Isolation and characterization of *ERD4* gene from *Brassica juncea* and its functional validation

By

Archana N. Rai
LIFE 01200804005
Homi Bhabha National Institute, Mumbai

A thesis submitted to the
Board of Studies in Life Sciences

In partial fulfillment of requirements
for the Degree of

DOCTOR OF PHILOSOPHY
of
HOMI BHABHA NATIONAL INSTITUTE

August, 2014

STATEMENT BY AUTHOR

This dissertation has been submitted in partial fulfillment of requirements for an advanced degree at Homi Bhabha National Institute (HBNI) and is deposited in the Library to be made available to borrowers under rules of the HBNI.

Brief quotations from this dissertation are allowable without special permission, provided that accurate acknowledgement of source is made. Requests for permission for extended quotation from or reproduction of this manuscript in whole or in part may be granted by the Competent Authority of HBNI when in his or her judgment the proposed use of the material is in the interests of scholarship. In all other instances, however, permission must be obtained from the author.

Archana N. Rai

DECLARATION

I, hereby declare that the investigation presented in the thesis has been carried out by me. The work is original and has not been submitted earlier as a whole or in part for a degree / diploma at this or any other Institution / University.

Archana N. Rai

LIST OF PUBLICATIONS ARISING FROM THE THESIS

Journal

Published: One

Membrane topology and predicted RNA-binding function of the 'Early Responsive to Dehydration (ERD4)' plant protein. **Rai A., Suprasanna P., D'Souza SF, Kumar V., *Plos One* 2012,** 7(3):e32658.

Under Communication: One

Novel chloroplast targeted RNA binding protein ERD4 improves salinity and drought tolerance.

Conference

1. Molecular characterization of abiotic stress responsive *Brassica juncea BjERD4* gene **Archana N. Rai** and P. Suprasanna. National Conference of Plant Physiology; National Research Centre on Groundnut, Junagadh; 11-13, 2013**.**

Others:

1. Gene Bank Submissions:

- Brassica juncea ERD4 protein (ERD4) gene, complete cds gi|161006797|gb|EU126607.2|[161006797]
- Brassica juncea early responsive to dehydration 4 protein-like mRNA, partial sequence. gi|183376723|gb|EU596450.1|[183376723].

Archana N. Rai

Dedicated to my beloved father and grandmother (Mama)

ACKNOWLEDGEMENTS

Completion of this doctoral thesis is an outcome of the affection and support of many good people who were always there to extend their whole hearted support. I would like to express my sincere gratitude to all of them. I would like to express my special appreciation and thanks to my guide Dr. P. Suprasanna. Sir you have been a tremendous mentor for me. I would like to thank you for encouraging my research and for allowing me to grow as a researcher. Your advice on both research as well as on my career has been priceless.

I would also like to thank my committee members, Dr. S.K. Apte (Chairman), Dr.A.Sharma (Member), Dr. N.Jawali (Member) and Dr. S .F. D'souza (ex-giude and member) for their support, critical comments and suggestions. I would especially like to thank Dr.Vinay Kumar for his scientific advice and many insightful discussions and suggestions during my bioinformatics and proteiomics studies. I am also very grateful to Dr. K.V. Rao, Professor, CPMB, Osmania University, Hyderabad for providing lab facilities for plant transgenic development and characterization.

I would also like to thank all of my colleagues Dr. Ashish, Dr. Sudhakar, R.K.Ji, Srinivas Sir Manish, Laxmi ma'm, Jincy, Survey Ji, Shriram and Ganesh, who supported me during experimentation, in writing, and incented me to strive towards my goal.. I extend my gratitude to Dr. Vikas Patade and Dr.Vinayak for their help and suggestions to perform some experiments in the early days of my Ph.D thesis work. I thank Dr.Hema Rajaram (MBD) for helping me in HBNI matters, and acknowledge the crucial help extended by Dr. Gagan (BARC) and Srinath (CPMB) in my research work. I also thank Sumit, Aman and Shyam for their support.

A special thanks to my family. Words cannot express how grateful I am to my mother-in law, father-in-law, my mother, and father who supported me in every possible way to see the completion of this work. I am immensely indebted to Sweta, Bausky, Chintu, Santu and Bantu for understanding me and providing the strength to achieve the goal. At the end I would like express appreciation to my beloved husband Neeraj who has been

always my support in the moments when there was no one to answer my queries. To my sweethearts Ananya and Arshia, I would like to express my thanks for being such 'wonderful girls' always cheering me up.

……Archana N. Rai

CONTENTS

SYNOPSIS

Homi Bhabha National Institute

Ph. D. PROGRAMME

1. **Name of the Student:** Archana Neeraj Rai

2. **Name of the Constituent Institution:** Bhabha Atomic Research Centre

3. **Enrolment No. :** LIFE 01200804005

4. **Title of the Thesis:** "Isolation and characterization of *ERD4* gene from *Brassica juncea* and its functional validation"

5. **Board of Studies:** Life Sciences

Drought and salinity are among the most serious challenges to crop production in the world. Both traditional breeding and genetic engineering methods are being utilized to improve crop plants for drought and salinity tolerance. Characterization of abiotic stress-responsive genes is essential for elucidating the responsive mechanisms by which plants can be improved for stress tolerance. Salt stress exposure modulates expression of certain genes in the root tissue within 15 min (Kawasaki et al, 2001). These early response genes are probably inducible transcriptional activators or upstream signal pathway components which may act as the fate dominators of salt tolerance (Ouyang et al, 2007). It is thus imperative to understand the complex molecular mechanisms of early signal perception, signaling and secondary messengers which regulate the early responses to salt or drought stress. *Brassica juncea* is an important oil seed crop but it's productivity is reduced under salinity and drought stress. Not much is known about the early responses to salt stress and the ensuing regulatory mechanisms controlling tolerance.

The Early Responsive to Dehydration (*ERD*) genes are defined as those genes that are activated in early hours during drought stress in Arabidopsis (Kiyosue et al, 1994a). *ERD4* is one of the genes expressed under dehydration stress during early hours (Kiyosue et al, 1994a) and down regulated during rehydration (Oono, 2003) in Arabidopsis. A sugarcane homologue of *ERD4* has been found to be activated by transcription factor *CBF4* (Mc Qualter and Dookun-Saumtally, 2007) However, in the protein data base ERD4 has been classified as a hypothetical transmembrane protein which have DUF (Domain of unknown function) 221. In view of this, the present investigations were performed to clone the *ERD4* gene from *Brassica juncea* and to understand the structure,

function, localization of *Brassica juncea* ERD4 (BjERD4) protein and to study the role in abiotic stress tolerance. The work is compiled into a thesis covering five chapters. Chapter I covers general details and literature survey related to the present work. Chapter II is on cloning and subcellular localization study of *ERD4* from *Brassica juncea*. Chapter III presents studies on ERD4 protein function. Expression analysis and validation of the role of *ERD4* in abiotic stress tolerance is detailed in chapter IV and Chapter V includes conclusions and future directions.

Chapter I: Introduction

In this introductory chapter, literature related to the topic of the thesis is described. Drought and salinity are the major abiotic stresses that adversely affect physiological and metabolic processes, leading to diminished growth and yield of crop plants. The effects of drought range from morphological to molecular levels and are evident at all phenological stages of plant growth. Salt inhibits plant growth first, by reducing the plant's ability to take up water, which leads to slower growth. Later, it may enter the transpiration stream and eventually injure cells in the transpiring leaves, further reducing growth. This is the salt-specific or ion-excess effect of salinity (Munns and Tester, 2008). In the last decade, studies on physiological and molecular mechanisms of abiotic stress tolerance have led to characterization of a number of genes associated with stress adaptation (Cushman and Bohnert, 2000; Fraire-Velázquez and Balderas-Hernández, 2013). Some of these genes are specific for a particular stress while others are shared between various stresses. These may play regulatory roles in response to developmental or environmental cues (Liere et al, 2011; Krasensky and Jonak, 2012.).

Gene expression varies with the time after the stress is applied (Kawasaki et al, 2001). Based on the time of induction dehydration stress-induced genes can be divided into two categories, i.e., responsive to dehydration and early responsive to dehydration (Shinozaki and Yamaguchi-Shinozaki, 1997). To date, a total of 16 complementary DNAs (cDNAs) for *ERD* genes have been isolated from 1-h-dehydrated *A. thaliana* (Kiyosue et al, 1994a). Most of these have been characterized and some of them found to be involved in abiotic stress tolerance.

Chapter II: Cloning and subcellular localization of *ERD4* from *Brassica juncea*.

For all the expreiments, *Brassica juncea* cv. Pusa bold was used. Primers for full length *ERD4* gene amplification were designed from Arabidopsis *ERD4* gene sequence (AT1G69450). *ERD4* like gene was amplified using *Brassica juncea* genomic and cDNA. These amplified products were cloned in TA cloning vector and sequenced. Blast results of genomic *ERD4* like gene from *Brassica juncea* showed 81% sequence identity with Arabidopsis *ERD4* gene. Based on the homology, *ERD4* like gene form *B. juncea* has been named as *BjERD4*. Comparative alignment analysis of the genomic and the cDNA sequence of *BjERD4* gene indicated the presence of five introns. The *BjERD4* cDNA encodes a protein of 723 amino acid residues with the predicted molecular weight of 80.93KDa and a theoretical PI of 9.23. The homologs of *B. juncea* ERD4 protein were identified in various plant lineages and were found to be conserved in all the plants for which proteome data is available. Transmembrane topology was also predicted using bioinformatic tools.

