistent with differential accumulation of proteins during various stages of seed development, variations have also been observed in transcript levels for different proteins during various stages of seed development. Out of the large number of diverse genes expressed during embryogenesis, most encode rare mRNAs of unknown function. A small set of genes, however, directs the synthesis of abundant mRNAs that encode seed storage proteins that are packed preferentially into protein bodies in the embryo, dicot cotyledons and monocot endosperms (Mandal and Mandal, 2000). In addition to storage proteins, several other proteins like lectins, proteinase inhibitors and late embryogenesis abundant (LEA) proteins too are synthesized in the embryo during seed development. Apart from possibly serving as storage proteins, these proteins have also been implied as defence proteins against several biotic and abiotic stresses (Gatehouse *et al.*, 1991). While the changes in transcriptome and proteome profiles with seed development have provided valuable clues linking seed development with gene networks (Gallardo *et al.*, 2003; 2007; Nakabayashi *et al.*, 2005; Hajduch *et al.*, 2005; Cadman *et al.*, 2006; Chibani *et al.*, 2006; Joosen *et al.*, 2007; Carrera *et al.*, 2008, Liu *et al.*, 2009; Houston *et al.*,

2009) many important unresolved questions associated with the regulatory network mechanisms involved in gene expression during seed development remain unresolved. Except for SSP genes, for which the structure of cis regulatory elements is well documented, little is known about the interactions between master regulator proteins and other putative target genes. The high diversity of seed storage compounds among plant species implies that distinct biosynthesis pathways and, therefore, distinct putative regulatory mechanisms are involved. Consequently, the use of non-model plant species might be important for identifying new regulators, as well as for highlighting the differences between conserved and non-conserved seed maturation processes. In order to better understand the seed maturation processes, many studies have been done in model plants such as *Arabidopsis thaliana* (Gallardo *et al.,* 2001; Charmont *et al.*, 2005; Cho *et al.*, 2007; Amme *et al.*, 2006) and *Lotus japonicus* (Dam *et al.*, 2009). However, the use of non model plant species might be important for identifying new gene regulator and also highlight the differences between conserved and non-conserved seed maturation processes.

Common buckwheat (*Fagopyrum esculentum* Moench) could be one such non model species for elucidation of processes involved in seed development in plants. Common buckwheat is a hermaphroditic, self-incompatible plant that contains one ovule per ovary and has an extremely high rate of seed abortion. The crop has also been identified by International Plant Genetic Resources Institute (IPGRI) and Consultative Group on International Agriculture (CGIAR) as one of the important underutilized nutraceutical crops which can be used as a genetic base for identification of genes coding for the seed proteins rich in essential amino acids. Even though several investigations have been carried out by researchers across the globe on various aspects of reproductive biology of buckwheat, information on the

molecular basis of seed development in the plant is scanty. The present work focuses on isolation and characterization of genes coding for abundantly expressed seed storage proteins during seed maturation in the crop. This may provide clues to molecular processes which might be unique to buckwheat species or family and cannot be elucidated via a model plant.

The present work focused on isolation and characterization of novel abundantly expressed seed storage protein genes in maturing grains of common buckwheat. The approaches followed were (i) PCR amplification of target gene(s) from RNA isolated from grains of common buckwheat at different stages of maturation by reverse-transcription (RT)-PCR and (ii) construction of cDNA library from RNA isolated from grains of common buckwheat at mid maturation stage of development and from grains at early stages of germination followed by amplification of target gene(s) by nested PCR using the cDNA library as the template and primers designed from nucleotide sequences of genes available in the gene bank databases.

Total RNA was isolated from a suitable mass of powdered tissues from grains harvested at different stages of development and used for RT-PCR as well as construction of developmental stage specific cDNA library. The isolated RNA could be detected on agarose gels as two fluorescing bands corresponding to a typical RNA profile of 28S rRNA and 18S rRNA with a band thickness and intensity of approximately 2:1 ratio.

PCR amplification of genes expressed abundantly in the maturing grains of common buckwheat was attempted by RT-PCR with cDNA prepared from RNA isolated from grains at different stages of maturation as the template and oligonucleotide primers designed from nucleotide sequences avaialable in

genebank data bases. amplification with primer pair CJ1F-CJ2R generated a 400 bp amplicon in all the reaction mixtures and an amplicon showing apparent molecular mass of 1.0 kb in the amplification mixture having cDNA synthesized from RNA isolated from grains harvested between 16 to 20 DAF. The 1.0 kb amplicon could not be detected in amplification mixtures containing cDNA prepared from RNA isolated from grains harvested at any other stage. This indicates the absence or low abundance of transcripts coding for the putative buckwheat dehydrin gene in the grains during other developmental stages. Analysis of the nucleotide sequence of the 1.0 kb amplicon revealed the presence of a domain comprising of the sequence "GTCGGAGGATGATGGACAAG GAGGAAGAAGAAAGAAAAAAGGGTTGAAAGAAAAGATAAA" between position 557 to 616. This domain was found to be conserved in genes coding for dehydrin like proteins from several plants including *Solanum lycopersicum* (AC215480) *Cornus sericea* (AF345988), *Helianthus annuus* (AJ438980), *Eriobotrya japonica* (FJ472835), *Panax ginseng* dehydrin (DQ487110), *Phoenix dactylifera* (DQ399792), *S.tuberosum* (X83597), *Tithonia rotundifolia* (AJ250127). BLASTn analysis of the nucleotide sequence revealed a maximum of 99% and 97% homology with *Coffea canephora* dehydrin genes DH2a (acc. no. DQ323989) and DH2b (DQ323990) respectivelyand 67% homology with a *Camellia sinensis* dehydrin gene (acc no. FJ436979). It could thus be assumed, that the DNA fragment amplified in the present study was a partial nucleotide sequence of a gene coding for a dehydrin like protein.

Our results on the detection of transcripts coding for dehydrin class of proteins in grains of buckwheat during mid maturation stage of development are in conformity with other reports indicating enhanced expression of genes coding

for dehydrin like proteins during the maturation phase of grain development (Close, 1996; Choi and Close, 2000; Nylander, et al., 2001; Puhakainen et al., 2004; Hinniger et al., 2006). On the basis of the pattern of accumulation of some of the *Lea* transcripts during seed development, with the highest level of transcripts reported at incipient desiccation, Gomez et al. (1988), Mundy and Chua (1988), Close et al. (1989) and Pages et al. (1995) have suggested that LEA polypeptides play a decisive role in the acquisition of desiccation tolerance during embryo maturation.

The deduced amino acid sequence for the 974 bases of the 1.0 kb amplicon comprised of 112 amino acid residues with a predicted isoelectric point (p*I*) of 5.87 and calculated molecular weight of 12 kDa. BLASTp analysis of the deduced amino acid sequence against non-redundant protein database identified the protein with the dehydrin family of proteins. Using an alignment that permitted maximum homology, the deduced amino acid sequence showed a maximum of 63% homology with *Coffea canephora* dehydrins DH2, DH2a and DH2b (acc. nos. DQ338457, DQ323990 & DQ323989). Dehydrin like proteins are known to possess the "K" segment comprising of the sequence "EKK(S/G)(M/I)(V/M)(E/D)KI(M/K)EKLPGHH" and the "Y" domain comprising of the sequence "D(E/Q)YGNP". The amino acid sequence of 112 residues identified in the present study showed three "Y" segments located between position 10-35 and one 'K" segment represented by the sequence "EKKSMVEKIMEKLPGHH" located between position 93-107. The presence of "K" and "Y" segments in the buckwheat amino acid sequence clearly indicates that the protein belongs to the dehydrin/ dehydrin-like protein family. While the conserved sequence of 15 residues viz. "EKKSMVEKIMEKLPG" in the "K"-

segment is a distinctive feature of group 2 LEA proteins (Close *et al.*, 1989, 1993; Rorat, 2006), the Y-segment comprised of the sequence "DE/QYGNP" has been reported as a common feature of all dehydrin like proteins (Close *et al.*, 1993; 1996; Campbell and Close, 1997. On the basis of presence of three "Y" segments, one "K" segment coupled with the absence of a "Ser" tract, the putative buckwheat dehydrin identified in the present study can be classified under the YK-subgroup with a Y_3K_1 architecture. A distinctive feature of the buckwheat amino acid sequence identified in the present study was the substitution of Serine tract with 'Asn-Phe-Arg' at position 66-68. While majority of dehydrins identified to date fall into the class that contains the "Ser" tract a number of dehydrins have been reported to lack the serine tract. Examples of dehydrins that lack a "Ser" tract include dehydrin/dehydrin-like proteins from spinach (Neven et al., 1993), wheat (Houde et al., 1992, Guo et al., 1992), barley (Close et al., 1995), pea (Roberton and Chandler, 1992). MOTIF SCAN on the deduced amino acid sequence identified a putative protein kinase C phosphorylation site, an N-glycosylation site, a kinase II phosphorylation site and two N-myristoylation sites in the sequence. While Roberton and Chandler, (1992), Levi (1999) and Caruso et al. (2002) have reported the presence of N-glycolysylation and myristoylation sites in dehydrins/dehydrin-like proteins, Mehta et al (2009) have reported a site for Casein kinase II in the AmDHNa ORF, Typically, dehydrins contain a high proportion of charged and polar amino acids with a low fraction of non-polar, hydrophobic residues and either few or no tryptophan and cysteine residues (Close *et al.*, 1989; Battaglia *et al.*, 2008). The buckwheat dehydrin type protein in the present study shares these features.

The phylogenetic tree generated from the alignment data of 112 deduced amino acids sequence in the present study with amino acid sequences of other dehydrin/dehydrin-like proteins available in the EMBL data base, revealed that the putative buckwheat (*Fagopyrum esculentum*) dehydrin identified in the present study emerged as a separate entity but was closer to the dehydrins from *Coffea canephora.* Dehydrins/dehydrin-like proteins from plants belonging to different groups/families are known to show low homologies. The phylogenetic tree of dehydrins/dehydrin-like proteins developed in the present study is one such example.

Total RNA isolated separately from grains of buckwheat harvested between 16-20 DAF as well as from germinating grains of common buckwheat was used as a template for construction of stage specific cDNA library using the Stratagene ZAP-cDNA Gigapack III Gold cloning kit as per the manufacturer's protocol. PCR amplification of cDNA library constructed from RNA isolated from grains of common buckwheat harvested between 16-20 Days after flowering as the template and 5' GGATCCGGATTGGAGCAAGCGTTCTGC 3' as the forward primer and 5' GAAACGCTCCCTCTCCTTCTCATC 3' as the reverse primer, generated an amplicon showing an apparent molecular mass of 0.5 Kb. This amplicon could not be detected in the lane having the amplification mixture containing the cDNA template prepared from RNA isolated from grains harvested at early stages of germination. BLASTn analysis of the nucleotide sequence comprising of 576 bases for the 0.5 Kb amplicon clearly identified the sequence with the nucleotide sequences of genes coding for legumin type seed storage proteins. The sequence showed 97% and 92% homology with nucleotide sequences of the genes coding for *Fagopyrum esculentum* legumin like proteins

(acc. nos. D87980 and D87982) with a sequence query coverage of 98% and 100%, respectively. Sequence alignment revealed a deletion of 3 bases "GGG" at position 528 in the nucleotide sequence of the DNA fragment amplified in the present study, a feature which was also detected in nucleotide sequences of genes coding for other buckwheat seed storage proteins used for generating the alignment matrix. However, a deletion of 12 bases "GAGGAGAAGGAG" detected at position 552 in the 576 bp nucleotide sequence was observed only in nucleotide sequences of legumin type seed storage proteins of *Fagopyrum esculentum* (acc. nos. AF152003 & D87980). These features distinguished the nucleotide sequence amplified in the present study from nucleotides sequences of genes coding for legumin type seed storage proteins of other plants as well as from the nucleotide sequence of gene coding for the allergenic protein of tartary buckwheat.

The amplification of nucleotide sequence for the gene coding for legumin type proteins from mid maturation stage specific cDNA library and not from seed germination stage cDNA library indicates the mid maturation developmental stage specific expression of the gene in grains of common buckwheat. Similar pattern of legumin gene expression during seed development has been reported by Panitz *et al* (1995), Darmency *et al.* (2005). Darmency *et al.* (2005) have detected the transcripts for legumin A and vicilin in the embryo cells of *M. truncatula* and *P. sativum* specifically at mid-embryogenesis. It was suggested that legumin A and vicilin expression patterns could be considered as suitable embryo-specific markers during histo-differentiation at mid-embryogenesis in *M. truncatula* and *P. sativum..* On the basis of changes in the expression pattern of genes coding for SSPs in *Vicia faba*, Panitz *et al.* (1995) have concluded that storage proteins of *Vicia faba*

accumulated transiently during early seed development and were used as nutritive reserves for the growing embryo. Even though seed storage protein gene expression has been shown to be rigorously tissue and developmental stage specific (Goldberg *et al.*, 1989; Perez-Grau and Goldberg, 1989; Guerche *et al.*, 1990; Sunderlikova and Wihelm, 2002) and several trans-acting factors that have been implicated in regulating the expression these genes, the role of specific sequences in seed protein gene expression is still a matter of discussion. It is therefore not surprising that analysis of individual elements comprising these ensembles has sometimes led to conflicting results on the role of specific elements in a regulatory ensemble (Baumlein *et al.*, 1991, 1992).

ORF finder tool from NCBI (http://www.ncbi.nlm.nih.gov) identified 5 open reading frames in the 576 bp nucleotide sequence. Sequence similarity analysis of the sequence with BLASTp against non-redundant protein database, identified the protein as belonging to the legumin family. The 7-element fingerprint, which has been universally identified as a signature for the 11-13S globulin family of proteins, was also detected in the deduced amino acid sequence of 191 residues. The 7 elements identified in the deduced amino acid sequence were, 'RQNVNRPSRADVFNPRAG' at P'12-29, 'DSNNLPILEFIQLSAQHVV LY' at P'35-55,'NAILGPRWNLNAHSALYVTRG' at P'57-79, 'VQVVGDEGR SVFDDNVQ' at P'81-97, 'GQILVVPQGFAVVLKA' at P'99-114, 'EGLEWVELKNDDNAITSPI' at P'117-135 and 'TSVLRAIPVEVLANSYDI' at P'139-156. This clearly established that the sequence in the present study belongs to globulin family.

The identification of a putative N-myristoylation site ' Gly-Leu-Glu-Gln-Ala-Phe' (P'1-6), 5 putative kinase II phosphorylation sites ' Ser-Arg-Ala-Asp'

(P'19-22), Thr-Arg-Gly-Glu' (P'75-78), Ser-Val-Phe-Asp' (P'90-93), Ser-Thr-Lys-Glu' (P'157-160), and Ser-Arg-Asp-Glu' (P'179-182) and a protein kinase C phosphorylation site represented by the sequence Gly-Leu-Glu-Gln-Ala-Phe' between residues 157-160 indicates that apart from being a seed storage protein, the deduced amino acid sequence might be involved in other metabolic activities during seed maturation. The software also identified a cupin domain between residues 13 to 162 predominantly represented by amino acids which would preferentially fold to form barrel shaped structures. Based on the hydropathic index of Kyte and Doolittle (1982), the major regions of hydrophilic nature detected in the sequence were between residues 3-43, 60-100, 114-134 and 152-189 indicating a predominantly hydrophilic character. This observation is in conformity with the hydropathic analysis of the 26 kDa basic subunit of buckwheat 13S globulin reported by Bharali, 2002. Statistical analysis of the sequence by SAPS (Brendel *et al.,* 1992) revealed that the charged residues were evenly distributed, with the sequence having 30.4% non polar residues, 21.5% polar uncharged residues and 25.6% polar charged residues. These results are in conformity with the predicted hydrophilic domains in the hydropathy analysis.

The ClustalW multiple alignment of the sequence with amino acid sequences of other seed storage proteins belonging to the legumin family showed >90% homology with the acidic chain of buckwheat 13S globulin (acc. no. Q9XFM4) and allergenic protein (acc. no. AAF34635). The percentage homology ranged from 43% - 53% with legumin genes from other plants including *Coffea arabica* (acc. no. AAC61881) and *Magnolia salicifolia* (CAA57846). It has been observed that most of the amino acid residues that are conserved amongst the legumin proteins are also present in the deduced amino acid sequence. The

conserved residues included P1'G, P7'C, P14'N, P18'P, P25'N, P26'P, P28'A, P29'G, P30'R, P40'P, P41'I, P42'L, P46'Q, P48'S, P54'L, P55'Y, P57'N, P58'A, P62'P, P64'W, P65'N, P67'N, P68'A, P69'H, P70'S, P73'Y, P77'G, P80'R, P82'Q, P84'V, P88'G, P92'F, P99'G, P100'Q, P105'P, P106'Q, P110'V, P118'G, P125'K, P137'G, P140'S, P146'P, P149'V, P157'S, P161'A, P164'L, P165'K, P168'R and P170'E. The globally conserved position of these amino acid residues indicates their importance in the secondary structure formation and stability of the protein.

MOTIF SCAN identified 5 putative kinase II phosphorylation sites, a putative N-myristoylation site and a protein kinase C phosphorylation site in the deduced amino acid sequence identified in the present study. Phosphorylation has been shown to play a major role in modulating the function and DNA-binding activity of many nuclear proteins, including transcription factors and proteins involved in chromatin organization (Dang *et al.*, 1994; Armstrong *et al.*, 1997; Hoffmann *et al.*, 1998). The identification of such sites in the deduced amino acid sequence is indicative of its role in signal transduction processes during seed maturation. Further, identification of a cupin domain in the deduced amino acid sequence indicated that the protein belongs to the cupin superfamily. The cupin superfamily is one of the largest, most functionally diverse families of proteins including 7S (vicilin) and 11S (legumin) seed storage proteins (Bäumlein *et al.*, 1995). These proteins have a characteristic signature domain comprising of two histidine containing motifs separated by an inter motif region of variable length. This domain consists of six ß strands within a conserved ß barrel structure.

When the amino acid residues of the sequence were plotted as a function of hydropathic index, the sequence showed a predominantly hydrophilic character.

Slightom *et al.* (1985) have also reported similar results on the hydrophilicity of *Phaseolus vulgaris* phaseolin. They have also reported the presence of two N-glycosyl recognition sites in the protein. While majority of the reports have described legumin type proteins as hydrophilic, Jain (2004) have reported the presence of two hydrophobic regions located within the first 76 residues (residues 6 to 21 and 67 to 76), each of which is long enough to span a membrane bilayer in peanut glycinin Gly-1.

Phylogenetic tree constructed from the alignment data of the deduced amino acid sequence with amino acid sequences of seed storage proteins available in EMBL database revealed that the deduced amino acid sequence of 191 residues clustered together into one clade with amino acid sequences of *Fagopyrum esculentum* 13S globulin (acc. no. Q9XFM4, O023880, AAF34635). Even though the amino acid sequences of allergenic protein from *F. tataricum* and the 22kDa seed storage protein from *F. gracilipes* clustered separately, they showed greater closeness with the group comprising the sequences bearing accession nos. Q9XFM4, O023880, AAF34635 and the sequence of 191 residues identified in the present study. Out of all the sequences analyzed in the present study, the amino acid sequence of *Coffea arabica* 11S globulin appeared to be the closest relative of buckwheat sequences.

Amplification of genomic DNA with primer pair 5' GGATCCGGATTGGAGCAAGCGTTCTGC 3' as the forward primer and 5' GAAACGCTCCCTCTCCTTCTCATC 3' as the reverse primer amplified a 0.7 Kb. DNA from the template genomic DNA. Pair wise alignment of the sequence with the nucleotide sequence of the amplicon generated with primer pair NKC5-NKC6 and cDNA library from grains harvested at 16-20 DAF as the template clearly

103994

revealed that the two sequences shared absolute homology. The alignment also revealed a gap of 96 bases corresponding to position between 142-238 bases in the 455bp nucleotide sequence of the amplicon generated with genomic DNA as the template and primer pair NKC5-NKC6. GENSCANW identified this region on the nucleotide sequence of the 0.7kb amplicon as an intron. Even though legumin genes are known to have a fairly fixed intron-exon architecture of three introns and four exons (Shewry, 1995), some of the legumin genes have been reported to have a two intron-four exon architecture (Shotwell and Larkins, 1989). While Vonder *et al.* (1988) have related legumin gene evolution to loss of introns, Shotwell and Larkins (1989) have suggested that the evolution of legumin gene has progressed towards addition of introns. Our result showing the presence of only one intron may be due to the partial sequence of the DNA amplified product. Sequence homology of the 107 deduced amino acid residues with BLASTp against non-redundant protein database clearly identified the protein as belonging to the legumin gene family. Northern hybridization against RNA isolated from grains harvested at 10-13 DAF, 16-20 DAF, 23-25 DAF and 30-35 DAF with [α-P^{32}]-dATP labelled 0.7 Kb PCR product generated a strong signal against RNA isolated from grains harvested at 16-20 DAF. Signals were also detected against RNA isolated from grains harvested between 23-25 DAF and 30-35 DAF but were of lesser intensity than the signal against RNA isolated from grains harvested between 16-20 DAF. Although most of the genes expressed during the later phases of embryogenesis are also active during the post-germination phase, all storage protein mRNAs have a particular period during which it reaches a maximum level and then declines (Yama *et al.,* 1999). During rice seed formation, Duan and Sun (2005) have observed that the genes encoding storage proteins glutelin, prolamin, globulin and albumin followed a

pattern of low initial expression at early developmental stage, reached its peak at mid maturation stage and maintained a significant expression profile till maturation except for glutelin gene which started declining towards maturation. Our result on the transcript level changes corresponding to the putative dehydrin gene as well as the legumin clearly indicate mid maturation specific expression of the two genes.

Several studies have clearly established the role of dehydrin genes in protective reactions thereby promoting maintenance of embryo structure under conditions of water loss during seed maturation (Puhakainen *et al.,* 2004; Choi and Close, 2000; Nylander, *et al.,* 2001), signalling cascade during embryo development (Goday *et al*., 1994; Mehta *et al*., 2009), protein-protein and protein-lipid interactions (Dure, 1993; Epand *et al*., 1995; Kovacs *et al*., 2008) as also during transition of perisperm to endosperm Hinniger *et al.* (2006) thereby indicating a prominent role for dehydrins in maturation phase of embryo development. On the other hand a similar role for seed storage proteins has not been identified definitively Even though extensive evidence has pointed towards a rigorous temporal and spatial control of seed storage protein gene expression during the maturation phase of seed development (Goldberg *et al.,* 1989; Perez-Grau and Goldberg, 1989; Guerche *et al*., 1990; Panitz *et al.* , 1995; Sunderlikova and Wihelm, 2002; Abirached-Darmency *et al*., 2005; Abirached-Darmency *et al*., 2005), the role of seed storage proteins in modulating developmental processes in seeds is still a matter of discussion. Majority of the legumin type seed storage proteins are hydrophilic in nature. Therefore, it is possible that besides acting as a major source of amino acids during germi[illegible]tial stages of seedling growth, these proteins may also be involved in conferring desiccation tolerance to the seeds during the maturation phase.

Acc No ......
Date 103994
Class ...
Sub Heading by 26-10-10
Checked by ......
Transcribed by ......

ISOLATION AND FUNCTIONAL CHARACTERIZATION OF NOVEL ABUNDANTLY EXPRESSED SEED STORAGE PROTEIN GENES IN MATURING GRAINS OF COMMON BUCKWHEAT

By

CRESSIDA JAMIR

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

NORTH EASTERN HILL UNIVERSITY
SHILLONG-793022, INDIA
2009

Thesis
LIBRARY 103994
Acc No ......
Acc By 26-10-10
Date ......
Class by ......
Sub.Heading by ......
Enter by ......
Transcribed by ......

DS
583.57
JAM

CONTENTS

Page

Declaration.....(I)

Acknowledgement.....(II-III)

Dedication.....(IV)

Abbreviation.....(V-VI)

List of table and figures.....(VII-IX)

CHAPTERS

1. INTRODUCTION.....1-7
2. REVIEW OF LITERATURE.....8-26
3. MATERIALS AND METHODS
   (i) Materials.....27
   (ii) Methods.....27-45
   (iii)Chemicals and Solutions.....45-48
4. RESULTS.....49-64
5. DISCUSSION.....65-81
6. GENERAL SUMMARY AND DISCUSSION.....82-97
7. REFERENCES.....98-118

The North Eastern Hill University

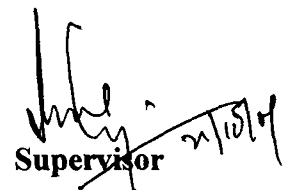
October, 2009

DECLARATION

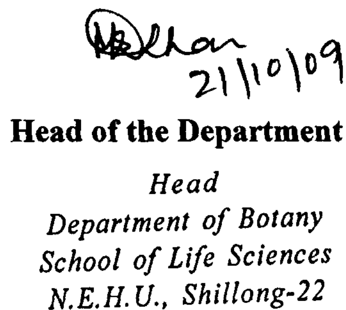
I, Cressida Jamir, hereby declare that the subject matter of this thesis entitled "Isolation and functional characterization of novel abundantly expressed seed storage protein genes in maturing grains of common buckwheat" is the record of work done by me. The contents of this thesis did not form any basis of the 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.

CRESSIDA JAMIR



Supervisor



Head of the Department

Head
Department of Botany
School of Life Sciences
N.E.H.U., Shillong-22

ACKNOWLEDGEMENTS

Dickens once wrote, "Reflect on your present blessings..." my heart and mind was instantly filled with thoughts of many wonderful people. My research work will be incomplete if I don't say "Thank You" - two small words to convey the infinite feeling in my heart to those people who gave directions and walked along with me throughout this long journey of mine.

I am forever grateful and indebted to my supervisor and mentor, Prof. N. K. Chrungoo, Department of Botany, NEHU. Completion of my thesis work would not have been possible without his support, patience and endless efforts. Thank you so much, sir.

I sincerely thank Prof. A. K. Misra, Department of Botany, NEHU, for readily allowing me to use the facilities in his laboratory whenever required. I also thank Prof. M. S. Dkhar, Head, Department of Botany for all the necessary help and support. My gratitude also goes to all the faculties of the department for their encouragement.

All the non-teaching staffs of Botany department deserve my special thanks for their ever ready helping hand.

I would like to specially thank Dr. S. K. Sopory, ICGEB, New Delhi for allowing me to work in his lab and Mr. Rabi who helped me out with cDNA library construction.

My labmates – Dr. Sangeeta Bharali, Dr. Anusuya Rout, Sanga, Tia, Shiny, Nabanita, Felly, Neeta, Ganga, Lashai and Jashmi – thank you for being great

companions. My friends from Molecular Genetics Lab. – Dr. Anindita Khan, Mhathung, David, Manik, Meena, Mipeshwaree, Insanara and Jubanri were always there to help me out; thanks a lot. Manoj deserves a special mention for always keeping things in place, letting my day start without a hitch.

'Thank you' to all my friends who were there for me through this rough yet fruitful journey of mine. I am thankful to Nabanita for being there in every way possible; David and Subarna for being my software engineer and chef; Ricky for being my resourceful friend; Nzan, a shoulder to lean on; Gautam for being my buffer; Ngabu, Julia and Hlantei for enduring my mood swings. Thank you all, for making my life colourful.

I thank the World Bank-Indian Council of Agricultural Research for the financial support during my tenure as SRF.

My love and gratitude to my parents, brothers, sisters-in-law, nieces and nephew for their encouragement and support during my research work.

Above all, I thank God, my 'Guiding Light'.

TO MY PARENTS

ABBREVIATIONS

Amp	: Ampicillin
BLAST	: Basic Local Alignment Search Tool
bp	: Base pair
CTAB	: Cetyl Trimethyl Ammonium Bromide
dATP	: 2' Deoxyadenosine 5'-triphosphate
dCTP	: 2' Deoxyadcytosine 5'-triphosphate
dGTP	: 2' Deoxyguanosine 5'-triphosphate
dTTP	: 2' Deoxythymidine 5'-triphosphate
dNTP	: 2' Deoxynucleotide 5'-triphosphate
DNA	: Deoxyribonucleic Acid
EDTA	: Ethylene Di-amine Tetra Acetic acid
Kb	: Kilo base
ME	: β-Mercaptoethanol
MOPS	: Morpholinopropane Sulfonic acid
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
UV	: Ultra Violet
μCi	: Micro Curie

DAF	: Days After Flowering
DEPC	: Diethyl Pyrocarbonate
M	: Molar
mM	: Milli Molar
cDNA	: Complementary DNA
cpm	: Counts Per Minute
EtBr	: Ethidium Bromide
IPTG	: Isopropyl thio-β-D-galactose
kDa	: Kilo Dalton
pI	: Isoelectric point
SSPE	: Saline Sodium Phosphate EDTA buffer
Tris	: Tris (hydroxymethyl) aminomethane
X-GAL	: 5-bromo-4-chloro-3-indolyl-β-D-galactoside
DTT	: Dithiothreitol
DMSO	: Dimethyl Sulfoxide

List of Table and Figures

Table: 3.1 : List of primers used for amplifying abundantly expressed seed storage protein genes during seed maturation.

Fig. 2.1 : Schematic representation of embryogenesis consisting of two phases: morphogenesis and maturation.

Fig. 3.1: : Close up of the common buckwheat plant (*fagopyrum esculentum* Moench) in flowering stage

Fig. 3.2 : Grains of common buckwheat harvested at different maturation stages of development.

Fig. 3.3 : Schematic representation of a typical three step cycle parameter for PCR amplification.

Fig. 4.1(a) : Electrophoresis profile of total RNA isolated from grains of common buckwheat (*Fagopyrum esculentum* Moench) harvested at different maturation stages.

Fig. 4.1(b) : Electrophoresis profile of second strand synthesis of total RNA isolated from grains of common buckwheat harvested at 16-20 days after flowering (DAF) using labeled [α^{32}P]dATP random hexamer.

Fig. 4.2 : Electrophoresis profile of amplification product with cDNA from grains of common buckwheat harvested at different maturation stages and early germination stage using primer pair CJ1F-CJ2R.

Fig. 4.3 : Nucleotide sequence of 974 bases for the 1.0 Kb fragment generated from cDNA of grains harvested at 16-20 DAF using primer pair CJ1F and CJ2R.

Fig. 4.4 : ClustalW (1.81) multi-alignment of the 974 bp nucleotide sequences (Buckwheat_DHN) for the 1.0 Kb amplicon generated with cDNA from 16-20 DAF seed stage as template and primer pair CJ1F - CJ2R with nucleotide sequences of other dehydrin/dehydrin like genes.

Fig. 4.5 : Deduced amino acid sequences of the coding region of 974 bp nucleotide sequences for the 1.0 Kb amplicon generated from 16- 20 DAF cDNA as the template and primer pair CJ1F and CJ2R.

Fig. 4.6(a) : Diagrammatic representation of the motif sequence patterns and its frequency in the 112 deduced amino acid sequences as predicted by MOTIF SCAN.

Fig. 4.6(b) : Diagrammatic representation of the Protein Kinase C phosphorylation site, Kinase II phosphorylation site, N-glycosylation site and N-

Myristoylation sites in the 112 deduced amino acid sequence as predicted by MOTIF SCAN.

Fig. 4.6(c) : Diagrammatic representation of the dehydrin domains positioned between 16 to 37 and 82-109 amino acid residues in the 112 deduced amino acid sequence as predicted by MOTIF SCAN.

Fig. 4.7 : ClustalW (1.81) multi-alignment of the deduced amino acid sequences (Buckwheat_DHN) for the 974 bp nucleotide sequence of cDNA from grains harvested at 16-20 DAF using primer pair CJ1F and CJ2R with amino acid sequences of dehydrin protein of other plants available in the database.

Fig. 4.8(a) : Hydropathy profile of the deduced 112 amino acid for the 1.0 Kb amplicon generated from 20 DAF buckwheat seed cDNA using CJ1F and CJ2R primer pair.

Fig. 4.8(b) : Flexibility analysis of the 974 bases for the 1.0 Kb amplicon generated from 20 DAF buckwheat seed cDNA using CJ1F and CJ2R primer pair.

Fig. 4.9 : Secondary structure prediction profile of the 112 deduced amino acid sequence for the 1.0 Kb amplicon generated with 16-20 DAF buckwheat seeds cDNA as the template and CJ1F-CJ2R as primer pair.

Fig. 4.10 : Phylogenetic tree based on the alignment matrix of the deduced 112 amino acid sequence for the 1.0 Kb amplicon fragment generated from 16-20 DAF seed cDNA using primer pair CJ1F-CJ2R with the amino acid sequences of dehydrin type proteins of various plants.

Fig. 4.11 : Electrophoresis profile of amplified fragment with cDNA library as template and primer pair NKC5 and NKC6;

Fig. 4.12 : Nucleotide sequence of 576 bp for the amplified fragment generated with cDNA of mid maturation seed stage and primer pair NKC5 and NKC6.

Fig. 4.13 : ClustalW (1.81) multi-alignment of 576 bp nucleotide sequence (BuckwheatSSP) of cDNA from mid maturation seed stage generated using primer pair NKC5 and NKC6 with nucleotide sequences of SSP genes of other plants derived from gene bank.

Fig. 4.14 : Deduced amino acid sequence of the 576 bp nucleotide sequence for the 0.5 Kb amplicon generated from cDNA of grains harvested at mid maturation stage of development and primer pair NKC5 and NKC6.

Fig. 4.15(a): *Diagrammatic representation of the motif patterns and frequency in* the 191 amino acid sequence as predicted by MOTIF SCAN.

Fig. 4.15(b): Diagrammatic representation of the N-myristoylation site and Kinase II phosphorylation sites in the 191 deduced amino acid sequence as predicted by MOTIF SCAN.

Fig. 4.15(c): Diagrammatic representation of the cupin domain positioned between 13 to 162 in the 191 deduced amino acid sequence as predicted by MOTIF SCAN.

Fig. 4.16 : Hydropathy profile of the 191 deduced amino acid sequence as predicted by Kyte and Doolittle method. The x-axis represents the residue number.

Fig.4.17 : Secondary structure prediction profile of the 191 deduced amino acid sequence using GOR4 method.

Fig. 4.18 : ClustalW (1.81) multi-alignment of the 191 deduced amino acid sequences (BuckwheatSSP) for the 576 bp nucleotide sequence (Buckwheat) of cDNA from mid matured seed stage generated using primer pair NKC5 and NKC6 with seed storage protein sequences of other plants derived from gene bank.

Fig. 4.19 : Phylogenetic tree based on the alignment matrix of the 191 deduced amino acid sequences with amino acid sequences of seed storage proteins of various plants available in the gene bank database.

Fig. 4.20(a): Total genomic DNA isolated from 15 days old seedling of *Fagopyrum esculentum* Moench.

Fig. 4.20(b): PCR amplification profile of buckwheat genomic DNA as template and NKC5 and NKC6 as the primer pair.

Fig. 4.21 : Nucleotide sequence of 455bp for the DNA fragment 0.7 Kb amplified from buckwheat genomic DNA using primer pair NKC5-NKC6.

Fig. 4.22 : ClustalW (1.81) pairwise alignment of the 576bp nucleotide sequence generated from 16-20 DAF cDNA template and 455bp nucleotide sequence generated from genomic DNA using primer pair NKC5-NKC6.

Fig. 4.23 : Diagramatic representation of the intron/exon architecture in the 455bp nucleotide sequence of the amplicon generated with genomic DNA as the template and NKC5-NKC6 as the primer pair.

Fig: 4.23 : Dot blot of total RNA isolated from grains of common buckwheat harvested at different maturation stages hybridised with [α-P^{32}]-dATP labelled 0.7 Kb fragment from genomic DNA using primers NKC5-NKC6.

CHAPTER : I
INTRODUCTION

INTRODUCTION

Seed formation is a complex process that involving fertilization of the egg cell, morphogenesis of the embryo and maturation phase which is characterized by accumulation of storage reserves and loss of water from the tissues (Harada, 1997). The maturation of embryos ends with a desiccation phase after which the embryo enters into a quiescent state, thereby permitting its maintenance and survival under a range of environmental conditions.

The sporophytic generation of higher plants is initiated with the double fertilization event that results in the formation of a single-celled zygote and a progenitor cell of the endosperm. Embryogenesis describes the subsequent period of development, during which the zygote undergoes a complex series of morphological and cellular changes resulting in the formation of a developmentally arrested mature embryo

comprised of an embryonic axis with shoot and root poles and cotyledon(s). Thus, the events that occur during embryonic development establish the organization of the plant body and prepare the embryo for both dormancy and germination. During embryogenesis, the single-celled zygote follows a defined pattern of cell division and differentiation to form the mature embryo. The asymmetry of zygote reflects the polar organization of the egg implying that embryonic polarity may be established during embryo sac development. Russell (1993) has ascribed the polar organization of the embryo to the redistribution of endoplasmic reticulum, plastids, and mitochondria after fertilization of the egg cell.

Development of the embryo occurs in two distinct phases - morphogenesis during which the basic body plan of the plant is established, followed by the maturation phase during which processes related to the embyro's entry into metabolic quiescence and subsequent germination occur. In many plants, the zygote elongates and then divides asymmetrically to form daughter cells of different sizes and cytoplasmic densities. The apical daughter cell after two rounds of longitudinal and one round of transverse divisions gives rise to an eight-cell embryo proper. At the eight-cell stage, four regions with different developmental fates can be recognized: (1) the apical embryo domain, (2) the central embryo domain, (3) the basal embryo domain or hypophysis; and (4) the extra embryonic suspensor. Clonal analyses confirm that the contribution of each cell to the seedling body plan is highly predictable (Reviewed by Laux *et al.,* 2004). However, rare variations in the cell division pattern do occur where each cell differentiates according to its final position, establishing that developing plant cells are

flexible and assume their fate corresponding to positional information (Poethig *et al.*, 1986; Saulsberry *et al.*, 2002). A dramatic transformation of embryo morphology occurs during the transition from the globular to the heart stage. The shift in embryo symmetry from radial at the globular stage to bilateral at the heart stage represents the initial delineation of the two major embryonic organ systems, the cotyledons and axis. Following their formation, the cotyledons and axis elongate rapidly as a result of cell division and cell expansion (Tykarska, 1979, 1980). While embryo morphogenesis occurs as a monophasic phenomenon in lower vascular plants with no distinct end to embryonic development or a definite beginning to postembryonic growth (Steeves and Sussex, 1989), morphogenesis of higher plant embryos is interrupted by a period of maturation during which storage reserve macromolecules, including storage proteins, lipids, and carbohydrates, accumulate in virtually all cells of the embryo (Walbot, 1978; Jenner, 1982; Slack and Browse, 1984; Casey *et al.*, 1986; Tykarska, 1987; Shotwell and Larkins, 1989).

Plant development is integrated by long-range signals, such as growth factors, which are transported along the shoot-root axis (Schiavone and Cooke, 1987; Schiavone and Racusen, 1990; Lyndon, 1990; Fischer and Neuhaus, 1996; Creelman and Mullet, 1997; Kende and Zeevaart, 1997). Abscisic acid (ABA) has been considered as the key hormone required throughout the seed maturation process (Nambara and Marion-Poll, 2003; Finch-Savage and Leubner-Metzger, 2006). ABA signaling has been shown to be intimately connected to the expression of four master regulator genes viz. ABSCISIC ACID INSENSITIVE 3 (*ABI3*), FUSCA3 (*FUS3*), LEAFY COTYLEDON 1 (*LEC1*)

and *LEC2* during the maturation phase (Finkelstein *et al.*, 2002). Cadman *et al.*, (2006) have also suggested that the induction of dormancy during late-maturation phase was associated with a high ratio of ABA to GA which was mainly regulated by the dynamic balance between ABA biosynthesis and degradation in the seeds. Studies involving mutant phenotypes as the test material have underlined the role of auxins in cell division and vascular differentiation in the basal domain which comprises hypocotyl, radical and primary root meristem (Berleth and Chatfield, 2002, Jenik and Barton, 2005; Leyser, 2006; Krogan and Berleth, 2007, Jenik *et al.*, 2007). Liu *et al.*, (1993) have emphasized that polar transport of auxins was required to direct cell divisions in the globular stage embryo but not in heart shaped stage embryo. Polar auxin transport, a defining feature of plants, has been reported to be mediated by a family of efflux carriers, the PIN proteins named after *Arabidopsis* PINFORMED1 (Gälweiler *et al.*, 1998; Paponov *et al.*, 2005). Spatial and temporal changes in metabolite profiles including metabolite concentration gradients have also been suggested to provide signals for the onset and regulation of the seed maturation. The embryo sucrose:hexose ratio has been proposed as the key element of the signaling pathway triggering the transition phase (reviewed by Weber *et al.*, 2005). Changes in ratio of sucrose to hexose in embryos have been suggested to be related to the formation of transfer cells, establishing an epidermis-localized sucrose uptake system (Offler *et al.*, 2003).

Seed embryogenesis is one of the major areas of interest and in recent times, researchers have started adopting functional genomics strategies to study its physiological process. Seed proteins comprise one of the major and important storage

materials, which accumulate in high amounts during the maturation phase of seed development. These proteins remain stable throughout the developmental arrest and are then specifically degraded to serve as a source of carbon and nitrogen at the initial stages of germination and seedling growth. Seed storage protein genes represent abundantly transcribed genes that are controlled in a precise developmental manner during various stages of seed development. Okamuro and Goldberg, (1989) and Goldberg *et al.* (1989) have put forth four features of seed protein gene expression during seed development which raises the question of genetic and molecular relationships between seed development and changes in protein pool:

(i) Seed protein gene expression is regulated temporally during embryogenesis;
(ii) Seed protein genes are expressed exclusively during embryogenesis or are expressed at very low levels in mature plant;
(iii) Seed protein gene expression is regulated spatially within embryogenic organ systems;
(iv) Seed protein mRNAs are localized within specific cells within the embryo cotyledon.

Out of the large number of diverse genes expressed during embryogenesis, most encode rare mRNAs of unknown function. A small set of genes, however, directs the synthesis of abundant mRNAs that encode seed storage proteins that are packed preferentially into protein bodies in the embryo, dicot cotyledons and monocot endosperms (Mandal and Mandal, 2000). In addition to storage proteins, several other proteins like lectins, proteinase inhibitors and late embryogenesis abundant (LEA) proteins too are synthesized in the embryo during seed development. Apart from possibly serving as storage proteins, these proteins have also been implied as defence proteins against several biotic and abiotic stresses. While the changes in transcriptome

and proteome profiles with seed development have provided valuable clues linking seed development with gene networks (Gallardo *et al.*, 2003; 2007; Nakabayashi *et al.*, 2005; Hajduch *et al.*, 2005; Cadman *et al.*, 2006; Chibani *et al.*, 2006; Joosen *et al.*, 2007; Carrera *et al.*, 2008; Houston *et al.*, 2009) many important unresolved questions associated with the regulatory network mechanisms involved in gene expression during seed development remain unresolved. Except for SSP genes, for which the structure of *cis* regulatory elements is well documented, little is known about the interactions between master regulator proteins and other putative target genes. The high diversity of seed storage compounds among plant species implies that distinct biosynthesis pathways and, therefore, distinct putative regulatory mechanisms are involved. Consequently, the use of non-model plant species might be important for identifying new regulators, as well as for highlighting the differences between conserved and non-conserved seed maturation processes. In order to better understand the seed maturation processes, many studies have been done in model plants such as *Arabidopsis thaliana* (Gallardo *et al.*, 2001; Charmont *et al.*, 2005; Cho *et al.*, 2007; Amme *et al.*, 2006) and *Lotus japonicus* (Dam *et al.*, 2009). However, the use of non model plant species might be important for identifying new gene regulator and also highlight the differences between conserved and non-conserved seed maturation processes.

Common buckwheat (*Fagopyrum esculentum* Moench) could be one such non model species for elucidation of processes involved in seed development in plants. Common buckwheat is a hermaphroditic, self-incompatible plant that contains one ovule per ovary and has an extremely high rate of seed abortion. The crop has also been

identified by International Plant Genetic Resources Institute (IPGRI) and Consultative Group on International Agriculture (CGIAR) as one of the important underutilized nutraceutical crops which can be used as a genetic base for identification of genes coding for the seed proteins rich in essential amino acids. Even though several investigations have been carried out by researchers across the globe on various aspects of reproductive biology of buckwheat, information on the molecular basis of seed development in the plant is scanty. The present work focuses on isolation and characterization of genes coding for abundantly expressed seed storage proteins during seed maturation in the crop. This may provide clues to molecular processes which might be unique to buckwheat species or family and cannot be elucidated via a model plant.

CHAPTER : II
REVIEW OF LITERATURE

Seed development is a crucial process in the lifecycle of higher plants, providing the link between two distinct sporophytic generations, and thereby, maintenance of the species. Seed development may be divided into four distinct stages: embryo patterning, embryo growth, seed filling and seed desiccation. Following completion of embryo growth, major changes include accumulation of storage products (e.g., proteins, oil, and starch), followed by desiccation, during which seeds acquire drought tolerance and primary dormancy is often induced (Bewley and Black, 1994). The maturation of embryo ends with a desiccation phase after which the embryo enters into a quiescent state, thereby permitting its maintenance and survival under a range of environmental *conditions. Embryogenesis describes the developmental period during which a single* cell differentiates into a mature embryo.

Over the past decade, several studies have provided evidence on the importance of the concerted action of a considerable number of signaling pathways in the control of the maturation processes, integrating information from genetic programs and from both

hormonal and metabolic signals (Weber *et al.*, 2005; Wobus and Weber, 1999). During the normal life cycle of higher plants, embryogenesis is initiated by sperm-egg contact at fertilization. The start of the developmental program is a consequence of changes due to fertilization events as well as activation of embryogenic events with well-defined timing (Epel, 1990). The sporophytic generation of higher plants is initiated with the double fertilization event that results in the formation of a single-celled zygote and a progenitor cell of the endosperm. Embryogenesis describes the subsequent period of development, during which the zygote undergoes a complex series of morphological and cellular changes resulting in the formation of a developmentally arrested mature embryo comprised of an embryonic axis with shoot and root poles and cotyledon(s). Thus, the events that occur during embryonic development establish the organization of the plant body and prepare the embryo for both dormancy and germination. Differentiation of totipotent embryogenic cells proceeds according to a pre-set program and terminally differentiated cells are formed. In contrast to other eukaryotes, the differentiation program in plants is flexible and almost any fully differentiated plant cell can become embryogenic under denned conditions.

Embryo development consists of two phases - morphogenesis and maturation phases. During the morphogenesis phase, the basic body plan of the plant is established with the specification of the shoot–root axis and the formation of embryonic organ and tissue systems. Maturation phase begins with a transition phase during which the switch from maternal to filial controls occurs (Weber *et al.,* 2005). During this phase the embryo growth ceases, storage products accumulate and acquires tolerance to desiccation leading to seed dormancy. Light and electron microscopy have provided

Fig. 2.1: Schematic representation of embryogenesis consisting of two phases: morphogenesis and maturation. During embryogenesis, the water content decreases and dry weight increases as a result of storage macromolecules accumulation including storage proteins, LEA proteins, lipids and carbohydrates. Maturation-phase seeds also possess a high ratio of ABA to GA that influences physiological processes.

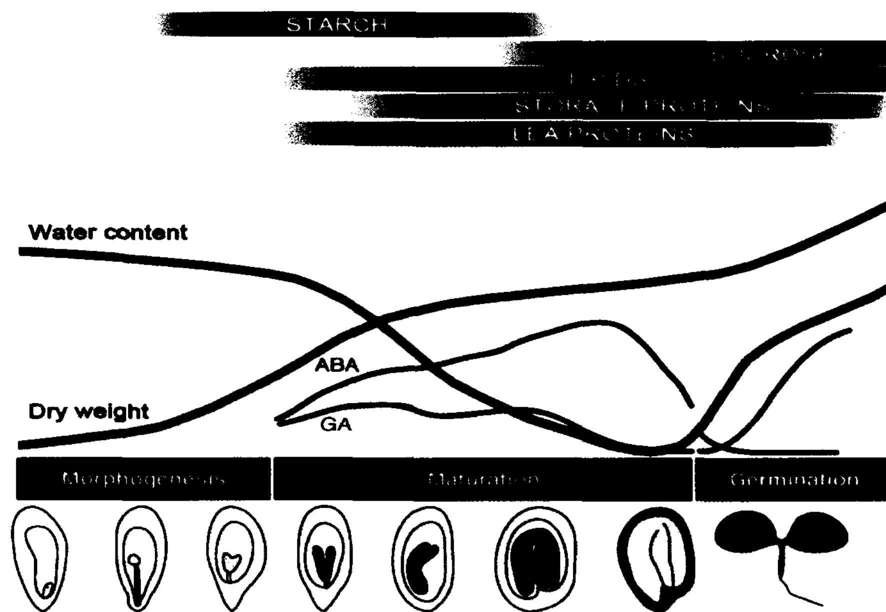


Fig. 2.1

detailed descriptions of the morphological and anatomical changes that characterize embryonic development (Maheshwari, 1950; Wardlaw, 1955; Natesh and Rau, 1984; Raghavan, 1986). Embryogenesis occurs with such consistency that the cleavage patterns have been used to classify embryo ontogenesis into six distinct types (Johansen, 1950; Maheshwari, 1950). In particular, the crucifers Arabidopsis, *Brassica napus* (oilseed rape), and *Capsella bursa-pastoris* represent plants in which the early divisions are virtually invariant (Schulz and Jensen, 1968a; Tykarska, 1979; Mansfield and Briarty, 1990). Other plants, such as cotton, have much more variable cleavage patterns (Pollock and Jensen, 1964). Thus, there is no universal pattern of early cell cleavages that describes embryonic development. In many angiosperms, the initial division of the zygote is transverse and asymmetric, generating a small, chalazally oriented apical cell and a large basal cell (Pritchard, 1964; Schulz and Jensen, 1968b; Tykarska, 1976; Mansfield and Briarty, 1990). Similarly, in cruciferous plants, the apical cell gives rise to the bulk of the embryo proper, including the cotyledons, shoot apex, and hypocotyl, whereas a part of the root apex and the suspensor originate from the basal cell (Schulz and Jensen, 1968a; Tykarska, 1976). In other plants, however, the first division of the zygote can be symmetrical, oblique, or longitudinal (Sivaramakrishna, 1978). Thus, a transverse and asymmetric division of the zygote is not a prerequisite for embryonic development in all plants. The apical cell undergoes two longitudinal divisions to produce a four-celled embryo proper which is followed by a transverse division to produce two tiers of cells called the octant stage embryo. The octant stage embryo shows distinct domains with a main axis of polarity and the linear formation of shoot and root meristems, cotyledons and hypocotyl. The division of the octant stage embryo

initiates the formation of the protoderm, which is the precursor of the epidermis. The delineation of the protoderm establishes the globular stage embryo, and during this stage, the embryo develops radial patterning through a series of cell divisions, with the outer layer of cells differentiating into the 'protoderm.' The embryo completes its growth during the subsequent torpedo and cotyledonary stages (Schulz and Jensen, 1968b; Tykarska, 1979; Mansfield and Briarty, 1990).

A dramatic transformation of embryo morphology occurs during the transition from the globular to the heart stage. Cell divisions parallel to the surface occur at specific regions of the lateral margins of the globular stage embryo, resulting in the emergence of the two lobes of the cotyledons. The shift in embryo symmetry from radial at the globular stage to bilateral at the heart stage represents the initial delineation of the two major embryonic organ systems, the cotyledons and axis. Following their formation, the cotyledons and axis elongate rapidly as a result of cell division and cell expansion (Tykarska, 1979, 1980). Other tissues and structures characteristic of postembryonic plants can be discerned in a heart stage embryo (Tykarska, 1979; Mansfield and Briarty, 1990). The latter stages of embryogenesis are concerned primarily with preparing the embryo for developmental arrest and germination (Crouch, 1987; Kermode, 1990; Galau *et al.*, 1991; Thomas, 1993). While embryo morphogenesis occurs as a monophasic phenomenon in lower vascular plants with no distinct end to embryonic development or a definite beginning to postembryonic growth (Steeves and Sussex, 1989), morphogenesis of higher plant embryos is interrupted by a period of maturation during which storage reserve macromolecules, including storage proteins, lipids, and carbohydrates, accumulate in virtually all cells of the embryo

(Walbot, 1978; Jenner, 1982; Slack and Browse, 1984; Casey et al., 1986; Tykarska, 1987; Shotwell and Larkins, 1989). These reserves are particularly prevalent in the embryonic cotyledons of plants that do not store substantial reserves in their endosperm, and are in large part responsible for a rapid increase in embryo mass and size (Mansfield and Briarty, 1992). These macromolecules subsequently serve as a nutrient source for the growing seedlings. During the final stages of embryogenesis, embryos also acquire the ability to withstand desiccation and eventually enter a period of metabolic quiescence (Kermode, 1990). The mature embryo remains dormant until it encounters conditions appropriate for germination.

A striking characteristic of plants is that tissue and organ formation are repetitive processes that occur continuously (Steeves and Sussex, 1989). The majority of morphogenetic events such as formation of leaves, stems, roots, and reproductive structures occur postembryonically. However, during embryonic development, the polar axis of the plant is established, domains that set up the organization of the plant body are defined, and the primary tissue and organ systems are delineated. Unfertilized egg cell, which is present in an asymmetrically organized embryo sac and ovule with the nucleus and much of the cytoplasm confined to the chalazal pole and a large vacuole towards the micropylar end, represents the first step in development of polarity in plants (Mansfield *et al.*, 1990; Russell, 1993; Reiser and Fischer, 1993). Russell (1993) has ascribed the polar organization of the embryo to the redistribution of endoplasmic reticulum, plastids, and mitochondria after fertilization of the egg cell. More dynamic examples of zygote polarization include *Papaver nudicaule* and maize where the positions of the nucleus and vacuole in the unfertilized egg are exchanged following

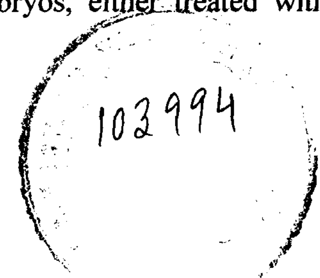
fertilization, with the nucleus becoming oriented toward the chalazal pole (Olson and Cass, 1981; van Lammeren, 1981). Because the asymmetry of zygote reflects the polar organization of the egg, it is assumed that embryonic polarity may be established during embryo sac development.

The polarity of the embryo defines an axis upon which the body plan of the plant is elaborated. Continued embryonic development may be viewed conceptually as a series of partitioning events that sequester increasingly more specialized regions. An early compartmentation step appears to involve the creation of three spatial domains *viz.* apical, central and basal along the longitudinal axis of the embryo (Jürgens *et al.*, 1991; Mayer *et al.*, 1991). The apical domain is composed of the cotyledons, shoot apex, and upper hypocotyl; the central domain includes the majority of the hypocotyl; and the basal domain consists primarily of the root. Initial evidence for these domains was derived from studies of Arabidopsis mutants defective in the specification of the plant's spatial organization (Mayer *et al.*, 1991). These included (i) the apical domain mutant, *gurke*, which lacks cotyledons and the shoot apical meristem, (ii) the central domain mutant, *fackel,* in which the central domain is lacking and the cotyledons appear to be fused directly to the roots (iii) the basal deletion mutant, *monopteros*, which lacks hypocotyl as well as root and (iv) the terminal mutant, *gnom*, which has deletions in both apical as well as basal domains. Although pattern formation mutants were originally identified as seedlings, the morphological consequences of the mutations can be first observed at an early embryonic stage, suggesting that the domains are established early in embryogenesis (Mayer *et al.*, 1991). This is evident from the non formation of cotyledon primordial in embryos of the apical deletion mutant *gurke.* On

the basis of analysis of the mutants defective in the specification of the plant's spatial organization, Mayer *et al.*, (1991) have suggested the involvement of 40 zygotically active genes in the establishment of body organization in Arabidopsis. However, because mutants were identified only at the seedling stage, this may be a minimal estimate if other mutants defective in pattern formation are unable to complete embryogenesis and produce viable seedlings. Although the mutant phenotypes *GURKE* (*GRK*) and *PASTICCINO* (*PAS*) (Torres Ruiz *et al.*, 1996; Faure *et al.*, 1998; Vittorioso *et al.*, 1998) have been reported to have impaired formation of cotyledons, these mutations are pleiotropic and the gene products are either unknown or cannot be easily designated in a signalling context. Berleth and Chatfield (2002) have identified two genes *viz.* *SHOOT MERISTEMLESS* (*STM*) and *WUSCHEL* (*WUS*) which have been suggested to act as permanent positive regulators of stem cell identity in the SAM. *STM* encodes a maize KNOTTED-type homeo-domain transcription factor and is expressed late during the globular stage in incipient SAM (Long *et al.*, 1996). Loss-of-function mutations in *STM* resulted in embryos, which displayed fused cotyledons and failed to produce SAM (Barton and Poethig, 1993).

The polar axis of the embryo is defined with the specification of the shoot and root apical meristems. Formation of the embryonic root apex occurs in a process that involves derivatives of both the basal and apical cells of a two-celled embryo. The hypophyseal region, derived from the topmost cell of the suspensor, is incorporated into the embryo proper, where it gives rise to part of the root cap and its initial cells and the ground meristem initial cells (Tykarska, 1979). By contrast, delineation of the embryonic shoot apex is much more cryptic. Clonal analyses in cotton and maize has

indicated that the embryonic region that gives rise to the shoot apical meristem, the epiphysis, is determined in an early globular stage embryo (Christianson, 1986; Poethig *et al.*, 1986). In relationship to the embryonic domains, epiphysis formation has been viewed as a partitioning event in the apical domain that segregates cells with the potential to become the shoot apical meristem (Fernandez *et al.*, 1991). Support for this view has been inferred from the analysis of 2S and 12S storage protein mRNA accumulation in oilseed rape embryos (Fernandez et al., 1991; Thomas, 1993). The non accumulation of transcripts for storage proteins in the epiphyseal cells and their abundance in cotyledonary cells has been interpreted to indicate that the two regions are segregated functionally within the apical domain (Thomas, 1993). Moreover, the lower boundary of the epiphysis is indicated by a distinct line of cell walls that separate adjacent cells with and without storage protein mRNA. This boundary transcends tissue types, extending through the ground tissue and into the protoderm, and coincides with the position of the O' line. This observation is consistent with the idea that the progenitors of the shoot apex are partitioned early in embryogenesis, because the first transverse cell division of the embryo proper that generates the O' line occurs before the cell cleavage that delineates the protoderm. Morphological changes during the globular stage to heart stage transition are the first visible sign of the formation of the two embryonic organ systems, the cotyledons and the axis. The emergence of the cotyledons from a radially symmetrical globular stage embryo represents another partitioning event in the apical region, indicating that groups of cells are induced to proliferate at specific sites. Liu *et al.*, (1993) have suggested the involvement of polar auxin transport in directing these localized cell divisions. Globular stage embryos, either treated with


103994

auxin transport inhibitors or derived from plants genetically defective in polar auxin transport, do not produce two cotyledonary lobes but, rather, form fused cotyledons that develop as ring-shaped structures above shoot apices. This requirement for polar auxin transport is specific to the globular stage because auxin transport inhibitors applied to heart-shaped embryos do not induce fused cotyledons. This implies that a conduit for auxin transport must be formed in the globular stage embryo to signal the site of cotyledon formation or that auxin is perceived differently by distinct groups of cells in the apical region.

Plant development is integrated by long-range signals, such as growth factors, which are transported along the shoot-root axis (Schiavone and Cooke, 1987; Lyndon, 1990; Schiavone and Racusen, 1990; Fischer and Neuhaus, 1996; Creelman and Mullet, 1997; Kende and Zeevaart, 1997). Abscisic acid (ABA) has been invoked as the key hormone which is required throughout the process of seed maturation (Nambara and Marion-Poll, 2003; Finch-Savage and Leubner-Metzger, 2006). The importance of its role in seed maturation had been shown earlier in immunomodulation experiments in which reductions in ABA activity in seeds resulted in a switch from the seed maturation program to a germination program (Phillips *et al.*, 1997). ABA signaling has been shown to be intimately connected to the expression of four master regulator genes viz. ABSCISIC ACID INSENSITIVE 3 (*ABI3*), FUSCA3 (*FUS3*), LEAFY COTYLEDON 1 (*LEC1*) and *LEC2* (Finkelstein, *et al.* 2002). Cadman, *et al.* (2006) have suggested that the induction of dormancy during later late-maturation phase was associated with a high ratio of ABA to GA which was mainly regulated by the dynamic balance between ABA biosynthesis and degradation in the seeds. Although the trigger for ABA synthesis

is yet to be identified, expression studies of genes involved in ABA biosynthesis (including ZEAXANTHIN EPOXIDASE [ZEP] and 9-cis-EPOXYCAROTENOID DIOXYGENASES [NCEDs]) and catabolism (such as CYP707As encoding ABSCISIC ACID 80-HYDROXYLASES) have highlighted their key roles in the spatial and temporal accumulation of ABA in seeds (Nambara and Marion-Poll, 2005). Nevertheless, ABA accumulation during seed maturation has not been found to be greatly affected by any alteration in ZEP expression, indicating that the regulatory control of ABA biosynthesis might occur downstream of epoxycarotenoid synthesis (Frey *et al.*, 2006). CYP707A genes are known to be major regulators of ABA degradation in *Arabidopsis* (Okamoto *et al.* 2006). CYP707A1 and CYP707A2 have been shown, respectively, to play roles in the reduction of ABA content in the embryo at mid-maturation and in both the embryo and endosperm during late maturation (Okamoto *et al.* 2006). The expression of CYP707A genes has also been found to be the major mechanism regulating ABA catabolism in the seeds of *Phaseolus vulgaris* (Yang and Zeevaart, 2006) and *Hordeum vulgare* (Millar *et al.*, 2006). While the barley, HvNCED2 gene has been shown to be responsible for a significant increase in ABA levels during grain development, HvCYP707A1 is responsible for a rapid subsequent decrease in ABA (Chono *et al.*, 2006). Interestingly, whereas HvNCED2 expression is regulated in a growth dependent manner, HvNCED1 and HvCYP707A1 are both affected by environmental conditions during grain development, highlighting the importance of environmental controls on ABA homeostasis. ABA has also been reported to suppress GA biosynthesis in developing seeds through the induction of AtGA2ox6, which reduces the levels of bioactive Gas (Seo, 2006). Moreover, ABA-

mediated suppression of GA biosynthesis occurs independently of cell type, suggesting that ABA itself or ABA signals are transmitted to other cells to regulate GA biosynthesis. Abscisic acid response mutants analysis revealed that the ABA signal transduction pathway involves protein kinases, phosphatases and transcription factors (Finkelstein *et al.*, 2002; Himmelbach *et al.*, 2003). Although a complete picture of the signaling mechanisms responsible for generating the apical–basal polarity of the embryo is lacking, several lines of evidence have implicated the hormone auxin as a key player in generating the apical–basal polarity in developing embryos (Jenik and Barton, 2005; Leyser, 2006; Krogan and Berleth, 2007). Studies on mutant phenotypes suggest that auxin signal transduction is required for cell division and vascular differentiation in the basal domain which comprises hypocotyl, radical and primary root meristem, (Berleth and Chatfield, 2002). Liu *et al.*, (1993) have emphasized that polar transport of auxins was required to direct cell divisions in the globular stage embryo but not in heart shaped stage embryo. Auxins have long been known to stimulate root formation and could therefore serve as signals for root initiation. Consistent with this assumption, the uppermost suspensor cell have been reported to fail in adopting 'hypophyseal' identity in embryos mutant for any one of the auxin signal transduction genes of the *MP*, *BDL* and *AXR6* class. These genes have been suggested to be involved in either regulation of cell differentiation through auxin signalling or auxin signal transduction within the uppermost suspensor cell or for both (Berleth and Chatfield, 2002). However, the fate of cells in root apical meristem (RAM) has been speculated to be largely controlled by positional cues. Scheres and Heidstra (1999) have identified one such positional cue, which originated from the Quiescent Center (QC) and conferred stem cell identity to

cells surrounding the QC. Sabatini *et al.* (1999) have, however, suggested that the positioning of QC was conferred by auxin which acted as a positional signal.

Several studies have identified spatial and temporal changes in metabolite profiles that occur during legume seed maturation (reviewed by Borisjuk, *et al.*, 2003). Metabolite concentration gradients provide signals for the onset of the seed maturation phase and its regulation. Among these metabolites, sugars, and notably the embryo sucrose:hexose ratio, are major components of the signaling pathway triggering the transition phase (reviewed by Weber *et al.*, 2005). This change in sugar balance has been suggested to be related to the formation of transfer cells, establishing an epidermis-localized sucrose uptake system (Offler *et al.*, 2003). Free hexose levels decrease dramatically in the embryo while a strong and transient increase in sucrose uptake occurs. Sucrose signals, which are known to have the potential to act on transcriptional and post-transcriptional regulation, control storage and differentiation processes through the regulation of metabolic enzymes, gene expression and activity (Gibson, 2005; Rolland *et al.*, 2006). The expression of the sucrose transporter AtSUC5 in the endosperm has been shown to play a major role in the progression into transition and storage phases in Arabidopsis (Baud *et al.*, 2005). Further, the loss-of-function apetala2 (ap2) mutation in the plant was shown to cause a gain in seed mass owing to increases in embryo cell size and number (Gomez *et al.*, 2006). This phenotype is the consequence of an increased hexose:sucrose ratio in ap2 seeds, which might promote an extended period of cell division. AP2, better known as a floral homeotic gene, might regulate the transition phase through its action on sugar signaling. Sugar signaling could thus be assumed to be a ubiquitous regulatory system involved in seed maturation. Seed

maturation is also controlled metabolically via nitrogen availability (Rolletschek *et al.*, 2005; Hernandez-Sebastia *et al.*, 2005). Fait *et al.* (2006) have identified a second metabolic switch that operates during the transition to the desiccation phase and in which seed metabolism fundamentally changes from the accumulation of oil and storage protein to increase in the content of free amino acids, sugars and degraded fatty acids. This metabolic switch could be involved in late-maturation control of seed development. Buitink *et al.* (2006) have shown that sucrose accumulation was a prerequisite for the acquisition of desiccation tolerance during this last developmental phase. In addition to providing the required energy for the desiccation phase the late metabolic switch has been suggested to trigger production of metabolites that can become rapidly available during imbibitions (Chia *et al.*, 2005). However, the underlying mechanisms regulating the role of metabolites during the embryo maturation phase are still not understood fully.

Over the years extensive work has been carried out on changes in protein profiles in seeds during different stages of development (Ruuska, *et al.*, 2002; Vensel *et al.*, 2005; Hajduch *et al.*, 2005; Giavalisco, *et al.*, 2005; Sheoran, *et al.*, 2006; Hajduch *et al.*, 2006; Yang *et al.*, 2007; Park *et al.*, 2007; Gallardo *et al.*, 2007; Mechin *et al.*, 2007; Agrawal *et al.*, 2008; Dam *et al.*, 2009; Balbuenaa *et al.*, 2009). Typically, proteins involved in biosynthetic pathways are expressed in the early stage of embryogenesis and start degrading during the seed filling stage when the proteins involved in seed maturation proteins like SSPs and desiccation tolerance proteins are synthesized. Proteins involved in seed maturation and desiccation tolerance usually begin to down regulate earlier than the SSPs during seed germination. Yang *et al.*,

(2007) have demonstrated differential accumulation of 148 proteins during seed germination in rice. While the accumulation of 69 proteins was up-regulated that of 63 proteins was down-regulated. The up-regulated proteins included α-amylase, UDP-glucose dehydrogenase, fructokinase, phosphoglucomutase and pyruvate decarboxylase. On the other hand, the down-regulated proteins included those associated with seed maturation like early embryogenesis proteins, late embryogenesis abundant proteins and proteins associated with desiccation tolerance. Gallardo *et al.* (2003) profiled and identified 84 proteins at various stages of seed development in *Medicago truncatula.* On the basis of dynamics of changes in proteome and transcriptome during seed development in *Medicago truncatula,* Gallardo *et al.*, (2007) have observed >2 fold changes in 77% of the proteins and 50% of transcripts during seed development in the plant. On the other hand the percentage of transcripts showing >2 fold upregulation between 5 to 13 DAF in Arabidopsis was only 35% (Ruuska *et al.*, 2002). Genes related to the biosynthesis of storage components showed several distinct temporal expression patterns. While a number of genes encoding core fatty acid synthesis enzymes displayed a bell-shaped pattern of expression between 5 and 13 days after flowering the expression of storage proteins and other known abscisic acid–regulated genes showed marked increases during later part of mid maturation stage. This suggests a more complex network of gene expression during the process of seed development.

In recent years, molecular biology approaches have also been used to investigate embryogenesis. Several genes that are related to cell differentiation, morphogenesis, desiccation tolerance and signal transduction are expressed during embryogenesis and function as part of the embryogenesis program (Ikeda *et al.,* 2006).

While a large proportion of the proteins found at early developmental stages in soybean seeds have been shown to belong to the 2S albumin fraction those accumulating during the maturation phase have been shown to belong to the 7.5S and 11.8S fractions (Eldridge *et al.*, 1966; Catsimpoolas *et al.,* 1969; Catsimpoolas and Leuthner, 1969). In developing cotyledons of *Pisum sativum,* the globulins have been shown to be synthesized till 27 DAF while the accumulation of albumins stopped after 24 DAF (Basha, 1974).

Consistent with differential accumulation of proteins during various stages of seed development, variations have also been observed in transcript levels for different proteins during various stages of seed development. Dong *et al.*, (2004) have demonstrated the presence of higher levels of transcripts coding for enzymes involved in biosynthetic pathways, structural proteins, regulatory factors, and transport or protective proteins at 15 DAP in embryos of *Brassica napus.* On the other hand, embryos at 25 DAP had higher levels of mRNAs for storage proteins such as napin, cruciferin and oleosin.

Seed development proceeds through a series of spatially and temporally regulated gene expression steps and these steps are usually characterized by specific molecular and metabolic events which control the course of physiological changes. Control of expression can either be as a 'switch' *i.e.* whether the gene is expressed or not expressed at all, or as a 'regulator'- i.e. whether the gene is synthesized at a high or low rate. This dual level process is because factors or conditions that turn a gene 'on' may not lead to high expression level of that gene (Gatehouse *et al.,* 1986). Transcript profiling of *Brassica napus* has lead to identification of two groups of genes. While the

genes of one group show diverse temporal expression patterns throughout seed development and are expressed only in seeds, the genes belonging to the other group do not show any tissue specific expression patterns (Dong *et al.,* 2004). Further, the seed specific genes showed developmental stage specific expression patterns with genes coding for vacuolar processing enzymes required for epidermal surface development expressing preferentially during very early stage of maturation viz. 10 DAP (Tanaka *et al.*, 2001; Gruis *et al.*, 2002) and those coding for SSPs expressing during early maturation stages viz. 15 DAP (Duan and Sun, 2005). While the genes encoding storage proteins including LEA proteins were found to have low initial of expression at very early stage, their expression increased with progressive development reached its peak at mid maturation stage (Oliveira *et al.*, 2007; Bies-Ethéve *et al.*, 2008; Hundertmark and Hincha, 2008). On the basis of sequence similarities, LEA proteins have been grouped into at least six groups out of which group 1, group 2 and group 3 constitute the major groups (Wise, 2003). LEA proteins belonging to Group 2 are also known as 'dehydrins. These proteins are found mainly in plants including algae. Dehydrins (DHNs) are known to accumulate in seeds during late maturation stage of embryo development or during dehydration stress (Svensson *et al.*, 2002). LEA proteins have also been detected in vegetative tissues of several plants (Bies-Ethéve *et al.*, 2008). Bies-Ethéve *et al.* (2008) has suggested that regulation of LEA genes with similar expression profile may not necessarily involve the same regulatory pathway. DHNs are encoded by a various multigene family and are differentially regulated. For instance, 13 *Dhn* genes identified in barley are dispersed over seven genetic map locations (Choi *et al.,* 1999; Sevensson *et al.*, 2002) and regulated variably by drought, low temperature

and embryo development (Tomassini *et al.*, 2008). DHNs have been found to be localized in cytosol (Roberts *et al.*, 1993), nucleus (Houde *et al.*, 1995), chloroplast (Artus *et al.*, 1996), vacuole (Heyden *et al.*, 2002), plasma membrane and protein bodies (Asghar *et al.*, 1994; Egerton-Warburton *et al.*, 1997; Puhakainen *et al.*, 2004). Zhu *et al.* (2000) and Nylander *et al.* (2001) have suggested that members of the *Dhn* multigene family have rather distinct biological functions including protection of the native proteins during tissue desiccation. Because of hydrophilicity, high content of GLY (>20%) and the absence of a defined 3 dimensional structure in pure form, DHNs have been referred to as "Intrinsically Disordered/Unstructured Proteins" (IUPs) or "hydrophilins" (Garay-Arroyo *et al.*, 2000; Tompa, 2005; Kovacs *et al.*, 2008). This disordered/ unstructured nature confers structural flexibility to adapt to protein environment changes such as water potential, pH, ionic strength.

Analysis of seed storage proteins gene expression have indicated that the expression of SSP genes and the accumulation of SSPs was restricted to either endosperm/embryo or cotyledons (Greenwood and Chrispeels, 1985; Sengupta-Gopalan *et al.*, 1985; Hall *et al.*, 1999; Fujino *et al.*, 2001; Milisavljevic *et al.*, 2004; Jain *et al.*, 2004). Developmental regulated expression of SSP genes has also been observed in *Phaseolus* (Hall *et al.*, 1999; Sengupta-Gopalan *et al.*, 1985), Brassica (Ellerstrom *et al.*, 1996; Stalberg *et al.*, 1993), *Arabidopsis* (Guerche *et al.*, 1990), Barley (Rahman *et al.*, 1983) and buckwheat (Fujino *et al.*, 2001). Mandal and Mandal (2000) have suggested that the spatial and temporal control of SSP accumulation was exerted at transcriptional rather than at post transcriptional level. Seed storage proteins (SSP) are one of the major storage materials which accumulate in high amounts during the maturation phase of

seed development. Further, the accumulation of SSP is preceded by the preparation of the embryo to differentiate into the embryonic axis and the cotyledons.

Although it was shown many years ago that master regulators interact to control the various features of seed maturation (Parcy *et al.*, 1997) the molecular basis of their complex genetic interaction was only recently discovered (To *et al.*, 2006). Analysis of ABI3, FUS3 and LEC2 expression in abi3, fus3, lec1 and lec2 single mutants and in several double and triple mutants has identified the multiple regulatory links involved in seed maturation. Furthermore, analysis of the results of ectopic expression of LEC1 has demonstrated that seed storage protein (SSP) gene expression is controlled by LEC1 through the regulation of ABI3 and FUS3 expression (Kagaya *et al.*, 2005a).

Although each master regulator activates a precise set of target genes during the maturation phase, overall, these are controlled in an overlapping, rather than a hierarchical manner. The ectopic expression of Viviparous (VP1), the ABI3 maize ortholog, in *Arabidopsis* plants, induced the expression of a large number of seed-specific genes in vegetative tissues, thereby indicating the importance of VP1 and ABI3 in expression of seed specific genes (Suzuki *et al.*, 2003). The activity of VP1, a monocot gene, in *Arabidopsis*, which is a dicot, indicates that ABI3-related regulatory mechanisms could be conserved across a wide range of plant species. Ectopic expression of LEC1, LEC2 or FUS3 has also been shown to activate seed maturation pathways in vegetative tissues (Casson and Lindsey, 2006; Santos Mendoza *et al.*, 2005; Kagaya *et al.*, 2005b). LEC2 has been shown to bind specifically to the conserved RY DNA motif (Braybrook *et al.*, 2006). Although FUS3 could also directly bind the promoter of some seed maturation specific genes (Kroj *et al.*, 2003; Monke *et al.*, 2004;

Curaba *et al*., 2004), its major function during seed maturation has been suggested to be limited to restriction of the expression domain of TRANSPARENT TESTA GLABRA1 (TTG1) TF (Tsuchiya *et al*., 2004). The existence of two incompatible theories, FUS3 acting either directly or indirectly, illustrates the complexity of the genetic interactions during seed maturation that still remain to be unraveled. Except for SSP genes, for which the structure of *cis* regulatory elements is well documented, little is known about the interactions between master regulator proteins and other putative target genes. The high diversity of seed storage compounds among plant species implies that distinct biosynthesis pathways and, therefore, distinct putative regulatory mechanisms are involved. Consequently, the use of non-model plant species might be important for identifying new regulators, as well as for highlighting the differences between conserved and non-conserved seed maturation processes.

CHAPTER : III

MATERIALS AND METHODS

1. MATERIALS:

1.1 Plant Material:

Grains of common buckwheat (*Fagopyrum esculentum* Moench) were obtained from the 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 (0.45 μm pore size) and membrane filters (0.2 μm pore size) from Fluka was used for Southern blotting and filter sterilizing the chemicals, respectively.

2. PROTOCOLS:

2.1 Isolation of Total RNA:

Grains of common buckwheat were harvested at different stages of development *viz.* 5-7 DAF, 10-13 DAF, 16-20 DAF and 23-25 DAF and surface sterilized in the

Fig. 3.1: Close up of the common buckwheat plant (*fagopyrum esculentum* Moench) in flowering stage

Fig. 3.2: Grains of common buckwheat harvested at different maturation stages of development.
5-7 DAF=Very early developmental stage; 10-13=Late early developmental stage; 16-20 Mid maturation developmental stage; 23-25 DAF=Late maturation stage; 25-30 DAF=Matured stage.

Fig.3.1

Fig.3.2

laboratory by treatment with 0.01% $HgCl_2$ solution for 5 minutes followed by washing with sterile distilled water. Mature grains of common buckwheat were also sterilized in the same manner by treatment with 0.01% $HgCl_2$ solution for 5 minutes and the sterilized mature grains were germinated on a clean germination paper in an incubator at 27°C. The germinated seedlings were harvested after 7 days of incubation and treated again with 0.01% $HgCl_2$ solution for 5 minutes for sterilization. The $HgCl_2$ treated grains/tissues were washed with sterile DEPC treated (0.2% DEPC) distilled water and crushed to a fine powder under liquid nitrogen in a sterile pestle and mortar. Total RNA was isolated from a suitable mass of the powdered tissue from the grains according to the method of Lopez-Gomez and Gomez-Lim (1992).