Subcellular localization of ERD4 protein was conducted using *in silico* as well as by using GFP protein tagging approach.

A. *In silico* studies: The analysis of *B. juncea* ERD4 by the ambiguous targeting predictor (ATP) suggested a score of 0.39, which weakly suggested dual targeting of the ERD4 protein. The analysis of ERD4 orthologs by the ATP, however, suggested wide variations in the confidence score with a low score of 0.19 for some ERD4 proteins that clearly indicated localization of ERD4 in only one compartment. In order to get detailed information, analysis was conducted to experimentally validate chloroplastic envelope proteins of *A. thaliana*. An overall amino acid composition and N-terminal sequence logo plots of the 123 selected proteins (ENV dataset) from Arabidopsis proteome (Kleffmann et al, 2004) were analyzed. The positional abundance of amino acids in sequence logos showed abundance of Ser residues and under representation of Arg residues in the ENV dataset. The amino acid composition analysis also showed much higher abundance of Ser, Ala and Leu residues in the N-terminal sixteen residues and sixteen residues as compared to the full-length proteins. The low percentages of the positively charged Arg/Lys residues and significantly higher percentage of Ser residues in the N-terminal sixteen residues of ERD4 proteins thus suggested the chloroplastic localization of ERD4.

B. Subcellular localization of ERD4 protein using GFP protein tagging approach

1. Plasmid construction and *Agrobacterium* transformation: The GFP coding sequence was amplified with PCR from pCAMBIA 1302. The full length *ERD4* was amplified from the *Brassica Juncea* cDNA and sub cloned in frame into the N-terminal side of *GFP*. The CaMV- BjERD4-GFP-OCS fragment was subcloned into *pART27* plant expression vector. The *gfp* gene was amplified from *pCAMBIA 1302* and cloned in *pART7* and then subcloned in pART27 plant expression vector which was used as positive control for sub cellular localization. The resulting transformation vectors

pART27-BjERD4-GFP and pART27-GFP were transformed into *Agrobacterium tumefaciens Eha105*.

2. Cellular localization of BjERD4 protein using tobacco leaf infiltration: Agro infiltration for transient expression was performed in tobacco leaves. Intact protoplasts were isolated from the GFP expressed leaf area. Confocal microscopic observation of transformed protoplasts demonstrated that GFP fluorescence was dispersed throughout the entire protoplast when transformed with control plasmid. In contrast using the BjERD4-GFP fusion construct, expression was exclusively observed in the chloroplast (plastid).

Chapter III: Characterization of BjERD4 protein

A. *In silico* studies: Using computational approaches, at least nine transmembrane helices and a globular domain of 165 amino acid residues (183–347) were detected in ERD4 protein. The structural-functional annotation of the globular domain was arrived at using fold recognition methods, which suggested in its sequence presence of two tandem RNA-recognition motif (RRM) domains each folded into $\beta\alpha\beta\beta\alpha\beta$ topology. The structure based sequence alignment with the known RNA-binding proteins revealed conservation of two non-canonical ribonucleoprotein sub-motifs in both the putative RNA-recognition domains of the ERD4 protein. The function of highly conserved ERD4 protein may thus be associated with its RNA-binding ability. Based upon multiple sequence alignment, RNA binding amino acid sequences was also predicted.

B. Functional validation of globular RRM motif of ERD4 protein

1. Cloning and expression of RRM motif of ERD4 protein: A 528 bp ERD4 fragment was cloned in pET28-His tag vector and transformed into *E.coli* BL21. The optimum conditions for expression of this protein in soluble fraction were optimized and the recombinant protein of approx. 20 KD was observed on SDS page in the inclusion bodies as well as in the soluble proteins. Purification of RRM domain of ERD4 protein was done using soluble fraction of cell lysate by using immobilized-metal (Ni^{2+}) affinity chromatography (IMAC).

2. RNA electrophoresis mobility shift assay: Purified RRM motif of ERD4 protein was incubated with *in vitro* transcribed RNA. The RNA products were separated on a 1% agarose gel. Bovine Serum Albumin (BSA) protein and Reverse Transcriptase (RT) protein treated with the same procedures served as a negative and positive control respectively. Electrophoresis mobility shift assay (EMSA) result showed that ERD4 – RRM protein domain can bind RNA efficiently.

Chapter IV: Expression analysis and functional validation of *ERD4* gene using knockdown and overexpression approaches.

1. Expression analysis: For constitutive expression analysis, basal expression of ERD4 gene was checked in different plant developmental stages which showed maximum expression in young leaves. Twenty one days old seedlings of Brassica cultivar "Pusa Bold" were treated, with 200 mM NaCl and 20 % PEG for 0 hr, 0.5, 1 hr, 2 hr, 4 hr, 8 hr and 16 hr. Mannitol, SA, ABA, cold and heat stress was also given for 1hr. Total RNA was isolated from the leaf and root samples. Real-time quantitative RT-PCR was carried

out for the target gene *ERD4*. The *Brassica juncea* β actin gene was amplified for gene expression normalization and to provide relative quantification.

In case of 200 mM NaCl treatment, significant increase in *BjERD4* gene expression was detected as early as within 0.5 hr (2.165 fold induction) which reached maximum after 4 hr (6 fold induction) in roots, while in shoots maximum expression was observed only after 2 hr (2.7 fold induction). In case of 20 % PEG, early induction of *BjERD4* gene was observed after 0.5 hr of treatment in root (2.75 fold). Significant increase in *BjERD4* gene expression was also detected when seedlings were treated with ABA (100µM) and other stress conditions.

2. Preparation of Knockdown and overexpression constructs and Arabidopsis transformation: A 442 bp of *ERD4* gene coding region was amplified from *Brassica juncea* cDNA and cloned in sense and antisense order into Hannibal (EHSA) and was further sub-cloned in pART27 plant expression vector. For overexpression construct full length *BjERD4* gene was cloned in pART7 vector and further subcloned in plant expression vector pART27. The knockdown and overexpression constructs were then transformed in *Agrobacterium Eha105*. Presence of constructs was confirmed by PCR using primers for *nptII* gene. For Arabidopsis transformation flower dip method was used. Transformed seeds were selected using kanamycin and plants were generated till T4 generation.

3. Molecular characterization of *ERD4* transformed lines: *ERD4* knockdown and overexpressed lines of Arabidopsis were identified based on kanamycin selection. Further these lines were confirmed by PCR using *nptII* primers. The Real time PCR was performed for all these lines using primers designed to amplify specific gene sequence

cloned in constructs. Significant reduction of transcript in two lines was observed in knockdown lines. Transcript reduction was upto 3.6 fold. Significant increases in transcript level in overexpressed plants were observed.

4. Assessment of the performance of Knockdown and Overexpressed lines under salinity and PEG treatment: Ten day old WT, RNAi L-2, RNAi L-5, EO-1 and EO-4 seedlings and were transferred to different treatment conditions (Control-MS, Salinity-MS+100mM NaCl and Drought- 10% PEG/ 150mM mannitol) for 7 days. These lines were assessed in terms of growth parameters, chlorophyll content, ROS production and lipid peroxidation. In the wild type, significantly high chlorophyll content and less ROS production and lipid peroxidation were observed as compared to knockdown lines under stress condition which suggested role of *ERD4* gene in abiotic stress tolerance. But overexpressed plant showed better tolerance under salinity and drought conditions.

Chapter V: Conclusions

In summary, an *Early Responsive to Dehydration (ERD4)* gene was cloned from *Brassica juncea* and induction of the gene expression was observed under drought and salinity treatments. The organ specific expression indicated that *BjERD4* may function in the normal programme of the plant growth and development suggesting that in addition to its role in stress response, *BjERD4* may also be involved in these processes. Based upon *in silico* studies, it was confirmed that the protein is localized in chloroplast membrane with at least nine transmembrane helices. Confocal microscopy of transformed protoplasts confirmed the chloroplastic (plastids) localization of ERD4. By fold-prediction algorithms, the presence of two RNA-recognition motifs was detected in ERD4 protein sequence. ERD4-RRM domain interaction with RNA was validated using RNA EMSA.

To further study the role of BjERD4, knockdown mutant of *ERD4* was generated in Arabidopsis. The performance of knockdown mutant lines under salinity and drought conditions showed significant increase in MDA content and ROS production in knockdown lines where as chlorophyll content was found to be decreased as compared to wild type.

Future directions: Elucidation of the stress responses of crop plants assumes great relevance in view of the challenges posed by abiotic stresses to crop productivity. Since, ERD4 protein function has been shown to have chloroplastic RNA binding, RNA which is binding to this protein can be identified. Arabidopsis database mining has shown nine more ERD4 like proteins whose role can also be depicted in relation to abiotic stress tolerance. Promoter can also be isolated and used for stress specific expression of *ERD4* gene. The role of *ERD4* in abiotic stress tolerance can be further assessed in crop plants using transgenic methods so that it gene can be used as a candidate genes for crop improvement. Overexpressed Arabidopsis lines can be useful in understanding role of *ERD4* in other abiotic and biotic stresses tolerance and in molecular cross talk.

References:

Cushman, J. C. and Bohnert, H. J. (2000) genomic approaches to plant stress tolerance. Curr. Opin. Plant Biol. 3(2):117–124

Fraire-Velázquez, S. and Emmanuel Balderas-Hernández, V. (2013) Abiotic Stress in Plants and Metabolic Responses. Abiotic Stress - Plant Responses and Applications in Agriculture. Edited by Vahdati K and Leslie C Intechopen: doi: 10.5772/54859. pp 25-48.

Horsch, R., Fry, J., Hoffman, N., Eichholtz, D., Rogers, S. and Fraley, R. (1985) A simple and general method for transferring genes into plants. Science. 227: 1229–1231.

Kawasaki, S., Borchert, C., Deyholos, M., Wang, H., Brazille, S., Kawai, K., Galbraith, D. and Bohnert, H.J. (2001) Gene expression profiles during the initial phase of salt stress in rice. Plant Cell. 13:889-905.

Kiyosue, T., Yamaguchi-Shinozaki, K. and Shinozaki, K. (1994a) Cloning of cDNAs for genes that are early-responsive to dehydration stress (ERDs) in Arabidopsis thaliana L.: identification of three ERDs as HSP cognate genes. Plant Mol. Biol. 25: 791-798.

Kleffmann, T., Russenberger, D., von Zychlinski, A., Christopher, W., Sjölander, K., Gruissem, W. and Baginsky, S. (2004) The Arabidopsis thaliana chloroplast proteome reveals pathway abundance and novel protein functions. Curr Biol. 14: 354–362.

Krasensky, J. and Jonak, C. 2012 Drought, salt, and temperature stress-induced metabolic rearrangements and regulatory networks. J Exp Bot. 63(4):1593-608.

Liere, K., Weihe, A. and Börner, T. (2011) The transcription machineries of plant mitochondria and chloroplasts: Composition, function, and regulation. Jour. Plant Physiology. 168: 1345– 1360.

Mc-Qualter, R. B. and Dookun-Saumtally, A. (2007) Expression profiling of abiotic-stress-inducible genes in sugarcane. XXVI Congress, International Society of Sugar Cane Technologists. 878-888.

Munns, R. and Tester, M. (2008) Mechanisms of salinity tolerance. Annu. Rev. Plant Biol. 2008. 59:651–681.

Oono, Y., Seki, M., Nanjo, T., Narusaka, M., Fujita, M., Satoh, R., Satou, M., Sakurai, T., Ishida, J., Akiyama, K., Iida, K., Maruyama, K., Satoh, S., Yamaguchi-Shinozaki, K. and Shinozaki, K. (2003) Monitoring expression profiles of *Arabidopsis* gene expression during rehydration process after dehydration using ca. 7000 full-length cDNA microarray. Plant J. 34**:** 868–887.

Ouyang, Bo., Yang, T., Li, H., Zhang, L., Zhang. Y., Zhang, J., Fei, Z. and Ye, Z. (2007) Identification of early salt stress response genes in tomato root by suppression subtractive hybridization and microarray analysis. Jour Exp Botany. 58: 507-520.

Shinozaki, K. and Yamaguchi-Shinozaki, K. (1997) Gene expression and signal transduction in water-stress response. Plant Physiol. 115: 327- 334.

LIST OF FIGURES

LIST OF TABLES

CHAPTER I: INTRODUCTION

Stress is defined as a sudden change in the environment that exceeds the organism's optimal and causes homeostatic imbalance, which must be compensated for. Plants growing under field conditions are exposed to various environmental factors, which constitute their macro and microenvironment. Any deviation in these factors from the optimum levels is deleterious to plants and leads to stress. Stress may be caused due to abiotic factors like strong light, UV, high and low temperatures, freezing, drought, salinity, heavy metals, hypoxia or the biotic factors like viruses, insects, nematodes, bacteria, fungi etc. At a given point of time, plant may have to face even a combination of more than one of above mentioned factors. Abiotic stress, in fact is the principal cause of crop failure worldwide, lowering the average production and productivity of most of the major crops by more than 60% [1], thus threatening sustainability and food security. Therefore, understanding abiotic stress responses and tolerance mechanism is relevant and very important.

1.1. Abiotic stress response and tolerance

Plant cells receive stress signals through various sensors (not yet known), and the signals are transduced by various signaling pathways in which many secondary messengers like including calcium, reactive oxygen species (ROS) and inositol phosphates [2]. These second messengers, such as inositol phosphates, further modulate the intracellular calcium level. This perturbation in cytosolic Ca^{2+} level is sensed by calcium binding proteins, also known as Ca^{2+} sensors. These sensors apparently lack any enzymatic activity and change their conformation in a calcium dependent manner. These sensory proteins then interact with their respective interacting partners often initiating a phosphorylation cascade and target the major stress responsive genes or the transcription

factors controlling these genes. Some stress-inducible genes encode functional proteins that are directly involved in stress tolerance. Other stress-inducible genes encode regulatory proteins, such as signal transducers, that presumably form positive and negative feedback loops to regulate stress responses. Stress induced changes in gene expression in turn may participate in the generation of hormones like ABA, salicylic acid and ethylene [3]. These molecules may amplify the initial signal and initiate a second round of signaling that may follow the same pathway or use altogether different components of signaling pathway. Certain molecules also known as accessory molecules may not directly participate in signaling but participate in the modification or assembly of signaling components [4]. These proteins include the protein modifiers, which may be added cotranslationally to the signaling proteins like enzymes for myristoylation, glycosylation, methylation and ubiquitination (Fig.1.1.1). The products of these stress genes ultimately lead to plant adaptation and help the plant to survive and surpass the unfavorable conditions. Thus, plant responds to stresses as individual cells and synergistically as a whole organism.

Drought Stress: Drought is one of the most serious world-wide problems for agriculture. The effects of drought range from morphological to molecular levels and are evident at all phenological stages of plant growth at whatever stage the water deficit takes place. Cell growth reduction is one of the most drought-sensitive physiological processes due to the reduction in turgor pressure [5]. Under severe water deficiency, cell elongation of higher plants can be inhibited by interruption of water flow from the xylem to the surrounding elongating cells [6]. Drought stress also trims down the photo-assimilation and metabolites required for cell division. Impaired mitosis, cell elongation and

expansion result in reduced plant height, leaf area and crop growth under drought [7, 8]. An important effect of water deficit is on the acquisition of nutrients by the root and their transport to shoots.

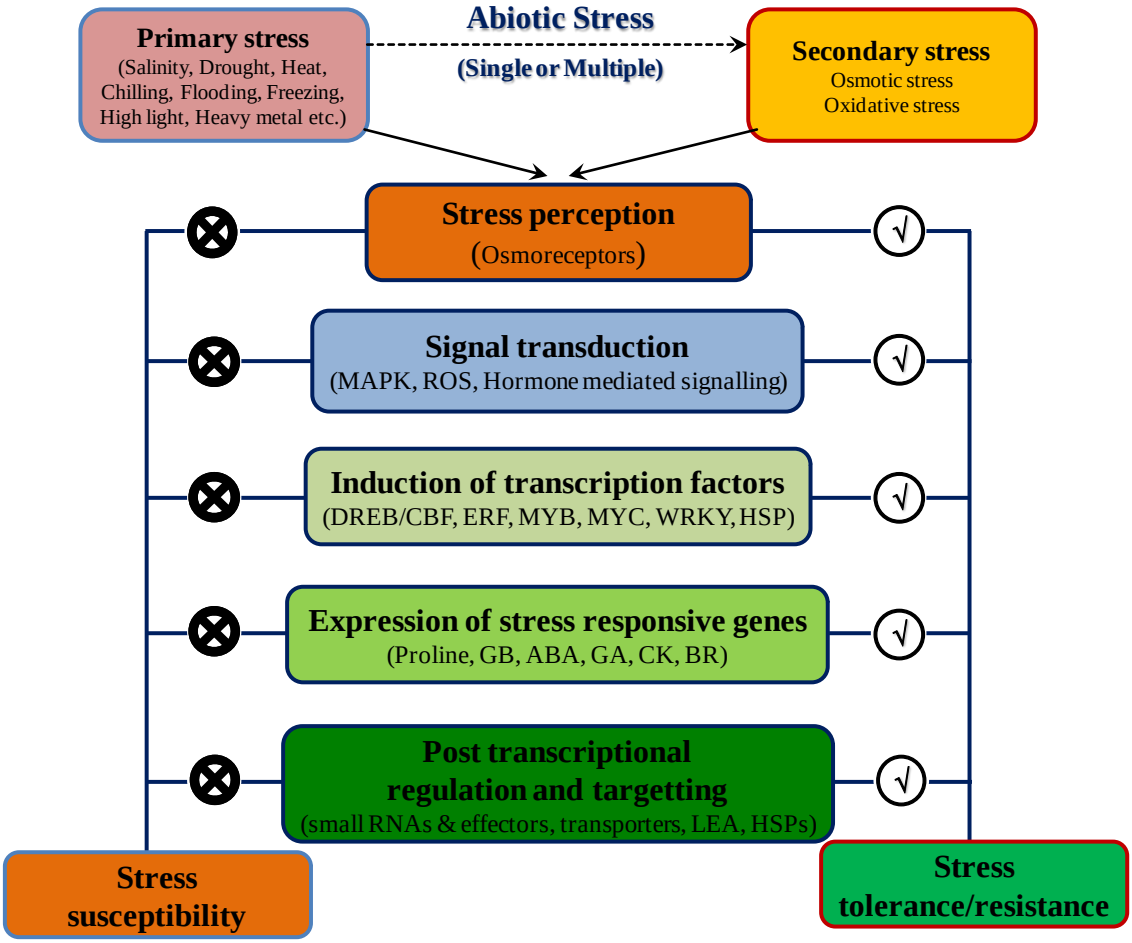


Fig. 1.1.1: Mechanism of abiotic stress response

Lowered absorption of the inorganic nutrients can result from interference in nutrient uptake and the unloading mechanism, and reduced transpirational flow [9, 10]. A major effect of drought is reduction in photosynthesis, which arises by a decrease in leaf expansion, impaired photosynthetic machinery, premature leaf senescence and associated reduction in food production [11]. Exposure of plants to drought stress leads to the generation of reactive oxygen species, including superoxide anion radicals (O^{-}_{2}),

hydroxyl radicals (OH), hydrogen peroxide (H_2O_2), alkoxy radicals (RO) and singlet oxygen (O_2^1) [12]. These reactive oxygen species enhanced peroxidation of membrane lipids and degradation of nucleic acids, and both structural and functional proteins. Various organelles including chloroplasts, mitochondria and peroxisomes are the first target of reactive oxygen species produced under drought stress [13, 14].