A suitable mass of the powdered tissue was transferred to a 1.5ml eppendorf tube and 2 volume of extraction buffer was added to it. The mixture was vortexed for 30 seconds to disperse the clumps of powdered mass formed during the addition of extraction buffer. Vortexing was followed by addition of β-mercaptoethanol to a final concentration of 1.0%, 0.2V absolute ethanol and 0.11V of 5M potassium acetate. The homogenate was vortexed again for 1 minute to mix the contents thoroughly. Vortexing was followed by addition of chloroform:isoamyl alcohol (49:1) in the ratio of 1:1 and vortexed for 1 minute to mix the contents. The mixture was centrifuged at 13000 rpm for 10 min at 4°C in a Heraeus (Biofuge) table top centrifuge to separate the aqueous phase from the organic phase. 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 3M was 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 pelletted by centrifugation at 13,000 rpm for 10 min. at 4°C. The pelletted RNA was washed twice with 80% ethanol, vacuum dried and dissolved in appropriate volume of DEPC treated autoclaved water. The isolated RNA was electrophoressed 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.

For electrophoresis under denaturing conditions, approximately 1μg of RNA was mixed with 1.5μl of 10X MOPS, 3μl of 37% formaldehyde and 7.5μl of deionised formamide and the solution was heated at 65°C for 10 minutes in dry heating bath to denature the RNA. The denatured RNA was cooled on ice for 10 min. after which a suitable aliquot of the sample was mixed with loading buffer in the ratio of 6:1. The mixture was loaded on the well of a 1% formaldehyde agarose gel and electrophoressed at 50V for 2 hr. in 1X MOPS buffer. After the electrophoresis was complete, the gel was removed from the tray and stained with 0.5μg/ml ethidium bromide for 15 minutes. The gel was subsequently destained by several washes in DEPC treated water to remove the excess stain and till the background fluorescence disappeared. The RNA was visualised by exposing the gel to UV light on a UV transilluminator.

The isolated RNA was quantified spectrophotometrically by measuring absorbance of the sample at 260nm in a Perkin Elmer Lamda35 UV-Vis

spectrophotometer. An absorbance value of 1 at 260 nm was considered equal to 40 μg/ml of the RNA. The concentration of RNA in the solution was calculated according to the formula: RNA (μg ml^{-1}) = A_{260} x dilution factor x 40. The purity of RNA in the sample was determined by measurement of the absorbance of the sample at 260 and 280 nm and determination of the A_{260}/A_{280} ratio.

All the plastic and glass wares used for RNA isolation were pretreated with 0.2% DEPC prior to surface sterilization by autoclaving at 15psi for 15 min. The glasswares used for isolation of RNA were then baked in an oven at 120° C for 12 hours before use in order to degrade any contaminating RNases. All the solutions used for RNA isolation, except Tris-HCl, were treated with 0.2% DEPC so as to inactivate the RNases. In case of Tris-HCl, the solution was filter sterilized using membrane filters with a pore size of < 0.2 μm.

2.2 Reverse Transcriptase-Polymerase Chain Reaction

Total RNA from grains harvested at different stages of development *viz.* 5-7 DAF, 10-13 DAF, 16-20 DAF, 23-25 DAF and also from grains at early stages of germination was used as a template for synthesis of first cDNA strand using the Stratagene's First Strand cDNA synthesis kit according to the manufacturer's protocol. The reaction mixture for synthesis of first strand of cDNA comprised of:

- Total RNA (5 μg) 11.5 μl
- Oligo$(dT)_{18}$ primer (0.5μg/μl) 1.0 μl
- 5X reaction buffer 4.0 μl
- Ribonuclease inhibitor (40U/μl) 0.5 μl
- 10 mM dNTP mix 2.0 μl
- M-MuLV reverse transcriptase (200U/μl) 1.0 μl

A suitable volume of the RNA solution representing 5μg of total RNA was mixed with 1μl (0.5μg) of Oligo$(dT)_{18}$ primer and incubated at 70°C for 5 min. in a circulatory

water bath followed by quick chilling over ice for 5 min. Appropriate volumes of 5X reaction buffer, ribonuclease inhibitor, and 10mM dNTP mix were added to the chilled mixture and the solutions incubated at 37°C for 5 min. Incubation at 37° C was followed by addition of appropriate volume of M-MuLV reverse transcriptase and the solution incubated again at 37°C for 1 hour. The reaction was stopped by heating at 70°C for 10 min. The reaction mixture for 2nd strand synthesis comprised of:

▪ Ist strand reaction mix.	30 μl
▪ Klenow's fragment buffer	20 μl
▪ DTT	5 μl
▪ RNase H buffer	5 μl
▪ dNTPs	4 μl
▪ RNase H	1 μl
▪ Random primer	2 μl
▪ Klenow's fragment	3 μl
▪ DEPC treated water	130 μl

The mixture was incubated at 16°C for 2 hr. in a circulatory water bath (Pharmacia Biotech). Double stranded cDNA was recovered from the reaction mixture by mixing with 1 volume of phenol:choloroform:isoamyl (25:24:1) and subsequent recovery of the aqueous phase after centrifugation of the mixture at 13000 rpm for 10 min at 4°C in a Heraeus (Biofuge) table top centrifuge. DNA was precipitated from the aqueous phase by addition of 1/10th volume 3M sodium acetate (pH 5.2) and two volumes of 100% ethanol and overnight incubation of the mixture at -20°C. 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 cDNA was used as the template for PCR amplication of the target DNA with primers designed from the nucleotide sequences available in the gene bank

databases (Table 3.1). The amplification programme comprised of one cycle of hot start at 94°C for 5 minutes, 35 cycles comprising of denaturation at 94°C for 1 min., annealing at 60°C for 1 min. and chain extension at 72°C for 1 min. followed by elongation at 72°C for 10 mins. The PCR cycle is diagrammatically represented in fig 3. 2. A typical 25µl reaction volume in a 0.2 ml GeneAmp reaction tube contained:

▪ Sterile H_2O	12.7 µl
▪ 10X Reaction buffer	2.0 µl
▪ 10mM dNTP mix	1.0 µl
▪ primer 1	0.5 µl
▪ primer 2	0.5 µl
▪ cDNA template (~0.5ng)	5.0 µl
▪ $MgCl_2$ (25mM)	2.0 µl
▪ Taq DNA polymerase (5units/(µl)	0.3 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, 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 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.

Table:3.1: List of primers used for amplifying abundantly expressed seed storage protein genes during seed maturation.

PRIMER	SEQUENCE	GC content (%)	Tm (°C)
CJ1F	5' CCGGGCAGGTATCTCGGCTAG 3'	66.7	68.4
CJ2R	5' ATGTCTCCGGGCATCCTGG 3'	63.2	64.5
NKC5	5' GGATCCGGATTGGAGCAAGCGTTCTGC 3'	55.5	71.0
NKC6	5' GAAACGCTCCCTCTCCTTCTCATC 3'	54.1	60.0

Fig. 3.3: Schematic representation of a typical three step cycle parameter for PCR amplification.

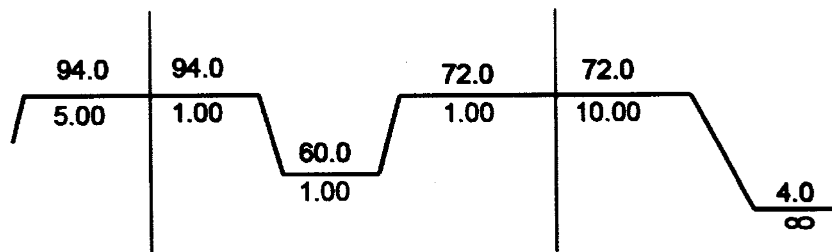


Fig.3.3

2.3 Construction of cDNA library

cDNA library was constructed from RNA isolated from grains of common buckwheat harvested at mid maturation stage *viz.* 16-20 DAF and from grains at early stages of germination (7 days of incubation) with Stratagene's cDNA Synthesis Kit as per the manufacturers' protocol. Prior to use the isolated RNA was cleaned up with Qiagen RNeasy Plant RNA clean up kit as per the manufacturer's protocol. The steps involved in cDNA library construction included (i) Isolation of mRNA from total RNA, (ii) cDNA synthesis and directional ligation into the Uni-Zap XR vector and, (iii) packaging into appropriate host cells.

Isolation of mRNA from total RNA: The mRNA was isolated from total RNA following the magnetic separation method based on affinity-binding to oligo $(dT)_{18}$. A suitable volume of the RNA solution representing 25µg of total RNA was mixed with 5µl (1µg/µl) of biotinylated oligo$(dT)_{18}$ in 1.5ml eppendorf tube containing 5µl of 10mM Tris-HCl pH 8.0 and 0.5M KCl. The mixture was heated at 70°C for 5 min. in order to uncoil the secondary structures thereby helping the primers to bind properly. This was followed by incubating the mixture containing the RNA sample at room temperature or 25 ± 2°C for 2 min. 500µl of streptavidin magnetic beads were transferred to another eppendorf tube and the beads separated from the aqueous solution by creating a magnetic field. The magnetic beads were washed twice with 1X PCR buffer and neutralized by two washes in a neutralizing solution comprising of 0.5M KCl in 10mM Tris HCl (pH8.0) and finally suspended in the same buffer. The solution of total RNA solution and the biotinylated oligo$(dT)_{18}$ was mixed with the paramagnetic beads solution and incubated at room temperature for 20 min. after

which mRNA was separated from the total RNA solution by keeping a magnet to the side of the tube containing the mixture thereby creating a magnetic field which would attract the paramagnetic beads with the adsorbed mRNA. The aqueous phase was discarded and the beads along with the adsorbed mRNA was washed thrice with 10mM Tris- HCl buffer (pH 8.0) containing 0.5M KCl followed by one wash with 10mM Tris-HCl buffer (pH 8.0) containing 0.25M KCl. The washed beads with the adsorbed mRNA was suspended in 50µl of DEPC treated autoclaved water and incubated at 70°C for 5 min. to dissociate the mRNA from oligo$(dT)_{18}$ + paramagnetic beads. The oligo$(dT)_{18}$ + streptavidin magnetic beads were then separated from the mRNA in the aqueous solution by applying a magnetic field with the help of a magnet and the aqueous phase containing the mRNA transferred to a fresh new eppendorf tube. The suspension of oligo$(dT)_{18}$ + streptavidin magnetic beads was washed again with 5µl of DEPC treated autoclaved water in the same manner as described above to rescue any traces of mRNA from the beads. The tubes containing the isolated mRNA solution were centrifuged at 5000 rpm for 1 minute and the supernatant transferred to a fresh tube. mRNA was precipitated from the solution with $1/10^{th}$ volumes of 3M sodium acetate (pH5.2) and 2.5 volumes of absolute alcohol after overnight storage at -20°C. The precipitated mRNA was pelletted by centrifugation for 15 minutes at 14000 rpm in a refrigerated microcentrifuge. The pellet was washed twice with 70% alcohol followed by drying under vacuum in a speed Vac concentrator. The dried pellet of mRNA was subsequently dissolved in 37.5µl of DEPC treated autoclaved water.

cDNA synthesis and library construction: The isolated mRNA was used as a template for cDNA synthesis using the Stratagene's cDNA synthesis kit as per the manufacturer's manual. The reaction mixture for synthesis of first strand of cDNA comprised of:

- mRNA (5µg) : 37.5 µl
- 10X first strand buffer : 5.0 µl
- First strand methyl nucleotide mixture : 3.0 µl
- Linker primer (1.4 µg/µl) : 2.0 µl
- RNase block Ribonuclease Inhibitor (40U/µl) : 1.0 µl

The mRNA solution was incubated at 70°C for 5 min. to ensure uncoiling of the secondary structures and then transferred to the tube containing other components of the reaction mixture. The eppendorf tube containing the reaction mixture for Ist strand synthesis was incubated at room temperature for 10 min. for allowing the linker primer to anneal to the mRNA template. 1.5µl Reverse Transcriptase was added to the reaction mixture and the solutions incubated at 42°C for 1 hour after which the tube containing the reaction mixture was incubated over ice for 5 min. The reaction mixture for 2nd strand synthesis comprised of:

- 10X second strand buffer : 20.0 µl
- dNTP mixture : 7.0 µl
- Sterile distilled water : 110.0 µl
- RNase H (1.5U/µl) : 2.0 µl
- DNA polymerase I (9U/µl) : 11.0 µl

The reaction mixture for second strand synthesis was gently vortexed and incubated at 16°C for 2.5 hr. in a circulatory water bath. The reaction was stopped by placing the tube immediately on ice. Double stranded cDNA was made blunt by incubating at 72°C for 30 min. after addition of 23µl of blunting dNTP mixture and 2µl (2.5U/µl) of cloned *pfu* DNA polymerase. The reaction mixture containing the blunt ended double stranded DNA was mixed with 1 volume of phenol:choloroform:isoamyl (25:24:1) and the

aqueous phase was recovered after centrifugation at 13000 rpm for 10 min at 4°C in a Heraeus (Biofuge) table top centrifuge. 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 followed by overnight incubation of the mixture at -20°C. 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 resuspended in 9µl of *Eco*RI adaptors. The tube containing the blunted cDNA and the *Eco*RI adaptors was incubated at 4°C for 30 mins. after which 1µl of 10X ligase buffer,1µl of 10mM rATP and 1µl of T4 DNA ligase (4U/µl) were added to it. The mixture was centrifuged at 1000 rpm for 30 sec. and then incubated overnight at 8°C in a circulatory water bath for ligating the *Eco*RI adaptors to the cDNA blunt ends. The reaction was stopped by heat inactivation at 70°C for 30 min. after which the reaction mixture was centrifuged again at 1000 rpm for 30 sec. The mixture was cooled at room temperature for 5 min. for subsequent phosphorylation of the *Eco*RI adaptor ends. The reaction mixture for phosphorylation of the *Eco*RI adaptor ends comprised of:

- Ligated cDNA : 12 µl
- 10X ligase buffer : 1 µl
- 10mM rATP : 2 µl
- Sterile water : 5 µl
- T4 polynucleotide kinase (5U/µl) : 2 µl

The mixture was incubated at 37°C for 30 min. followed by heat inactivation of the reaction by incubation at 70°C for 30 min. The cDNA ligated with *Eco*RI adaptors were digested with *Xho*I and the digested cDNA was size fractionated through a drip column containing Sephacryl S-400-HR. The cDNA fractions collected were extracted with phenol:chloroform (1:1) and recovered by precipitating with 100% ethanol. Recovered cDNA inserts were ligated into Uni-ZAP XR vector. The ligation reaction consisted of:

- cDNA insert (~100ng) : 9.0 μl
- 10X ligase buffer : 0.5 μl
- 10mM rATP (pH7.5) : 0.5 μl
- Predigested Uni-ZAP XR vector (1μg) : 1.0 μl
- T4 DNA ligase (4U/μl) : 0.5 μl

The reaction tube containing the above mixture was incubated overnight at 12°C and the library packaged into $RecA^-$ *E. coli* host strain XL1-Blue MRF' cells. After the packaging reaction, the primary library was checked by Blue-white colour selection method with IPTG and X-gal as per the cDNA synthesis kit manual. The primary library was amplified in NZY agar as per the manufacturers' protocol; the amplified library was stored at -80°C as aliquots in 7 % (v/v) DMSO for further use.

The cDNA library constructed from grains of common buckwheat harvested at mid maturation stage *viz.* 16-20 DAF and early stages of germination was used as the template for PCR amplication of the target DNA with primers designed from the nucleotide sequences available in the gene bank databases (Table 3.1). The amplification reactions were performed in a thermal cycler (Perkin Elmer) with the amplification programme comprising of one cycle of hot start at 94°C for 5 min; 35 cycles of denaturation at 94°C for 1.0 min, annealing at 60°C for 1.0 min, and polymerization/primer extension at 72°C for 1.0 min followed by one cycle of chain elongation at 72°C for 10 min. A typical 25μl reaction volume in a 0.2 ml GeneAmp reaction tube contained the following components:

- Sterile H2O : 17.2 μl
- 10X Reaction buffer : 2.0 μl
- 10mM dNTP mix : 0.5 μl
- primer 1 : 0.5 μl
- primer 2 : 0.5 μl
- cDNA library (~0.5ng) : 2.0 μl
- $Mgcl_2$ (25mM) : 2.0 μl
- Taq DNA polymerase (3 units/(μl) : 0.3 μl

Positive (-primers) and negative (-template cDNA) controls were also maintained along with the other reaction mixtures. After the amplification, the entire reaction mixture was electrophoressed on 1.2% agarose gel at 50 volts for 5 hours. The DNA on the agarose gel was visualized under UV light on a transilluminator after staining the gel with $0.5gml^{-1}$ ethidium bromide staining solution. The region of agarose gel having the amplified DNA was excised from the gel with a sterile blade and the DNA was eluted with QIAEX-II gel extraction kit as per the manufacturer's protocol. The eluted DNA was lyophilized and stored at -80°C till it was used.

2.4 Northern hybridization of RNA by dot blotting

Total RNA isolated from grains harvested at 10-13 DAF, 16-20 DAF, 23-25 DAF and 30-35 DAF were dot blotted on a positively charged nylon membrane as per Sambrook *et al.*, (1989). RNA samples were denatured in a reaction mixture consisting of:

- RNA (2.0 µg) : 10.0 µl
- 20XSSC : 10.0 µl
- Formaldehyde (37%) : 7.0 µl
- Formamide : 20.0 µl

The mixture was incubated at 65°C for 10 minutes in a circulatory water bath (Pharmacia Biotech) followed by rapid cooling in ice. For dot blotting, the positively charged nylon membrane was soaked in 10X SSC buffer treated with DEPC for 10 minutes and placed in the Trans-Vac TE 80 vacuum blotter (Hoefer). The suction was switched on and 10X SSC buffer was allowed to filter through the membrane. Prior to loading the denatured RNA sample, 2 volume of ice cold 20X SSC buffer was added and the sample was loaded on the slots. The solution containing the denatured RNA was

allowed to filter through under a negative pressure of 20 mbars. Thereafter, 1ml of 10X SSC buffer was allowed to filter through each slot. The membrane was removed, air dried and the transferred RNA was immobilised by UV crosslinking in a UV-croslinker. The membrane with the denatured total RNA was used for hybridizing with appropriate radiolabelled genomic DNA probes.

2.5 Preparation of radio labelled probe:

The DNA probe used for hybridization was radiolabelled with α-P^{32}[dATP] 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's 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 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, 1.0µl each of random hexamer, dTTP, dCTP, dGTP, 5.0µl each of nuclease free water and [α-P^{32}]dATP (~50µCi) and 1.0µl of klenow's fragment were added 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 1ml disposable syringe for collection of the eluant 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 P^{32}-labelled DNA with a specific activity of $5\text{-}6x10^{-7}$cpu μg^{-1} was obtained.

2.6 Northern hybridization:

Prior to hybridization with the radiolabelled probe, the membrane carrying the RNA was incubated with 10ml 5X SSPE buffer containing 100mg 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 65°C to allow the probe to bind to the target RNA. 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 RNA.

2.7 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, 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. 500mg each of the powdered tissue was transferred to a 1.5ml microcentrifuge tubes containing 500µl 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 1/10th 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, 5μl 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 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. 10μl of DNA solution was mixed with 990μl of ultrapure water and the absorbance of the solution recorded at 260nm 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.

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 bank. The primers were designed using Primer3 online tool (Rozen and Skaletsky, 2000). Amplification reactions were carried out 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 (55-60°C, 1 min), and polymerization/primer extension (72°C, 1 min) and one cycle of chain elongation (72°C, 15 min). A typical 25μl reaction volume (in 0.2 ml reaction tube) contained the following components:

Component	Volume
▪ 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 (0.5μg 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 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.

2.8 Nucleotide sequencing and sequence analysis:

Nucleotide sequencing of the amplified DNA was carried out by automated sequencing service rendered by Bangalore Genei and TCGA, India and M/s Microsynth GmBH, Switzerland. For this purpose, the amplified DNA was gel purified, lyophilized and sent for sequencing. 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 sequences were subjected to BLAST (http://www.ncbi.nlm.nih.gov/Blast) analysis to determine homology with other sequences in the data bank and ORF FINDER of NCBI to predict the coding region and the deduced amino acid sequence. Secondary structure prediction was done using GOR IV method software (Garnier *et al.*,

1996) (http://npsa-pbil.ibcp.fr/) and hydrophilicity and flexibility analysis was done by MacVector software. The sequences were aligned to each other with CLUSTALW (www.ebi.ac.uk/clustalw). MOTIF SCAN software (http://myhits.isb-sib.ch/cgi-bin/motif_scan) was used for motif search on amino acid sequence.

3. CHEMICALS AND SOLUTIONS

3.1 Chemicals:

Tris, tri-sodium citrate, sodium chloride, magnesium chloride, calcium chloride, magnesium sulphate, lithium chloride, sodium acetate, potassium acetate, Triton X-100, SDS, ammonium acetate, sucrose, sarkosyl, agarose, ethidium bromide (EtBr), CTAB, β-mercaptoethanol, EDTA, MOPS, formamide, formaldehdye were purchased from Himedia Laboratories Ltd. DiethylenePyrocarbonate (DEPC) for isolating total RNA was procured from Sigma-Aldrich, USA.

Isopropanol, dehydrated alcohol, phenol, chloroform, formaldehyde solution, formamide, methanol and other chemicals of routine use were purchased from Sisco Research Laboratories Pvt. Ltd. India. X-Ray developer and fixer were purchased from Allied Photographics India Limited.

Streptavidin magnetic beads and biotinylated oligo$(dT)_{18}$ were procured from New England Biolabs, Inc.

3.2 Membranes and Filters

Positively charged nylon membrane (0.45 μm pore size) and membrane filters (0.2 μm pore size) from Fluka was used for Northern blotting and filter sterilizing the chemicals, respectively.

3.3 Radiochemicals:

[α-^{32}P]-dATP was purchased from Bhabha Atomic Research Centre, Mumbai.

3.4 Molecular weight markers:

DNA molecular weight markers which included λDNA *Eco*R1/*Hind*III double digest and standard 100 bp ladder were purchased from M/S Bangalore Genei Pvt. Ltd., India.

3.5 Enzymes and reagent kits:

Restriction endonucleases (with reaction buffers), ribonuclease A and High Prime DNA labeling kit were from Bangalore Genei, India. QIAEX II Gel Extraction kit and RNeasy Plant Mini kit were purchased from Qiagen GmbH, Germany. First Strand cDNA Synthesis Kit was procured from Stratagene.

3.6 Oligonucleotide Primers:

The oligonucleotide primers used for amplification were synthesized at Sigma-Aldrich, USA and Bangalore Genei, India. They were dissolved separately in ultra pure water to make 100 μM stock solutions an aliquoted into 5μM working solutions. The aliquots were stored at -20°C.

3.7 Deoxyribonucleotide (dNTP) solutions:

Deoxyribonucleotides mix was procured from Bangalore Genei, India. The four dNTP was obtained as 100 mM stock which was diluted with nuclease free water to obtain a 5 mM dNTP mix working solution.

3.8 *Taq* DNA polymerase and Assay Buffer:

Taq DNA polymerase (3U/μl) and the buffer were purchased from Bangalore Genei.

3.9 Buffers and solutions:

RNA extraction buffer
2% SDS
50 mM EDTA
150 mM Tris pH 7.5 (adjusted with 1M Boric acid)

Tris-EDTA (TE) buffer, pH8.0:
10 mM Tris-Cl, pH 8.0
1 mM EDTA pH 8.0

5X Tris-borate (TBE) buffer (pH 7.5):
800ml water
20ml 0.5M EDTA, pH8.0
54.0gm Tris base
27.5gm boric acid

DNA Gel loading buffer (5X):
25mg bromophenol blue,
30% glycerol,
Adjusted to pH 7.0 and made up the volume to 100 ml with sterile water. The solution was filter sterilized and stored at -20° C.

RNA Gel loading buffer:
50% glycerol
1mM EDTA pH 8.0,
0.25% bromophenol blue,
0.25% xylene cyanole FF
The solution was made to 100 ml, filter sterilized, aliquoted into 1.5 ml eppendorf tubes and stored in -20° C.

10X MOPS buffer:
0.4M MOPS pH7.0,
0.1M sodium acetate,
0.01M EDTA pH7.0
pH was adjusted with 1N NaOH

20X SSC:
0.3M sodium citrate,
3M NaCl
Adjusted pH to 7.0 with 1M HCl.

2X CTAB buffer:
100mM Tris-HCl, pH8.0
20mM EDTA
5M NaCl
2% CTAB
1% PVP
β-mercaptoethanol
PVP and β-mercaptoethanol were added just before use.

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

5M Potassium acetate solution, pH 4.8
60ml 5M potassium acetate solution
11.5ml glacial acetic acid
28.5ml water

Isopropyl thio-β-D-galactoside (IPTG)
100mg IPTG
1ml of sterile autoclaved water
Filter sterilise

5-bromo-4 chloro-3-indolyl-β-D-galactoside (X-GAL)
50mg X-gal
1ml dimethyl formamide
Filter sterilize

Tris equilibrated phenol:
Liquefied phenol, redistilled
50mM Tris base (adjusted to ~pH 10)
TE buffer, pH 8.0

Distilled phenol was mixed with equal volume of 50mM Tris and kept stirring on a magnetic stirrer for 4 hours. The phases were allowed to separate in room temperature and the top aqueous phase was gently removed. This was repeated till the pH was brought to 8.0. The upper aqueous phase was then removed, $1/10^{th}$ volume of TE buffer was added and stored at 4°C in brown bottle wrapped with aluminum foil.

CHAPTER : IV
RESULTS

Grains of common buckwheat *(Fagopyrum esculentum* Moench) were procured from the North Eastern Regional Station of National Bureau of Plant Genetics Resources, Shillong and multiplied in the experimental garden of Botany Department, North Eastern Hill University, Shillong.

The present work focused on isolation and characterization of novel abundantly expressed seed storage protein genes in maturing grains of common buckwheat. The approaches followed were (i) PCR amplification of target gene(s) from RNA isolated from grains of common buckwheat at different stages of maturation by reverse-transcription (RT)-PCR and (ii) construction of cDNA library from RNA isolated from grains of common buckwheat at mid maturation stage of development and from grains at early stages of germination followed by amplification of target gene(s) by nested PCR using the cDNA library as the template and primers designed from nucleotide sequences of genes available in the gene bank databases.

REVERSE TRINCRIPTASE AMPLIFICATION:

Experimental

Grains of common buckwheat were harvested at different stages of maturation and sterilized by immersing for 5 minutes in 0.01% $HgCl_2$ solution for 5 minutes. The $HgCl_2$ treated grains were washed with sterile DEPC treated water and crushed to a fine powder under liquid nitrogen in a sterile DEPC treated pestle and mortar. Total RNA was isolated from a suitable mass of the powdered tissue from grains harvested at different stages of development *viz.* 5-7 DAF, 10-13 DAF, 16-20 DAF and 23-25 DAF according to the method of Lopez-Gomez and Gomez-Lim (1992). Amplification of target gene(s) from the isolated RNA was carried out with Stratagene RT-PCR kit according to the manufacturer's protocol.

Results

The Lopez-Gomez and Gomez-Lim method used in the present investigation yielded fairly good quality RNA from all the tissue samples used in the present investigation. Under UV light, the isolated RNA was detected on the agarose gel as two fluorescing bands corresponding to a typical RNA profile of 28S rRNA and 18S rRNA with a band thickness and intensity of approximately 2:1 ratio (Fig.4.1). There was no fluorescence due to the presence of DNA in the electrophoressed RNA sample. The electrophoresis profile of the isolated RNA did not reveal any significant degradation of the RNA. The ratio of A_{260}/A_{280} in the isolated RNA samples ranged between 1.7 to 1.8. The concentration of RNA in the preparation was measured as a function of absorbance shown by the sample at 260nm using the formula: A_{260} x dilution factor x 40 = µg/ml

Fig. 4.1(a) Electrophoresis profile of total RNA isolated from grains of common buckwheat (*Fagopyrum esculentum* Moench) harvested at different maturation stages.
L1- 5-7 DAF, L2- 10-13 DAF, L3- 16-20 DAF, L4- 23-25 DAF.

(b): Electrophoresis profile of second strand synthesis of total RNA isolated from grains of common buckwheat harvested at 16-20 days after flowering (DAF) using labeled [α^{32}P]dATP random hexamer.

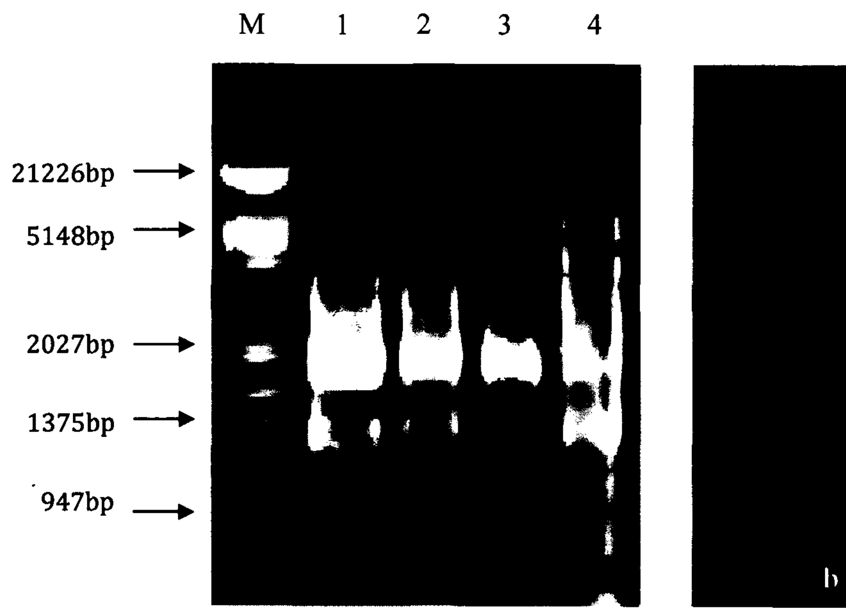


Fig. 4.1

RNA. The yield of RNA isolated from grains at different stages of development ranged between 0.5 μg gm^{-1} to 0.75 μg gm^{-1} powdered mass.

PCR amplification of genes expressed abundantly in the maturing grains of common buckwheat was attempted by RT-PCR with cDNA prepared from RNA isolated from grains at different stages of maturation as the template and oligonucleotide primers designed from nucleotide sequences avaialable in genebank data bases. The amplification programme comprised of hot start at 94°C for 5 minutes, 35 cycles comprising of denaturation at 94°C for 1 min., annealing at 60°C for 1 min. and chain extension at 72°C for 1 min. followed by elongation at 72°C for 10 mins. PCR amplification with primer pair CJ1F-CJ2R generated a 400 bp amplicon in all the reaction mixtures and an amplicon showing apparent molecular mass of 1.0 kb in the amplification mixture having cDNA synthesized from RNA isolated from grains harvested between 16 to 20 DAF (Fig. 4.2). While the 0.4 kb amplicon was amplified in all the reaction mixtures containing cDNA templates synthesized from RNA isolated from grains at different stages of development and primer pair CJ1F-CJ2R, the 1.0 kb amplicon was amplified only in the amplification mixture containing cDNA prepared from RNA isolated from grains harvested between 16 to 20 DAF. The 1.0 kb amplicon was not detected in amplification mixtures containing cDNA prepared from RNA isolated from grains harvested at any other stage. The nucleotide sequences comprising of 974 bases for the 1.0 kb amplicon is presented in fig. 4.3. A significant feature of the sequence detected was the presence of a domain comprising of the sequence "GTCGGAGGATGATGGACAAGGAGGAAGAAGAAAGAAAAAAGGGTTGAAA GAAAAGATAAA" between position 557 to 616. This domain was found to be

Fig. 4.2: Electrophoresis profile of amplification product with cDNA from grains of common buckwheat harvested at different maturation stages and early germination stage using primer pair CJ1F-CJ2R.
M_1- λDNA Hind III/EcoRI digest marker, L1- cDNA of 5-7 DAF, L2- cDNA of 10-13 DAF, L3- cDNA of 16-20 DAF, L4- cDNA of 23-25 DAF, L6- Seedling stage, M_2- 100bp ladder.

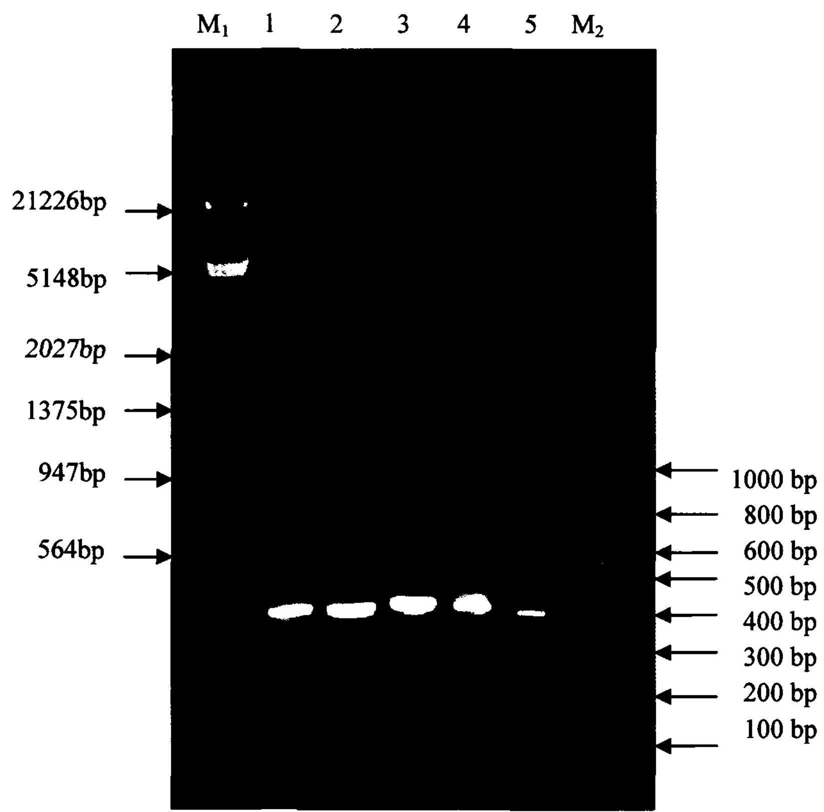


Fig. 4.2

Fig.4.3: Nucleotide sequence of 974 bases for the 1.0 Kb fragment generated from cDNA of grains harvested at 16-20 DAF using primer pair CJ1F and CJ2R. The underlined segment represents the conserved domain for dehydrin genes.

```
1    CTAAAATTCG  TCAACCCCAA  GTCTCAGGCT  ACCTTAATTT  CAGTGCCCTT
51   TTTCTTTATT  TTTTTCTAAT  AACAGGAGTC  CTGGAAAATG  GCTGACTTGG
101  GTGATGAATA  TGGAAATCCT  ATGCAGTTGA  CCGACCAGTA  TGGCAACCCG
151  GTTCAGCTCA  AGGACGAGTA  TGGCAACCCA  ATGCAGCTTA  GCGGTGTAGC
201  TATCACCGCC  CTTGTGGGGA  CGGCTAGTGC  TGTCCATTCT  ACTGGAACCG
251  GACCAACTCG  TCCCTGCTGC  CACTGGAACC  CAGCAACTTC  AGGAGCAGCT
301  TCATCGGTCT  AGCAGCTCAA  GCTCTGGCTC  GGTGAGATAC  TTGCCAAGTT
351  ACAATGTGTG  TGTCTGTGTG  TGTATAATGC  GCCATCATAA  TTGTTTGCTT
401  GACAGATCCT  GTTAATAATG  AACCGTAATT  TGACGTAAAG  TGTACACGTT
451  TTGTTTTTCT  GGGACTTACA  TAATATCGAA  TCAGGCTCCT  GTTGAATTTG
501  AATGTTGTTG  GGGAAAAGAA  AATTTTGGTC  CGCTGAGTTG  TTGAATTTGG
551  TTTATGGTCG  GAGGATGATG  GACAAGGAGG  AAGAAGAAAG  AAAAAAGGGT
601  TGAAAGAAAA  GATAAAGGAG  AAACTAACGG  GCGGGAGGCA  CAAGGACAGA
651  GACGATCAGG  AGCACATCGA  TGATCAGCAC  GGCGCACAGC  GCCTCTCCTC
701  CAACAACCAC  CACTGGCAGC  GGGACGTCTA  CTACAGTCGG  GGGTCAGCAG
751  CATGAAAAGA  AGAGCATGGT  GGAGAAGATT  ATGGAAAAGC  TCCCTGGCCA
801  TCACGACACC  CGCTAGTTAC  CTACCACAAC  ATACTGTGAT  CATCGTGTAA
851  AATCTCTCCT  GATGCCTAGG  AAATCTAGAT  TATGTTAGGC  ATTTTGTTTG
901  GTATGTATGT  GTGATTAAGA  CCTTGTTGTG  CGCTTGAATC  TTGAACGTGC
951  ATGGGATTTG  CTTGGTTTGA  AAAA
```

Fig. 4.3

conserved in genes coding for dehydrin like proteins from several plants including *Solanum lycopersicum,* (AC215480) *S. tuberosum* (X83597), *Cornus sericea* (AF345988), *Helianthus annuus* (AJ438980), *Eriobotrya japonica* (FJ472835), *Panax ginseng* (DQ487110), *Phoenix dactylifera* (DQ399792), *Tithonia rotundifolia* (AJ250127). Fig. 4.4 shows the alignment of the nucleotide sequence of 1.0 kb amplicon generated in the present study with nucleotide sequences of genes coding for dehydrin/dehydrin like proteins from other plants. Using an alignment that permitted maximum homology the nucleotide sequence showed 99% homology with *Coffea canephora* DH2b (acc. no. DQ323990) and 96% homology with *Coffea canephora* DH2 (acc. no. DQ338457). A statistical evaluation of the alignment revealed that the sequence homologies were highly significant. Compared with the nucleotide sequences of dehydrin genes DH2, DH2a and DH2b, the nucleotide sequence of the putative buckwheat dehydrin like gene amplified in the present study showed an insertion of six bases "CTTGTG" at position 217, seven bases "CGTCCCT" at position 258 and a single nucleotide "G" at position 682. These features distinguished the buckwheat nucleotide sequence from nucleotide sequences of *Coffea canephora* dehydrin like genes.

The deduced amino acid sequence for the 974 bases of the 1.0 kb amplicon comprised of 112 amino acid residues with a predicted isoelectric point (p*I*) of 5.87 and calculated molecular weight of 12 kDa (Fig. 4.5). BLASTp analysis of the deduced amino acid sequence against non-redundant protein database identified the protein with the dehydrin family of proteins. Dehydrin like proteins are known to possess the "K" segment comprising of the sequence "EKK(S/G)(M/I)(V/M)(E/D)KI(M/K)EKLPGHH"

Fig. 4.4: ClustalW (1.81) multi-alignment of the 974 bp nucleotide sequences (Buckwheat_DHN) for the 1.0 Kb amplicon generated with cDNA from 16-20 DAF seed stage as template and primer pair CJ1F - CJ2R with nucleotide sequences of other dehydrin/dehydrin like genes available in the database.'*' denotes conserved residues.