Plants respond, adapt to and survive under drought stress by the induction of various morphological, biochemical and physiological responses [15]. Plants have adopted escape and avoidance mechanisms at morphological level and phenotypic difference in root and leaves structure also help in plant survival during stress [16]. One of the physiological mechanisms is osmotic adjustment which allows the cell to decrease osmotic potential and, as a consequence, increases the gradient for water influx and maintenance of turgor [17]. Osmotic adjustment is accomplished with the accumulation of compatible solutes like proline, glycine betaine [18]. Another mechanism which helps in stress tolerance is antioxidant defense system constitutes both enzymatic and non-enzymatic components. Enzymatic components include superoxide dismutase, catalase, peroxidase, ascorbate peroxidase and glutathione reductase. Non-enzymatic components contain cystein, reduced glutathione and ascorbic acid [19]. Plant cellular water deficit condition triggers the changes in gene expression (up- and down-regulation). Gene expression may take place directly by the stress conditions or result from secondary stresses and/or injury responses. Some of the genes like Aquaporin, *DREB* transcription factor, dehyrins, *RD29* have been cloned and validated for drought stress tolerance [20]. Drought stress gene expression can be triggered directly or through ABA, a plant stress hormone [21].

Salinity stress: Salinity stress is one of the most severe environmental stresses, impairing crop production by at least 20% of irrigated land worldwide. In addition, increased salinity of arable land is expected to have devastating global effects, resulting in up to 50% land loss by the middle of the twenty-first century [22]. Many crop species are very sensitive to soil salinity, are referred to as glycophytes, and cannot grow at 100 mM NaCl whereas halophytes being salt-tolerant plants can grow at salinity over 250 mM. The growth limitation due to salinity stress may be due to the osmotic effect of salt in the earlier phase or the ionic stress which impacts growth much later and with a less effect than the osmotic stress, especially at low to moderate stress [23]. The two effects give rise to a two-phase growth response to salinity. Thus, early responses to water and salt stress are essentially identical. Salt-specific effects occur mainly in old leaves where salt brought in with the transpiration stream accumulates to high levels over time. Ionic stress results in premature senescence of older leaves. Toxicity symptoms (chlorosis, necrosis) of ionic stress in mature leaves is because of high Na+ which affects plants by disrupting protein synthesis and interfering with enzyme activity [24, 25, 26]. The similarity between water and salt stress also apply to most metabolic processes: all processes apart from those relating to ion transport. Hormonal responses are also similar; for instance, abscisic acid levels rise within 1 h of a imposition of water stress [27] and salt stress [28].

High salinity stress causes an imbalance in sodium ions (Na^+) homeostasis, which is maintained by the coordinated action of various pumps, ions, Ca^{2+} sensors, and its downstream interacting partners, which ultimately results in the efflux of excess Na^+ ions. Certain channels show more selectivity to K^+ over Na^+. These include the K inward-rectifying channel, which mediates the influx of K^+ upon plasma membrane

hyperpolarization and selectively accumulates K^+ over Na^+ ions. The histidine kinase transporter (HKT) is a low-affinity Na^+ ion transporter, which blocks the entry of Na^+ ions into the cytosol [29]. The nonspecific cation channel is a voltage-independent channel, which acts as a gate for the entry of Na^+ into plant cells. Moreover, there is the K^+ outward rectifying channel, which opens during the depolarization of the plasma membrane and mediates the efflux of K^+ and the influx of Na^+ ions, leading to Na^+ accumulation in the cytosol. The vacuolar Na^+/H^+ exchanger (NHX) helps push excess Na^+ ions into vacuoles. Na^+ extrusion from plant cells is powered by the electrochemical gradient generated by H^+-ATPases, which permit the NHX to couple the passive movement of H^+ inside along the electrochemical gradient and extrusion of Na^+ out of the cytosol. Another pump, the H^+ /Ca^{2+} antiporter (CAX1), helps in Ca^{2+} homeostasis [30, 31].

1.2. Strategies for finding abiotic stress tolerance determinants in plants

There are various approaches that are commonly used to identify plant genes associated with abiotic stress tolerance. Conventional breeding has been based on empirical selection for yield [32]. However, this approach is far from being optimal, since yield is a quantitative trait and characterized by a low heritability and a high genotype-x-environment interaction [33]. It is strongly believed that understanding of a physiological and molecular basis may help target the key traits that limit yield. Such an approach may complement conventional breeding programs and hasten yield improvement [34]. The genetic approach utilizes natural or induced variation in stress tolerance or stress gene regulation. Gene mutations are induced by using chemical or physical mutagens.

Therefore, one can search for plant mutants that are hypersensitive to stress and these mutants can help in the identification of tolerant gene(s) [35].

The biochemical approach depends on the previous knowledge of particular enzyme or biochemical pathway relevant to salt tolerance. The quantitative and qualitative changes in protein under salt stress is mainly evidenced through employing western analysis, enzyme kinetics, fraction isolation, PAGE, HPLC, mass spectrometry etc [36].

In the gene expression profiling approach, the expression under stress condition is compared to that in the absence of stress. Genes that show increased expression under stress are cloned by using a variety of differential or subtractive screening techniques. In contrast, down-regulation of gene expression by stress is relatively understudied. Identification of the down-regulated genes is an important part of understanding tolerance mechanisms, as these may be important negative determinants of stress tolerance. Gene expression profiling using microarrays or gene chip is a novel approach to identify higher number of transcripts [37]. A major limitation of this approach is that many genes, critical for plant tolerance are not induced by stress however many induced genes are not important for tolerance as their induced expression may be a consequence of stress injury. Eventually, the functional analysis approach (through transgenic, over or under expression or through reverse genetics to identify knockout mutant) is necessary to establish the importance of the induced or suppressed gene(s) for plant salt tolerance. Approaches used for identification of stress tolerance determinants are illustrated in Fig. 1.2.1.

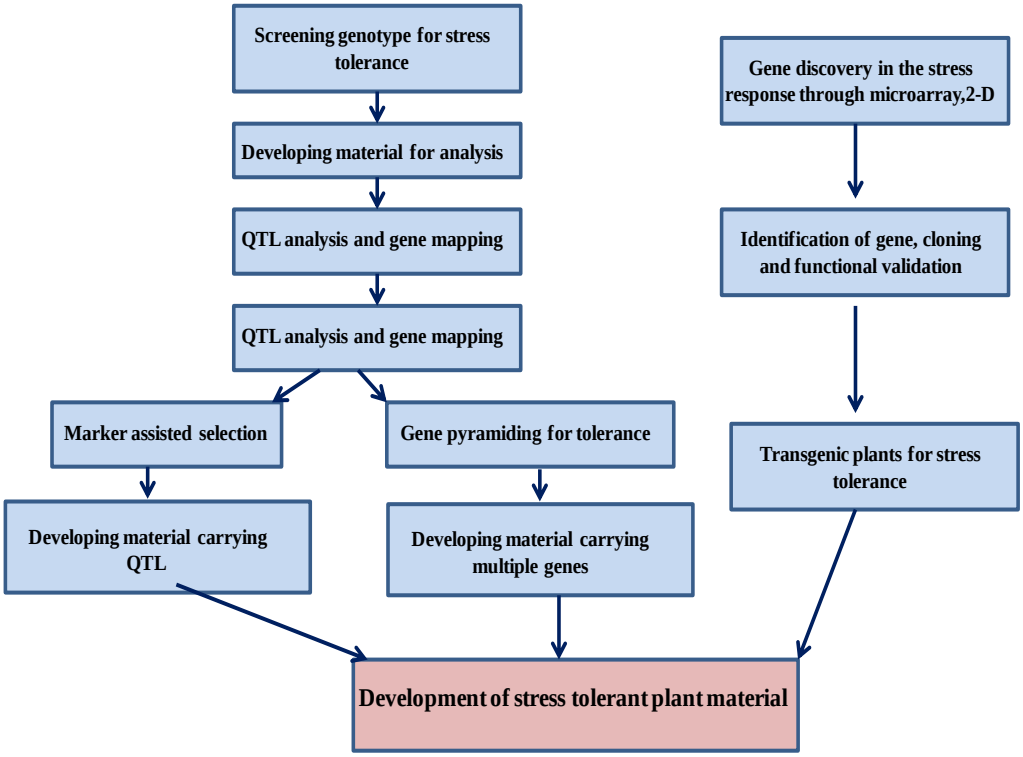


Fig. 1.2.1. Approaches used for identification of stress tolerance determinants

1.3. Genes involved in drought and salinity stress tolerance

A number of genes that respond to drought, cold and high-salinity stresses at the transcriptional level have been described [38]. The products of the stress inducible genes are classified into, those that directly protect against environmental stresses and those that regulate gene expression and signal transduction in the stress response. It is important to analyze the functions of stress-inducible genes not only to understand the molecular mechanisms of stress tolerance and the responses of higher plants but also to improve the stress tolerance of crops by gene manipulation.

A variety of genes are induced by drought stress, and functions of their gene products have been predicted from sequence homology with known proteins [39]. Based upon function, these gene products are classified into two groups. The first group

includes proteins that probably function in stress tolerance like water channel proteins involved in the movement of water through membranes, the enzymes required for the biosynthesis of various osmoprotectants (sugars, proline, and betaine), proteins that may protect macromolecules and membranes (LEA protein, osmotin, antifreeze protein, chaperon, and mRNA binding proteins), proteases for protein turnover (thiol proteases, Clp protease, and ubiquitin) and the detoxification enzymes (glutathione S-transferase, soluble epoxide hydrolase, catalase, superoxide dismutase, and ascorbate peroxidase). The second group contains protein factors involved in further regulation of signal transduction and gene expression that probably function in stress response: protein kinases, transcription factors and enzymes in phospholipid metabolism [20].

These stress responsive genes can also be categorized as early and late induced genes [22, 40]. Early genes are induced within minutes of stress signal perception and often express transiently. Various transcription factors are included in the list of early genes as the induction of these genes does not require synthesis of new proteins and signaling components are already primed. In contrast, most of the other genes, which are activated by stress more slowly, i.e. after hours of stress perception are included in the late induced category. The expression of these genes is often sustained. These genes include the major stress responsive genes such as RD (responsive to dehydration)/ KIN (cold induced)/COR (cold responsive), which encodes and modulate the proteins needed for synthesis, for example LEA-like proteins (late embryogenesis abundant), antioxidants, membrane stabilizing proteins and synthesis of osmolytes. Since many abiotic stresses can impinge upon a plant simultaneously and require more than one response for tolerance to occur. This could be done by altering levels or patterns of

expression of transcript involved in the early responses to stresses, as has been described to some extent in Arabidopsis [41].

ABA serves as an endogenous messenger in response to biotic and abiotic stress in plants. Drought results in production of high levels of ABA, accompanied by a major shift in global gene expression in plant cells and, consequently, an adaptive physiological response to the stress [42]. In addition to stress, ABA also controls other important and finely regulated processes such as growth and development, structure and regulation of stomatal function and seed dormancy [43]. During regulation of plant development, ABA also acts in intricate cross-communication with other important phytohormones, such as gibberellic acid, ethylene, auxin and brassinosteroids [43].

1.4. The Early Responsive to Dehydration (*ERD*) genes and their functional diversity

The early perception of water deprivation is critical for recruitment of genes that promote plant tolerance. The *ERD* genes are defined as genes that are rapidly activated during drought stress. The encoded proteins show a great structural and functional diversity and constitute the first line of defense against drought stress in plants.

ERD1 encodes a chloroplast ATP-dependent protease [44] and *ERD2* encodes a, HSP70 [44] *ERD3* encodes a methyltransferase in the pMT21 family [45], *ERD4* encodes a membrane protein [46], *ERD5* and *ERD6* encode a mitochondrial dehydrogenase proline protein and a carbohydrates carrier protein, respectively [47, 48]. *ERD7* encodes a protein related to senescence and dehydration [49], *ERD8* encodes a hsp81-family protein [44], *ERD9*, 11 and 13 belong to the family of glutathione-S-transferase [50]. ERD13/AtGSTF10, a plant phi specific class GST (Glutathione-S-transferase) is an

interacttion protein with BAK1 (BRI1 Associated receptor Kinase 1). BAK1 is a co-receptor, which forms a receptor complex with BRI1 (brassinosteroid (BR) receptor) to regulate brassinosteroid signaling in Arabidopsis. *ERD10* and 14 belong to the LEA protein family [44], ERD15 was first classified as a hydrophilic protein [51], which has a PAM2 interaction domain which interacts with poly-A tail binding proteins (PABP) [52]. ERD15 from Arabidopsis has been functionally characterized as a common regulator of the abscisic acid (ABA) response and salicylic acid (SA)-dependent defense pathway [52]. *ERD16* encodes a ubiquitination extension protein [44] (Table 1.4.1).

Table 1.4.1. List of *ERD* (Early Responsive to Dehydration) genes and their function in different plant systems

Gene / GenBank accession no.	Function	Reference
ERD1/D17582	ClpA/B ATP-dependent protease	[44]
ERD2/M23105	Heat shock protein (hsp70-i)	[44]
ERD3/NP_567575.1	Methyltransferase PMT21	[45]
ERD4/NP_564354.1	Integral membrane protein	[46]
ERD5/D83025	Precursor of proline dehydrogenase	[47]
ERD6/D89051	Sugar transporter	[48]
ERD7/NP_179374.1	Senescence/ dehydration related protein	[49]
ERD8/Y11827	Heat shock protein hsp81-2)	[44]
ERD9/NP_172508.4	Glutathione-S-transferase	[50]
ERD10/D17714	Group II LEA protein (lti29/lti45)	[51]
ERD11/D17672	Glutathione-S-transferase	[50]
ERD12/NP_189204.1	Allene oxide cyclase	[53]
ERD13/D17673	Glutathione-S-transferase	[50]
ERD14/D17715	Group II LEA protein	[51]
ERD15/D30719	Hydrophilic protein	[52]
ERD16/J05507	Ubiquitin extension protein	[51]

With respect to expression controlled by phytohormones, *ERD* genes present varied functions and responses in ABA signaling, some being sensitive to ABA during germination and development [54] and are involved in stress tolerance [2]. Other genes

are induced in response to more than one phytohormone [52]. Early Responsive to Dehydration 15 (*ERD15*) was characterized as a negative regulator of ABA and is induced by ABA, SA, injury and pathogen infection [52]. ABA application increases the expression of some members of the ERD group including *ERD10* and *ERD14* [51] while causing no effect on others, such as *ERD2*, 8 and 16 [44].

In Soybean, comparative Deep Super SAGE libraries, including one control and a bulk of six stress times imposed (from 25 to 150 min of root dehydration) for drought-tolerant and sensitive soybean accessions, enabled identification of new molecular targets for drought tolerance. A total of 1,127 unitags were up-regulated only in the tolerant accession, whereas 1,557 were up-regulated in both as compared to their controls. An expression profile concerning the most representative Gene Ontology (GO) categories for the tolerant accession revealed the expression "protein binding" as the most represented for "Molecular Function", whereas CDPK and CBL were the most up-regulated protein families in this category. Furthermore, particular genes expressed different isoforms according to the accession, showing the potential to operate in the distinction of physiological behaviors. Candidate genes related to "hormone response" (*LOX*, *ERF1b*, *XET*), "water response" (*PUB*, *BMY*), "salt stress response" (*WRKY*, *MYB*) and "oxidative stress response" (*PER*) figured among the most promising molecular targets. Additionally, nine transcripts (*HMGR*, *XET*, *WRKY20*, *RAP2*-4, *EREBP*, *NAC3*, *PER*, *GPX5* and *MYB*) validated by RT-qPCR (four different time points) confirmed their differential expression and pointed that after 25 minutes a transcriptional reorganization started in response to the new condition, with important differences between both accessions.

For identification of early responsive genes against salt stress, SSH libraries were constructed for the root tissue of two cultivated tomato (*Solanum lycopersicum*) genotypes: LA2711, a salt-tolerant cultivar, and ZS-5, a salt-sensitive cultivar [55]. A subset of clones from these SSH libraries were used to construct a tomato cDNA array and microarray analysis was carried out to verify the expression changes of this set of clones upon a high concentration of salt treatment at various time points compared to the corresponding non-treatment controls. A total of 201 non-redundant genes that were differentially expressed upon 30 min of severe salt stress either in LA2711 or ZS-5 were identified from microarray. In particular, genes involved in the metabolic pathways of nitrogen reduction and fixation, and methionine biosynthesis were significantly affected by salt stress [55]. Among salt-modulated genes, several heat shock proteins (HSPs) were identified. HSPs, which act as molecular chaperones, play a crucial role in protecting plants against stress by reestablishing normal protein conformations, and thus, maintain cellular homeostasis [56]. The diversity of the putative functions of identified genes indicated that salt stress resulted in a complex response in tomato plants.

To summarize, under abiotic stress conditions plants adopt several intricate strategies to overcome the environmental changes that threaten their growth and development. First plants allocate more metabolic energy for defending themselves. Then cell walls are reinforced and the concentration of several protective solutes in the cytoplasm increase. A large number of early response genes regulated by drought and salt stress identified by various studies encode unknown proteins, indicating that there is still a great deal to discover with regard to the mechanism of the drought and salt tolerance in plant species.