ClustalW (1.81) multi-alignment

```
Buckwheat_DHN       ---CTAAAATTCGTCAACCCCAAGTCTCAGGCTACCTTAATTTCAGTGCCCTTTTTCTTT 57
Coffea_DHN_DH2b     ---CTAAAATTCGTCAACCCCAAGTCTCAGGCTACCTTAATTTCAGTGCCCTTTTTCTTT
Coffea_DHN_DH2      CAGCTAAAATTCGTCAACCCCAAGTCTCAGGCTACCTTAATTTCAGTGCCCTTTTTCTTT
Coffea_DHN_DH2a     ---CTAAAATTCGTCAACCCCAAGTCTCAGGCTACCTTAATTTCAGTGCCCTTTTTCTTT
                       *********************************************************

Buckwheat_DHN       ATTTTTTTCTAATAACAGGAGTCCTGGAAAATGGCTGACTTGGGTGATGAATATGGAAAT 117
Coffea_DHN_DH2b     ATTTTTTTCTAATAACAGGAGTCCTGGAAAATGGCTGACTTGCGTGATGAATATGGAAAT
Coffea_DHN_DH2      ATTTTTTTCTAATAACAGGAGTCCTGGAAAATGGCTGACTTGCGTGATGAATATGGAAAT
Coffea_DHN_DH2a     ATTTTTTTCTAATAACAGGAGTCCTGGAAAATGGCTGACTTGCGTGATGAATATGGAAAT
                    ****************************************** *****************

Buckwheat_DHN       CCTATGCAGTTGACCGACCAGTATGGCAACCCGGTTCAGCTCAAGGACGAGTATGGCAAC 177
Coffea_DHN_DH2b     CCTATGCAGTTGACCGACCAGTATGGCAACCCGGTTCAGCTCAAGGACGAGTATGGCAAC
Coffea_DHN_DH2      CCTATGCAGTTGACCGACCAGTATGGCAACCCGGTTCAGCTCAAGGACGAGTATGGCAAC
Coffea_DHN_DH2a     CCTATGCAGTTGACCGACCAGTATGGCAACCCGGTTCAGCTCAAGGACGAGTATGGCAAC
                    ************************************************************

Buckwheat_DHN       CCAATGCAGCTTAGCGGTGTAGCTATCACCGCCCTTGTGGGGACGGCTAGTGCTGTCCAT 237
Coffea_DHN_DH2b     CCAATGCAGCTTAGCGGTGTAGCTATCACCGCC------GGGACGGCTAGTGCTGTCCAT
Coffea_DHN_DH2      CCAATGCAGCTTAGCGGTGTAGCTATCACCGCC------GGGACGGCTAGTACTGTCCAT
Coffea_DHN_DH2a     CCAATGCAGCTTAGCGGTGTAGCTATCACCGCC------GGGACGGCTAGTGCTGTCCAT
                    *********************************      ************ ********

Buckwheat_DHN       TCTACTGGAACCGGACCAACTCGTCCCTGCTGCCACTGGAACCCAGCAACTTCAGGAGCA 297
Coffea_DHN_DH2b     TCTACTGGAACCGGACCAACT-------GCTGCCACTGGAACCCAGCAACTTCAGGAGCA
Coffea_DHN_DH2      TCTACTGGAACCGGACCAACT-------GCTGCCACTGGAACCCAGCAACATCAGGAGCA
Coffea_DHN_DH2a     TCTACTGGAACCGGACCAACT-------GCTGCCACTGGAACCCAGCAACATCAGGAGCA
                    *********************       ********************** *********

Buckwheat_DHN       GCTTCATCGGTCTAGCAGCTCAAGCTCTGGCTCGGTGAGATACTTGCCAAGTTACAATGT 357
Coffea_DHN_DH2b     GCTTCATCGGTCTAGCAGCTCAAGCTCTGGCTCGGTGAGATACTTGCCAAGTTACAATGT
Coffea_DHN_DH2      GCTTCATCGGTCTAGCAGCTCAAGCTCTGGCTCGGTGAGATACTTGCCAAGTTACAATGT
Coffea_DHN_DH2a     GCTTCATCGGTCTAGCAGCTCAAGCTCTGGCTCG--------------------------
                    **********************************

Buckwheat_DHN       GTGTGTCTGTGTGTGTATAATGCGCCATCATAATTGTTTGCTTGACAGATCCTGTTAATA 417
Coffea_DHN_DH2b     GTGTGTCTGTGTGTGTATAATGCGCCATCATAATTGTTTGCTTGACAGATCCTGTTAATA
Coffea_DHN_DH2      GTGTGTCTGTGTGTGTATAATGCGCCATCATAATTGTTTGCTTGACAGATCCTGTTAATA
Coffea_DHN_DH2a     ------------------------------------------------------------

Buckwheat_DHN       ATGAACCGTAATTTGACGTAAAGTGTACACGTTTTGTTTTTCTGGGACTTACATAATATC 477
Coffea_DHN_DH2b     ATGAACCGTAATTTGACGTAAAGTGTACACGTTTTGTTTTTCTGGGACTTACATAATATC
Coffea_DHN_DH2      ATGAACCGTAATTTGACGTAAAGTGTACACGTTTTGTTTTTCTGGGACTAACATAATATC
Coffea_DHN_DH2a     ------------------------------------------------------------

Buckwheat_DHN       GAATCAGGCTCCTGTTGAATTTGAATGTTGTTGGGGAAAAGAAAATTTTGGTCCGCTGAG 537
Coffea_DHN_DH2b     GAATCAGGCTCCTGTTGAATTTGAATGTTGTTAGCTAAAAGAAAATTTTGGTG-GCTGAG
Coffea_DHN_DH2      GAATCAGGCTCCTGTTGAATTTGAATGTTGTTAGCTAAAAGAAAATTTTGGTG-GCTGAG
Coffea_DHN_DH2a     ------------------------------------------------------------

Buckwheat_DHN       TTGTTGAATTTGGTTTATGGTCGGAGGATGATGGACAAGGAGGAAGAAGAAAGAAAAAAG 597
Coffea_DHN_DH2b     TTGTTGAATTTGGTTTATGGTCGGAGGATGATGGACAAGGAGGAAGAAGAAAGAAAAAAG
Coffea_DHN_DH2      TTGTTGAATTTGGTTTATAGACGGAGGATGATGGACAAGGAGGAAGAAGAAAGAAAAAAG
Coffea_DHN_DH2a     --------------------TCGGAGGATGATGGACAAGGAGGAAGAAGAAAGAAAAAAG
                                         ***************************************

Buckwheat_DHN       GGTTGAAAGAAAAGATAAAGGAGAAACTAACGGGCGGGAGGCACAAGGACAGAGACGATC 657
Coffea_DHN_DH2b     GGTTGAAAGAAAAGATAAAGGAGAAACTAACGGGCGGGAGGCACAAGGACAGAGACGATC
Coffea_DHN_DH2      GGTTGAAAGAAAAGATAAAGGAGAAACTAACGGGCGGTAGGCACAAGGACAGAGACGATC
Coffea_DHN_DH2a     GGTTGAAAGAAAAGATAAAGGAGAAACTAACGGGCGGGAGGCACAAGGACAGAGACGATC
                    ************************************* **********************
```

```
Buckwheat_DHN       AGGAGCACATCGATGATCAGCACGGCGCACAGCGCCTCTCCTCCAACAACCACCACTGGC 717
Coffea_DHN_DH2b     AGGAGCACATCGATGATCAGCACG-CGCACAGCGCCTCTCCTCCAACAACCACCACTGGC
Coffea_DHN_DH2      AGGAGCACATCGATGATCAGCACG-CGCACAGCGCCTCTCCTCCAACAACCACCACTGGC
Coffea_DHN_DH2a     AGGAGCACATCGATGATCAGCACG-CGCACAGCGCCTCTCCTCCAACAACCACCACTGGC
                    ************************ ***********************************

Buckwheat_DHN       AGCGGGACGTCTACTACAGTCGGGGGTCAGCAGCATGAAAAGAAGAGCATGGTGGAGAAG 777
Coffea_DHN_DH2b     AGCGGGACGTCTACTACAGTCGGGGGTCAGCAGCATGAAAAGAAGAGCATGGTGGAGAAG
Coffea_DHN_DH2      AGCGGGACGTCTACTACAGTCGGGGGTCAGCAGCATGAAAAGAAGAGCATGGTGGAGAAG
Coffea_DHN_DH2a     AGCGGGACGTCTACTACAGTCGGGGGTCAGCAGCATGAAAAGAAGAGCATGGTGGAGAAG
                    ************************************************************

Buckwheat_DHN       ATTATGGAAAAGCTCCCTGGCCATCACGACACCCGCTAGTTACCTACCACAACATACTGT 837
Coffea_DHN_DH2b     ATTATGGAAAAGCTCCCTGGCCATCACGACACCCGCTAGTTACCTACCACAACATACTGT
Coffea_DHN_DH2      ATTATGGAAAAGCTCCCTGGCCATCACGACACCCGCTAGTTACCTACCACAACATACTGT
Coffea_DHN_DH2a     ATTATGGAAAAGCTCCCTGGCCATCACGACACCCGCTAGTTACCTACCACAACATACTGT
                    ************************************************************

Buckwheat_DHN       GATCATCGTGTAAAATCTCTCCTGATGCCTAGGAAATCTAGATTATGTTAGGCATTTTGT 897
Coffea_DHN_DH2b     GATCATCGTGTAAAATCTCTCCTGATGCCTAGGAAATCTAGATTATGTTAGGCATTTTGT
Coffea_DHN_DH2      GATCATCGTGTAAAATCTCTCCTGATGCCTAGGAAATCTAGATTATGTTAGGCATTTTGT
Coffea_DHN_DH2a     GATCATCGTGTAAAATCTCTCCTGATGCCTAGGAAATCTAGATTATGTTAGGCATTTTGT
                    ************************************************************

Buckwheat_DHN       TTGGTATGTATGTGTGATTAAGACCTTGTTGTGCGCTTGAATCTTGAACGTGCATGGGAT 957
Coffea_DHN_DH2b     TTGGTATGTATGTGTGATTAAGACCTTGTTGTGCGCTTGAATCTTGAACGTGCATGGGAT
Coffea_DHN_DH2      TTGGTATGTATGTGTGATTAAGACCTTGTTGTGCGCTTGAATCTTGAACGTGCATGGGAT
Coffea_DHN_DH2a     TTGGTATGTATGTGTGATTAAGACCTTGTTGTGCGCTTGAATCTTGAACGTGCATGGGAT
                    ************************************************************

Buckwheat_DHN       TTGCTTGGTTTGAAAAA--------------------------- 974
Coffea_DHN_DH2b     TTGCTTGGTTTGAAAAA---------------------------
Coffea_DHN_DH2      TTGCTTGGTTTGATTTGATTTGG---------------------
Coffea_DHN_DH2a     TTGCTTGGTTTGATTTGATTTGGTGAAATAAGTTGTACTAAAAA
                    *************
```

Fig. 4.4

and the “Y” domain comprising of the sequence “D(E/Q)YGNP”. The amino acid sequence of 112 residues identified in the present study showed both the segments with the “K” segment represented by the sequence “EKKSMVEKIMEKLPGHH” located between 93-107 bases. The sequence showed three “Y” segments located between 10-35 bases. While the Ist and IIIrd “Y” segments had the sequence DEYGNP, the IInd “Y” segment was represented by the sequence DQYGNP (Fig. 4.5). Motif search on the deduced amino acid sequence with MOTIF SCAN (http://myhits.isb-sib.ch/cgi-bin/motif_scan) identified a putative kinase C phosphorylation site “SWK” between residue 2-4, a N-glycosylation site “Asn-Arg-Thr-Asn” between residue 53-56, a kinase II phosphorylation site “Ser-Met-Val-Glu” between residue 96-99 and two N-myristoylation sites “Gly-Val-Ala-Ile-Thr-Ala” between residues 40-45 and “Gly-Thr-Ser-Thr-Thr-Val” between 82-87(Fig.4.6a, b). MOTIF SCAN also detected the presence of two dehydrin domains comprising of the sequences “MQLTDQYGNPVQLKDEYG NPMQ” and “GTSTTVGGQQHEKKSMVEKIMEKLPGHH” between position 16-37 and 82-109, respectively (Fig.4.6c).

The deduced amino acid sequence was compared with amino acid sequences of other dehydrin/dehydrin-like proteins available in the data bases using ClustalW. Using an alignment that permitted maximum homology, the deduced amino acid sequence showed a maximum of 63% homology with predicted amino acid sequences of dehydrin like proteins DH2, DH2a and DH2b (acc. nos. DQ338457, DQ323990, DQ323989) from *Coffea canephora.* The percentage homology ranged from 42 – 46% with deduced amino acid sequences of dehydrin genes from other plants including *Vitis vinifera* (acc. no. XP002283605) and *Raphanus sativus* (acc. no. P21298). ClustalW multiple

Fig. 4.5: Deduced amino acid sequences of the coding region of 974 bp nucleotide sequences for the 1.0 Kb amplicon generated from 16- 20 DAF cDNA as the template and primer pair CJ1F and CJ2R. "__" indicates conserved "Y" segments, "_ _" indicates conserved "K" segment. The "□" indicates the motif patterns and segments highlighted with yellow colour indicates the dehydrin domains as predicted by MOTIF SCAN.

Fig. 4.6(a) Diagrammatic representation of the motif sequence patterns and its frequency in the 112 deduced amino acid sequences as predicted by MOTIF SCAN.

(b) Diagrammatic representation of the Protein Kinase C phosphorylation site, Kinase II phosphorylation site, N-glycosylation site and N-Myristoylation sites in the 112 deduced amino acid sequence as predicted by MOTIF SCAN.

(c) Diagrammatic representation of the dehydrin domains positioned between 16 to 37 and 82-109 amino acid residues in the 112 deduced amino acid sequence as predicted by MOTIF SCAN.

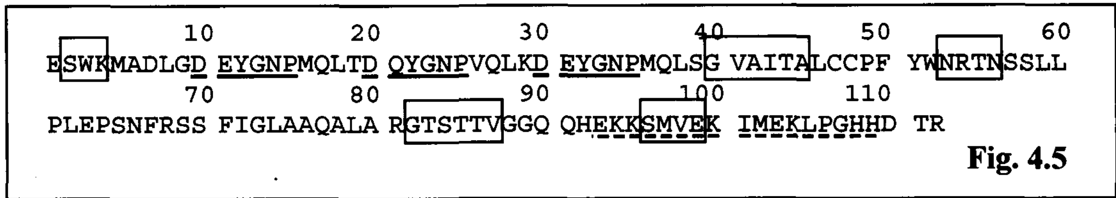
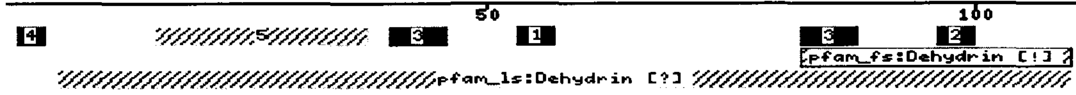
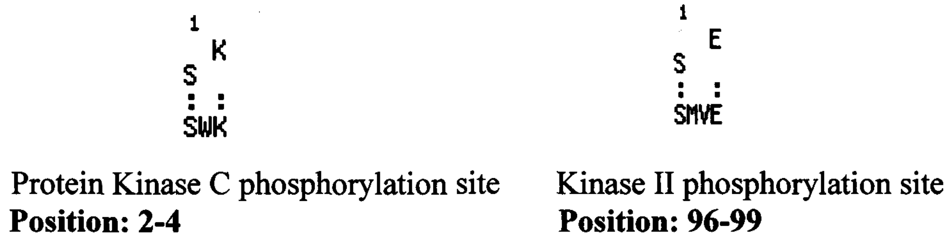


Fig. 4.5



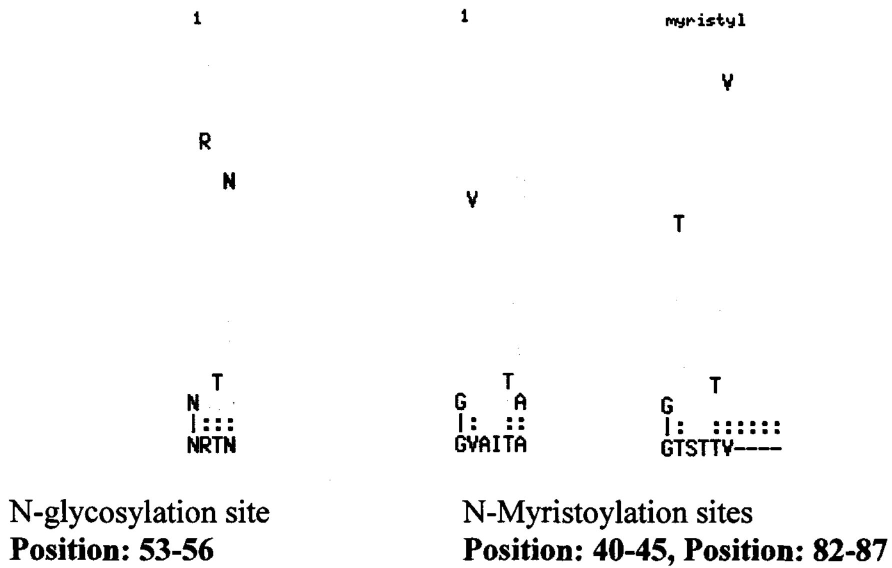
1: N glycosylation site, 2: Casein Kinase II phosphorylation site, 3: N- Myristoylation site, 4: Protein Kinase C phosphorylation site, 5: Dehydrin

Fig.4.6a



Protein Kinase C phosphorylation site
Position: 2-4

Kinase II phosphorylation site
Position: 96-99



N-glycosylation site
Position: 53-56

N-Myristoylation sites
Position: 40-45, Position: 82-87

Fig.4.6b

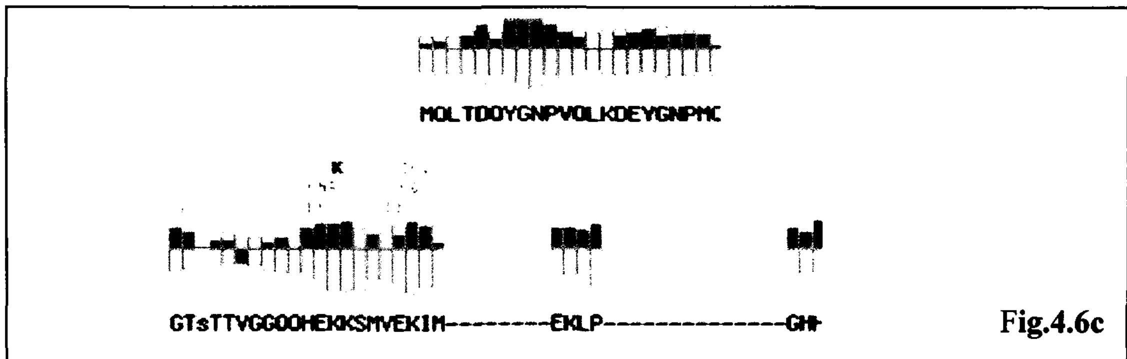


Fig.4.6c

alignment of the sequence with deduced amino acid sequences of other dehydrin like proteins showed two segments positioned between residue 1-38 and 90-109 with the highest level of conserved residues (Fig. 4.7). ClustalW analysis also revealed that the buckwheat protein had greater similarity with dehydrin like proteins of *Coffea canephora* than that from *Arabidopsis thaliana*, *Raphanus sativa* or *Vitis vinifera*. Significantly, some of the residues within the dehydrin domains in the deduced amino acid sequence of putative buckwheat dehydrin like protein identified in the present study were found to be represented by amino acids with similar functional groups. While the amino acid sequences of *Coffea arabica* DH2, DH2a, DH2b and putative dehydrin like protein identified in the present study had Methionine at position 16 and Valine at position 98, the dehydrin from *Arabidopsis thaliana* (acc.no. NP195624), *Raphanus sativus* (acc. no. P21298) and the putative hypothetical protein from *Vitis vinifera* (acc. no. XP00283505) had isoleucine at position 16 and Leucine at position 98. Similarly, Glutamine at position 17 and Methionine at position 97, which were found to be invariant in dehydrins of *Coffea arabica* DH2, DH2a, DH2b (acc. no. ABC67157, ABC55672 & ABC55673), *Vitis vinifera* (acc. no. XP00283505) and putative dehydrin like protein identified in the present study, were represented by Histidine and Isoleucine, respectively in dehydrins of *Arabidopsis thaliana* (acc.no. NP195624) and *Raphanus sativus* (acc. no. P21298). Similarly, Lysine at P_{29} in *Coffea arabica* dehydrins DH2, DH2a, DH2b and the putative dehydrin like protein identified in the present study was represented by Threonine in *Arabidopsis thaliana* dehydrin (acc.no. NP195624) and *Vitis vinifera* dehydrin (acc. no. XP00283505) and by Serine in *Raphanus sativus dehydrin* (acc. no. P21298). Some of the other significant features of

Fig. 4.7: ClustalW (1.81) multi-alignment of the deduced amino acid sequences (Buckwheat_DHN) for the 974 bp nucleotide sequence of cDNA from grains harvested at 16-20 DAF using primer pair CJ1F and CJ2R with amino acid sequences of dehydrin protein of other plants available in the database. '*' denotes conserved residues and ':' denotes invariant or similar residues.

```
                 CLUSTAL (1.81) multiple sequence alignment


Coffea_DH2         ----MADLRDEYGNPMQLTDQYGNP-VQLKDEYGNPMQLSGVA-----------------
Coffea_DH2a        ----MADLRDEYGNPMQLTDQYGNP-VQLKDEYGNPMQLSGVA-----------------
Coffea_DH2b        ----MADLRDEYGNPMQLTDQYGNP-VQLKDEYGNPMQLSGVA-----------------
Buckwheat_DHN      ESWKMADLGDEYGNPMQLTDQYGNP-VQLKDEYGNPMQLSGVA-----------------
Arabidopsis_DHN    ----MADLRDEKGNPIHLTDTQGNPIVDLTDEHGNPMYLTGVVSSTPQHKESTTSDIAEH
Raphanus_DHN       ----MADLKDERGNPIHLTDAYGNP-VQLSDEFGNPMHITGVASSAPQYKDSVTGNIAEY
Vitis_hypo.        ----MADLRDEHGNPIQLTDQHGNP-VQLTDEHGNPMHLTGVASTHPITTTTAPVHDADA
                       **** ** ***::***   *** *:*.**.****  ::**.

Coffea_DH2         ------------ITAGTASTVHSTGTGPT--------AATGTQQHQEQLHRSSSSSSGST
Coffea_DH2a        ------------ITAGTASAVHSTGTGPT--------AATGTQQHQEQLHRSSSSSSGSS
Coffea_DH2b        ------------ITAGTASAVHSTGTGPT--------AATGTQQLQEQLHRSSSSSSGSS
Buckwheat_DHN      ------------ITA-LCCPFYWNRTNSS--------LLP--------LEPSNFRSS---
Arabidopsis_DHN    PTSTVGETHPAAAPAGAGAATAATATGVSAGTGATTTGQQHHGSLEEHLRRSGSSSSSSS
Raphanus_DHN       PT----EAPPAGVAAGTGAA-ATTAAGVT--TSETTTGQEHHGSLGEHLRRSGSSSSSSS
Vitis_hypo.        LKTT-GESHPP--TSGHGIADQAVHGGAP----VAAEPAEGGGGEVHHEKRESSSSSSSS
                               .:    .      . .                 . ..   **

Coffea_DH2         EDDGQGGRRKKKGLKEKIKEKLTGGRHKDRDDQEHIDDQHAHSASPPTTTTGSGTSTT--
Coffea_DH2a        EDDGQGGRRKKKGLKEKIKEKLTGGRHKDRDDQEHIDDQHAHSASPPTTTTGSGTSTT--
Coffea_DH2b        EDDGQGGRRKKKGLKEKIKEKLTGGRHKDRDDQEHIDDQHAHSASPPTTTTGSGTSTT--
Buckwheat_DHN      ---------------------FIG--------------------LAAQALARGTSTT--
Arabidopsis_DHN    EDDGQGGRRKK-SIKEKIKEKFGSGKHKDEQ-------------TPATATTTGPATT--
Raphanus_DHN       EDDGQGGRRKK-SIKDKIKDKLGGGKHKDEQ-------------TPTTATTTGPTTTTT
Vitis_hypo.        EDDGQGGRRKK-GLKEKIKEKLTGGKHKEEQGHAP-----TTGVAITTTTTTTGSATTP-
                                         : .                 .: :    *.:**

Coffea_DH2         ----VGGQQHEKKSMVEKIMEKLPGHHDTR-- 162
Coffea_DH2a        ----VGGQQHEKKSMVEKIMEKLPGHHDTR-- 162
Coffea_DH2b        ----VGGQQHEKKSMVEKIMEKLPGHHDTR-- 162
Buckwheat_DHN      ----VGGQQHEKKSMVEKIMEKLPGHHDTR-- 112
Arabidopsis_DHN    ------DQPHEKKGILEKIKDKLPGHHNHNHP 185
Raphanus_DHN       TTGAAADQHHEKKGILEKIKEKLPGHHNH-HP 184
Vitis_hypo.        VTVQHHHEQHEKKGMLEKIKDKLPGHHTS-H- 191
                          : ****.::*** :******
```

Fig. 4.7

the alignment were the identification of conserved segments present invariably in all the sequences of dehydrin/dehydrin-like proteins analysed in the present study. These included the sequences 'Met-Ala-Asp-Leu' at position 5-8, 'Asp-Glu' at position 10-11 'Gly-Asn-Pro' at position 13-15 and 'Gly-Val' at 41-42. The Gly-Asn-Pro' between residues 13-15 and Gly-Val' between 41-42 flanked the dehydrin domain between residues 16-37. These segments were comprised of predominantly aliphatic type of amino acids. A distinctive feature of the buckwheat amino acid sequence in the present study was the substitution of Serine tract with 'Asn-Phe-Arg' at position 66-68 (Fig. 4.7).

Hydropathy analysis of the deduced amino acid sequence by MacVector software using the Kyte and Doolittle method (1982) identified two strong hydrophilic domains in the sequence. While one of the domains comprised of residues 1-40, the other comprised of residues 80-112. The flexibility analysis of the amino acid sequence carried out by MacVector indicated that the sequence had a relatively high degree of flexibility (Fig.4.8a, b). SAPS software (Brendel *et al.,* 1992) analysis of the amino acid sequence identified 23.2% non polar group, 25.8% uncharged polar group and 19.6% charged polar group. The regions having high concentration of charged residues were in correlation with the hydrophilic regions deduced by hydropathy analysis.

Secondary structure prediction of the amino acid sequence carried out by GOR4 (Garnier *et al.*, 1996) indicated a 20.54% alpha helix, 54.46% random coil and 25% extended strand for the deduced amino acid sequence (Fig. 4.9). Statistical analysis of the sequence revealed that the N-terminal region comprising of 70 residues had

Fig. 4.8(a) Hydropathy profile of the deduced 112 amino acid for the 1.0 Kb amplicon generated from 20 DAF buckwheat seed cDNA using CJ1F and CJ2R primer pair. The x-axis represents the residue number.

(b) Flexibility analysis of the 974 bases for the 1.0 Kb amplicon generated from 20 DAF buckwheat seed cDNA using CJ1F and CJ2R primer pair.

Fig. 4.9: Secondary structure prediction profile of the 112 deduced amino acid sequence for the 1.0 Kb amplicon generated with 16-20 DAF buckwheat seeds cDNA as the template and CJ1F-CJ2R as primer pair.

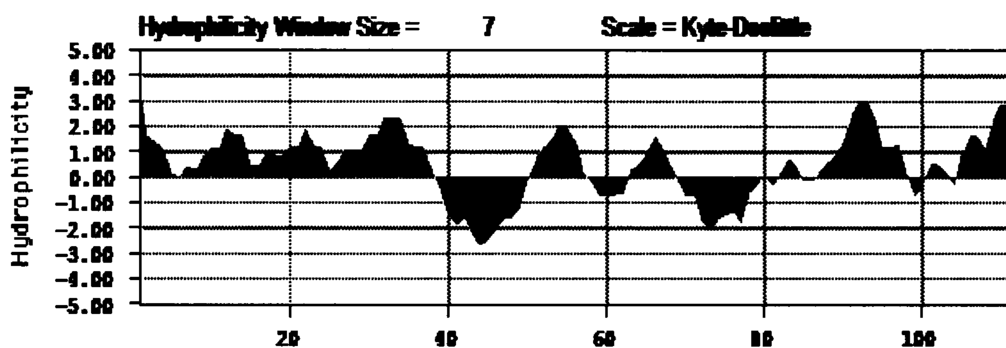


Fig.4.8a

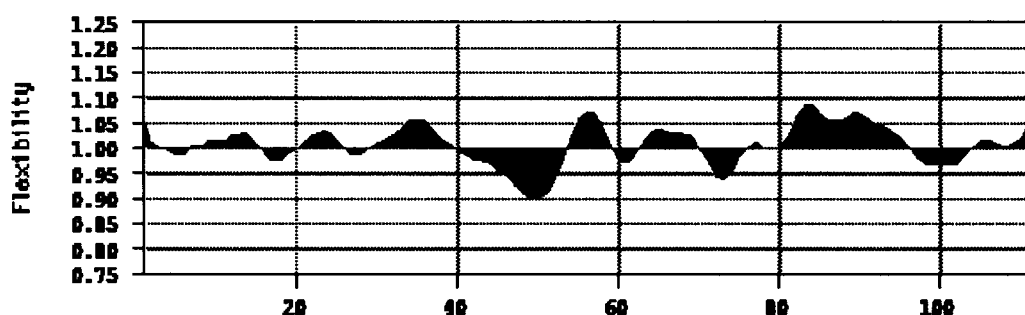


Fig.4.8b

Sequence length : 112

GOR4 :

Alpha helix	(Hh)	:	23	is	20.54%
3_{10} helix	(Gg)	:	0	is	0.00%
Pi helix	(Ii)	:	0	is	0.00%
Beta bridge	(Bb)	:	0	is	0.00%
Extended strand	(Ee)	:	28	is	25.00%
Beta turn	(Tt)	:	0	is	0.00%
Bend region	( )	:	0	is	0.00%
Random coil	( )	:	61	is	54.46%
Ambigous states	(?)	:	0	is	0.00%
Other states		:	0	is	0.00%

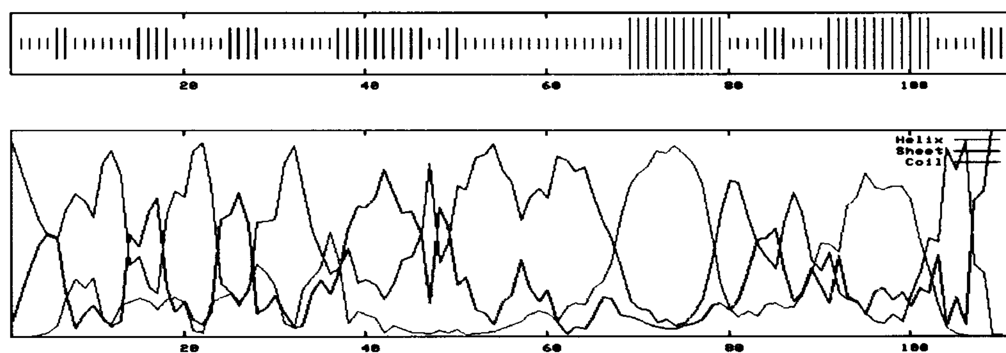


Fig.4.9

predominantly random coil and extended strand secondary structure while the C-terminal region of 42 residues were primarily of α-helical configuration.

A phylogenetic tree was constructed using maximum parsimony method, with the alignment data of deduced amino acid sequence in the present study with amino acid sequences of other dehydrin/dehydrin-like proteins available in the EMBL data base, describing the relationship of the deduced amino acid sequence with amino acid sequences of dehydrin/dehydrin-like proteins from other plants. The amino acid sequence from common buckwheat (*Fagopyrum esculentum*), identified in the present study formed one clade with dehydrins from *Coffea canephora* having low diverging branches. The sequences from *Raphanus sativa, Arabidopsis thaliana* and *Hordeum vulgare* formed another clade. Dehydrin like protein from *Camellia sinensis*, *Panax ginseng*, *Helianthus annuus* and *Cornus sericea* emerged as distinctly different groups within the tree (Fig.4.10).

CONSTRUCTION OF DEVELOPMENTAL STAGE SPECIFIC cDNA LIBRARY AND PCR SCREENING:

Experimental:

Grains of common buckwheat were harvested at mid maturation stage of development (16-20 DAF) and surface sterilized in the laboratory by treatment with 0.01% $HgCl_2$ solution for 5 minutes. Mature grains of common buckwheat were also sterilized in the same manner by treatment with 0.01% $HgCl_2$ solution for 5 minutes. The sterilized grains were germinated on a clean germination paper in an incubator at 27°C. The germinated seedling were harvested after 7 days of incubation and treated again with 0.01% $HgCl_2$ solution for 5 minutes for sterilization. Total RNA was

Fig. 4.10: Phylogenetic tree based on the alignment matrix of the deduced 112 amino acid sequence for the 1.0 Kb amplicon fragment generated from 16-20 DAF seed cDNA using primer pair CJ1F-CJ2R with the amino acid sequences of dehydrin type proteins of various plants. Accession nos. of the sequences selected for the analysis is indicated against each branch.

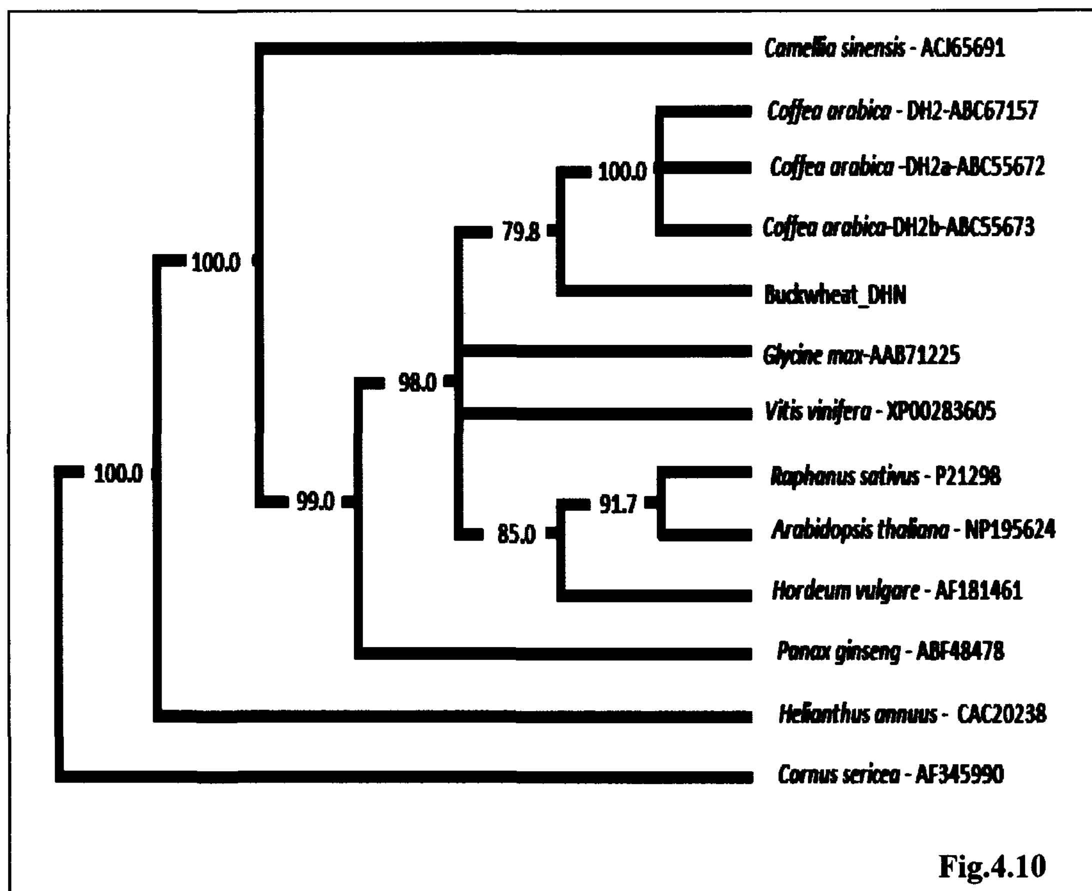


Fig.4.10

isolated separately from grains of buckwheat harvested between 16-20 DAF as well as from germinating grains of common buckwheat. The $HgCl_2$ treated grains/tissues were washed with sterile DEPC treated (0.2 % DEPC) distilled water and crushed to a fine powder under liquid nitrogen in a sterile pestle and mortar. Total RNA was isolated from a suitable mass of the powdered tissue according to the method of Lopez-Gomez and Gomez-Lim (1992). The isolated RNA was used as a template for construction of stage specific cDNA library using the Stratagene ZAP-cDNA Gigapack III Gold cloning kit as per the manufacturer's protocol. The libraries were screened by polymerase chain reaction with primers designed from conserved regions of nucleotide sequences of mid maturation stage specific abundantly expressed proteins available in gene bank data bases. The amplification programme consisted of hot start at 94°C for 5 minutes, 35 cycles comprising of denaturation at 94°C for 1 min., annealing at 60°C for 1 min. and chain extension at 72°C for 1 min. followed by elongation at 72°C for 10 min.

Results

PCR amplification of cDNA library constructed from RNA isolated from grains of common buckwheat harvested between 16-20 Days after flowering *viz.* mid maturation stage of development as the template and primer pair NKC5-NKC6, generated an amplicon showing an apparent molecular mass of 0.5 Kb. The amplified product appeared as a single band corresponding to a molecular mass of 0.5 kb on the agarose gel after electrophoresis (Fig. 4.11). The band could not be detected in the lane having the amplification mixture containing the cDNA template prepared from RNA isolated from grains harvested at early stages of germination (Fig. 4.11). The

Fig. 4.11: Electrophoresis profile of amplified fragment with cDNA library as template and primer pair NKC5 and NKC6;
L1- cDNA library from grains harvested at 16-20 DAF as template,
L2- cDNA library from early germination stage as template.

Fig. 4.12: Nucleotide sequence of 576 bp for the amplified fragment generated with cDNA of mid maturation seed stage and primer pair NKC5 and NKC6.

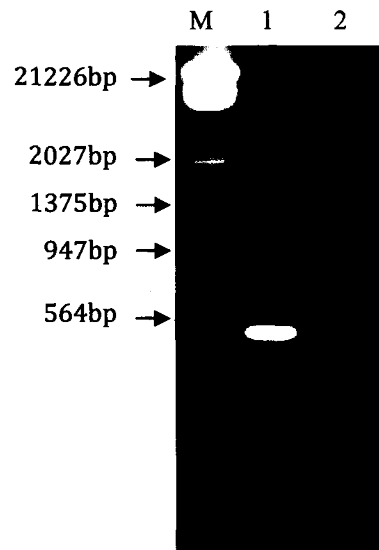


Fig. 4.11

1	GGATTGGAGC	AAGCGTTCTG	CAACCTGAAA	TTCAGGCAAA	ATGTTAACAG
51	GCCTTCTCGC	GCCGACGTCT	TCAACCCACG	CGCCGGTCGT	ATCAACACCG
101	TAGACAGCAA	CAATCTCCCG	ATCCTCGAAT	TCATCCAACT	TAGCGCCCAG
151	CACGTCGTCC	TCTACAAGAA	TGCGATCCTC	GGACCGAGAT	GGAACTTGAA
201	CGCGCACAGC	GCACTGTACG	TGACGAGAGG	AGAAGGAAGA	GTCCAGGTTG
251	TCGGAGATGA	AGGAAGAAGT	GTGTTCGACG	ACAACGTGCA	GCGAGGACAG
301	ATCCTTGTGG	TCCCACAGGG	ATTCGCAGTG	GTGTTGAAGG	CAGGAAGAGA
351	AGGACTGGAG	TGGGTGGAGT	TGAAGAACGA	CGACAACGCC	ATAACCAGCC
401	CGATTGCCGG	GAAGACTTCG	GTGTTGAGGG	CGATCCCTGT	GGAGGTTCTT
451	GCCAACTCCT	ACGATATCTC	GACGAAGGAA	GCGTTCAGAT	TGAAGAATGG
501	GAGGCAGGAG	GTTGAGGTCT	TCCGACCATT	CCAGTCCCGA	GATGAGAAGG
551	AGAGGGAGCG	TTTCTCCATA	GTTTAA		

Fig.4.12

nucleotide sequence comprising of 576 bases for the 0.5 Kb amplicon is presented in fig.4.12. BLASTn analysis clearly identified the sequence with the nucleotide sequences of genes coding for legumin type seed storage proteins. The sequence showed 97% and 92% homology with nucleotide sequences of the genes coding for *Fagopyrum esculentum* legumin like proteins (acc. nos. D87980 and D87982) with a sequence query coverage of 98% and 100%, respectively. The sequence showed 91% homology with nucleotide sequence of the gene coding for *Fagopyrum tataricum* allergenic protein (acc. no. DQ849083) with a sequence query coverage of 100%. BLASTn analysis of the sequence also revealed 66% homology with gene coding for *Ficus pumila*11S globulin precursor 2B (acc. no. EF091696) and 57% sequence homology with the nucleotide sequences of genes coding for "citrin", the seed storage protein of *Citrus sinensis* (acc. no. U38914) and *Gossypium hirsutum* β-globulinB (acc. no. M16936). ClustalW alignment of the 576 bp nucleotide sequence with nucleotide sequences of seed storage protein genes from other plants is presented in fig.4.13. Sequence alignment of the nucleotide sequence of 576 bp DNA fragment amplified in the present study with nucleotide sequences of seed storage protein genes from other plants revealed a deletion of 3 bases "GGG" at position 528 in the nucleotide sequence of the DNA fragment amplified in the present study, a feature which was also detected in nucleotide sequences of genes coding for other buckwheat seed storage proteins used for generation of alignment matrix for the present study. However, a deletion of 12 bases "GAGGAGAAGGAG" detected at position 552 in the 576 bp nucleotide sequence was observed only in nucleotide sequences of legumin type seed storage proteins of *Fagopyrum esculentum* (acc. nos. AF152003 & D87980).

Fig. 4.13: ClustalW (1.81) multi-alignment of 576 bp nucleotide sequence (BuckwheatSSP) of cDNA from mid maturation seed stage generated using primer pair NKC5 and NKC6 with nucleotide sequences of SSP genes of other plants derived from gene bank.(* denotes conserved residues).

clustalW (1.81) multi-alignment

```
F.esculentum_87982        AATGGAGTCGAGCAAGGGTTCTGCAATCTTAAATTCAGGCGAAATTTTAA 998
F.esculentum_87980        AATGGATTGGAGCAAGCGTTCTGCAACCTGAAATTCAAGCAAAATGTTAA 1199
F_tataricum             . AATGGATTGGAGCAAGCGTTCTGTAACCTAAAATTCAGGCAAAATGTTAA 1037
BuckwheatSSP              ---GGATTGGAGCAAGCGTTCTGCAACCTGAAATTCAGGCAAAATGTTAA 47
Citrus_citrin             AATGGCTTCGAGGAAACTATCTGTACAATGAAACTAAGGCACAACATCGA 1008
Gossypium_β-globulinB     AACGGCTTAGAAGAAACATTCTGCTCAATGAGACTGAAACACAGGACCCC 1052
Ficus_11Sglobulin         AACGGCTTGGAAGAGACATTCTGCACTCTGAGGATGAGGCACAACATCGA 1006
                            **  * **  *     ****     * *   * *  *  *

F.esculentum_87982        CACGCCTACTAACACCTACGTCTTCAACCCACGCGCCGGTCGTATCAACA 1048
F.esculentum_87980        CAGGCCTTCTCGCGCCGACGTCTTCAACCCACGCGCCGGTCGTATCAACA 1249
F_tataricum               CAGGCCTTCTCACGCCGACGTCTTCAACCCACGCGCCGGACGTATCAACA 1087
BuckwheatSSP              CAGGCCTTCTCGCGCCGACGTCTTCAACCCACGCGCCGGTCGTATCAACA 97
Citrus_citrin             TAAACCATCACACGCTGATGTCTACAACCCCCGGGCCGGACGTGTCACCA 1058
Gossypium_β-globulinB     TGCTTCCTC---TGCTGATGTTTTCAACCCACGAGGTGGTCGCATCACCA 1099
Ficus_11Sglobulin         CCGCCCTTCTCAGGCCGACATCTTCAACCCCCGCGGCGGCCGCTTCACCA 1056
                               *  *     *  *  * * ****** ** *  ** **  *** **

F.esculentum_87982        CCGTGAACAGCAACAGTCTCCCAATCCTCGAATTCCTCCAACTTAGCGCC 1098
F.esculentum_87980        CTGTAAACAGCAACAATCTCCCAATCCTCGAATTCATCCAGCTTAGCGCC 1299
F_tataricum               CCGTCAACAGTAACAATCTCCCAATCCTCGAATTCCTCCAACTTAGCGCC 1137
BuckwheatSSP              CCGTAGACAGCAACAATCTCCCGATCCTCGAATTCATCCAACTTAGCGCC 147
Citrus_citrin             CCGTCAACAGATTCAACCTTCCTATCCTTCGAGACCTCCAGCTTAGTGCT 1108
Gossypium_β-globulinB     CAGTTAACAGTTTCAATCTTCCCATTCTCCAATACCTCCAACTCAGCGCC 1149
Ficus_11Sglobulin         CCGTCAACAACTTCAACCTCCCCATCCTCCGTTTCCTCCGCCTCACCGCC 1106
                          * **  ***     **  ** ** ** **       * ***  ** *  **

F.esculentum_87982        CAGCACGTCGTCCTCTACAAGAATGCGATCATCGGACCGAGATGGAACTT 1148
F.esculentum_87980        CAGCACGTCGTCCTCTACAAGAATGCGATCCTCGGACCGAGATGGAACTT 1349
F_tataricum               CAACACGTCGTCCTCTACAAGAATGCGATCATCGGACCGAGATGGAACTT 1187
BuckwheatSSP              CAGCACGTCGTCCTCTACAAGAATGCGATCCTCGGACCGAGATGGAACTT 197
Citrus_citrin             GAGAAAGGAAACCTTTACCCGAATGCCCTGTTGGCGCCACAGTGGAACTT 1158
Gossypium_β-globulinB     GAGAGGGGAGTCCTTTACAATAATGCTATCTACGCTCCTCACTGGAACAT 1199
Ficus_11Sglobulin         GAGAGAGGTGTCCTCTACAAGAACGCTATGATGGCACCACACTTCAACTT 1156
                           *    *   *** ***   ** **  *     *  **     *  *** *

F.esculentum_87982        GAACGCGCACAGCGCACTGTACGTGACGAGAGGAGAAGGAAGAGTCCAAG 1198
F.esculentum_87980        GAACGCGCACAGCGCACTGTACGTGACGAGAGGAGAAGGAAGAGTCCAAG 1399
F_tataricum               GAACGCACACAGCGCACTGTACGTGACAAGAGGAGAAGGAAGAGTCCAGG 1237
BuckwheatSSP              GAACGCGCACAGCGCACTGTACGTGACGAGAGGAGAAGGAAGAGTCCAGG 247
Citrus_citrin             GAATGCCCACAGCATAGTCTACGTAACAAGGGGCAACGGCAGGATGCAAA 1208
Gossypium_β-globulinB     GAATGCCCACAGCATTGTTTACATCACAAGGGGAAATGGAAGGATTCAAA 1249
Ficus_11Sglobulin         GAACAGCCACAGCGTGTTCTACGTCACCAGGGGAAGCGGCCGATGCCAGA 1206
                          ***    ******    * *** * ** ** **     **  *    **

F.esculentum_87982        TTGTTGGAGACGAAGGAAAGAGTGTGTTCGACGACAAGGTGCAGCGAGGA 1248
F.esculentum_87980        TTGTGGGAGATGAAGGAAGAAGTGTGTTCGACGACAACGTGCAGCGAGGA 1449
F_tataricum               TTGTTGGAGACGAAGGAAAGAGTGTATTCGACGACAACGTGCAGCGAGGA 1287
BuckwheatSSP              TTGTCGGAGATGAAGGAAGAAGTGTGTTCGACGACAACGTGCAGCGAGGA 297
Citrus_citrin             TTGTAGCGGAGAACGGGGAGAATGTGTTCGACGGTCAAATCCGGGAGGGT 1258
Gossypium_β-globulinB     TTGTGTCGGAAAATGGAGAGGCGATATTCGATGAGCAGGTTGAGAGGGGT 1299
Ficus_11Sglobulin         TCGTCGACGACTTCGGGCGGAACGTGTTCGATGGTGAGGTCCAGGAGGGG 1256
                          * **     **     **        * ***** *   *   *   *   **

F.esculentum_87982        CAGATCCTTGTGGTGCCACAGGGATTCGCTGTGGTGTTGAAGGCAGGAAG 1298
F.esculentum_87980        CAGATCCTTGTAGTCCCACAGGGATTCGCAGTGGTGTTGAAGGCAGGAAG 1499
F_tataricum               CAGATCCTTGTGGTGCCACAGGGATTCGCAGTGGTGGTGAAGGCAGGAAG 1337
BuckwheatSSP              CAGATCCTTGTGGTCCCACAGGGATTCGCAGTGGTGTTGAAGGCAGGAAG 347
Citrus_citrin             CAGCTGATCGTTGTTCCGCAGGGCTTCGCCGTCGTGAAGAGGGCAGGTAA 1308
Gossypium_β-globulinB     CAGGTCATAACCGTACCCCAGAATCATGCAGTGGTGAAAAAAGCAGGAAG 1349
Ficus_11Sglobulin         CAGCTGTTGGTAGTGCCACAGAACTACGCCGTGGCGAAGCAAGCCAGCAA 1306
                          *** *  *     ** ** ***        ** ** * *        **  * *


F.esculentum_87982        AGAAGGATTGGAGTGGGTGGAGTTGAAGAACAGCGGCAACGCCATAACCA 1348
F.esculentum_87980 AGAAGGACTGGAGTGGGTGGAGTTGAAGAACGACGACAACGCCATAACCA 1549
F_tataricum               ACAAGGATTGGAGTGGGTGGAGTTGAAGAACAACGATAACGCCATAACCA 1387
BuckwheatSSP              AGAAGGACTGGAGTGGGTGGAGTTGAAGAACGACGACAACGCCATAACCA 397
Citrus_citrin             CCGTGGACTGGAGTGGATATCATTCAAGACAAACGACGTCGCCATGACAA 1358
Gossypium_β-globulinB     GCGAGGGTTTGAATGGATAGCATTCAAGACAAATGCCAATGCTAAGATTA 1399
Ficus_11Sglobulin         CCGCGGATTCGAGTGGATCGCCATCAAGACCAACGACAACGCCATGAGAA 1356
                             **  * ** *** *     * ****      *      ** *  *  *
```

```
F.esculentum_87982         GTCCGATCGGCGGGAGGACATCGGTGTTGAGGGCGATCCCTGTGGAGGTT 1398
F.esculentum_87980         GTCCGATTGCCGGGAAGACTTCGGTGTTGAGGGCGATCCCTGTGGAGGTT 1599
F_tataricum                GTCCGATTGCCGGTAGGACTTCGGTGTTGAGGGCGATCCCTGTGGAGGTA 1437
BuckwheatSSP               GCCCGATTGCCGGGAAGACTTCGGTGTTGAGGGCGATCCCTGTGGAGGTT 447
Citrus_citrin              GCCAGCTGGCCGGAAGGGCTTCGGTGTTAAGAGGGCTTCCGTTGGACGTT 1408
Gossypium_β-globulinB      GTCAGATTGCTGGACGTGTCTCCATTATGCGAGGATTGCCGGTGGACGTG 1449
Ficus_11Sglobulin          ACCCGCTCGCCGGAAGGATTTCGGCCATCCGAGCAATGCCGGAGGACCTG 1406
                             * * * *  **       **     *  * *   * **   ***  *

F.esculentum_87982         CTTGCCAACTCGTATGATATCTCGACGAAGGAAGCATATAAATTGAAGAA 1448
F.esculentum_87980         CTTGCCAACTCATACGATATCTCGACGAAGGAAGCTTTCAGATTGAAGAA 1649
F_tataricum                CTGGCCAACTCGTATGATATCTCGACGGAGGAAGCATACAAATTGAAGAA 1487
BuckwheatSSP               CTTGCCAACTCCTACGATATCTCGACGAAGGAAGCGTTCAGATTGAAGAA 497
Citrus_citrin              ATCCAGAACTCGTTCCAAGTGTCGAGGGATGAAGCTCAGAGGTTGAAGTA 1458
Gossypium_β-globulinB      CTGGCCAACTCCTTTGGTATATCCCGGGAGGAGGCCATGAGGTTGAAGCA 1499
Ficus_11Sglobulin          CTATTTAACGCATTCCGCATTTCGAGAGAGCAAGCCAGGAACTTGAAGAA 1456
                            *    *** * *      * **     *  * **    *  ****** *

F.esculentum_87982         TGGGAGGCAGGAGGTTGAGGTCTTCCGACCA---TTCCAGTCCCGAGATG 1495
F.esculentum_87980         TGGGAGGCAGGAGGTTGAGGTCTTCCTACCA---TTCCAGTCCCGAGATG 1696
F_tataricum                TGGGAGGCAGGAGGTTGAGGTCTTCCGACCA---TTCCAGTCCCGATATG 1534
BuckwheatSSP               TGGGAGGCAGGAGGTTGAGGTCTTCCGACCA---TTCCAGTCCCGAGATG 544
Citrus_citrin              CAACAGGCAGGAGCTGACTGTGTTTACTCCAGGGCCTAGGTCGCAGTGG- 1507
Gossypium_β-globulinB      TAACAGACAGGAGGTGTCGGTTTTTAGCCCAAGGCAAGGCTCGCAACAGT 1549
Ficus_11Sglobulin          CAACAGGGAAGAGGTCACTGTTTTCAGCACCCAGTTTTTTTAACAAGGCA 1506
                               **  * *** *    ** **     *          *  *

F.esculentum_87982         AGAAGGAG------------AGGGAGCGTTTCTCCATAGTTTAAGAGGAA 1533
F.esculentum_87980         AGAAGGAG------------AGGGAGCGTTTCTA-AGAG----AGACAAA 1729
F_tataricum                AGAAGGAGGAGGAGAAGGAGAGGGAACGTTTCTCCATAGTTTAAGAGAGA 1584
BuckwheatSSP               AGAAGGAG------------AGGGAGCGTTTCTCCATAGTTTAA------ 576
Citrus_citrin              GGCTTAACCG--TAG-CCTAAGAGAAATTATGTAAATG--GGTACGTAAT 1552
Gossypium_β-globulinB      AGATCAAACAACTAA-TCTAGGTGTACGTACGTACGTA--ATTGCATCAA 1596
Ficus_11Sglobulin          GGAAGTACTAATAAGCTCCTGTTTCAAACGAGTAAAAC-GTTACCCACC 1555
                            *    *                         *
```

Fig.4.13

This deletion was not detected in the nucleotide sequence of gene coding for the allergenic protein (acc. no. DQ849083) from grains of tartary buckwheat (*Fagopyrum tataricum*). These features distinguished the nucleotide sequence amplified in the present study from nucleotides sequences of genes coding for legumin type seed storage proteins of other plants as well as from the nucleotide sequence of gene coding for the allergenic protein of tartary buckwheat.

ORF finder tool from NCBI (http://www.ncbi.nlm.nih.gov) identified 5 open reading frames in the 576 bp nucleotide sequence. The predicted amino acid sequence consisting of 191 residues is presented in fig. 4.14. Sequence similarity analysis of the sequence with BLASTp against non-redundant protein database, identified the protein as belonging to the legumin family. The 7-element fingerprint, which has been universally identified as a signature for the 11-13S globulin family of proteins, was also detected in the deduced amino acid sequence of 191 residues. The 7 elements identified in the deduced amino acid sequence were, 'RQNVNRPSRADVFNPRAG' at P'12-29, 'DSNNLPILEFIQLSAQHVVLY' at P'35-55,'NAILGPRWNLNAHSALYVTRG' at P'57-79, 'VQVVGDEGRSVFDDNVQ' at P'81-97, 'GQILVVPQGFAVVLKA' at P'99-114, 'EGLEWVELKNDDNAITSPI' at P'117-135 and 'TSVLRAIPVEVLANSY DI' at P'139-156.

The sequence was subjected to MOTIF SCAN (http://myhits.isb-sib.ch/cgi-bin/motif_scan) for motif/domain search for identification of motifs within the sequence. Motif search on the deduced amino acid sequence with MOTIF SCAN identified a putative N-myristoylation site represented by the amino acid sequence Gly-Leu-Glu-Gln-Ala-Phe between residues 1-6, 5 putative kinase II phosphorylation sites

Fig. 4.14: Deduced amino acid sequence of the 576 bp nucleotide sequence for the 0.5 Kb amplicon generated from cDNA of grains harvested at mid maturation stage of development and primer pair NKC5 and NKC6. The " ☐" indicates the motif patterns and segments highlighted with yellow colour indicates the 7-element fingerprint domains.

Fig. 4.15(a): Diagrammatic representation of the motif patterns and frequency in the 191 amino acid sequence as predicted by MOTIF SCAN.

(b): Diagrammatic representation of the N-myristoylation site and Kinase II phosphorylation sites in the 191 deduced amino acid sequence as predicted by MOTIF SCAN.

(c): Diagrammatic representation of the cupin domain positioned between 13 to 162 in the 191 deduced amino acid sequence as predicted by MOTIF SCAN.

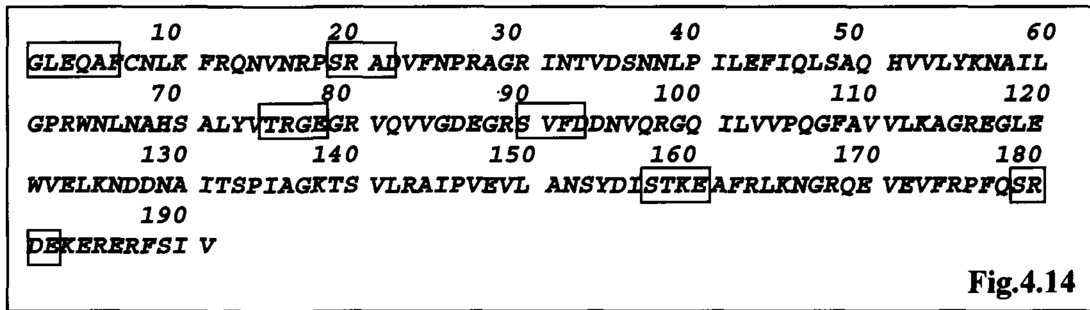
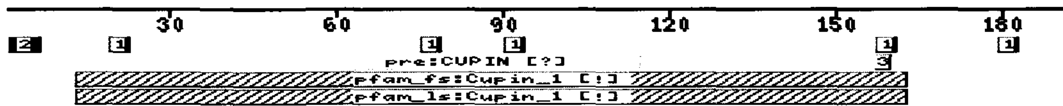
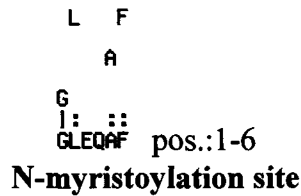


Fig.4.14



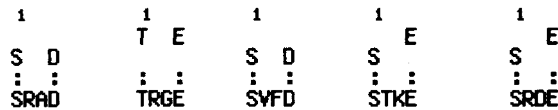
1: Kinase 2 phosphorylation site, 2: N-myristoylation site, 3: Kinase C phosphorylation site

Fig.4.15a



pos.:1-6

N-myristoylation site



Positions, 19-22, 75-78, 90-93, 157-160, 179-182

Kinase II phosphorylation sites

Fig.4.15b

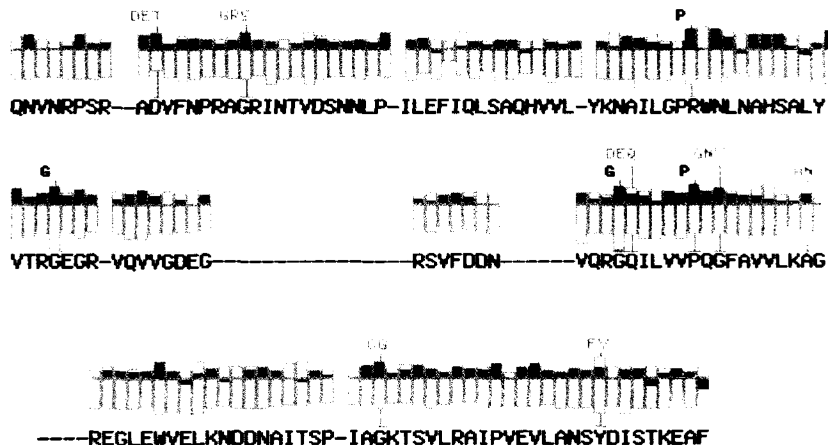


Fig.4.15c

represented by the sequences Ser-Arg-Ala-Asp' (P'19-22), Thr-Arg-Gly-Glu' (P'75-78), Ser-Val-Phe-Asp' (P'90-93), Ser-Thr-Lys-Glu' (P'157-160), and Ser-Arg-Asp-Glu' (P'179-182) and a protein kinase C phosphorylation site represented by the sequence Gly-Leu-Glu-Gln-Ala-Phe' between residues 157-160 in the deduced amino acid residue (Fig. 4.15a,b). The software also identified a cupin domain predominantly represented by amino acids which would preferentially fold to form barrel shaped structures between residues 13 to 162 in the deduced amino acid sequence (Fig.4.15c). When the amino acid residues of the sequence were plotted as a function of hydropathic index, the sequence showed a predominantly hydrophilic character (Fig.4.16). Based on the hydropathic index of Kyte and Doolittle (1982), the major regions of hydrophilic nature detected in the sequence were between residues 3-43, 60-100, 114-134 and 152-189. Statistical analysis by SAPS (Brendel *et al.,* 1992) revealed that the charged residues were evenly distributed in the sequence with the sequence having 30.4% non polar residues, 21.5% polar uncharged residues and 25.6% polar charged residues. Regions of high charge residue concentration in the deduced protein sequence correlated with the predicted hydrophilic domains in the hydropathy analysis.

The secondary structure of the sequence was predicted using GOR4 (Garnier *et al.*, 1996). The software estimated the protein to have 28.27% α-helix, 18.32% extended strand and 53.40% random coil structure. Statistical analysis indicated that residues between 1-40 and 57-103 were predominantly random coils whereas residues between 41-56 and 108-186 had predominantly α-helix configuration (Fig. 4.17).

The ClustalW multiple alignment of the sequence with amino acid sequences of other seed storage proteins belonging to the legumin family is presented in fig. 4.18.

Fig. 4.16: Hydropathy profile of the 191 deduced amino acid sequence as predicted by Kyte and Doolittle method. The x-axis represents the residue number.

Fig.4.17: Secondary structure prediction profile of the 191 deduced amino acid sequence using GOR4 method.

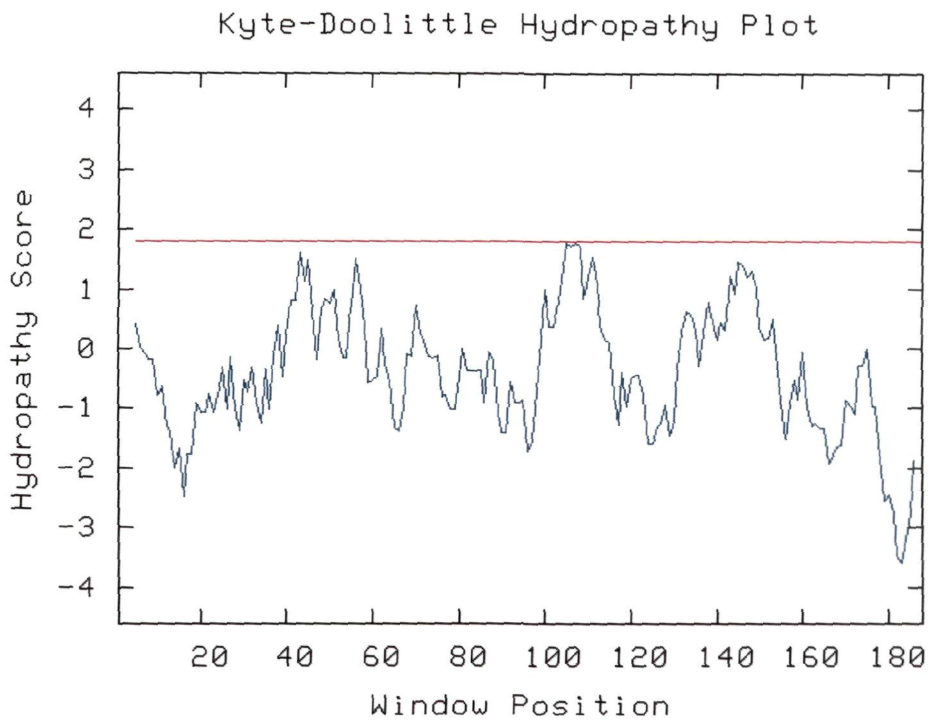


Fig.4.16

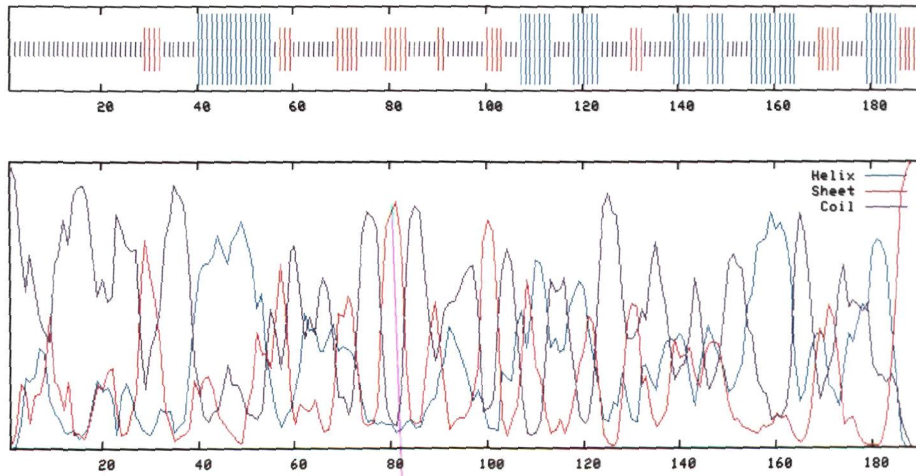


Fig.4.17

Using an alignment that permitted maximum homology, the deduced amino acid sequence of 191 residues showed >90% homology with the acidic chain of buckwheat 13S globulin (acc. no. Q9XFM4) and allergenic protein (acc. no. AAF34635). The percentage homology ranged from 43% - 53% with legumin genes from other plants including *Coffea arabica* (acc. no. AAC61881) and *Magnolia salicifolia* (CAA57846). The 7-element fingerprint, detected in the deduced amino acid sequence of 191 residues shared 100% homology with *Fagopyrum esculentum* 13S globulin and allergenic proteins (acc. no. Q9XFM4 and AAF34635). The "motif 4" comprising of the sequence "VQVVGDEGRSVFDDNVQ" and "motif 7" comprising of the sequence "TSVLRAIPVEVLANSYDI" were observed to be highly conserved in all the seed storage proteins from *Fagopyrum esculentum* as well as *F. tataricum* (acc. nos. Q9XFM4, AAF34635, O23880 and AAK97787). Out of the 191 residues compared, the position of 49 residues was found to be conserved amongst all the legumin like seed storage proteins analyzed in the present study. The conserved residues included P1'G, P7'C, P14'N, P18'P, P25'N, P26'P, P28'A, P29'G, P30'R, P40'P, P41'I, P42'L, P46'Q, P48'S, P54'L, P55'Y, P57'N, P58'A, P62'P, P64'W, P65'N, P67'N, P68'A, P69'H, P70'S, P73'Y, P77'G, P80'R, P82'Q, P84'V, P88'G, P92'F, P99'G, P100'Q, P105'P, P106'Q, P110'V, P118'G, P125'K, P137'G, P140'S, P146'P, P149'V, P157'S, P161'A, P164'L, P165'K, P168'R, P170'E. Other residues which were represented by amino acids with invariant/similar functional groups included P2' L/I/V, P3' E/D, P8' N/T/S, P10' K/R/E, P15' V/I/F, P23' V/F/I/T, P24' F/Y, P31' I/L/V, P32' N/T, P34' V/L, P39' L/I/F, P45' I/V/L, P47' L/M, P49' A/V/I, P51' H/R/K, P52' V/G/A, P59' I/L, P66'L/M/I/V, P71' A/I/L/V, P72' L/I/R/V, P74' V/I/A/M, P79' G/A, P81' V/I, P83'

Fig. 4.18: ClustalW (1.81) multi-alignment of the 191 deduced amino acid sequences (BuckwheatSSP) for the 576 bp nucleotide sequence (Buckwheat) of cDNA from mid matured seed stage generated using primer pair NKC5 and NKC6 with seed storage protein sequences of other plants derived from gene bank. ‘*’ denotes conserved residues.

```
CLUSTAL W (1.81) multiple sequence alignment

                                                                                 60
Populus_protein        GIEETFCTARLKHNINDPERADFFNPRAGRLTTVNSLNLPILRSVQLSVERGVLYPNALM
Ricinus_SSP            GLEETFCTRRLRHNINKPSEADIYNPRAGRVTSVNSHNLPILRYLQLSIQKAVLYKNALM
Buckwheat_SSP          GLEQAFCNLKFRQNVNRPSRADVFNPRAGRINTVDSNNLPILEFIQLSAQHVVLYKNAIL
F.escul_13S1glob.      GLEQAFCNLKFRQNVNRPSRADVFNPRAGRINTVDSNNLPILEFIQLSAQHVVLYKNAIL
F.escul_allergenic     GLEQAFCNLKFRQNVNRPSRADVFNPRAGRINTVDSNNLPILEFIQLSAQHVVLYKNAIL
F.tarta_allergenic     GLEQAFCNLKFRQNVNRPSHADVFNPRAGRINTVNSNNLPILEFLQLSAQHVVLYKNAII
F.escul_12S2glob.      GVEQGFCNLKFRRNFNTPTNTYVFNPRAGRINTVNSNSLPILEFLQLSAQHVVLYKNAII
F.gracilipes_22kDa     GLEQAFCNLKFRQNVNNPARADVFNPQAGRINTVDSNTLPILDFLQLSAQHVVLYKNAIL
Coffea_11S             GLEETLCTVKLSENIGLPQEADVFNPRAGRITTVNSQKIPILSSLQLSAERGFLYSNAIF
Magnolia_leg_PRCSR     GLEEIQCSSKLTYNIADPTRADVYNPQAGRITSLNSQKLPILNVLQLSAERGVLYRNALL
Avena_11S              GLEENFCSLEARKNIENPQHADTYNPRAGRITRLNSKNFPILNIVQMSATRVNLYQNAIL
Oryza_hypo_protein     GLDETFCTMRVRQNIDNPNLADTYNPRAGRITYLNGQKFPILNLVQMSAVKVNLYQNALL
                       *:::  *. .   *.  *  :  :**:***:.::.  .:***  :*:*:    ** **::

                                                                                120
Populus_pred.          SPHWNMNAHSIIYITRGNGRIQIVGDNGQTIFDGEVREGQVVTAPQSFAVVKKAGSQGFE
Ricinus_SSP            TPHWNINAHSIRYITRGSGRVQIVNENGDSVFDGQVQRGQMFTVPQNFVVITKASNEGLE
Buckwheat_SSP          GPRWNLNAHSALYVTRGEGRVQVVGDEGRSVFDDNVQRGQILVVPQGFAVVLKAGREGLE
F.escul_13S1glob.      GPRWNLNAHSALYVTRGEGRVQVVGDEGRSVFDDNVQRGQILVVPQGFAVVLKAGREGLE
F.escul_allergenic     GPRWNLNAHSALYVTRGEGRVQVVGDEGRSVFDDNVQRGQILVVPQGFAVVLKAGREGLE
F.tarta_allergenic     GPRWNLNAHSALYVTRGEGRVQVVGDEGKSVFDDNVQRGQILVVPQGFAVVVKAGRQGLE
F.escul_12S2glob.      GPRWNLNAHSALYVTRGEGRVQVVGDEGKSVFDDKVQRGQILVVPQGFAVVLKAGREGLE
F.gracilipes_22kDa     GPRWNLNAHSALYVTRGEGRVQVVGDEGNAVFDDVVQRGQILVVPQGFAVVLRKRREGLV
Coffea_11S             APHWNINAHSALYVIRGNARIQVVDHKGNKVFDDEVKQGQLIIVPQYFAVIKKAGNEGFE
Magnolia_leg_PRCSR     APQWNVNAHSLVYATRGNGRVQIVGEQGRPVFDGELREGQLVVVPQSFAVVKKAGNEGFE
Avena_11S              SPFWNINAHSVIYMIQGHARVQVVNNNGQTVFSDILHRGQLLIVPQHFVVLKNAEREGCQ
Oryza_hypo_protein     SPFWNINAHSVVYITQGRARVQVVNNNGKTVFDGELRRGQLLIIPQHHVVIKKAQREGCS
                        * **:****  *  :* .*:*:*..:*  :*.. ::.**:.  ** ..*: .   :*

                                                                                 180
Populus_pred.          WVSFKTNDNAQVSELAGRVSTIRGLPVEVVANSFQISREDARRLKNNR-EEVSVFSPSQS
Ricinus_SSP            WVSFKTNDNAKINQLAGRVSAIRSMPEEVVANAFQVSVEDARRLKDNR-QEVTLLSPGS-
Buckwheat_SSP          WVELKNDDNAITSPIAGKTSVLRAIPVEVLANSYDISTKEAFRLKNGR-QEVEVFRPFQS
F.escul_13S1glob.      WVELKNDDNAITSPIAGKTSVLRAIPVEVLANSYDISTKEAFRLKNGR-QEVEVFRPFQS
F.escul_allergenic     WVELKNDDNAITSPIAGKTSVLRAIPVEVLANSYDISTKEAFRLKNGR-QEVEVFRPFQS
F.tarta_allergenic     WVELKNNDNAITSPIAGRTSVLRAIPVEVLANSYDISTEEAYKLKNGR-QEVEVFRPFQS
F.escul_12S2glob.      WVELKNSGNAITSPIGGRTSVLRAIPVEVLANSYDISTKEAYKLKNGR-QEVEVFRPFQS
F.gracilipes_22kDa     -VELKNSDNAVTSPIAGKTSVLNAIPVDVLATAYDISKPEAFKLKNGRRQEIEVFRPFQS
Coffea_11S             YVAFKTNDNAMINPLVGRLSALRAIPEEVLRSSFQISSEEAEELKYGR-QEALLLS-EQS
Magnolia_leg_PRCSR     YVAFKTNDNAMNSPLVGKTSVIRAMPEDVLINSYRISREEARRLKYNR-EEIAVFAPRFS
Avena_11S              YISFKTNPNSMVSHIAGKTSILRALPIDVLANAYRISRQEARNLKNNRGEEFGAFTPKLT
Oryza_hypo_protein     YIALKTNPDSMVSHMAGKNSIFRALPDDVVANAYRISREEARRLKHNRGDELGVFTP---
                        : :*.. ::  . : *: * :.. :* :*: .:: :*  :* .** .* :*   :

Populus_pred.         G------RSEEIA-----------
Ricinus_SSP            -------RST--------------
Buckwheat_SSP          R----DEKERERFSIV--------
F.escul_13S1glob.      R----DEKERERFSIV--------
F.escul_allergenic     R----DEKERERFSIV--------
F.tarta_allergenic     RYEKEEEKERERFSIV--------
F.escul_12S2glob.      R----DEKERERFSIV--------
F.gracilipes_22kDa     R----DEKERERFSIV--------
Coffea_11S             Q-----QGKREVA-----------
Magnolia_leg_PRCSR     S------QQRAAA-----------
Avena_11S              QTGFQSYQDIEEASS---------
Oryza_hypo_protein     SHAYKSYQDISISRAPEQHGELDK
                               .
```

Fig.4.18

V/I, P90' V/I, P95'V/L, P100' I/V/L/M, P101' L/V/F/I, P102' V/T/I, P103' V/A/I, P108' A/V, P110' V/I/L, P121' W/Y, P122' V/I, P125' N/T, P132' S/N, P134' I/L/M, P135' A/G/V, P137' K/R, P140' V/T/A/I, P141' L/I/F, P144' I/L/M, P147' E/D, P149' L/V, P151' N/T/S, P153' Y/F, P155' I/V, P159' E/D, P170' V/A/I/F/L, P172' V/L/A, P178' S/T.

A phylogenetic tree constructed from the alignment data of the deduced amino acid sequence with amino acid sequences of seed storage proteins available in EMBL database is presented in fig.4.19. The deduced amino acid sequence of 191 residues identified in the present study clustered together into one clade with amino acid sequences of *Fagopyrum esculentum* 13S globulin (acc. no. Q9XFM4, O023880, AAF34635). Even though the amino acid sequences of allergenic protein from *F. tataricum* and the 22kDa seed storage protein from *F. gracilipes* clustered separately, they showed greater closeness with the group comprising the sequences bearing accession nos. Q9XFM4, O023880, AAF34635 and the sequence of 191 residues identified in the present study. Out of all the sequences analyzed in the present study, the amino acid sequence of *Coffea arabica* 11S globulin appeared to be the closest relative of buckwheat sequences. While the amino acid sequences of *Ricinus communis* (acc. no. AAF73008) and P*opulus trichocarpa* (acc.no. XP002329472) formed one group, the globulin type seed storage proteins of *Magnolia salicifolia* (acc. no. CAA57846), *Avena sativa* (acc. no. CAA52763) and *Oryza sativa* hypothetical protein (acc.no. EEC75527) emerged as distinctly separate clades.