1.5. Hypothetical proteins and abiotic stress tolerance:

More than 40% of known proteins lack any annotation within public databases and are usually referred to as hypothetical proteins despite most of them being real and many being evolutionarily conserved and thus are expected to play important biological roles [57]. “Hypothetical proteins” have been grouped into families, and the latest release of PFAM contains 2,156 families annotated as domains of unknown function (95% of families of unknown function in PFAM are called DUFs). Classifying DUF families into superfamilies and clans is more problematic, as such classification often depends on additional information, such as three dimensional structures and/or protein function, and such information is not obviously available [58].

To date, comparative genomics in eukaryotes has focused largely on genes that encode proteins with experimentally defined domains or motifs (proteins with defined features (PDFs) [59]. Because the analysis of PDFs. revealed a high degree of similarity among different species, it has been accepted widely that the uniqueness of a particular species was driven by changes in regulatory genes or elements [60], as opposed to the divergence of established coding sequences or the creation of new genes. This has led to a wide-spread perspective that just a few model organisms can provide the experimental foundation to assign functions to nearly every eukaryotic gene. However on average, 20% to 40% of all eukaryotic genomes sequenced to date contain genes that encode for proteins of unknown function [61]. Those protein which does not show any similarity to previously defined domains or motifs of hidden Markov model protein family (HMMPFAM) defined as proteins with obscure features (POFs) and were distinguished

from proteins with defined features (PDFs), which contained at least one previously defined domain or motif [61, 62].

These are often chosen because of their general relevance to fundamental questions in a broad group of organisms or because they exhibit strong evolutionary conservation. The functional characterization of genes with unknown function might provide an insight into role of unknown proteins in different organisms in the basic or specialized processes, new and undiscovered pathways [61, 62].

Since the primary and secondary structure do not give any information regarding structure and function of DUF group protein, the identification of geometric relationships between protein structures, by the use of structural alignment methods, offers a powerful approach in identifying structural and functional relationships between highly divergent proteins [63]. It is well established that proteins evolve partly through rearrangements of larger fragments, typically domains, and nature of these fragments determine biological function of proteins [64]. The analysis of proteins at individual domain levels can facilitate functional annotation of uncharacterized genes and proteins [65].

Recently, function of a large number of proteins of DUF families has been proposed based on the structural homology of experimentally determined structures to functionally annotated proteins [57]. The functional domains can also be identified reliably by computational analysis such as prediction of the secondary structure, transmembrane segments, and by fold-recognition [66, 67]. An atomic model of the identified domain can further be obtained from the sequence alone by identifying homologs using sequence-sequence comparison or by fold assignment using structure-sequence alignment [68]. With the available computational tools, it is also possible to

identify residues involved in the biological function based on the structure-structure comparison. The utility of these approaches can be extended for predicted structural models of uncharacterized proteins enabling functional annotation of related proteins. Such a strategy is particularly useful for membrane proteins as their experimental structure-function determination is a difficult task.

Role of some of the Arabidopsis specific POF has been shown in oxidative stress tolerance using transgenic approach [69]. Transgenic plants were tested for their tolerance to oxidative stress imposed by paraquat or t-butyl hydroperoxide, or were subjected to osmotic, salinity, cold, and heat stresses. More than 70% of all expressed proteins conferred tolerance to oxidative stress. In contrast, 90% of the expressed proteins did not confer enhanced tolerance to the other abiotic stresses tested, and approximately 50% rendered plants more susceptible to osmotic or salinity stress. Two Arabidopsis-specific POFs, and an Arabidopsis and Brassica-specific protein of unknown function, conferred enhanced tolerance to oxidative stress. Based upon result species specific pathways were suggested for cellular repair and/or protection against oxidative stress [69].

1.6. RNA binding protein (RBP) and stress tolerance

RBPs have been shown to function as central regulators in the post-transcriptional regulation of RNA metabolism during diverse cellular processes, including growth, development, and stress responses. RNA metabolism includes post-transcriptional processes, such as RNA processing, pre-mRNA splicing, mRNA export, localization, turnover, and translational control. The complexity of the post-transcriptional regulation of eukaryotic gene expression is reflected in the diversity of the RBP family. Typical

RBPs contain one or more RNA recognition motifs (RRM, also known as RBD or RNP domain) or K homology (KH) domains [70]. In addition to these well-conserved domains, various auxiliary domains or motifs, such as glycine-rich, arginine-rich, arginine-glycine (RGG) repeat, serine-arginine (SR)- repeat, arginine-aspartate (RD)-repeat, and zinc finger motifs, are frequently found in RBPs [71]. The highly conserved RRM sequences are involved in the recognition of precursor-mRNAs, mature-mRNAs, and small nuclear RNAs, and in protein–protein interactions, leading to the formation of ribonucleoprotein (RNP) complexes [72]. The auxiliary domains are also involved in protein–protein interactions and dictate RNA-binding specificity [73].

Plant RBPs have been widely demonstrated to be regulatory factors controlling flower development, circadian rhythms, absscisic acid signaling, stress responses, and chromatin modification [74]. Since the first gene encoding GRP was identified in maize, cDNA encoding homologous proteins have been found in diverse plant species, including alfalfa, Arabidopsis, barley, Brassica, rice and tobacco. The biological functions of plant GRPs in the responses of plants to changing environmental conditions have been investigated and the involvement of GRPs in diverse biological and biochemical processes are being uncovered [75]. In particular, the functions of plant GRPs during cold acclimation were elucidated based on the fact that they are significantly induced by low temperature. The biological roles of several GRPs in plants under stress conditions have been characterized (Table 1.6.1). It was reported that AtGRP2, AtGRP4 and AtGRP7, three AtGRPs of the eight AtGRP family members in Arabidopsis, have different impacts on seed germination, seedling growth and stress tolerance of Arabidopsis plants under diverse stress conditions.

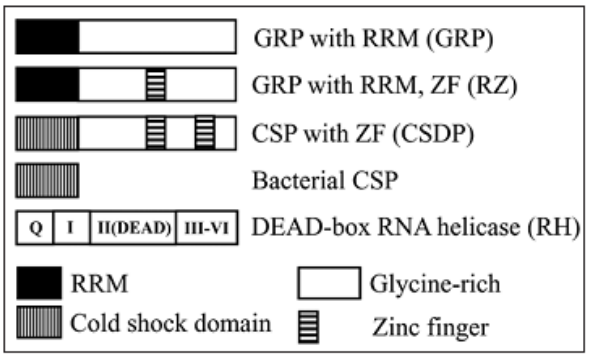


Fig. 1.6.1. Schematic presentation of the domain structures of RNA-binding proteins discussed in this chapter. Glycine-rich RNA-binding proteins (GRPs) contain an RNA-recognition motif RRM) at their N-terminus and a glycine-rich region at their C-terminus. RZs harbor an RRM at their N-terminus and a glycine-rich region interspersed with a CCHC-type zinc finger at their C-terminus. Cold shock domain proteins (CSDPs) contain a cold shock domain (CSD) in their N-terminal half as well as additional glycine-rich regions interspersed by CCHC-type zinc fingers in their C-terminal half. DEAD-box RNA helicases (RHs) contain motifs called Q, I, II (DEAD), III, IV, V and VI domains.

The mechanism by which these RBPs mediate their activity is the regulation of RNA folding in cells. RNA molecules have the tendency to fold into alternative secondary structures [76]. These alternative misfolded structures can then interfere with the normal function of RNA molecules and thus have to be resolved. It has been suggested that formation of alternative misfolded structures is prevented or reversed by the action of proteins generally known as RNA chaperones. The roles of RNA chaperones are more prominent when cells are exposed to low temperatures, as misfolded RNA molecules become stabilized and cannot assume native conformation without the help of RNA chaperones [77]. It was determined that bacterial CSPs function as RNA chaperones by destabilizing the overstabilized secondary structures in mRNAs for efficient translation at low temperatures. Since it was found that cyanobacteria lack CSPs

Table 1.6.1. Overview of plant RNA– binding proteins mentioned in the text and their roles in stress response

Protein name	Plant source	Gene ID	Expression/ function	Reference
Glycine rich RNA-Binding protein (GRP)				
AtGRP2	Arabidopsis	At4g13850	Cold/freezing tolerance	[78]
AtGRP4	Arabidopsis	At3g23830	-	[79]
Atgrp7	Arabidopsis	At2g21660	Cold/ freezing tolerance Negative role to salt/ drought	[80]
OsGRP1	Rice	Os01g68790	Cold/ freezing tolerance	[81]
OsGRP4	Rice	Os04g33810	Cold/ freezing tolerance	[81]
OsGRP6	Rice	Os12g31800	Cold/ freezing tolerance	[81]
Zinc finger GRP (RZ)				
AtRZ-1a	Arabidopsis	At3g26420	Cold/ Freezing tolerance Negative role to salt/ drought	[78, 82]
AtRZ-1b	Arabidopsis	At1g60650	-	[81]
AtRZ-1c	Arabidopsis	At5g04280	-	[81]
OsRZ-1	Rice	Os03g61990	-	[81]
OsRZ-2	Rice	Os07g08960	Cold/ freezing tolerance	[81]
OsRZ-3	Rice	Os03g47800		[81]
Cold shock domain protein (CSDP)				
AtCSDP1	Arabidopsis	At4g36020	Cold/ freezing tolerance Negative role to salt/ drought	[83]
AtCSDP2	Arabidopsis	At4g38680	Responds to cold, Positive role to salt	[81, 84]
AtCSDP3	Arabidopsis	At2g17870	Cold/ Freezing tolerance	[85]
AtCSDP4	Arabidopsis	At2g21060	-	
OsCSDP1	Rice	Os02g02870	Responds to cold	[86]
OsCSDP	Rice	Os08g03520	Responds to cold	[86]
DEAD-box RNA helicase (RH)				
AtRH9	Arabidopsis	At3g22310	Cold/Freezing tolerance	[80]
AtRH25	Arabidopsis	At5g08620	Cold/Freezing tolerance	[80]
LOS4	Arabidopsis	At3g53110	Cold/Freezing tolerance	[19]
STRS1	Arabidopsis	At1g31970	Supressor of salt, osmotic, heat	[87]
STRS2	Arabidopsis	At5g08620	Supressor of salt, osmotic, heat	[87]

but contain a cold‐induced RRM protein instead, it was hypothesized that RRM proteins may actually substitute for the function of CSPs in cyanobacteria.