The CTAB method for isolation of genomic DNA, used in the present study, yielded a fairly good DNA with the yield of DNA ranging from 15-20µg DNA gm^{-1}

Fig. 4.19: Phylogenetic tree based on the alignment matrix of the 191 deduced amino acid sequences with amino acid sequences of seed storage proteins of various plants available in the gene bank database.

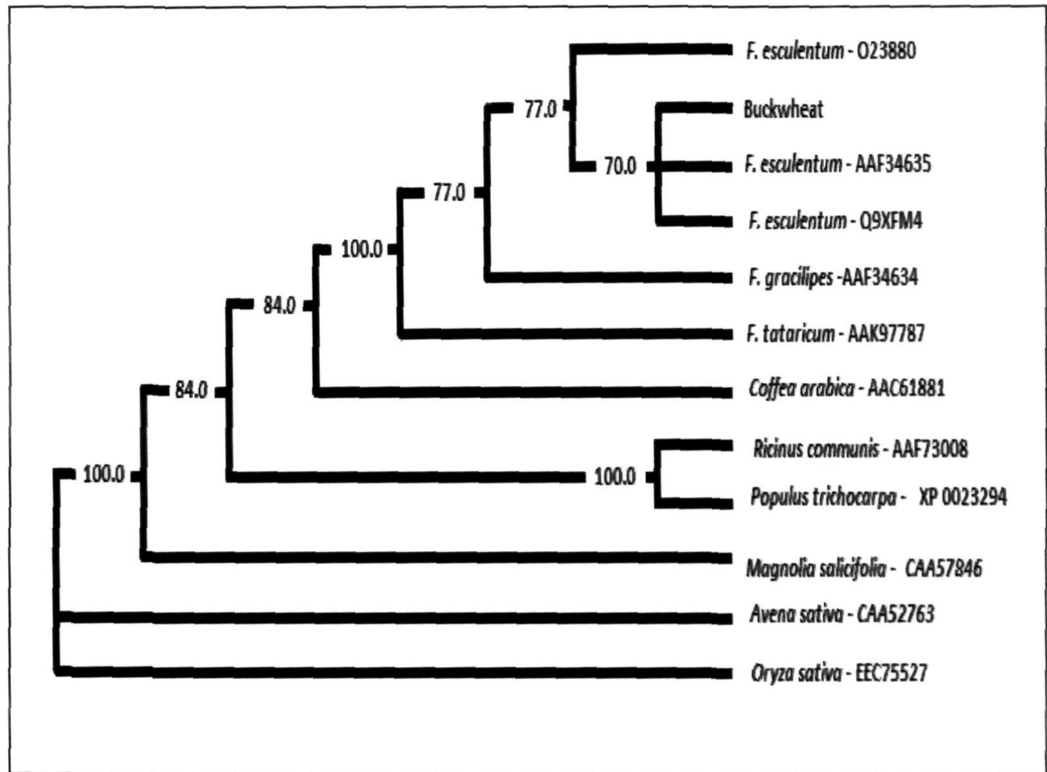


Fig.4.19

powdered tissue. Under UV light, the agarose gel was detected as a single high molecular band, corresponding to a molecular mass of around 21 kb (Fig. 4.20a). Amplification with oligonucleotide primer pair NKC5-NKC6 and amplification program comprising of hot start at 94°C for 5 minutes, 35 cycles comprising of denaturation at 94°C for 1 min., annealing at 60°C for 1 min. and chain extension at 72°C for 1 min. followed by elongation at 72°C for 10 min. amplified a 0.7 Kb DNA from the template genomic DNA (Fig. 4.20b). The partial nucleotide sequence of 455 bases for the amplified DNA is presented in fig. 4.21. Pairwise alignment of the sequence with the nucleotide sequence of the amplicon generated with primer pair NKC5-NKC6 and cDNA library from grains harvested at 16-20 DAF as the template clearly revealed that the two sequences shared a high degree of homology. ClustalW pairwise alignment between the nucleotide sequences of the 0.5 kb amplicon generated from mid maturation stage specific cDNA library and the 0.7 kb amplicon generated from genomic DNA as the template revealed a gap of 96 bases corresponding to position 142-238 bases in the 455bp nucleotide sequence of the amplicon generated from genomic DNA as the template and primer pair NKC5-NKC6 (Fig.4.22). GENSCANW (http://genes.mit.edu/GENSCAN.html) identified this region of the nucleotide sequence of the 0.7kb amplicon as an intron (Fig. 23). The 0.7 Kb amplicon was used as a probe to assess changes in the transcript level corresponding to the gene during different stages of development in grains of common buckwheat. 2μg each of the total RNA sample extracted from grains harvested between 10-13 DAF, 16-20 DAF, 23-25 DAF and 30-35 DAF were dot blotted on positively charged nylon membranes. Northern hybridization with the [α-P^{32}]-dATP labelled 0.7 Kb PCR product, amplified

Fig.4.20(a): Total genomic DNA isolated from 15 days old seedling of *Fagopyrum esculentum* Moench;
M = EcoRI/HindIII double digest λ DNA,
L1–L3 = Genomic DNA

(b): PCR amplification profile of buckwheat genomic DNA as template and NKC5 and NKC6 as the primer pair.

Fig. 4.21: Nucleotide sequence of 455bp for the DNA fragment 0.7 Kb amplified from buckwheat genomic DNA using primer pair NKC5-NKC6.

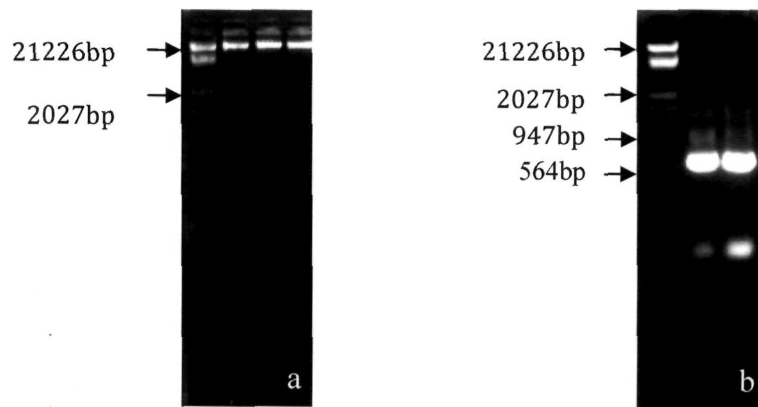


Fig.4.20

```
1    CAACTGAATA  GCAATGTTAC  AGGCCTTCTC  GCGCCGACGT  CTTCAACCCA
51   CGCGCCGGGT  CGTATCAACA  CCGGTAAACA  GCAACAATCT  CCCAATCCTC
101  GAATTCATCC  AGCTTAGCGC  CCAGCACGTC  GTCCTCTACA  AGGTAATTTA
151  AACCCCAGTT  GAAGACCTAA  AACCGACAAA  TCTGGAAAGT  GAAGCTTGAT
201  TAACTTAAGA  TTGACGATAA  TTCGTTGAAT  TTATGCAGAA  TGCGATCCTC
251  GGACCGAGAT  GGAACTTGAA  CGCGCACAGC  GCACTGTACG  TGACCAGAGG
301  AGAAGGAAGA  GTCCAGGTTG  TGGGAGATGA  AGGAAGAAGT  GTGTTCGACG
351  ACTACGTGCA  GCGAGGACAG  ATCCTTGTGG  TCCCACAGGG  ATTCGCAGTG
401  GTGTTGAAGG  CAGGAAAGGA  AGGACTGGAG  TGGGTGGAGT  TGAAGAAGGG
451  GATAA
```

Fig.4.21

Fig. 4.22: ClustalW (1.81) pairwise alignment of the 576bp nucleotide sequence generated from 16-20 DAF cDNA template and 455bp nucleotide sequence generated from genomic DNA using primer pair NKC5-NKC6 (* denotes conserved residues).

```
cDNA576bp   GGATTGGAGCAAGCGTTCTGCAACCTGAAATTCAGGCAAAATGTTAACAGGCCTTCTCGC 60
gDNA455bp   ---------------------------CAACTGAATAGCAATGTTA-CAGGCCTTCTCGC 32
                                            ** * *     ******* *************

cDNA576bp   GCCGACGTCTTCAACCCACGCGCCGG-TCGTATCAACACCG-TAGACAGCAACAATCTCC 118
gDNA455bp   GCCGACGTCTTCAACCCACGCGCCGGGTCGTATCAACACCGGTAAACAGCAACAATCTCC 92
            ************************** ************** ** ***************

cDNA576bp   CGATCCTCGAATTCATCCAACTTAGCGCCCAGCACGTCGTCCTCTACAAG---------- 168
gDNA455bp   CAATCCTCGAATTCATCCAGCTTAGCGCCCAGCACGTCGTCCTCTACAAGGTAATTTAAA 152
            * ***************** ******************************

cDNA576bp   ------------------------------------------------------------
gDNA455bp   CCCCAGTTGAAGACCTAAAACCGACAAATCTGGAAAGTGAAGCTTGATTAACTTAAGATT 212

cDNA576bp   --------------------------AATGCGATCCTCGGACCGAGATGGAACTTGAACG 202
gDNA455bp   GACGATAATTCGTTGAATTTATGCAGAATGCGATCCTCGGACCGAGATGGAACTTGAACG 272
                                      **********************************

cDNA576bp   CGCACAGCGCACTGTACGTGACGAGAGGAGAAGGAAGAGTCCAGGTTGTCGGAGATGAAG 262
gDNA455bp   CGCACAGCGCACTGTACGTGACCAGAGGAGAAGGAAGAGTCCAGGTTGTGGGAGATGAAG 332
            ********************** ************************** **********

cDNA576bp   GAAGAAGTGTGTTCGACGACAACGTGCAGCGAGGACAGATCCTTGTGGTCCCACAGGGAT 322
gDNA455bp   GAAGAAGTGTGTTCGACGACTACGTGCAGCGAGGACAGATCCTTGTGGTCCCACAGGGAT 392
            ******************** ***************************************

cDNA576bp   TCGCAGTGGTGTTGAAGGCAGGAAGAGAAGGACTGGAGTGGGTGGAGTTGAAGAACGACG 382
gDNA455bp   TCGCAGTGGTGTTGAAGGCAGGAAAGGAAGGACTGGAGTGGGTGGAGTTGAAGAA-GGGG 451
            ************************  ***************************** *  *

cDNA576bp   ACAACGCCATAACCAGCCCGATTGCCGGGAAGACTTCGGTGTTGAGGGCGATCCCTGTGG 442
gDNA455bp   ATAA-------------------------------------------------------- 455
            * **

cDNA576bp   AGGTTCTTGCCAACTCCTACGATATCTCGACGAAGGAAGCGTTCAGATTGAAGAATGGGA 502
gDNA455bp   ------------------------------------------------------------

cDNA576bp   GGCAGGAGGTTGAGGTCTTCCGACCATTCCAGTCCCGAGATGAGAAGGAGAGGGAGCGTT 562
gDNA455bp   ------------------------------------------------------------

cDNA576bp   TCTCCATAGTTTAA 576
gDNA455bp   --------------
```

Fig.4.22

Fig. 4.23: Diagramatic representation of the intron/exon architecture in the 455bp nucleotide sequence of the amplicon generated with genomic DNA as the template and NKC5-NKC6 as the primer pair.

Fig: 4.24: Dot blot of total RNA isolated from grains of common buckwheat harvested at different maturation stages hybridised with [α-P^{32}]-dATP labelled 0.7 Kb fragment from genomic DNA using primers NKC5-NKC6.
E=10-13 DAF, M=16-20 DAF, L=23-25 DAF, MA=30-35 DAF

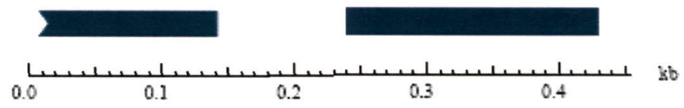


Fig.4.23

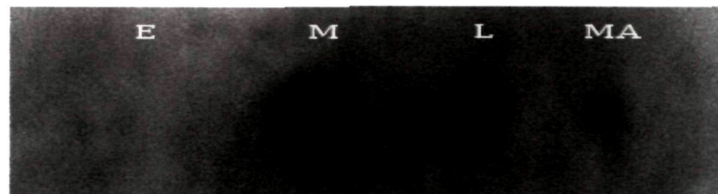


Fig.4.24

in the present study, generated a strong signal against RNA isolated from grains harvested at 16-20 DAF. While the blots did not show any signal against RNA isolated from grains harvested at 10-13 DAF, weak signals could be detected against RNA isolated from grains harvested between 23-25 DAF and 30-35 DAF. Compared with the intensity of signals against RNA isolated from grains harvested between 23-25 DAF and 30-35 DAF, the intensity of signal against RNA isolated from grains harvested between16-20 DAF was much stronger (Fig. 4.24).

CHAPTER : V
DISCUSSION

Seed development proceeds through a series of spatially and temporally regulated steps, which are usually characterized by specific molecular and metabolic events, which control the course of developmental processes. Key metabolic events associated with distinct stages of seed development are reflected by changes in the mRNA subsets within the developing seed. In particular, storage-protein synthesis which is prevalent during the expansion stage, declines markedly during later stages of maturation and does not occur during germination. The changes also include degradation of residual mRNAs for storage proteins present in the dry seed (Dure 1985; Bewley and Marcus 1990; Kermode 1990). Work carried out during the present investigation involved isolation and characterization of novel abundantly expressed seed storage protein genes in maturing grains of common buckwheat.

Standard RNA extraction methods using GITC-phenol-chloroform (Chomczynski and Sacchi. 1987)), RNeasy kit, or TRIzol reagent failed to produce good quality RNA from starch-rich seed species such as buckwheat. The Lopez-

Gomez and Gomez-Lim (1992) protocol used in the present investigation yielded fairly good quality RNA from the grains of common buckwheat. Under UV light, the isolated RNA was detected on the agarose gel as two fluorescent bands corresponding to a typical RNA profile of 28S rRNA and 18S rRNA. The ratio of band thickness as well as the intensity of fluorescence between the bands corresponding to 28S rRNA and 18S rRNA was nearly 2:1. Electrophoresis profile of the isolated RNA did not reveal any significant degradation of the RNA nor did it reveal contamination of the sample with DNA. The ratio of A_{260}/A_{280} for the isolated RNA samples ranged between 1.7 to 1.8 and the yield of RNA from grains at different stages of development ranged between 0.5μg gm^{-1} to 0.75μg gm^{-1} powdered mass.

Isolation of high-quality RNA from plant seeds is very critical for seed-specific gene analysis. Seed endosperm contains very high levels of starch, which causes the solidification of samples in the guanidine isothiocyanate (GITC)-based RNA extraction buffers. Tissues with a high level of starch can also hinder resuspension of precipitated RNA or contaminate the RNA pellet by co-precipitation (Wilkins and Smart. 1996). The presence of two sharp bands corresponding to 28S and 18S rRNA, without any significant degradation, and the high A_{260}/A_{280} ratio clearly indicated that the isolation method used in the present study yielded good quality RNA.

The isolated RNA could be successfully used for Reverse transcriptase polymerase chain reaction and also for construction of a cDNA library. Reverse Transcriptase- PCR with cDNA, prepared from RNA isolated from grains at different stages of maturation, as the template and primer pair CJ1F-CJ2R amplified a 1.0 kb DNA fragment from cDNA synthesized from RNA isolated from grains harvested

between 16 to 20 DAF. The amplicon was not detected in amplification mixtures containing cDNA prepared from RNA isolated from grains harvested at any other stage of development. BLASTn analysis of the nucleotide sequence of the amplicon clearly identified it with gene sequences coding for dehydrin/ derhydrin-like proteins. The sequence showed 99% and 97% homology with *Coffea canephora* dehydrin genes DH2a (acc. no. DQ323989) and DH2b (DQ323990), respectively. The nucleotide sequence showed 67% homology with a *Camellia sinensis* dehydrin gene (acc no. FJ436979). The amplification of the DNA fragment from RNA isolated from mid maturation stage of seed development and not from any other stage. This indicates the absence or low abundance of transcripts coding for the putative buckwheat dehydrin gene in the grains during other developmental stages. These results indicate the mid maturation developmental stage specific expression of the gene in grains of common buckwheat. Similar pattern of LEA gene expression has been reported by Gomez *et al.* (1988), Mundy and Chua (1988), Close *et al.* (1989), Han *et al.* (1997), Choi and Close (2000), Nylander *et al.* (2001) and Hinniger *et al.* (2006).

Over the last few years, several reports have linked higher expression of dehydrins with seed maturation as well as protection of tissues against osmotic stress (Puhakainen *et al.,* 2004; Choi and Close, 2000; Nylander, *et al.,* 2001). Our results on the detection of transcripts coding for dehydrin class of proteins in grains of common buckwheat during mid maturation stage of development are in conformity with other reports indicating enhanced expression of genes coding for dehydrin like proteins during the maturation phase of grain development (Close, 1996; Hinniger *et al.,* 2006). The pattern of accumulation of some of the *Lea* transcripts during seed development, with the highest level of transcripts reported at incipient desiccation,

have led to the suggestion that LEA polypeptides play a decisive role in the acquisition of desiccation tolerance during embryo maturation (Gomez *et al.*, 1988; Mundy and Chua 1988; Close *et al.*, 1989; Pages *et al.*, 1995). Han *et al.* (1996) have observed differential pattern of *Lea* transcript accumulation and decline within the endosperm tissues of castor bean seeds during their development. While the transcripts for class I and II *Lea* proteins started to accumulate upto 35 DAP, distinct differences were observed in *Lea* gene expression during later stages of development with Class I *Lea* mRNAs declining substantially during maturation. These observations indicate a stage specific role of different classes of LEA proteins in seed development. Similar observations have been made by Hinniger *et al.* (2006) on *Coffea canephora*. Hinniger *et al*. (2006) also reported an unusual expression pattern of LEA gene (*Cc*LEA1) in *Coffea canephora* during grain development. The transcripts for *Cc*LEA1 were detected only during one relatively short period of grain development in the large green grain stage and not in either the small green or yellow stages of grain development. Since this period of grain development spans the period when the perisperm tissue undergoes a substantial size reduction and the endosperm expands significantly it was suggested that *Cc*LEA1 proteins might have a specific role during perisperm/endosperm transition.

Compared with the nucleotide sequences of dehydrin genes DH2, DH2a and DH2b, the nucleotide sequence of putative buckwheat dehydrin like gene, amplified in the present study, showed an insertion of six bases "CTTGTG" at position 217, seven bases "CGTCCCT" at position 258 and a single base "G" at position 682. These features distinguished the buckwheat nucleotide sequence from nucleotide sequences of *Coffea canephora* dehydrin like genes DH2, DH2a and DH2b. A significant feature of the sequence was the presence of a domain comprising of the

sequence "GTCGGAGGATGATGGACAAGGAGGAAGAAGAAAGAAAAAAG GGTTGAAAGAAAAGATAAA" between position 557 to 616. This domain has been reported as a conserved segment present in genes coding for dehydrin like proteins from several plants including *Solanum lycopersicum* (AC215480), *S. tuberosum* (X83597), *Cornus sericea* (AF345988), *Helianthus annuus* (AJ438980), *Eriobotrya japonica* (FJ472835), *Panax ginseng* dehydrin (DQ487110), *Phoenix dactylifera* (DQ399792), *Tithonia rotundifolia* (AJ250127).

The deduced amino acid sequence of the 1.0 kb amplicon, amplified in the present study, comprised of 112 amino acids with a predicted isoelectric point (p*I*) of 5.87 and calculated molecular weight of 12kDa. Using an alignment that permitted maximum homology, the deduced amino acid sequence showed a maximum of 63% homology with *Coffea canephora* dehydrins DH2, DH2a and DH2b (acc. nos. DQ338457, DQ323990 & DQ323989).The homology percentage ranged from 42 – 46% with amino acid sequences of dehydrin/dehydrin like proteins from other plants including *Vitis vinifera* (acc. no. XP002283605) and *Raphanus sativus* (acc. no. P21298). The alignment also revealed the presence of two segments, positioned between residue 1-38 and 90-109, with the highest level of conserved residues. While the N-terminal region between position 1 and 38 contained three "Y" segments represented by the sequence "D(E/Q)YGNP" located between position 10-35, the C-terminal region between position 90-109 contained the "K" segment represented by the sequence "EKKSMVEKIMEKLPGHH" located between position 93-107. The presence of "K" and "Y" segments in the buckwheat amino acid sequence indicates that the protein belongs to the dehydrin/ dehydrin-like protein family.

Dehydrin like proteins are known to possess one or more lysine-rich stretches of 15 amino acids, called the "K" motifs, that are predicted to form class A amphipathic α-helices (Dure, 1993; Close, 1996, 1997) and two other motifs, an N-terminal "Y" segment and a serine-rich "S" segment which can be phosphorylated and is thought to participate in nuclear localization (Godoy *et al.*, 1994; Close, 1997). The "K" segment comprises of the consensensus sequence "EKK(S/G)(M/I)(V/M)(E/D)KI(M/K)EKLPGHH" and the "Y" domain comprises of the sequence "D(E/Q)YGNP". While the conserved sequence of 15 residues viz. "EKKSMVEKIMEKLPG" in the "K"-segment is a distinctive feature of group 2 LEA proteins (Close *et al.*, 1989, 1993; Rorat, 2006), the "Y" segment comprised of the sequence "DE/QYGNP" has been reported as a common feature of all dehydrin like proteins (Close *et al.*, 1993; 1996; Campbell and Close, 1997). Close (1996) and Svenssen *et al.* (2002) have emphasised the importance of the "K" segment in the functioning of dehydrins in response to dehydration/ dehydration related stresses. Close (1996) and Koag *et al.* (2003) have proposed that the short amphipathic "K" segments of dehydrin polypeptides interact with solvent-exposed hydrophobic patches on proteins undergoing partial denaturation thereby protecting the proteins against aggregate formation. An alternative proposal for at least part of the protective effect of dehydrins is the ability of these very stable, but relatively unstructured proteins, to tightly bind and organize water molecules (Soulages *et al.*, 2003). This may help in slowing down the rate of water loss from cells thereby improving the stability of certain macromolecules by the development of dehydrin based regions of more tightly bound 'ordered' water around these molecules. A distinctive feature of the buckwheat amino acid sequence identified in the present study was the substitution of Serine tract with 'Asn-Phe-Arg' at position 66-68. While majority of

dehydrins identified to date fall into the class that contains the "Ser" tract, a number of dehydrins have been reported to lack the serine tract. Examples of dehydrins that lack a "Ser" tract include dehydrin/dehydrin-like proteins from spinach (Neven *et al.*, 1993), wheat (Houde *et al.*, 1992, Guo *et al.*, 1992), barley (Close *et al.*, 1995), pea (Robertson and Chandler, 1992), *Glycine max* (Momma *et al.*, 2003, Nylander *et al.*, 2001) *Vigna unguiculata* (Ismail *et al.*, 1999a). It is possible that the serine residues on the "Ser' tract may be phosphorylated and that the phosphorylation may contribute towards binding of nuclear localization signal peptides for translocation of the protein. Similar observations have been made by Goday *et al.* (1994) and Mehta *et al.* (2009).

On the basis of the presence and arrangement of different motifs in a single polypeptide, Campbell and Close (1997) have classified the group 2 LEA proteins into five subgroups. Amongst these, the proteins having both "Y" as well as "K" segments have been classified under the "YK"-subgroup (Campbell and Close, 1997). On the basis of presence of three "Y" segments, one "K" segment coupled with the absence of the a "Ser" tract, the putative buckwheat dehydrin identified in the present study can be classified under the "YK" subgroup with "Y_3K_1" architecture. PCA60, a *Prunus persica* dehydrin belonging to the "YK" subgroup has been reported to exhibit cryoprotective activity towards low temperature-sensitive enzymes in the plant species (Wisniewski *et al.,* 1999)

MOTIF SCAN of the deduced amino acid sequence identified a putative protein kinase C phosphorylation site between residue 2-4, a N-glycosylation site between residue 53-56, a kinase II phosphorylation site between residue 96-99 and two N-myristoylation sites between residues 40-45 and 82-87. Post- and co-translational modifications of proteins such as phosphorylation, glycosylation, and

myristoylation occur on many proteins involved in signal transduction. In most cases, these modifications are essential for protein function to mediate membrane association or protein–protein interaction (Ishitani *et al.*, 2000). Mehta *et al.* (2009) have reported similar putative phosphorylation sites for Casein kinase II in theAmDHNa ORF. Similarly, N-glycolysylation and myristoylation sites have been reported from dehydrins/dehydrin-like proteins from other plants (Levi, 1999; Caruso *et al.,* 2002; Robertson and Chandler, 1992).

MOTIF SCAN also detected the presence of two dehydrin domains comprising of the sequences "MQLTDQYGNPVQLKDEYGNPMQ" and "GTSTTVGGQQHEKKSMVEKIMEKLPGHH" between position 16-37 and 82-109, respectively. Analysis of the 112 amino acid sequences showed that the protein is highly hydrophilic sharing the characteristic of all dehydrins. Typically, dehydrins contain a high proportion of charged and polar amino acids with a low fraction of non-polar, hydrophobic residues and either few or no tryptophan and cysteine residues (Close *et al.*, 1989; Battaglia *et al.*, 2008). The deduced amino acid sequence of the 1.0 kb amplicon, amplified in the present study, shares these features. SAPS software predicted the amino acid sequence to transform into a secondary structure with 20.54% alpha helix, 54.46% random coil and 25% extended strand. The K-segment of dehydrins is deduced to form α-helix (Baker *et al.*, 1988; Ismail *et al.*, 1999b), involved in protein-protein and protein-lipid interactions (Dure, 1993; Epand *et al.*, 1995; Kovacs *et al.*, 2008). The absence of any transmembrane helices in the deduced amino acid sequence of the putative buckwheat dehydrin like protein indicates that the protein may not have any function in the membranes. Presumably it functions within the cytosolic or nuclear compartment. This assumption is supported by the detection of putative sites for post translational

modification in the amino acid sequence of the putative buckwheat dehydrin like protein identified in the present study. Similar observations have been made by Pulla *et al.* (2007) for dehydrin gene from *Codonopsis lanceolata.*

The phylogenetic tree generated from the alignment data of the sequence 112 deduced amino acids deduced in the present study with amino acid sequences of other dehydrin/dehydrin-like proteins available in the EMBL data base, revealed that even though the putative buckwheat (*Fagopyrum esculentum*) dehydrin/ dehydrin like protein, identified in the present study emerged as a separate group, it had greater closeness with dehydrins from *Coffea canephora.* Dehydrins/dehydrin-like proteins from plants belonging to different groups/families are known to show low homologies. The phylogenetic tree of dehydrins/dehydrin-like proteins developed in the present study is one such example.

Seed filling is a dynamic, temporally regulated phase of seed development that determines the composition of storage reserves in mature seeds. Programmes of gene expression in developing seeds comprise of distinct classes of genes that are coordinately regulated (Hughes and Galau, 1989; Parcy *et al.*, 1994). The MAT (maturation) class of genes includes major *SSP* genes (like 2S albumins and 12S globulins) which are expressed at early and mid-maturation phases, whereas the LEA (late embryogenesis abundant) class includes primarily genes involved in the acquisition of desiccation tolerance. These genes are reported to be expressed at later stages of maturation (Wobus and Weber, 1999; Hoekstra *et al.,* 2001). MAT-specific or Class II genes encode major seed storage proteins, the expression of which starts from 9 days after fertilization, peaks at 15, and declines about 20 days after fertilization (Parcy *et al.*, 1994). Genes coding for seed proteins, one of the major and important storage materials, represent abundantly transcribed genes that

are controlled in a precise developmental manner. Expression of such genes is specific to a development of seeds during the mid to late stages of embryo maturation.

PCR screening of the mid maturation stage cDNA library of common buckwheat with the forward primer bearing the sequence 5'GGATCCGGATTGGAGCAAGCGTTCTGC3' and the reverse primer having the sequence 5'GAAACGCTCCCTCTCCTTCTCATC3' amplified a DNA fragment showing an apparent molecular mass of 0.5 Kb. No such amplicon could be amplified from cDNA library constructed from germinating seeds of common buckwheat. The absence of any amplification with seed germination stage cDNA library as the template indicates that the amplicon generated from the mid maturation stage specific cDNA library was specific to mid maturation stage of seed development. BLASTn analysis of the nucleotide sequence of the amplicon revealed a high degree of homology with nucleotide sequences of genes coding for legumin type proteins. Using an alignment that permitted maximum homology, the sequence showed 96% and 92% homology with *Fagopyrum esculentum* legumin like protein genes (acc. no. D87980 and D87982) and 93% against *Fagopyrum tataricum* allergenic protein gene (acc. no. DQ849083) respectively. The alignment also revealed 58% homology against *Ficus pumila* 11S globulin precusor 2B (acc. no. EF091696) and 57% homology against *Citrus sinensis* seed storage protein citrin (acc. no. U38914) and *Gossypium hirsutum* beta-globulin B (acc. no. M16936). A feature that distinguished the nucleotide sequence of the DNA fragment amplifies in the present study and the nucleotide sequences of other legumin genes from *Fagopyrum* species was a 3 base deletion "GGG" at position 528. However, a 12 base deletion "GAGGAGAAGGAG" detected at position 552 in the nucleotide

sequence of the DNA fragment amplified in the present study and *Fagopyrum esculentum* nucleotide sequences (acc. nos. D87980 & D87982) was not detected in *Fagopyrum tataricum* allergenic protein genes (acc. no. DQ849083) and seed storage protein genes of various plants. This feature distinguished the nucleotide sequence identified in the present study from seed storage protein genes of *Fagopyrum tataricum* and other plants.

Seed storage proteins are thought to be accumulated exclusively in the cell-expansion phase of embryogenesis and metabolized during germination and seedling growth. The peak of storage protein gene expression is known to be much earlier during seed development, when cells rapidly expand and accumulate reserves (White *et al.*, 2000). The amplification of nucleotide sequence for the gene coding for legumin type proteins from mid maturation stage specific cDNA library and not from seed germination stage cDNA library indicates the mid maturation developmental stage specific expression of the gene in grains of common buckwheat. Similar pattern of legumin gene expression during seed development has been reported by Panitz *et al.* (1995), Abirached-Darmency *et al.* (2005). Abirached-Darmency *et al.* (2005) have detected the transcripts for legumin A and vicilin in the embryo cells of *Medicago truncatula* and *Pisum sativum* specifically at mid-embryogenesis. They suggested that legumin A and vicilin expression patterns could be considered as suitable embryo-specific markers during histo-differentiation at mid-embryogenesis in *M. truncatula* and *P. sativum.* On the basis of changes in the expression pattern of genes coding for SSPs in *Vicia faba*, Panitz ***et al.*** (1995) have concluded that storage proteins of *Vicia faba* accumulated transiently during early seed development and were used as nutritive reserves for the growing embryo. Sunderlikova and Wihelm (2002) have reported differential expression of putative legumin, *Em*- and

two dehydrin-like homologues (designated as *Dhn*1a and *Dhn*1b) in developing embryos of *Quercus robur*. While the expression level of these genes was substantially high during early and mid maturation stages of development, it declined markedly during the late mid maturation stage to mature stage when, dessication was complete. Sunderlikova and Wihelm (2002) have suggested a strong developmental control of expression of these genes. Even though seed storage protein gene expression has been shown to be rigorously tissue and developmental stage specific (Goldberg *et al.,* 1989; Perez-Grau and Goldberg, 1989; Guerche *et al.*, 1990) and several trans-acting factors that have been implicated in regulating the expression these genes, the role of specific sequences in seed protein gene expression is still a matter of discussion. Two major factors have contributed to this conundrum. First, regulatory ensembles of seed protein genes are extensive, often including more than a kilobase of upstream sequence, making it difficult to identify DNA sequences for detailed functional analysis. Second, seed protein gene regulatory ensembles are the result of combinatorial interactions of multiple DNA elements. Therefore, it is not surprising that analysis of individual elements comprising these ensembles has sometimes led to conflicting results on the role of specific elements in a regulatory ensemble (Baumlein *et al.*, 1991, 1992).

ORF finder tool identified 5 open reading frames in the nucleotide sequence of the 0.5 kb DNA amplified in the present study. The ORF, starting at position 1, consisted of 191 amino acids which showed 100% homology with the acidic chain of *Fagopyrum esculentum* 13S globulin (acc. no. Q9XFM4) and *Fagopyrum esculentum* allergenic protein (acc. no. AAF34635). The percentage homology ranged from 43% - 53% with amino acid sequences of legumin type proteins from other plants including *Coffea arabica* (acc. no. AAC61881) and

Magnolia salicifolia (CAA57846). The 7-element fingerprint signature for the 11S globulin family observed in the deduced amino acid sequence of 191 residues also shared 100% homology with *Fagopyrum esculentum* 13S globulin and allergenic protein (acc. no. Q9XFM4 and AAF34635). Further, motif 4 "VQVVGDEGRSVFDDNVQ" and motif 7 "TSVLRAIPVEVLANSYDI" present in the 191 amino acid sequence were observed to be highly conserved in both *Fagopyrum esculentum* as well as *F. tataricum* seed storage proteins. This clearly indicates that the deduced amino acid sequence of 191 residues identified in the present study belongs to legumin family. MOTIF SCAN identified a putative N-myristoylation site beginning from the 1st residue, 5 putative kinase II phosphorylation sites at positions 19-22', 75-78', 90-93', 157-160', and 179-182', and a protein kinase C phosphorylation site at position 157-160' in the deduced amino acid residue. Similar post-translational processing sites have been reported for peanut glycinin Gly-1 (Jain, 2004). Phosphorylation has been shown to play a major role in modulating the function and DNA-binding activity of many nuclear proteins, including transcription factors and proteins involved in chromatin organization (Dang *et al.*, 1994; Armstrong *et al.*, 1997; Hoffmann *et al.*, 1998). The identification of such sites in the deduced amino acid sequence is indicative of its role in signal transduction processes during seed maturation. Protein phosphorylation is a ubiquitous form of regulation that influences many aspects of dynamic cellular behaviour in plant biology. Identification of 70 non-redundant phosphoproteins and their classification into 10 functional categories suggested that protein phosphorylation might be involved in the regulation of storage reserve synthesis (Agrawal and Thelen, 2006). Further, identification of a cupin domain in the deduced amino acid sequence of 191 residues starting from 13 to 162 indicated that

the protein belongs to the cupin superfamily. The cupin superfamily is one of the largest, most functionally diverse families of proteins including the 7S (vicilin) and 11S (legumin) seed storage proteins (Bäumlein *et al.,* 1995). These proteins have a characteristic signature domain comprising of two histidine containing motifs separated by an inter- motif region of variable length. This domain consists of six ß strands within a conserved ß barrel structure. The cupin superfamily of proteins in plants has been found to share a small number of globally conserved residues with 7S (vicilin) and 11S (legumin) seed storage proteins (Baumlein *et al.* 1995) and are generally known as germin-like proteins (GLPs). The expression of genes coding for some of the proteins belonging to this family has been correlated with specific stages of development such as embryogenesis (Domon *et al.* 1995; Neutelings, 1998), floral induction (Heintzen *et al.*, 1994; Staiger *et al.*, 1999), stress (Hurkman and Tanaka, 1996; Thordahl-Christensen et al. 1997; Hamel *et al.*, 1998; Vallelian-Bindschedler *et al.*, 1998; Schweizer *et al.*, 1999).

When the amino acid residues of the sequence were plotted as a function of hydropathic index, the sequence showed a predominantly hydrophilic character. Based on the hydropathic index of Kyte and Doolittle (1982), the major regions of hydrophilic nature detected in the sequence were between residues 3-43, 60-100, 114-134 and 152-189. Slightom *et al.* (1985) have also reported similar results on the hydrophilicity of *Phaseolus vulgaris* phaseolin. They have also reported the presence of two N-glycosyl recognition sites in the protein. While majority of the reports have described legumin type proteins as hydrophilic, Jain (2004) have reported the presence of two hydrophobic regions located within the first 76 residues (residues 6 to 21 and 67 to 76), each of which is long enough to span a membrane bilayer in peanut glycinin Gly-1.

A phylogenetic tree constructed using neighbor joining distance method revealed that the deduced amino acid sequence of 191 residues generated in the present study formed one clade along with *Fagopyrum esculentum* seed storage proteins. Even though the seed storage proteins of *F. tataricum* and *F. gracilipes* emerged as separate clades, they showed low level of divergence from the *Fagopyrum esculentum* group . *Coffea arabica* 11S globulin appeared as the closest relative of SSPs from *Fagopyrum* species.

Amplification reaction of genomic DNA with primer pair 5' GGATCCGGATTGGAGCAAGCGTTCTGC 3' as the forward primer and 5' GAAACGCTCCCTCTCCTTCTCATC 3' as the reverse primer amplified a 0.7 Kb. DNA from the template genomic DNA. Pair wise alignment of the sequence with the nucleotide sequence of the amplicon generated with primer pair NKC5-NKC6 and cDNA library from grains harvested at 16-20 DAF as the template clearly revealed that the two sequences shared absolute homology. The alignment also revealed a gap of 96 bases corresponding to position between 142-238 bases in the 455bp nucleotide sequence of the amplicon generated with genomic DNA as the template and primer pair NKC5-NKC6. GENSCANW identified this region on the nucleotide sequence of the 0.7kb amplicon as an intron. Even though legumin genes are known to have a fairly fixed intron-exon architecture of three introns and four exons (Shewry, 1995), some of the legumin genes have been reported to have a two intron-four exon architecture (Shotwell and Larkins, 1989). While Vonder *et al.* (1988) have related legumin gene evolution to loss of introns, Shotwell and Larkins (1989) have suggested that the evolution of legumin gene has progressed towards addition of introns. Our result showing the presence of only one intron may be due to the partial sequence of the DNA amplified product.

Northern hybridization against RNA isolated from grains harvested at 10-13 DAF, 16-20 DAF, 23-25 DAF and 30-35 DAF with [α-P^{32}]-dATP labelled 0.7 Kb PCR product generated a strong signal against RNA isolated from grains harvested at 16-20 DAF. The blots did not show any signal against RNA isolated from grains harvested at 10-13 DAF, whereas weak signals could be detected against RNA isolated from grains harvested between 23-25 DAF and 30-35 DAF which was of lesser intensity than the signal against RNA isolated from grains harvested between 16-20 DAF. Although most of the genes expressed during the later phases of embryogenesis are also active during the post-germination phase, all storage protein mRNAs have a particular period during which it reaches a maximum level and then declines. During rice seed formation, Duan and Sun (2005) have observed that the genes encoding storage proteins glutelin, prolamin, globulin and albumin followed a pattern of low initial expression at early developmental stage, reached its peak at mid maturation stage and maintained a significant expression profile till maturation except for glutelin gene which started declining towards maturation. Our result on the transcript level changes corresponding to the putative dehydrin gene as well as the legumin clearly indicate mid maturation specific expression of the two genes.

Several studies have clearly established the role of dehydrin genes in protective reactions thereby promoting maintenance of embryo structure under conditions of water loss during seed maturation (Puhakainen *et al.,* 2004; Choi and Close, 2000; Nylander, *et al.,* 2001), signalling cascade during embryo development (Goday *et al.*, 1994; Mehta *et al.*, 2009), protein-protein and protein-lipid interactions (Dure, 1993; Epand *et al.*, 1995; Kovacs *et al.*, 2008) as also during transition of perisperm to endosperm Hinniger *et al.* (2006) thereby indicating a prominent role for dehydrins in maturation phase of embryo development. On the

other hand, a similar role for seed storage proteins has not been identified definitively. Even though extensive evidence has pointed towards a rigorous temporal and spatial control of seed storage protein gene expression during the maturation phase of seed development (Goldberg *et al.,* 1989; Perez-Grau and Goldberg, 1989; Guerche *et al.*, 1990; Panitz *et al.*, 1995; Sunderlikova and Wihelm, 2002; Abirached-Darmency *et al.*, 2005; Abirached-Darmency *et al.*, 2005), the role of seed storage proteins in modulating developmental processes in seeds is still a matter of discussion. Majority of the legumin type seed storage proteins are hydrophilic in nature. Therefore, it is possible that besides acting as a major source of amino acids during germination and initial stages of seedling growth, these proteins may also be involved in conferring desiccation tolerance to the seeds during the maturation phase.

CHAPTER : VI
GENERAL SUMMARYAND DISCUSSION

Introduction

Seed development is a crucial process in the lifecycle of higher plants, providing the link between two distinct sporophytic generations, and thereby, maintenance of the species. Seed development may be divided into four distinct stages: (i) embryo patterning, (ii) embryo growth, (iii) seed filling, and (iv) seed desiccation. Following the completion of embryo growth, major changes occurring during the process of seed development include accumulation of storage products (e.g., proteins, oil, and starch), followed by desiccation, during which seeds acquire drought tolerance and primary dormancy is often induced (Bewley and Black, 1994). The maturation of embryos ends with a desiccation phase after which the embryo enters into a quiescent state, thereby permitting its maintenance and survival under a range of environmental conditions. Embryogenesis describes the developmental period during which a single cell differentiates into a mature embryo.

During embryogenesis, the zygote elongates and then divides asymmetrically to form daughter cells of different sizes and cytoplasmic densities. The apical daughter cell, after two rounds of longitudinal and one round of transverse divisions, gives rise to an eight-cell embryo proper. At the eight-cell stage, four regions with different developmental fates can be recognized: (i) the apical embryo domain, (ii) the central embryo domain, (iii) the basal embryo domain or hypophysis and (iv) the extra embryonic suspensor. Clonal analyses have confirmed that the contribution of each cell to the seedling body plan is highly predictable (Reviewed by Laux *et al.,* 2004). However, rare variations in the cell division pattern do occur where each cell differentiates according to its final position, establishing that developing plant cells are flexible and assume their fate corresponding to positional information (Poethig *et al.,* 1986; Saulsberry *et al.,* 2002).

Seed embryogenesis is one of the major areas of interest and in recent times, researchers have started adopting functional genomics strategies to study its physiological process. Seed proteins comprise one of the major and important storage materials, which accumulate in high amounts during the maturation phase of seed development. These proteins remain stable throughout the developmental arrest and are then specifically degraded to serve as a source of carbon and nitrogen at the initial stages of germination and seedling growth. Seed storage protein genes represent abundantly transcribed genes that are controlled in a precise developmental manner during various stages of seed development. Okamuro and Goldberg, (1989) and Goldberg *et al.* (1989) have put forth four features of seed protein gene expression during seed development which raises the question of genetic and molecular relationships between seed development and changes in protein pool:

(i) Seed protein gene expression is regulated temporally during embryogenesis;

(ii) Seed protein genes are expressed exclusively during embryogenesis

(iii) Seed protein gene expression is regulated spatially

(iv) Seed protein mRNAs are localized within specific cells within the embryo cotyledon.

In recent years, several genes, which are directly related with cell differentiation, morphogenesis, and desiccation tolerance have been identified (Ikeda *et al.,* 2006). Many of these genes are specifically expressed during embryogenesis and function as part of the embryogenesis program. Thus, a large number of proteins found at early developmental stages in soybean seeds have been shown to belong to the 2S albumin fraction while those accumulating during the maturation phase have been shown to belong to the 7.5S and 11.8S fractions (Eldridge *et al.*, 1966; Catsimpoolas *et al.,* 1969; Catsimpoolas and Leuthner, 1969). In developing cotyledons of *Pisum sativum,* the globulins have been shown to be synthesized till 27 DAF while the accumulation of albumins stopped after 24 DAF (Basha, 1974) Consistent with differential accumulation of proteins during various stages of seed development, variations have also been observed in transcript levels for different proteins during various stages of seed development. Dong *et al.*, (2004) have demonstrated the presence of higher levels of transcripts coding for enzymes involved in biosynthetic pathways, structural proteins, regulatory factors, and transport or protective proteins at 15 DAP in embryos of *Brassica napus.* On the other hand, embryos at 25 DAP had higher levels of mRNAs for storage proteins such as napin, cruciferin and oleosin.

Out of the large number of diverse genes expressed during embryogenesis, most encode rare mRNAs of unknown function. A small set of genes, however,

directs the synthesis of abundant mRNAs that encode seed storage proteins that are packed preferentially into protein bodies in the embryo, dicot cotyledons and monocot endosperms (Mandal and Mandal, 2000). In addition to storage proteins, several other proteins like lectins, proteinase inhibitors and late embryogenesis abundant (LEA) proteins too are synthesized in the embryo during seed development. Apart from possibly serving as storage proteins, these proteins have been implied as defense proteins against several biotic and abiotic stresses (Gatehouse *et al.*, 1991). While the changes in transcriptome and proteome profiles with seed development have provided valuable clues linking seed development with gene networks (Gallardo *et al.*, 2003; 2007; Nakabayashi *et al.*, 2005; Hajduch *et al.*, 2005; Cadman *et al.*, 2006; Chibani *et al.*, 2006; Joosen *et al.*, 2007; Carrera *et al.*, 2008, Liu *et al.*, 2009; Houston *et al.*, 2009), many important unresolved questions associated with the regulatory network mechanisms involved in gene expression during seed development remain unresolved. Except for SSP genes, for which the structure of cis regulatory elements is well documented, little is known about the interactions between master regulator proteins and other putative target genes. The high diversity of seed storage compounds among plant species implies that distinct biosynthesis pathways and, therefore, distinct putative regulatory mechanisms are involved. Consequently, the use of non-model plant species might be important for identifying new regulators, as well as for highlighting the differences between conserved and non-conserved seed maturation processes. In order to better understand the seed maturation processes, many studies have been done in model plants such as *Arabidopsis thaliana* (Gallardo *et al.*, 2001; Charmont *et al.*, 2005; Cho *et al.*, 2007; Amme *et al.*, 2006) and *Lotus japonicus* (Dam *et al.*, 2009). However, the use of non model plant species might be important for

identifying new gene regulator and also highlight the differences between conserved and non-conserved seed maturation processes.

The present work focused on isolation and characterization of novel abundantly expressed seed storage protein genes in maturing grains of common buckwheat. The approaches followed were (i) PCR amplification of target gene(s) from RNA isolated from grains of common buckwheat at different stages of maturation by reverse-transcription (RT)-PCR and (ii) construction of cDNA library from RNA isolated from grains of common buckwheat at mid maturation stage of development and from grains at early stages of germination followed by amplification of target gene(s) by nested PCR using the cDNA library as the template and primers designed from nucleotide sequences of genes available in the gene bank databases.

Results

The Lopez-Gomez and Gomez-Lim (1992) protocol used in the present study for isolation of total RNA yielded fairly good quality RNA from all the tissue samples. Under UV illumination, the isolated RNA was detected on the agarose gel as two fluorescing bands corresponding to a typical RNA profile of 28S rRNA and 18S rRNA. The ratio of band thickness and intensity of fluorescence between of 28S rRNA and 18S rRNA was approximately 2:1. Reverse Transcriptase-PCR with cDNA prepared from RNA isolated from grains harvested 5-7 DAF, 10-13 DAF, 16-20 DAF, 23-25 DAF and at early gemination stage as the template and primer pair with the sequence 5' CCGGGCAGGTATCTCGGCTAG3' and 5'ATGTCTCCGGGCATCCTGG3' amplified a 1.0 kb DNA fragment from cDNA template prepared from RNA isolated from grains harvested between 16-20 DAF. The amplicon was not detected in amplification mixtures containing cDNA prepared

from RNA isolated from grains harvested at any other stage. BLASTn analysis of the nucleotide sequence of the amplicon identified it as a dehydrin gene. The analysis showed the presence of a domain represented by the sequence "GTCGGAGGATGATGGACAAGGAGGAAGAAGAAAGAAAAAAGGGTTGA AAGAAAAGATAAA" between position 557 to 616 in the nucleotide sequence of the 1.0 kb amplicon. The sequence was identified as a conserved domain present genes coding for dehydrin like proteins from several plants including *Solanum lycopersicum* (AC215480) *Cornus sericea* (AF345988), *Helianthus annuus* (AJ438980), *Eriobotrya japonica* (FJ472835), *Panax ginseng* dehydrin (DQ487110), *Phoenix dactylifera* (DQ399792), *S.tuberosum* (X83597), *Tithonia rotundifolia* (AJ250127). Using an alignment that permitted maximum homology, the nucleotide sequence showed 99 % and 97% homology with *Coffea canephora* dehydrin genes DH2a (acc. no. DQ323989) and DH2b (DQ323990) respectively.

Amplification of the DNA fragment from RNA isolated from mid maturation stage of seed development and not from any other stage indicates the presence of transcripts coding for the putative buckwheat dehydrin gene specifically in the grains at mid maturation stage. No such amplicon could be detected in the amplification mixtures with cDNA templates from other stages of development. This indicates the absence or low abundance of transcripts coding for the putative buckwheat dehydrin gene in the grains during other developmental stages. These results indicate the mid maturation developmental stage specific expression of the gene in grains of common buckwheat. Similar pattern of LEA gene expression has been reported by Gomez *et al.* (1988), Mundy and Chua (1988), Close *et al.* (1989), Han *et al.* (1996), Choi and Close (2000), Nylander *et al.* (2001), Hininger *et al.* (2006). Our results on the detection of transcripts coding for dehydrin class of proteins in grains of buckwheat

during mid maturation stage of development are in conformity with other reports indicating enhanced expression of genes coding for dehydrin like proteins during the maturation phase of grain development (Close, 1996; Hinniger *et al.*, 2006). This, pattern of accumulation of some of the *Lea* transcripts during seed development, with the highest level of transcripts reported at incipient desiccation, have led to the suggestion that LEA polypeptides play a decisive role in the acquisition of desiccation tolerance during embryo maturation (Gomez *et al.* 1988; Mundy and Chua 1988; Close *et al.* 1989; Pages *et al.* 1995). Compared with the nucleotide sequences of dehydrin genes DH2, DH2a and DH2b, the nucleotide sequence of the putative buckwheat dehydrin like gene, amplified in the present study, showed an insertion of six bases "CTTGTG" at position 217, seven bases "CGTCCCT" at position 258 and a single base "G" at position 682. These features distinguished the buckwheat nucleotide sequence from nucleotide sequences of *Coffea canephora* dehydrin like genes.

BLASTp identified the deduced 112 amino acid sequence of the 1.0 kb amplicon, amplified in the present study, as a dehydrin like protein showing 63% homology with *Coffea canephora* dehydrins DH2, DH2a and DH2b (acc. nos. DQ338457, DQ323990 & DQ323989). The homology percentage ranged from 42 – 46% with amino acid sequences of dehydrin/dehydrin like proteins from other plants including *Vitis vinifera* (acc. no. XP002283605) and *Raphanus sativus* (acc. no. P21298). The alignment also revealed the presence of two segments, positioned between residue 1-38 and 90-109, with the highest level of conserved residues. While the N-terminal region between position 1 and 38 contained three "Y" segments represented by the sequence "D(E/Q)YGNP" located between position 10-35, the C-terminal region between position 90-109 contained the "K" segment

represented by the sequence "EKKSMVEKIMEKLPGHH" located between position 93-107. Dehydrin like proteins are known to possess one or more lysine-rich stretches of 15 amino acids, called the "K" motifs, that are predicted to form class A amphipathic α-helices (Dure, 1993; Close, 1996, 1997) and two other motifs, an N-terminal "Y" segment and a serine-rich "S" segment which can be phosphorylated and is thought to participate in nuclear localization (Godoy *et al.*, 1994; Close, 1997). The presence of "K" and "Y" segments in the buckwheat amino acid sequence indicates that the protein belongs to the dehydrin/dehydrin-like protein family. Close (1996) and Svenssen *et al.* (2002) have emphasized the importance of the "K" segment in the functioning of dehydrins in response to dehydration related stresses. Close (1996) and Koag *et al.* (2003) have proposed that the short amphipathic "K" segments of dehydrin polypeptides interact with solvent-exposed hydrophobic patches on proteins undergoing partial denaturation thereby protecting the proteins against aggregate formation. An alternative proposal for at least part of the protective effect of dehydrins is the ability of these very stable, but relatively unstructured proteins, to tightly bind and organize water molecules (Soulages *et al.*, 2003). A distinctive feature of the buckwheat amino acid sequence identified in the present study was the substitution of Serine tract with 'Asn-Phe-Arg' at position 66-68. While majority of dehydrins identified to date fall into the class that contains the "Ser" tract, a number of dehydrins have been reported to lack the serine tract. Examples of dehydrins that lack a "Ser" tract include dehydrin/dehydrin-like proteins from spinach (Neven *et al.*, 1993), wheat (Houde *et al.*, 1992, Guo *et al.*, 1992), barley (Close *et al.*, 1995), pea (Robertson and Chandler, 1992), *Glycine max* (Momma *et al.,* 2003; Nylander *et al.,* 2001) and *Vigna unguiculata* (Ismail *et al.*, 1999a). Goday *et al.* (1994) and Mehta *et al.* (2009) have postulated that the serine

residues on the "Ser' tract could be the site of phosphorylation and that the phosphorylation may contribute towards binding of nuclear localization signal peptides for translocation of the protein. Campbell and Close (1997) have classified the group 2 LEA proteins into five subgroups depending on the presence and arrangement of different motifs in a single polypeptide. Amongst these, the proteins having both "Y" as well as "K" segments have been classified under the "YK" subgroup (Campbell and Close, 1997). On the basis of presence of three "Y" segments, one "K" segment coupled with the absence of a "Ser" tract, the putative buckwheat dehydrin identified in the present study can be classified under the "YK"-subgroup with "Y_3K_1" architecture.

MOTIF SCAN on the deduced amino acid sequence identified a putative protein kinase C phosphorylation site between residue 2-4, a N-glycosylation site between residue 53-56, a kinase II phosphorylation site between residue 96-99 and two N-myristoylation sites between residues 40-45 and 82-87. Similar sites for post translational modifications have also been reported in other dehydrin like proteins (Caruso *et al.*, 2002; Roberton and Chandler, 1992; Mehta *et al.*, 2009). The software also detected the presence of two dehydrin domains comprising of the sequences "MQLTDQYGNPVQLKDEYGNPMQ" and "GTSTTVGGQQHEKKSMVEKIME KLPGHH" between position 16-37 and 82-109, respectively. Hydropathy analysis of the deduced amino acid sequence identified two strong hydrophilic domains in the sequence, between residues 1-40 and 80-112. SAPS software analysis also identified 23.2% non polar group, 25.8% uncharged polar group and 19.6% charged polar group in the amino acid sequence. Typically, dehydrins contain a high proportion of polar amino acids with a low fraction of non-polar, hydrophobic residues and either few or no tryptophan and cysteine residues (Close *et al.*, 1989; Battaglia *et al.*, 2008).

The deduced amino acid sequence of the 1.0 kb amplicon, amplified in the present study, shares these features. The software predicted the amino acid sequence to transform into a secondary structure with 20.54% alpha helix, 54.46% random coil and 25% extended strand. Statistical analysis of the sequence revealed that the N-terminal region consisting of 70 residues had predominantly random coil and extended strand secondary structure while the C-terminal region of 42 residues were primarily of α-helical configuration. The "K" segment of dehydrin is deduced to form α- helix (Baker *et al.*, 1988; Ismail *et al.*, 1999b), involved in protein-protein and protein-lipid interactions (Dure, 1993; Epand *et al.*, 1995; Kovacs *et al.*, 2008).

The phylogenetic tree constructed using maximum parsimony method, with the alignment data of deduced amino acid sequence generated in the present study with amino acid sequences of other dehydrin/dehydrin-like proteins available in the EMBL data base, revealed that even though the putative buckwheat (*Fagopyrum esculentum*) dehydrin/ dehydrin like protein, identified in the present study emerged as a separate group, it had greater closeness with dehydrins from *Coffea canephora.* Dehydrins/dehydrin-like proteins from plants belonging to different groups/families are known to show low homologies. The phylogenetic tree of dehydrins/dehydrin-like proteins developed in the present study is one such example.

PCR screening of the mid maturation stage and seed germination stage cDNA libraries of common buckwheat with the forward primer bearing the sequence 5'GGATCCGGATTGGAGCAAGCGTTCTGC3' and the reverse primer having the sequence 5'GAAACGCTCCCTCTCCTTCTCATC3' amplified a 0.5 Kb DNA fragment from mid maturation stage cDNA library. No such amplicon could be amplified from cDNA library constructed from germinating seeds of common buckwheat. Absence of the amplicon in the reaction mixture containing cDNA

library prepared from RNA isolated from germinating seeds indicated that the amplicon generated from the mid maturation stage specific cDNA library was stage specific. BLASTn analysis of the nucleotide sequence of the amplicon revealed a high degree of homology with nucleotide sequences of genes coding for legumin type proteins. Using an alignment that permitted maximum homology, the sequence showed 96% and 92% homology with *Fagopyrum esculentum* legumin like protein genes (acc. no. D87980 and D87982) and 93% against *Fagopyrum tataricum* allergenic protein gene (acc. no. DQ849083) respectively. The alignment also revealed 58% homology against *Ficus pumila* 11S globulin precusor 2B (acc. no. EF091696) and 57% homology against *Citrus sinensis* seed storage protein citrin (acc. no. U38914) and *Gossypium hirsutum* *β*-globulin (acc. no. M16936). A feature that distinguished, the nucleotide sequence of the 0.5 kb amplicon from other nucleotide sequences of legumin/legumin like genes from *Fagopyrum* spp. species was a 3 base deletion "GGG" at position 528 in the nucleotide sequence of the 0.5 kb amplicon. However, a 12 base deletion "GAGGAGAAGGAG" detected at position 552 in the nucleotide sequence of the DNA fragment amplified in the present study and *Fagopyrum esculentum* nucleotide sequences (acc. nos. D87980 & D87982) was not detected in *Fagopyrum tataricum* allergenic protein genes (acc. no. DQ849083) and seed storage protein genes of various plants. This feature distinguished the nucleotide sequence identified in the present study from seed storage protein genes of *Fagopyrum tataricum* and other plants.

The amplification of nucleotide sequence for the gene coding for legumin type proteins from mid maturation stage specific cDNA library and not from seed germination stage cDNA library indicates the mid maturation developmental stage specific expression of the gene in grains of common buckwheat. Similar pattern of

legumin gene expression during seed development has been reported by Panitz *et al* (1995), Darmency *et al.* (2005). Darmency *et al.* (2005) have detected the transcripts for legumin A and vicilin in the embryo cells of *M. truncatula* and *P. sativum* specifically at mid-embryogenesis. It was suggested that legumin A and vicilin expression patterns could be considered as suitable embryo-specific markers during histo-differentiation at mid-embryogenesis in *M. truncatula* and *P. sativum.*. On the basis of changes in the expression pattern of genes coding for SSPs in *Vicia faba,* Panitz *et al.* (1995) have concluded that storage proteins of *Vicia faba* accumulated transiently during early seed development and were used as nutritive reserves for the growing embryo. Even though seed storage protein gene expression has been shown to be rigorously tissue and developmental stage specific (Goldberg *et al.*, 1989; Perez-Grau and Goldberg, 1989; Guerche *et al.*, 1990; Sunderlikova and Wihelm, 2002) and several trans-acting factors that have been implicated in regulating the expression these genes, the role of specific sequences in seed protein gene expression is still a matter of discussion. It is therefore not surprising that analysis of individual elements comprising these ensembles has sometimes led to conflicting results on the role of specific elements in a regulatory ensemble (Baumlein *et al.*, 1991, 1992).

ORF finder tool identified 5 open reading frames in the nucleotide sequence of the 0.5 kb DNA amplified in the present study. The ORF, starting at position 1, consisted of 191 amino acids which showed 100% homology with the acidic chain of *Fagopyrum esculentum* 13S globulin (Q9XFM4) and *Fagopyrum esculentum* allergenic protein (acc. no. AAF34635). The percentage homology ranged from 43% - 53% with amino acid sequences of legumin type proteins from other plants including *Coffea arabica* (acc. no. AAC61881) and *Magnolia salicifolia* (CAA57846). The 7-element fingerprint signature for the 11S globulin family

observed in the in the deduced amino acid sequence of 191 residues also shared 100% homology with *Fagopyrum esculentum* 13S globulin and allergenic protein (acc. no. Q9XFM4 and AAF34635). Further, motif 4 "VQVVGDEGRSVFDDNVQ" and motif 7 "TSVLRAIPVEVLANSYDI" were observed to be highly conserved in both *Fagopyrum esculentum* as well as *F. tataricum* seed storage proteins. This clearly indicates that the deduced sequence of 191 amino acids, identified in the present study, belonged to legumin family.

MOTIF SCAN identified 5 putative kinase II phosphorylation sites, a putative N-myristoylation site and a protein kinase C phosphorylation site in the deduced amino acid sequence identified in the present study. Phosphorylation has been shown to play a major role in modulating the function and DNA-binding activity of many nuclear proteins, including transcription factors and proteins involved in chromatin organization (Dang *et al.*, 1994; Armstrong *et al.*, 1997; Hoffmann *et al.*, 1998). The identification of such sites in the deduced amino acid sequence is indicative of its role in signal transduction processes during seed maturation. Further, identification of a cupin domain in the deduced amino acid sequence indicated that the protein belongs to the cupin superfamily. The cupin superfamily is one of the largest, most functionally diverse families of proteins including 7S (vicilin) and 11S (legumin) seed storage proteins (Bäumlein *et al.*, 1995). These proteins have a characteristic signature domain comprising of two histidine containing motifs separated by an inter motif region of variable length. This domain consists of six ß strands within a conserved ß barrel structure.

When the amino acid residues of the sequence were plotted as a function of hydropathic index, the sequence showed a predominantly hydrophilic character. Slightom *et al.* (1985) have also reported similar results on the hydrophilicity of

Phaseolus vulgaris phaseolin. They have also reported the presence of two N-glycosyl recognition sites in the protein. While majority of the reports have described legumin type proteins as hydrophilic, Jain (2004) have reported the presence of two hydrophobic regions located within the first 76 residues (residues 6 to 21 and 67 to 76), each of which is long enough to span a membrane bilayer in peanut glycinin Gly-1.

A phylogenetic tree constructed using neighbor joining distance method revealed that the deduced amino acid sequence of 191 residues generated in the present study formed one clade with together *Fagopyrum esculentum* seed storage proteins. Even though the seed storage proteins of *F. tataricum* and *F. gracilipes* emerged as separate clades, they showed low level of divergence from the *Fagopyrum esculentum* group. *Coffea arabica* 11S globulin appeared as the closest relative of SSPs from *Fagopyrum* species .

Amplification of genomic DNA with primer pair 5' GGATCCGGATTGGAGCAAGCGTTCTGC 3' as the forward primer and 5' GAAACGCTCCCTCTCCTTCTCATC 3' as the reverse primer amplified a 0.7 Kb. DNA from the template genomic DNA. Pair wise alignment of the sequence with the nucleotide sequence of the amplicon generated with primer pair NKC5-NKC6 and cDNA library from grains harvested at 16-20 DAF as the template clearly revealed that the two sequences shared absolute homology. The alignment also revealed a gap of 96 bases corresponding to position between 142-238 bases in the 455bp nucleotide sequence of the amplicon generated with genomic DNA as the template and primer pair NKC5-NKC6. GENSCANW identified this region on the nucleotide sequence of the 0.7kb amplicon as an intron. Even though legumin genes are known to have a fairly fixed intron-exon architecture of three introns and four exons (Shewry, 1995),

some of the legumin genes have been reported to have a two intron-four exon architecture (Shotwell and Larkins, 1989). While Vonder *et al.* (1988) have related legumin gene evolution to loss of introns, Shotwell and Larkins (1989) have suggested that the evolution of legumin gene has progressed towards addition of introns. Our result showing the presence of only one intron may be due to the partial sequence of the DNA amplified product. Sequence homology of the 107 deduced amino acid residues with BLASTp against non-redundant protein database clearly identified the protein as belonging to the legumin gene family. Northern hybridization against RNA isolated from grains harvested at 10-13 DAF, 16-20 DAF, 23-25 DAF and 30-35 DAF with [α-P^{32}]-dATP labelled 0.7 Kb PCR product generated a strong signal against RNA isolated from grains harvested at 16-20 DAF. Signals were also detected against RNA isolated from grains harvested between 23-25 DAF and 30-35 DAF but were of lesser intensity than the signal against RNA isolated from grains harvested between 16-20 DAF. Although most of the genes expressed during the later phases of embryogenesis are also active during the post-germination phase, all storage protein mRNAs have a particular period during which it reaches a maximum level and then declines. During rice seed formation, Duan and Sun (2005) have observed that the genes encoding storage proteins glutelin, prolamin, globulin and albumin followed a pattern of low initial expression at early developmental stage, reached its peak at mid maturation stage and maintained a significant expression profile till maturation except for glutelin gene which started declining towards maturation. Our result on the transcript level changes corresponding to the putative dehydrin gene as well as the legumin clearly indicate mid maturation specific expression of the two genes.

Several studies have clearly established the role of dehydrin genes in protective reactions thereby promoting maintenance of embryo structure under conditions of water loss during seed maturation (Puhakainen *et al.,* 2004; Choi and Close, 2000; Nylander, *et al.,* 2001), signalling cascade during embryo development (Goday *et al.*, 1994; Mehta *et al.*, 2009), protein-protein and protein-lipid interactions (Dure, 1993; Epand *et al.*, 1995; Kovacs *et al.*, 2008) as also during transition of perisperm to endosperm Hinniger *et al.* (2006) thereby indicating a prominent role for dehydrins in maturation phase of embryo development. On the other hand a similar role for seed storage proteins has not been identified definitively Even though extensive evidence has pointed towards a rigorous temporal and spatial control of seed storage protein gene expression during the maturation phase of seed development (Goldberg *et al.,* 1989; Perez-Grau and Goldberg, 1989; Guerche *et al.*, 1990; Panitz *et al.* , 1995; Sunderlikova and Wihelm, 2002; Abirached-Darmency *et al.*, 2005; Abirached-Darmency *et al.*, 2005), the role of seed storage proteins in modulating developmental processes in seeds is still a matter of discussion. Majority of the legumin type seed storage proteins are hydrophilic in nature. Therefore, it is possible that besides acting as a major source of amino acids during germination and initial stages of seedling growth, these proteins may also be involved in conferring desiccation tolerance to the seeds during the maturation phase.

CHAPTER : VII
REFERENCES

Abirached-Darmency, M., Abdel-Gawwad, M. R., Conejero, G., Verdeil, J. L. and Thompson, R. (2005) *In situ* expression of two storage protein genes in relation to histo-differentiation at mid-embryogenesis in *Medicago truncatula* and *Pisum sativum* seeds J. Experimental Botany, 56(418): 2019-2028.

Agrawal, G. K. and Thelen, J. J. (2006) Large Scale Identification and Quantitative Profiling of Phosphoproteins Expressed during Seed Filling in Oilseed Rape. Molecular & Cellular Proteomics, 5: 2044–2059.

Agrawal, G. K., Hajduch, M., Graham, K., Thelen, J. J. (2008) In-depth investigation of the soybean seed-filling proteome and comparison with a parallel study of rapeseed. Plant Physiol., 148: 504–518.

Amme, S., Matros, A., Schlesier, B. and Mock, H. P. (2006) Proteome analysis of cold stress response in Arabidopsis thaliana using DIGE-technology. J. Exp. Bot., 1-10.

Armstrong, S. A., Barry, D. A., Leggett, R. W. and Mueller, C. R. (1997) Casein kinase II-mediated phosphorylation of the C terminus of Sp1 decreases its DNA binding activity. J. Biol. Chem., 272: 13489–13495.

Artus, N. N., Uemura, M., Steponkus, P. L., Gilmore, S. J., Lin, C. and Thomashow, M. F. (1996) Constitutive expression of the cold-regulated *Arabidopsis thaliana COR15a* gene affects both chloroplast and protoplast freezing tolerance. Proc. Natl. Acad. Sci., 93: 13404–13409.

Asghar, R., Fenton, R. D., DeMason, D. A. and Close, T. J. (1994) Nuclear and cytoplasmic localization of maize embryo and aleurone dehydrin. Protoplasma, 177: 87-94.

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

Balbuenaa, T. S., Silveirab, V., Junqueirac, M., Diasa, L. L.C., Santa-Catarinad, C., Shevchenkoc, A., Floha, E. I. S. (2009) Changes in the 2-DE protein profile during zygotic embryogenesis in the Brazilian Pine (*Araucaria angustifolia*). J. proteomics, 72337 – 352.

Barton, M. K. and Poethig, R. S. (1993) Formation of the shoot apical meristem in *Arabidopsis thaliana*: an analysis of development in the wild type and in the shoot meristemless mutant. Development, 119: 823-831.

Basha, S. M. M. (1974) Protein metabolism in the cotyledons of *Pisum sativum* L. during seed development and germination. Ph.D. thesis, Grad. Coll. Univ. Oklahoma, Norman, pp.79.

Battaglia, M., Olvera-Carrillo, Y., Garciarrubio, A., Campos, F. and Covarrubias A. A. (2008) The Enigmatic LEA Proteins and Other Hydrophilins. Plant Physiol., 148: 6–24.

Baud, S., Wuillème, S., Lemoine, R., Kronenberger, J., Caboche, M., Lepiniec, L. and Rochat, C. (2005) The AtSUC5 sucrose transporter specifically expressed in the endosperm is involved in early seed development in Arabidopsis. Plant J., 43: 824–836.

Bäumlein, H., Boerjan, W., Nagy, I., Panitz, R. and Wobus, U. (1991) Upstream sequences regulating legumin gene expression in heterologous transgenic plants. MOI. Gen. Genet., 225: 121-128.

Baumlein, H., Braun, H., Kakhovskaya, I. A. and Shutov, A. D. (1995) Seed storage proteins of sper,matophytes share a common ancestor with dessication proteins of fungi. J. Mol. Evol., 41: 1070-1075.

Bäumlein, H., Nagy, I., Villaroel, R., Inzb, 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 tissuespecific expression of a legumin gene. Plant J., 2: 233-239.

Berleth, T. and Chatfield, S. (2002) Embryogenesis: Pattern Formation from a Single Cell. The Arabidopsis Book, 1-27.

Bewley, J. D. and Black, M. (1994) Seeds. Physiology of development and germination. Plenum Press, New York, USA.

Bewley, J. D. and Marcus, A. (1990) Gene expression in seed development and germination. In: Progress in Nucleic Acid Research and Molecular Biology (W. E. Cohn and K. Moldave, eds.) Academic Press, Orlando, FL, Vol. 38. pp 165-193.

Bies, N., Aspart, L., Carles, C., Gallois, P. and Delseny, M. (1998) Accumulation and degradation of Em proteins in *Arabidopsis thaliana*: evidence for post-transcriptional controls. J. Exp. Bot., 49: 1925 – 1933.

Bies-Ethéve, N., Gaubier-Comella, P., Debures, A., Lasserre, E., Jobet, E., Raynal, M., Cooke, R. and Delseny, M. (2008) Inventory, evolution and expression profiling diversity of the LEA (late embryogenesis abundant) protein gene family in *Arabidopsis thaliana*. Plant Mol. Biol., 67: 107–124.

Borisjuk, L. Rolletschek, H., Weber, H., Wobus, U., Weber, H. (2003) Differentiation of legume cotyledons as related to metabolic gradients and assimilate transport into seeds. J. Exp. Bot., 54: 503–512.

Braybrook, S. A., Stone, S. L., Park, S., Bui, A. Q., Le, B. H., Fischer, R. L., Goldberg, R. B. and Harada, J. J. (2006) Genes directly regulated by LEAFY COTYLEDON2 provide insight into the control of embryo maturation and somatic embryogenesis. Proc. Natl. Acad. Sci. U.S. A., 103: 3468–3473.

Brendel, V., Bucher, P., Nourbakhsh, I., Blaisdell, B. E. and Karlin, S. (1992) Mathods and algorithms for statistical analysis of protein sequences. Proc. Natl. Acad. Sci. USA, 89: 2002-2006.

Buitink, J., Leger, J. J., Guisle, I., Vu, B. L., Wuillème, S., Lamirault, G., Le Bars, A., Le Meur, N., Becker, A., Küster, H. and Leprince, O. (2006) Transcriptome profiling uncovers metabolic and regulatory processes occurring during the transition from desiccationsensitive to desiccation-tolerant stages in Medicago truncatula seeds. Plant J., 47: 735–750.

Cadman, C. S. Toorop, P. E., Hilhorst, H. W., Finch-Savage, W. E. (2006) Gene expression profiles of Arabidopsis Cvi seeds during dormancy cycling indicate a common underlying dormancy control mechanism. Plant J., 46: 805–822.

Campbell, S. A. and Close, T. J. (1997) Dehydrins: genes, proteins, and associations with phenotypic traits. New Phytol, 137: 61–74.

Carrera, E., Holman, T., Medhurst, A., Dietrich, D., Footitt, S., Theodoulou, F. L., Holdsworth, M. J. (2008) Seed after-ripening is a discrete developmental pathway associated with specific gene networks in Arabidopsis. Plant Journal, 53: 214–224.

Caruso, A., Morabito, D., Delmotte, F., Kahlem G. and Carpin, S. (2002) Dehydrin induction during drought and osmotic stress in *Populus*. Plant Physiol Biochem., 40: 1033–1042.

Casey, R., Domoney, C. and Ellis, N. (1986) Legume storage proteins and their genes. Plant Mol. Cell. Biol., 3: 1-11.

Casson, S. A. and Lindsey, K. (2006) The turnip mutant of Arabidopsis reveals that LEAFY COTYLEDON1 expression mediates the effects of auxin and sugars to promote embryonic cell identity. Plant Physiol., 142: 526–541.

Catsimpoolas, N. and Leuthner, E. (1969) The major pH 4.5 soluble proteins conformational transitions in proteins and their inhibition by stabilizers. Biophys J., 65: 661-671.

Catsimpoolas, N., Ekenstam, C. and Meyer, E. W. (1969) Separation of soybean whey proteins by isoelectric focusing. Cereal Chem., 46: 357-369.

Charmont, S., Jamet, E., Pont-Lezica, R. and Canut, H. (2005) Proteomic analysis of secreted proteins from *Arabidopsis thaliana* seedlings: improved recovery following removal of phenolic compounds. Phytochemistry, 66: 453-461.

Chia, T. Y., Pike, M. J. and Rawsthorne, S. (2005) Storage oil breakdown during embryo development of *Brassica napus* (L.). J. Exp. Bot., 56: 1285–1296.

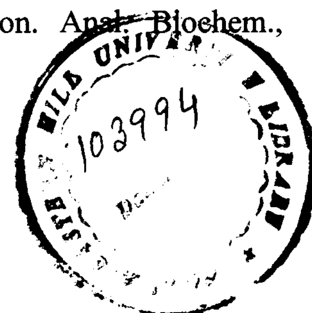
Chibani, K., Ali-Rachedi, S., Job, C., Job, D., Jullien, M. and Grappin, P. (2006) Proteome analysis of seed dormancy in Arabidopsis. Plant Physiol., 142: 1493-1510.

Cho, M. R., Lee, K. H., Hyun, Y. B., Lee, I. and Kim, H. J. (2007) Proteome Analysis of Vernalization-Treated *Arabidopsis thaliana* by Matrix-Assisted Laser Desorption/Ionization Time-of-Flight Mass Spectrometry. Bull. Korean Chem. Soc., 28: (3) 427-431.

Choi, D. W. and Close, T. J. (2000) A newly identified barley gene, Dhn12 encoding a YSK2 DHN, is located on chromosome 6H and has embryo-specific expression. Theoretical and Applied Genetics, 100: 1274–1278.

Choi, D-W., Zhu, B. and Close, T. J. (1999) The barley (*Hordeum vulgare* L.) dehydrin multigene family: sequences, allelic types, chromosome assignments, and expression characteristics of 11 *Dhn* genes of cv. Dicktoo. Theor. Appl. Genet., 98: 1234-1247.

Chomczynski, P. and Sacchi, N. (1987) Single-step method of RNA isolation by acid guanidinium thiocyanate-phenol-chloroform extraction. Anal. Biochem., 162:156-159.



Chono, M., Honda, I., Shinoda, S., Kushiro, T., Kamiya, Y., Nambara, E., Kawakami, N., Kaneko, S. and Watanabe, Y. (2006) Field studies on the regulation of abscisic acid content and germinability during grain development of barley: molecular and chemical analysis of pre-harvest sprouting. J. Exp. Bot., 57: 2421–2434.

Christianson, M. L. (1986) Fate map of the organizing shoot apex in Gossypium. Am. J. Bot., 73: 947-958.

Close, T. J. (1996) Dehydrins: emergence of a biochemical role of a family of plant dehydration proteins. Physiol. Plant, 97: 795–803.

Close, T. J. (1997) Dehydrins: a commonality in the response of plants to dehydration and low temperature. Physiologia Plantarum, 100: 291–296.

Close, T. J., Fenton, R. D., Yang, A., Asghar, R., DeMason, D. A., Crione, D., Meyer, N. C. and Moonan, F. (1993) Dehydrin: the protein. In: Plant Responses to Cellular Dehydration during Environmental Stress (T. J. Close and E. A. Bray, eds.), American Society of Plant Physiologists, Rockville, MD, pp. 104–118.

Close, T. J., Kortt, A. A. and Chandler, P. M. (1989) A cDNA-based comparison of dehydration-induced proteins (dehydrins) in barley and corn. Plant Mol. Biol., 13: 95–108.

Close, T. J., Meyer, N. C. and Radik, J. (1995) Nucleotide Sequence of a Gene Encoding a 58.5-Kilodalton Barley Dehydrin That Lacks a Serine Tract. Plant Physiol., 107: 289-290.

Creelman, R. A. and Mullet, J. E. (1997) Oligosaccharins, brassinolides and jasmonates: Non-traditional regulators of plant growth, development, and gene expression. Plant Cell, 9: 1211-1223.

Crouch, M. L. (1987) Regulation of gene expression during seed development in flowering plants. In: Developmental Biology: A Comprehensive Synthesis, Vol. 5, (L.W. Browder, ed.) Plenum Press, New York, pp. 367-404.