Indeed, GRPs do contain similar structural features to RRM proteins found in cyanobacteria. Therefore, it is highly likely that GRPs as well as CSDPs function as RNA

chaperones during the cold adaptation process in plants. This consideration is supported by a series of recent findings that demonstrated that GRPs, CSDPs and RHs harbor RNA chaperone activity. Arabidopsis AtGRP2 and AtGRP7, which confer cold and freezing tolerance in plants, complement the cold sensitivity of BX04 mutant *E. coli*, which lacks four CSPs and is highly sensitive to cold stress [80]. The RNA chaperone activities of several CSDPs and RHs have also been determined. Arabidopsis AtCSDP1 comprising 299 amino acids with seven CCHC‑type zinc fingers at the C‑terminus possesses RNA chaperone activity, whereas AtCSDP2 comprising 204 amino acids with two CCHC‑type zinc fingers at the C‑terminus does not have RNA chaperone activity. Domain swapping and deletion experiments have shown that, although the CSD itself harbors RNA chaperone activity, the number and length of zinc finger glycine‑rich domains of CSDPs are crucial to the full activity of the RNA chaperones [78].

1.7. Abiotic stress tolerance studies in Brassica spp.

Brassica juncea (L.) Czern. belongs to the Cruciferae (Brassicaceae) plant family, commonly known as the mustard family. The name crucifer is derived from the shape of the flowers that have four diagonally opposed petals in the form of a cross. *B. juncea* has pale green foliage, with a few hairs on the first leaves and leaf blades that terminate well up the petiole. Mature *B. juncea* plants grow to a height of one to two meters. The lower leaves are deeply lobed, while the upper leaves are narrow and entire. *B. juncea* is distinct from its close relatives *B. napus* and *B. rapa* in that the upper leaves of *B. juncea* are not clasping. The inflorescence is an elongated raceme and the flowers are pale yellow and open progressively upwards from the base of the raceme. The seed pods are slightly apprised and 2.5 to 5 cm in length exclusive of the

beak. The beak is 0.5 to 1 cm long. Seeds are round and can be yellow or brown. Mustard Oil is one of the major edible oils in India, the fixed oil content of rai varying between 28.6% and 45.7%. Oil is also used for hair oil, lubricants and, in Russia, as a substitute for olive oil. Adding 1.1–2.2% mustard oil to fresh apple cider retards fermentation. Seed residue is used as cattle feed and in fertilizers [88]. There are both vegetable and oilseed varieties of *B. juncea* that possibly have different origins [89]. Both types are considered to be natural amphidiploids (AABB genome, 2n=36) of *B. rapa* (AA genome, 2n=20) by *B. nigra* (BB genome, 2n=16) crosses.

Brassica juncea (L.) Czern. is a moderately drought and salinity tolerant oilseed and contains good resistance to pod shattering and blackleg disease [90, 91, 92]. Most abiotic stress studies have focused on the impact and tolerance of plants to individual stresses [93, 94]. Several researchers have examined the interaction of several stresses such as salinity and alkalinity [95], but further knowledge of the genetic mechanisms underpinning the shared and unique responses to these stresses is required for improved breeding for multi-stress tolerance.

The characterization of abiotic stress-responsive genes is essential for elucidating the responsive mechanisms, by which plants can be adapted to the stresses. Recently, transcriptome studies using microarray, 2D electrophoresis for proteome analyses are accelerating to identify massive key genes in developmental and different environmental conditions [96]. For most of the studies, Arabidopsis transcriptome information has been used for Brassica species because both species belong to the mustard family (*Brassicaceae*) and have evolutionary close relationship. The Brassica genome is

generally believed to have triplicated following divergence from Arabidopsis [97, 98] and thus consists of approximately 46,000 genes [99].

Table 1.7.1: Genes cloned and characterized from *Brassica juncea*

Host Plant	Transfor-med plant	Stress	Gene	Protein	Reference
B. juncea	Tobacco	Drought, Salt	*BjDREB1B*	DREB protein	[100]
B. juncea	Tobacco	ABA, ethephon, salicylic acid, and methyl jasmonate, NaCl, Mannitol or PEG	*AnnBj1*	annexin	[101]
Arabidopsis	*B. juncea*	Salt, metal, osmotic	*γ-tocopherol methyl transferase*	Tocopherol	[102]
B. juncea	Rice	Salinity	*gamma-glutamyl-cysteine synthetase (BrECS1 and BrECS2)*	Glutathione	[103]
B. juncea	*B. napus* and *B. carinata*	Salt, oxidative and drought	*PLEA1:BcZF1*	ZF protein	[104]
B. juncea	Tobacco	Methylglyoxal and high salt,	*glyoxalase I*	Glyoxalase I	[105]
B. juncea	Arabidopsis	Salt, heavy metal, Osmoticum	*γ-TMT*		[106]
B. juncea	Tobacco	Dehydration, salt, Heavy metal, Oxidative stress	*AnnBj1*	Annexin	[107]
B. juncea	*Arabidopsis and Tobacco*	Cadmium	*BjCdR15*	BjCdR15	[108]
B. juncea	*S. cerevisiae*	Cadmium	*BjGT1*	Glutathione	[109]

Furthermore, physiologically and morphologically, Brassica species are different from Arabidopsis. These indicate that all Brassica genes might not be represented in the Arabidopsis genome and not be regulated in a same manner of Arabidopsis. However, Arabidopsis transcriptome or proteome may provide some information in the gene regulation of Brassica plants and some of the genes related to abiotic stress tolerance has been cloned and its characteristics has been studied (Table. 1.7.1).

1.8. Subcellular localization of RNA binding protein and abiotic stress tolerance

Proteins are sorted into different cellular compartments such as cytoplasm, nuclear region, mitochondrion, etc. or may be secreted out of the cell, and their proper functioning relies on this precise process of subcellular localization. Subcellular localization of any protein can be detected either sequence prediction methods using different bioinformatics tools or experimental methods like organelle protein fraction or using GFP tagged protein method [110]. Extensive studies on subcellular localization prediction have led to the development of several methods, which can be classified as follows.

1. Amino acid composition-based methods: These methods utilize machine learning techniques, including neural networks [111] and support vector machines (SVM) [112]. This category includes methods like P-CLASSIFIER [113] and CELLO [114], which utilize n-peptide composition-based SVM approaches.

2. Methods that integrate various protein characteristics: Several methods including expert systems [115], k-nearest neighbor [116], SVM [117] support vector data description [118] and Bayesian networks [119], integrate various biological features that influence localization. The features that characterize a protein can be extracted from

biological literature, public databases, and related prediction systems. Both PSORTb [120] and PSLpred [121] integrate different analytical modules and demonstrate that the hybrid approaches perform better than each individual module.

3. Sequence homology-based methods: It has been suggested that protein subcellular localization is an evolutionary conserved trait. Efforts to address the relationship between evolutionary information and localization identity have relied heavily on exploiting sequence similarity to infer PSL. Such methods include phylogenetic profiling [122], domain projection [123] and a sequence homology based method [124]. Several other methods, such as PSORTb and PSLpred, also incorporate such sequence homology-based components in their analyses.

4. GFP tagging approach: GFP (green fluorescent protein) coding sequences are fused at either the 5′ or 3′ end of the coding region of a DNA sequence of interest, leading to production of N-terminal or C-terminal fusions of GFP. Such chimeric genes have been incorporated into stable transgenic plants, or introduced into plant cells for transient assays [110]. One of the advantages of GFP as a fluorescent probe is its lack of a requirement for an exogenous cofactor [125]. GFP can be expressed within intact tissues and processes can be monitored without the interruption caused by introduction of any other chemical reagents. Because of the impermeability of the plant cell to many stains and dyes readily taken up by animal cells, GFP technology has proved to be a more important tool for plant cell biology.