Curaba, J. Moritz, T., Blervaque, R., Parcy, F., Raz, V., Herzog, M. and Vachon, G. (2004) AtGA3ox2, a key gene responsible for bioactive gibberellin biosynthesis, is regulated during embryogenesis by LEAFY COTYLEDON2 and FUSCA3 in Arabidopsis. Plant Physiol., 136: 3660–3669.

Dam, S., Laursen, B. S., Ørnfelt, J. H., Jochimsen, B., Stærfeldt, H. H., Friis, C., Nielsen, K., Goffard, N., Besenbacher, S., Krusell, L., Sato, S., Tabata, S., Thøgersen, I. B., Enghild, J. J. and Stougaard, J. (2009) The proteome of

seed development in the model legume *Lotus japonicus*. Plant Physiol., 149: 1325–1340.

Dang, Q., Alghisi, G. C. and Gasser, S. (1994) Phosphorylation of the C-terminal domain of yeast topoisomerase II by casein kinase II affects DNA–protein interaction. J. Mol. Biol., 243:10–24.

Domon, J. M., Dumas, B., Lainé, E., Meyer, Y., David, A. and David, H. (1995) Three glycosylated polypeptides secreted by several embryogenic cell lines of pine show highly specific serological affinity to antibodies directed against the wheat germin apoprotein monomer. Plant Physiology, 108(1): 141-148.

Dong, J., Keller, W. A., Yan, W. and Georges, F. (2004) Gene expression at early stages of *Brassica napus* seed development as revealed by transcript profiling of seed-abundant cDNAs. Planta, 218: 483–491.

Duan, M. and Sun, S. S. M. (2005) Profiling the expression of genes controlling rice grain quality. Plant Molecular Biology 59: 165-178.

Dure, L. III (1993) Structural motifs in Lea protein. In: Plant response to cellular dehydration during environmental stress, vol 10. (T. J. Close and E. A. Bray, eds.) American Society of Plant Physiologists, Rockville, pp. 91–103.

Dure, L. S. III (1985) Embryogenesis and gene expression during seed formation. Oxford Surv. Plant Mol. Cell Biol., 2: 179-188.

Egerton-Warburton, L. M., Balsamo, R. A. and Close, T. J. (1997) Temporal accumulation and ultrastructural localization of dehydrins in *Zea mays* L. Physiol. Plant., 101: 545-555.

Eldridge, A. C., Anderson, R. L. and Wolf, W. J. (1966) Polyacrylamide gel electrophoresis of soybean whey proteins and trypsin inhibitors. Arch. Biochem. Biophys., 115: 495-504.

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.

Epand, R. M., Shai, Y., Segrest, J. P. and Anantharamaiah, G. M. (1995) Mechanisms for the modulation of membrane bilayer properties by amphipathic helical peptides. Biopolymers, 37: 319-338.

Epel, D. (1990) The initiation of development at fertilization. Cell Differ. Dev., 29: 1-12.

Fait, A. Angelovici, R., Less, H., Ohad, I., Urbanczyk-Wochniak, E., Fernie, A. R. and Galili, G. (2006) Arabidopsis seed development and germination is

associated with temporally distinct metabolic switches. Plant Physiol., 142: 839–854.

Faure, J. D., Vittorioso, P., Santoni, V., Fraisier, V., Prinsen, E., Barlier, I., Van Onckelen, H., Caboche, M., and Bellini, C. (1998) The *PASTICCINO* genes of *Arabidopsis thaliana* are involved in the control of cell division and differentiation. Development, 125: 909-918.

Fernandez, D. E., Turner, F. R. and Crouch, M. L. (1991) *In situ* localization of storage protein mRNAs in developing meristems of *Brassica napus* embryos. Development, 111: 299-313.

Finch-Savage, W. E. and Leubner-Metzger, G. (2006) Seed dormancy and the control of germination. New Phytol., 171: 501–523.

Finkelstein, R. R., Gampala, S. S. L. and Rock, C. D. (2002) Abscisic acid signaling in seeds and seedlings. Plant Cell, 14: S15–S45.

Fischer, C. and Neuhaus, G. (1996) Influence of auxin on the establishment of bilateral symmetry in monocots. Plant J., 9: 659-669.

Frey, A. Boutin, J-P., Sotta, B., Mercier, R. and Marion-Poll, A. (2006) Regulation of carotenoid and ABA accumulation during the development and germination of *Nicotiana plumbaginifolia* seeds. Planta, 224: 622–632.

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.

Galau, G. A., Jakobsen, K. S. and Hughes, D. W. (1991) The controls of late dicot embryogenesis and early germination. Physiol. Plant, 81: 280-288.

Gallardo, K., Firnhaber, C., Zuber, H., Héricher, D., Belghazi, M., Henry, C., Küster, H. and Thompson, R. (2007) A combined proteome and transcriptome analysis of developing *Medicago truncatula* seeds. Molecular & Cellular Proteomics 6.12: 2165-2179.

Gallardo, K., Job, C., Groot, S. P. C., Puype, M., Demol, H., Vandekerckhove, J. and Job, D. (2001) Proteomic analysis of *Arabidopsis* seed germination and priming. Plant Physiol., 126: 835-848.

Gallardo, K., Le Signor, C., Vandekerckhove, J., Thompson, R. D. and Burstin, J. (2003) Proteomics of *Medicago truncatula* seed development establishes the time frame of diverse metabolic processes related to reserve accumulation. Plant Physiology, 133: 664–682.

Gälweiler, L., Guan, C., Müller, A., Wisman, E., Mendgen, K., Yephremov, A. and Palme, K. (1998) Regulation of polar auxin transport by AtPIN1 in *Arabidopsis* vascular tissue. Science, 282:2226–30.

Garay-Arroyo, A., Colmenero-Flores, J. M., Garciarrubio, A., Covarrubias, A. A. (2000) Highly hydrophilic proteins in prokaryotes and eukaryotes are common during conditions of water deficit. J. Biol. Chem., 275: 5668–5674.

Garnier, J., Gibrat, J-F., Robson, B. (1996) GOR method for predicting protein secondary structure from amino acid sequence. Meth. Enzymol., 266: 540-543.

Gatehouse, J. A., Evans, I. M., Croy, R. R. D. and Boulter, D. (1986) Differential expression of genes during legume seed development. Phil. Trans. R. Soc. Lond., B314: 367-384.

Giavalisco, P., Wilson, D., Kreitler, T., Lehrach, H., Klose, J., Gobom, J. and Fucini, P. (2005) High heterogeneity within the ribosomal proteins of the *Arabidopsis thaliana* 80S ribosome. Plant Mol. Biol., 57: 577–591.

Gibson, S. I. (2005) Control of plant development and gene expression by sugar signaling. Curr. Opin. Plant Biol., 8: 93–102.

Goday, A., Jensen, A. B., Culianezmacia, F. A., Alba, M. M., Figueras, M., Serratosa, J., Torrent, M. and Pages, M. (1994) The maize abscisic acid-responsive protein RAB17 is located in the nucleus and interacts with nuclearlocalization signals. Plant Cell, 6: 351-360.

Godoy, J. A., Lunar, R., Torres-Schumann, S., Moreno, J., Rodrigo, M. and Pintor-Toro, J. A. (1994). Expression, tissue distribution and subcellular localization of dehydrin TAS14 in salt-stressed tomato plants. Plant Molecular Biology 26: 1921–1934.

Goldberg, R. B., Barker, S. J. and Perez-Grau, L. (1989) Regulation of gene expression during plant embryogenesis. Cell, 56: 149-160.

Gomez, J., Sfinchez-Martinez, D., Stiefel, V., Rigau, J., Puigdoménech P, Pagés M (1988) A gene induced by the plant hormone abscisic acid in response to water stress encodes a glycine-rich protein. Nature 334:262-264.

Gomez, L. D., Baud, S., Gilday, A., Li, Y., Graham, I. A. (2006) Delayed embryo development in the ARABIDOPSIS TREHALOSE-6-PHOSPHATE SYNTHASE 1 mutant is associated with altered cell wall structure, decreased cell division and starch accumulation. Plant J., 46: 69–84.

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

Gruis, D. F., Selinger, D. A., Curran, J. M., and Jung, R. (2002) Redundant proteolytic mechanisms process seed storage proteins in the absence of

seed-type members of the vacuolar processing enzyme family of cysteine proteases. Plant Cell, 14: 2863–2882.

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

Guerche, P., Tire, C., Grossi de Sa, F., De Clercq, A., Van Montagu, M. and Krebbers, E. (1990a) Differential expression of the *Arabidopsis* 2S albumin genes and the effect of increasing gene family size. Plant Cell, 2: 469-478.

Guo, W., Ward, R. W. and Thomashow, M. F. (1992) Characterization of a cold-regulated wheat gene related to *Arubidopsis: Cor47.* Plant Physiol., 100: 915-922.

Hajduch, M., Casteel, J. E., Hurrelmeyer, K. E., Song, Z., Agrawal, G. K. and Thelen, J. J. (2006) Proteomic analysis of seed filling in *Brassica napus*: developmental characterization of metabolic isozymes using high-resolution two-dimensional gel electrophoresis. Plant Physiol., 141: 32–46.

Hajduch, M., Ganapathy, A., Stein, J. W. and Thelen, J. J. (2005) A systematic proteomic study of seed filling in soybean: establishment of high-resolution two-dimensional reference maps, expression profiles, and an interactive proteome database. Plant Physiol., 137: 1397–1419.

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

Hamel, F., Breton, C. and Houde, M. (1998) Isolation and characterization of wheat aluminum-regulated genes: possible involvement of aluminum as a pathogenesis response elicitor. Planta, 205(4): 531-538.

Han, B., Hughes, D. W., Galau, G. A., Bewley, J. D. and Kermode, A. R. (1997) Changes in late-embryogenesis-abundant (LEA) messenger RNAs and dehydrins during maturation and premature drying of *Ricinus communis* L. seeds. Planta, 201:27-35

Harada, J. J. (1997) Seed maturation and control of germination. In: Cellular and Molecular Biology of Plant Seed Development. (Larkins, B. A. and Vasil, I. K. eds.) Kluwer Academic Publishers, Dordrecht, pp. 545-592.

Heintzen, C., Fischer, R., Melzer, S., Kappeler, K., Apel, K. and Staiger, D. (1994) Circadian oscillations of a transcript encoding a germin-like protein that is associated with cell walls in young leaves of the long-day plant *sinapis alba* l. Plant physiology, 106(3): 905-915.

Hernandez-Sebastia, C., Marsolais, F., Saravitz, C., Israel, D., Dewey, R. E. and Huber, S. C. (2005) Free amino acid profiles suggest a possible role for asparagine in the control of storage-product accumulation in developing seeds of low- and high-protein soybean lines. J. Exp. Bot., 56: 1951–1963.

Heyden, B. J, Alsheikh, M. K., Smith, E. A., Torvik, C. F., Seals, D. F. and Randall, S. K. (2002) The calcium-binding activity of a vacuole-associated, dehydrin-like protein is regulated by phosphorylation. Plant Physiol., 130: 675–687.

Himmelbach, A., Yang, Y. and Grill, E. (2003) Relay and control of abscisic acid signaling. Current Opinions in Plant Biology, 6: 470–479.

Hinniger, C., Caillet, V., Michoux, F., Amor, M. B., Tanksley, S., Lin, C. and Mccarthy, J. (2006) Isolation and characterization of cdna encoding three dehydrins Expressed during coffea canephora (robusta) grain development. Annals of botany, 1-11.

Hoekstra, F. A., Golovina, E. A., Tetteroo, F. A., and Wolkers, W. F. (2001) Mechanisms of plant desiccation tolerance. Trends in plant science, 6(9): 431-8.

Hoffmann, R., Craik, D. J., Pierens, G., Bolger, R. E. and Otvos Jr., L. (1998) Phosphorylation of the C-terminal sites of human p53 reduces non-sequence-specific DNA binding as modeled with synthetic peptides. Biochemistry, 37: 13755–13764.

Houde, M., Daniel, C., Lachapelle, M., Allard, F., Laliberté, S. and Sarhan, F. (1995) Immunolocalization of freezing-tolerance-associated proteins in the cytoplasm and nucleoplasm of wheat crown tissue. Plant J., 8: 583-593.

Houde, M., Danyluk, J., Laliberte, J. F., Rassart, E., Dhindsa, R. S. and Sarhan, F. (1992) Cloning, characterization, and expression of a cDNA encoding a 50-kilodalton protein specifically induced by cold- acclimation in wheat. Plant Physiol., 99: 1381-1387.

Houston, N. L., Hajduch, M. and Thelen, J. J. (2009) Quantitative Proteomics of Seed Filling in Castor: Comparison with Soybean and Rapeseed Reveals Differences between Photosynthetic and Non-photosynthetic Seed Metabolism.

Hughes, D. W. and Galau G. A. (1989) Temporally modular gene expression during cotyledon development. Genes Dev., 3(3): 358–369.

Hundertmark, M., Hincha, D. K. (2008) LEA (Late Embryogenesis Abundant) proteins and their encoding genes in *Arabidopsis thaliana*. BMC Genomics, 9: 118–139.

Hurkman, W. J. and Tanaka, C. K. (1996) Effect of salt stress on germin gene expression in barley roots. Plant physiol., 110(3): 971-977.

Ikeda, M., Umehara, M. and Kamada, H. (2006) Embryogenesis-related genes; its expression and roles during somatic and zygotic embryogenesis in carrot and Arabidopsis. Plant Biotechnology, 23: 153–161.

Ishitani, M., Liu J., Halfter,U., Kim, C-S., Shi, W. and Zhu, J-K. (2000) SOS3 Function in Plant Salt Tolerance Requires N-Myristoylation and Calcium Binding. The Plant Cell, 12: 1667–1677.

Ismail, A. M., Hall, A. E. and Close, T. J. (1999a) Purification and partial characterization of a dehydrin involved in chilling tolerance during seedling emergence of cowpea. Plant Physiol., 120 237-244.

Ismail, A. M., Hall, A. E. and Close, T. J. (1999b) Allelic variation of a dehydrin gene co-segregates with chilling tolerance during seedling emergence. Proc. Natl. Acad. Sci. USA, 23: 13569–13573.

Jain, A. K. (2004) Cloning and structural analysis of a cDNA clone encoding glycinin (*Gly-1*) seed storage protein of peanut. Journal of Biotechnology, 7(3): 221-231

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-localised protein homologous to ribosomal protein L13 from *Oryza sativa* indica. Biochemica et Biophysica Acta., 1676: 182-192.

Jenik, P. D. and Barton, M. K. (2005) Surge and destroy: the role of auxin in plant embryogenesis. Development, 132: 3577–85.

Jenik, P. D., Gillmor, C. S. and Lukowitz, W. (2007) Embryonic Patterning in *Arabidopsis thaliana.* Annu. Rev. Cell Dev. Biol., 23: 207–36.

Jenner, C. F. (1982) Storage of starch. In: Encyclopedia of Plant Physiology (NS): Plant Carbohydrates I, lntracellular Carbohydrates, Vol. 13A, (F. A. Loewus and W. Tanner, eds) Berlin: Springer-Verlag, pp. 700-747.

Johansen, D. A. (1950). Plant Embryology: Embryology *of* the spermatophyte. Chronica Botanica Co, Waltham, MA.

Joosen, R., Cordewener, J., Supena, E. D. J., Vorst, O., Lammers, M., Maliepaard C., Zeilmaker, T., Miki, B., America, T., Custers, J. and Boutilier, K. (2007) Combined Transcriptome and Proteome Analysis Identifies Pathways and Markers Associated with the Establishment of Rapeseed Microspore-Derived Embryo Development. Plant Physiol., 144: 155–172.

Jürgens, G., Mayer, U., Torres Ruiz, R. A., Berleth, T. and Misbra, S. (1991) Genetic analysis of pattern formation in the Arabidopsis embryo. Development, 1: 27-38.

Kagaya, Y. Okuda, R., Ban, A., Toyoshima, R., Tsutsumida, K., Usui, H., Yamamoto, A. and Hattori, T. (2005b) Indirect ABA-dependent regulation of seed storage protein genes by FUSCA3 transcription factor in Arabidopsis. Plant Cell Physiol., 46: 300–311.

Kagaya, Y., Toyoshima, R., Okuda, R., Usui, H., Yamamoto, A. and Hattori, T. (2005a) LEAFY COTYLEDON1 controls seed storage protein genes through its regulation of FUSCA3 and ABSCISIC ACID INSENSITIVE3. Plant Cell Physiol., 46: 399–406.

Kende, H. and Zeevaart, J. A. D. (1997) The five "classical" plant hormones. Plant Cell, 9: 1197-1210.

Kermode, A. R. (1990) Regulatory mechanisms involved in the transition from seed development to germination. CRC Crit. Rev. Plant Sci., 81: 280-288.

Koag, M. C., Fenton, R. D., Wilkens, S. and Close, T. J. (2003) The binding of maize DHN1 to lipid vesicles: gain of structure and lipid specificity. Plant Physiol., 131: 309–316.

Kovacs, D., Kalmar, E., Torok, Z. and Tompa, P. (2008) Chaperone activity of ERD10 and ERD14, two disordered stress-related plant proteins. Plant Physiol., 147: 381–390.

Krogan, N. T. and Berleth, T. (2007) From genes to patterns: Auxin distribution and auxin-dependent gene regulation in plant pattern formation. Can. J. Bot., 85(4): 353–368.

Kroj, T. Savino, G., Valon, C., Giraudat, J. and Parcy, F. (2003) Regulation of storage protein gene expression in Arabidopsis. Development, 130: 6065–6073.

Kyte, J.and Doolittle, R. F. (1982) A simple method for displaying the hydropathic character of a protein. J. Mol. Biol., 157(1):105-32.

Laux, T., Würschum, T. and Breuninger H. (2004) Supplement genetic regulation of embryonic pattern formation. The Plant Cell, 16: S190–S202.

Levi, A., Panta, G. R., Parmentier, C. M., Muthalif, M. M. Arora, R., Shanker, S. and Rowland, L. J. (1999) Complementary DNA cloning, sequencing and expression of an unusual dehydrin from blueberry floral buds. Physiol. Plant., 107: 98-109.

Leyser, O. (2006) Dynamic integration of auxin transport and signaling. Curr. Biol., 16: 424–33.

Liu, C-M., Xu, Z-H. and Chua, N-H. (1993) Auxin polar transport is essential for the establishment of bilateral symmetry during early plant embryogenesis. Plant Cell, 5: 621-630.

Long, J. A., Moan, E. I., Medford, J. I. and Barton, M. K. (1996) A member of the KNOTTED class of homeodomain proteins encoded by the *STM* gene of *Arabidopsis*. Nature, 379: 66-69.

Lopez-Gomez, R. and Gomez-lim, M. A. (1992) A method for extracting intact RNA from fruits rich in polysaccharides using ripe mango mesocarp. HortScience, 27: 440–442.

Lyndon, R. F. (1990) Plant Development: The Cellular Basis. Topics in Plant Physiology: 3 (M. Black and J. Chapman, eds.) Unwin Hyman. London.

Maheshwari, P. (1950) An lntroduction to the Embryology of Angiosperms, New York: McGraw-Hill.

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

Mansfield, S. G. and Briarty, L. G. (1990) Early embryogenesis in Arabidopsis thaliana. 11. The developing embryo. Can. J. Bot., 69: 461-476.

Mansfield, S. G. and Briarty, L. G. (1992) Cotyledon cell-development in *Arabidopsis thaliana* during reserve deposition. Can. J. Bot., 70: 151–164.

Mansfield, S. G., Briarty, L. G., and Erni, S. (1990) Early embryogenesis in *Arabidopsis thaliana.* I. The mature embryosac. Can. J. Bot., 69: 447-460.

Mayer, U., Torres Ruiz, R. A., Berleth, T., Misbra, S. and Jürgens, G. (1991) Mutations affecting body organization in the Arabidopsis embryo. Nature, 353: 402-407.

Mechin, V., Thevenot, C., Li Guilloux, M., Prioul, J. L. and Damerval, C. (2007) Developmental analysis of maize endosperm proteome suggests a pivotal role for pyruvate orthophosphate dikinase. Plant Physiol., 143: 1203-1219.

Mehta, P. A., Rebala, K. C. Venkataraman, G. and Parida, A. (2009) A diurnally regulation dehydrin from *Avicennia marina* that shows nucleo-cytoplasmic localization and is phosphorylated by Casein kinase II *in vitro*. Pl. Physiol. Biochem., 47(8): 701-709.

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

Millar, A. A., Jacobsen, J. V., Ross, J. J., Helliwell, C. A., Poole, A. T., Scofield, G., Reid, J. B. and Gubler, F. (2006) Seed dormancy and ABA metabolism in Arabidopsis and barley: the role of ABA 80-hydroxylase. Plant J., 45: 942–954.

Momma, M., Kaneko, S., Haraguchi, K. and Matsukura, U. Peptide mapping and assessment of cryoprotective activity of 26/27-kDa dehydrin from soybean seeds. Biosci. Biotechnol. Biochem. 67 (2003) 1832-1835.

Monke, G., Altschmied, L., Tewes, A., Reidt, W., Mock, H. P., Baumlein, H. and Conrad, U. (2004) Seed-specific transcription factors ABI3 and FUS3: molecular interaction with DNA. Planta, 219: 158–166.

Mundy, J. and Chua, N-H. (1988) Abscisic acid and water-stress induce the expression of a novel rice gene. EMBO J., 7:2279-2286.

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

Nakabayashi, K., Okamoto, M., Koshiba, T., Kamiya, Y. and Nambara, E. (2005) Genome-wide profiling of stored mRNA in *Arabidopsis thaliana* seed germination: epigenetic and genetic regulation of transcription in seed. Plant J., 41(5): 697-709.

Nambara, E. and Marion-Poll, A. (2003) ABA action and interactions in seeds. Trends in Plant Science, 8: 213–217.

Nambara, E. and Marion-Poll, A. (2005) Abscisic acid biosynthesis and catabolism. Annu. Rev. Plant Biol., 56: 165–185.

Natesh, S. and Rau, M. A. (1984) The embryo. In: Embryology of angiosperms (B. M. Johri, ed.) Berlin: Springer-Verlag. pp. 377-443.

Neutelings, G., Domon, J. M., Membre, N., Bernier, F., Meyer, Y., David, A. and David, H. (1998) Characterization of a germin-like protein gene expressed in somatic and zygotic embryos of pine. Plant Molecular Biology, 38(6): 1179-1190.

Neven, L., Haskell, G. D. W., Hofig, A., Li, Q. B. and Guy, C. L. (1993) Characterization of a spinach gene responsive to low-temperature and water-stress. Plant Mol. Biol., 21: 291-305.

Nylander, M., Svensson, J. E., Palva, T. and Welin, B. V. (2001) Stress-induced accumulation and tissue-specific localization of dehydrins in *Arabidopsis thaliana*. Plant Mol. Biol., 45: 263–279.

Offler, C. E., McCurdy, D. W., Patrick, J. W. and Talbot, M. J. (2003) Transfer cells: cells specialized for a special purpose. Annu. Rev. Plant Biol., 54: 431–454.

Okamoto, M. Kuwahara, A., Seo, M., Kushiro, T., Asami, T., Hirai, N., Kamiya, Y., Koshiba, T. and Nambara, E. (2006) CYP707A1 and CYP707A2, which encode abscisic acid 80-hydroxylases, are indispensable for proper control of seed dormancy and germination in Arabidopsis. Plant Physiol., 141: 97–107.

Okamuro, J. K. and Goldberg, R. B. (1989) Regulation of plant gene expression: general principles. In: The Biochemistry of Plants Vol.15. (P. K. Stumpf and E. E. Conn, eds.) Academic Press, New York, pp. 1-82.

Oliveira, E., Amara, I., Bellido, D., Odena, M. A., Domınguez, E., Pagés, M. and Goday, A. (2007) LC-MSMS identification of Arabidopsis thaliana heatstable seed proteins: enriching for LEA-type proteins by acid treatment. J. Mass Spectrom., 42: 1485–1495.

Olson, A. R. and Cass, D. D. (1981) Changes in megagametophyte structure in *Papaver nudicaule* following in vitro placental pollination. Am. J. Bot., 68: 1338-1341.

Páges, M., Vilardell, J., Jensen, A. B., Alba, M. M., Torrent, M. and Goday, A. (1995) Molecular biological responses to drought in maize. In: Interacting stresses on plants in a changing climate. (M. Jackson and C. Black, eds) NATO/ASI, Springer-Verlag, Berlin.

Panitz, R., Borisjuk, L., Manteuffel, R. and Wobus, U. (1995) Transient expression of storage-protein genes during early embryogenesis of*Vicia faba*: synthesis and metabolization of vicilin and legumin in the embryo, suspensor and endosperm. Planta, 196: 765–774.

Paponov, I. A., Teale, W. D., Trebar, M., Blilou, I. and Palme, K. (2005) The PIN auxin efflux facilitators: evolutionary and functional perspectives. Trends Plant Sci., 10:170–77.

Parcy, F., Valon, C., Raynal, M., Gaubier-Comella, P., Delseny, M. and Giraudat J. (1994) Regulation of gene expression programs during Arabidopsis seed development: roles of the ABI3 locus and of endogenous abscisic acid. Plant Cell, 6(11):1567–1582.

Parcy, F., Valon, C., Kohara, A., Misera, S. and Giraudat, J. (1997) The ABSCISIC ACID-INSENSITIVE3, FUSCA3, and LEAFY COTYLEDON1 loci act in

concert to control multiple aspects of Arabidopsis seed development. Plant Cell, 9: 1265–1277.

Park, M-H., Choa, S-W., Yuna, K-N., Chung, K-Y., Choi, J-S., Park, C-H., Kima, H-S., Jong, S-K., Campbell, C., Taijif, A. and Woo, S-H. (2007) Proteomic approach to seed developing proteins in buckwheat. Proceedings of the 10th International Symposium on Buckwheat, 25-30.

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

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.

Poethig, R. S., Coe, E. H. and Johri, M. M. (1986) Cell lineage patterns in maize embryogenesis: A clonal analysis. Dev. Biol., 117: 392–404.

Pollock, E. G. and Jensen, W. A. (1964) Cell development during early embryogenesis in Capsella and *Gossypium.* Am. J. Bot., 51: 915-921.

Pritchard, H. N. (1964) A cytochemical study of embryo development in Stellaria media. Am. J. Bot., 51: 472-479.

Puhakainen, T., Hess, M. W., Pirjo Makela, P., Svensson, J., Heino, P. and Palva, E. T. (2004) Overexpression of multiple dehydrin genes enhances tolerance to freezing stress in Arabidopsis. Plant Mol. Biol., 54: 743–753.

Pulla, R. K., Kim, Y-J., Kim, M. K., Senthil, K. S., In, J-G. and Yang, D.-C. (2007) Isolation of a novel dehydrin gene from *Codonopsis lanceolate* and analysis of its response to abiotic stresses. BMB reports, 338-343.

Raghavan, V. (1986) Embryogenesis in Angiosperms: A Developmental and Experimental Study. Cambridge University Press, Cambridge.

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.

Reiser, L. and Fischer, R.L. (1993) The ovule and the embryo sac. Plant Cell, 5: 1291-1301.

Roberts, J. K., DeSimone, N. A., Lingle, W. L. and Dure, L. (1993) Cellular concentrations and uniformity of cell type accumulation of two Lea proteins in cotton embryos. Plant Cell, 5: 769-780.

Robertson, M. and Chandler. P. M. (1992) Pea dehydrins: Identification, characterization and expression. Plant Mol. Biol., 19: 1031-1044.

Rolland, F., Baena-gonzalez, E. and Sheen, J. (2006) Sugar sensing and signaling in plants: conserved and novel mechanisms. Annu. Rev. Plant Biol., 57: 675–709.

Rolletschek, H., Koch, K., Wobus, U. and Borisjuk, L. (2005) Positional cues for the starch/lipid balance in maize kernels and resource partitioning to the embryo. Plant J., 42: 69–83.

Rorat, T., Szabala, B. M., Grygorowicz, W. J., Wojtowicz, B., Yin, Z. and Rey, P. (2006) Expression of SK3-type dehydrin in transporting organs is associated with cold acclimation in *Solanum* species. Planta, 224: 205-221.

Rozen, S. and Skaletsky H. (2000) Primer3 on the WWW for general users and for biologist programmers. In: Bioinformatics Methods and Protocols in the series Methods in Molecular Biology (S. Krawetz and S. Misener, eds.). Humana Press, Totowa, NJ, pp. 365-386.

Russell, S. D. (1993) The egg cell: Development and role in fertilization and early embryogenesis. Plant Cell, 5: 1349-1359.

Ruuska, S. A., Girke, T., Benning, C. and Ohlrogge, J. B. (2002) Contrapuntal networks of gene expression during Arabidopsis seed filling. Plant Cell, 14: 1191–1206.

Sabatini, S., Beis, D., Wolkenfelt, H., Murfett, J., Guilfoyle, T., Malamy, J., Benfey, P., Leyser, O., Bechtold, N., Weisbeek, P. and Scheres, B. (1999) An auxin-dependent distal organizer of pattern and polarity in the *Arabidopsis* root. Cell, 99: 463-472.

Sambrook, J., Fritsch, E. F. and Maniatis, T. (1989) Molecular Cloning: A Laboratory Manual (2nd Edition) Cold Spring Harbor Lab., Cold Spring Harbor, NY.

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.

Saulsberry, A., Martin, P. R., O'Brien, T., Sieburth, L. E., and Pickett, F. B. (2002) The induced sector Arabidopsis apical embryonic fate map. Development, 129: 3403–3410.

Scheres, B. and Heidstra, R. (1999) Digging out roots: Pattern formation, cell division and morphogenesis in plants. Curr. Top. Dev. Biol., 45: 207-247.

Schiavone, F. M. and Cooke, T. J. (1987) Unusual patterns of somatic embryogenesis in the domesticated carrot: Developmental effects of exogenous auxins and auxin transport inhibitors. Cell Differ., 21: 53-62.

Schiavone, F. M. and Racusen, R. H. (1990) Microsurgery reveals regional capabilities for pattern reestablishment in somatic carrot embryos. Dev. Biol., 141: 211-219.

Schulz, R. and Jensen, W. A. (1968a) Capsella embryogenesis: The early embryo. J. Ultrastruct. Res., 22: 376-392.

Schulz, R. and Jensen, W. A. (1968b) Capsella embryogenesis: The egg, zygote, and young embryo. Am. J. Bot., 55: 807-819.

Schweizer, P.; Christoffel, A. and Dudler, R. (1999) Transient expression of members of the germin-like gene family in epidermal cells of wheat confers disease resistance. Plant journal, 20(5): 541-552.

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

Seo, M., Hanada, A., Kuwahara, A., Endo, A., Okamoto, M., Yamauchi, Y., North, H., Marion-Poll, A., Sun, T., Koshiba, T., Kamiya, Y., Yamaguchi, S. and Nambara, E. (2006) Regulation of hormone metabolism in Arabidopsis seeds: phytochrome regulation of abscisic acid metabolism and abscisic acid regulation of gibberellin metabolism. Plant J., 48: 354–366.

Sheoran, I. S., Sproule, K. A., Olson, D. J. H., Ross, A. R. S. and Sawhney, V. K. (2006) Proteome profile and functional classification of proteins in *Arabidopsis thaliana* (Landsberg *erecta*) mature pollen. Sexual Plant Reproduction, 19: 185–196.

Shewry, P. R. (1995) Plant storage proteins Biological Reviews of the Cambridge Philosophical Society, 70(3): 375-426.

Shotwell, M. A. and Larkins, B. A. (1989) Molecular Biology and Biochemistry of seed storage proteins. In: The Biochemistry of Plants. A Comprehensive Treatise, Vol 15, Academic Press, N.Y. pp. 297-345.

Sivaramakrishna, D. (1978) Size relationships of apical cell and basal cell in two-celled embryos in angiosperms. Can. J. Bot., 56: 1434-1438.

Slack, C. R. and Browse, J. A. (1984) Synthesis of storage lipids in developing seeds. In: Seed Physiology, (D. R. Murray, ed.) Academic Press, Sydney, pp. 209-244.

Slightom, J. L., Drong, R. F., Klassy, C. and Hoffman, L. M. (1985) Nucleotide sequences from pbaseolln cDNA clones: the major storage proteins from Phaseolus vulgaris are encoded by two unique gene families. Nucleic Acids Research, 13(18): 6483-6498.

Soulages, J. L, Kim, K., Arrese, E. L., Walters, C. and Cushman, J. C. (2003) Conformation of a group 2 late embryogenesis abundant protein from soybean: evidence of poly (L-proline)-type II structure. Plant Physiology, 131: 963–975.

Staiger, D., Apel, K. and Trepp, G. (1999) The *atger3* promoter confers circadian clock-regulated transcription with peak expression at the beginning of the night. Pl. Mol. Biol., 40(5): 873-882.

Stalberg, K., Ellerstrom, 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.

Steeves, T. A. and Sussex, I. M. (1989) Patterns in plant development. Cambridge University Press, New York.

Sunderlikova, V. and Wilhelm, E. (2002) High accumulation of legumin and Lea-like mRNAs during maturation is associated with increased conversion frequency of somatic embryos from pedunculate oak (*Quercus robur* L.). Protoplasma, 220:97–103.

Suzuki, M., Ketterling, M. G., Li, Q-B. and McCarty, D. R. (2003) Viviparous1 alters global gene expression patterns through regulation of abscisic acid signaling. Plant Physiol., 132: 1664–1677.

Svensson, J., Ismail, A., Palva, E. T. and Close, T. J. (2002) Dehydrins. In: Cell and molecular responses to stress. (K. B. Storey and J. B. Storey, eds.) Elsevier Science, Netherlands, pp. 155-171.

Tanaka, H. H., Ononchi, M. Kondo, I. Hara-Nishimura, N., Nishmura, C., Machida, and Y. Machida (2001) A subtilisin-like serine protease is required for epidermal surface formation in *Arabidopsis* embryos ad juvenile plants. Dev., 128: 4681-4689.

Thomas, T. L. (1993) Gene expression during plant embryogenesis and germination: An overview. Plant Cell, 5: 1401-1410.

Thordahl-Christensen, H., Zhang, Z., Wei, Y. and Collinge, D. B. (1997) subcellular localization of H_2O_2 in plants: H_2O_2 accumulation in papillae and hypersensitive response during powdery-mildew interaction. Plant journal, 11(6): 1187-1194.

To, A., Valon, C., Savino, G., Guilleminot, J., Devic, M., Giraudat, J. and Parcy, F. (2006) A network of local and redundant gene regulation governs Arabidopsis seed maturation. Plant Cell, 18: 1642–1651.

Tommasini, L., Svensson, J. T., Rodriguez, E. M., Wahlid, A., Malatrasi, M., Kato, K., Wanamaker, S., Resnik, J. and Close, T. J. (2008) Dehydrin gene expression provides an indicator of low temperature and drought stress: transcriptome-based analysis of barley (*Hordeum vulgare* L.). Funct. Integr. Genomics, 8: 387–405.

Tompa, P. (2005) The interplay between structure and function in intrinsically unstructured proteins. FEBS Lett. 579: 3346-3354.

Torres Ruiz, R. A., Lohner, A. and Jürgens, G. (1996) The *GURKE* gene is required for normal organization of the apical region in the *Arabidopsis* embryo. Plant J., 10: 1005-1016.

Tsuchiya, Y., Nambara, E., Naito, S. and McCourt, P. (2004) The FUS3 transcription factor functions through the epidermal regulator TTG1 during embryogenesis in Arabidopsis. Plant J., 37: 73–81.

Tykarska, T. (1976) Rape embryogenesis. I. The proembryo development. Acta. Soc. Bot. Poloniae., 45: 3-16.

Tykarska, T. (1979) Rape embryogenesis. II. Development of embryo proper. Acta. Soc. Bot. Poloniae., 48, 391-421.

Tykarska, T. (1980) Rape embryogenesis. 111. Embryo development in time. Acta. Soc. Bot. Poloniae., 49: 369-385.

Tykarska, T. (1987) Rape embryogenesis. V. Accumulation of lipid bodies. Acta. Soc. Bot. Poloniae., 56: 573-584.

Vallelian-Bindschedler, l.; Mosinger, E.; Metraux, J. P. and Schweizer, P. (1998) Structure, expression and localization of a germin-like protein in barley (*hordeum vulgare* l.) that is insolubilized in stressed leaves. Plant Mol., Biol., 37(2): 297-308.

van Lammeren, A. A. M. (1981) Early events during embryogenesis in *Zea mays* L. Acta Soc. Bot. Poloniae., 50: 289-290.

Vensel, W. H., Tanaka, C. K., Cai, N., Wong, J. H., Buchanan, B. B. and Hurkman, W. J. (2005) Developmental changes in the metabolic pattern profiles of wheat endosperm. Proteomics, 5: 1594-1611.

Vittorioso, P., Cowling, R., Faure, J. D., Caboche, M. and Bellini, C. (1998) Mutation in the *Arabidopsis PASTICCINO1* gene, which encodes a new

103994
26-10-10

FK506-binding protein-like protein, has a dramatic effect on plant development. Mol. Cell Biol., 18: 3034-3043.

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(2): 433-443.

Walbot, V. (1978) Control mechanisms for plant embryogeny. In: Dormancy and Developmental Arrest, (M. E. Clutter, ed.) Academic Press, New York, pp. 113-166.

Wardlaw, C. W. (1955) Embryogenesis in Plants. Methuen and Co., Ltd., London.

Weber, H., Borisjuk, L. and Wobus, U. (2005) Molecular physiology of legume seed development. Annu. Rev. Plant Biol. 56: 253–279.

White, J. A., Todd, T., Newman, T., Focks, N., Girke, T., de Ilarduya, O. M., Jaworski, J. G., Ohlrogge, J. B. and Benning, C. (2000) A new set of Arabidopsis expressed sequence tags from developing seeds: the metabolic pathway from carbohydrates to seed oil. Plant Physiol., 124: 1582–1594.

Wilkins, T. A. and Smart, L. B. (1996) Isolation of RNA from plant tissue. p. 21-41. In P.A. Krieg (Ed.), A Laboratory Guide to RNA: Isolation, Analysis, and Synthesis. Wiley-Liss, New York.

Wise, M. J. (2003) LEAping to conclusions: a computational reanalysis of late embryogenesis abundant proteins and their possible roles. BMC Bioinform., 4: 52.

Wisniewski, M., Webb, R., Balsamo, R., Close, T. J., Yu, X. M. and Griffith, M. (1999) Purification, immunolocalization, cryoprotective, and antifreeze activity of PCA60: A dehydrin from peach (*Prunus persica*). Physiol. Plant. 105: 600-608.

Wobus, U. and Weber, H. (1999) Seed maturation: genetic programmes and control signals. Curr. Opin. Plant Biol., 2: 33–38.

Yang, P., Li, X., Wang, X., Chen, H., Chen, F. and Shen, S. (2007) Proteomic analysis of rice (*Oryza sativa)* seeds during germination. Proteomics, 7: 3358-3368.

Yang, S. H. and Zeevaart, J. A. (2006) Expression of ABA 80-hydroxylases in relation to leaf water relations and seed development in bean. Plant J., 47: 675–686.

Zhu, B., Choi, D-W., Fenton, R. and Close, T. J. (2000) Expression of the barley dehydrin multigene family and the development of freezing tolerance. Mol. Gen. Genet., 264: 145-153.