1.9. Early responsive to dehydration 4 (*ERD4*)

Early responsive to dehydration (*ERD*) genes could be rapidly induced to respond to dehydration and to various other abiotic stresses. The maize *ERD* gene (*ZmERD4*)

cloned by rapid amplification of complementary DNA (cDNA) ends. The *ZmERD4* cDNA had a total length of 2,536 bp with an open reading frame of 2,196 bp, 5′-untranslated region (UTR) of 48 bp, and 3′-UTR of 292 bp. The gene encoded a predicted polypeptide of 732 amino acids. A reverse-transcription polymerase chain reaction analysis revealed that *ZmERD4* was constitutively expressed in different tissues. RNA gel blot showed that *ZmERD4* could be induced by both drought and salt stress and also responded to abscisic acid treatment, but it was not induced by low temperature (4°C). Moreover, *35S::ZmERD4* transgenic plants displayed enhanced tolerance to water deficit and high salinity when compared to wild-type plants [126]. In sugarcane in response to abiotic stress treatments i.e. cold, salt and drought stress, changes in the expression levels of ERD4 were evaluated by real-time PCR analysis [127]. The results showed ERD4 induction most strongly during drought stress. Localization of ERD4 was reported in some of the studies like in plasma membrane [128], chloroplast and mitochondria [129]. But no report is available regarding ERD4 actual protein function in plants and there is ambiguity on exact localization of this protein.

In the backdrop of the above, the study was taken up to understand structure function and localization ERD4 as well as to elucidate its role in stress tolerance, with the following objectives:

1. Isolation and cloning of *ERD4* gene from *Brassica juncea*
2. Transcript profiling of *ERD4* gene and its functional validation using knockdown approach
3. Sub-cellular localization of ERD4 protein
4. Functional characterization of ERD4 protein

CHAPTER II:

CLONING AND SUBCELLULAR LOCALIZATION OF *ERD4* FROM *Brassica juncea*

2.1. Introduction:

Advances in molecular biology, genomics, proteomics and metabolomics have provided insight into plant gene regulatory network system, comprising of inducible-genes, expression programming and regulatory elements, corresponding biochemical pathways and diverse signalling molecules [96]. With the rapid progress in genome sequencing projects, plant scientists are stepping up the pace of gene function studies and for this, the gene of interest usually needs to be cloned first. Cloning can be performed either by 'map-based' (positional) or by 'homology based' methods.

Map-based approach: The first step in the map-based or positional cloning is to identify a molecular marker that lies close to the gene of interest [130]. Then the region has to be saturated with other molecular marker. The next step is to screen a large insert genomic library (BAC or YAC) with the closest marker to isolate clones. Once the initial markers that are flanking to the gene of interest identified, then by using chromosome walking gene of interest can be isolated and can further cloned.

Homology based approach: If the genes in two organisms are conserved, nucleic acid probes made from the already isolated gene may be able to hybridize specifically with the target gene [131]. Thus, if the genes from several organisms are available, those from the most closely related should be used. The screening specificity may be enhanced by determining the regions of most conserved sequence. Probes made from those regions have a greater likelihood of hybridizing at reasonable stringencies. Further, degenerate primers can be used for amplification of homologous gene [131].

Once the gene is cloned then its evolutionary history should be depicted to get more information regarding the conservation of the gene. For example, whether the genes

under investigation belong to the members of a single well-defined clade, all members of which appear to descend from a recent common ancestor as a direct result of speciation (orthologous genes), or do the sequences represent one or more ancient duplications (paralogous genes) [132]. A phylogenetic analysis also provides the basis for comparative genomics [133] and helps in predicting the structure and function of the homologous protein.

The eukaryotic cell can be divided into various morphologically and functionally distinct compartments. Proteins must be targeted to their appropriate compartment to ensure proper function. Understanding subcellular localization of protein is importance to help understand how the plant cell is functionally organized. It is necessary to know where enzymes and regulatory proteins are located in a specific plant cell at particular time of development and under particular environmental condition. It determines the environments in which proteins operate. As such, subcellular localization influences protein function by controlling access to and availability of all types of molecular interaction partners. Thus, knowledge of protein localization often plays a significant role in characterizing the cellular function of any hypothetical and newly discovered proteins.

The study of molecular mechanisms on how the final localization site of a protein is recognized and transported (often called protein sorting) is one of the central themes in modern cell biology. The most important principle of protein sorting is that each protein has the information of its final localization site as a part of its amino acid sequence. In many cases, proteins are first synthesized as precursors having an extra stretch of polypeptide which function as a “sorting signal”. They are specifically recognized and transported with some molecular machinery. After they are localized at their final

destination, these sorting signals are often cleaved off. Therefore, it should be possible to predict the subcellular localization site of a protein if we could specifically recognize its sorting signal, as the cellular machinery does. This attempt is still challenging, however, because our knowledge is incomplete and the sorting signal hypothesis allows some exceptions. That is, some proteins do not have sorting signals within their amino acid sequences but they are localized by binding with another protein that has the information.

Plant cells also harbour many proteins whose intracellular location is still unknown. Predictions of intracellular location from DNA sequence information by current computational methods are helpful in this context. Numerous software suites have been released in this field, based on various biological concepts and computational methods. Presently, there are four leading methods those are commonly used. The first uses the overall protein amino acid composition. For example, 'SubLoc' predicts protein localization based on the fact that proteins with different subcellular localizations usually have different amino acid compositions [102]. The second type of method utilizes known targeting sequences. One of the most important principles of the protein sorting mechanism is the existence of a targeting signal in the amino acid sequence that leads proteins to different organelles or out of the cell. Hence, several computational approaches focus on predicting the presence of certain targeting motifs in protein sequences, e.g. signal peptides (SPs), the mitochondrial targeting peptide (mTP), nuclear localization signals (NLS) and transmembrane alpha helices [101, 134, 135]. A third approach uses sequence homology and/or motifs. For example, the Proteome Analyst Subcellular Localization Server (PA-SUB) which utilizes keywords from the protein databases SWISS-PROT and the annotation of homologous proteins [119]. Finally, a

combination of the information obtained from the three categories described above has been used in prediction tools such as WoLF-PSORT (updated version of PSORT II) and the SherLoc2 [136, 137]. Due to their automated and high-throughput nature, computational methods are appealing for the large-scale assignment of protein subcellular locations.

Regardless of the algorithm used, however, computational predictions have always been based on available biological knowledge, which is far from complete. The enormous complexity of the protein sorting process, the existence of alternative transportation pathways and the lack of complete data for every organelle still limit the application of computational methods. To resolve this ambiguity in subcellular localization tradition methods like cell fractionation and protein purification methods can be used but these are technically challenging. Another method is immunofluorescence microscopy but the antibody production for immune-detection of a protein in sectioned tissues can be time consuming and laborious. Fusion of GFP (Green Florescence Protein) coding sequence to the target protein coding region of gene of unknown location is extremely valuable tool for determining where the protein, and its associated biochemical or regulatory processes resides within the plant cell. The use of transient expression through infiltration of *Agrobacterium tumefaciens* (agroinfiltration) harboring the transgene and promoter of interest should substantially decrease the time required to test candidate genes and might p1rovide a better platform to assess the potential of these gene products.

The present study has been carried out on the cloning of *Early Responsive to Dehydration4 (ERD4)* gene from *Brassica juncea* using the homology based cloning

approach. *ERD4* sequence analysis and phylogenetic analysis were performed to understand the gene as well as protein structure. Further, study on subcellular localization was conducted using *in silico* as well as GFP tagging approaches.

2.2. Material and methods:

2.2.1. Plant material:

The seedlings of Brassica (*B. juncea*) cultivar "Pusa Bold" were grown in a plant growth chamber with a 14-h photoperiod at 22±2°C. Twenty one days old seedlings were used for DNA and RNA isolation and subsequent gene cloning.

2.2.2. DNA isolation and PCR amplification:

All the DNA manipulations were carried out as described by Sambrook et al. [138]. All the DNA modifying enzymes including restriction enzymes were purchased from New England Biolabs, MA, USA. DNA ligase, Taq polymerase and high fidelity polymerase were purchased from (Invitrogen, CA, USA). DNA was isolated from *B. juncea* cv pusa bold leaf tissues using CTAB method [138]. Primers for full length *ERD4* gene amplification were designed from Arabidopsis *ERD4* gene sequence (AT1G69450) (NC_003070.9|:10715665-10718997 *Arabidopsis thaliana* chromosome 1) (forward primer: AGCCTGTCTTCTCTCACCTG, reverse primer: GTTACAACAACCATGGAA CACATAC). The PCR programme used as follows: 94°C for 5 minutes, 35 cycles of denaturation at 94°C for 1 min, annealing at 55°C for 1 min, elongation at 72°C for 1 min followed by 10 min extension at 72°C.

2.2.3. Full length *ERD4*cDNA cloning from *Brassica juncea*:

Total RNA and mRNA was isolated from 1 hr PEG (20%) treated *Brassica* plant root. cDNA was prepared using affinity script reverse transcriptase enzyme. Based on

BjERD4 genomic sequence, primers for full length (including initiation and termination codons) (Forward: GATCACGAATTCATGGAGTTTGCATCGTTTC, reverse: ACTGATGGTACCTTAA GCAGCTGATATTGC) were designed. Amplification, cloning and sequencing of the full length of cDNA of *BjERD4* were done as given below.

2.2.4. Ligation and transformation:

PCR product was ligated with TA cloning vector (pTZ57R/T, supplied by Fermentas) as follows: 3 µl (0.15 µg) vector, 7 µl (0.54 µg) gel purified PCR product, 3 µl of 5X ligation buffer, 1 µl of ATP, and 1 µl of T4 DNA ligase (Fermentas) were mixed and incubated at16 °C for overnight.

Ligated product was transformed into $CaCl_2$ competent *E.coliDH5α* cells. Ligated product was added to 100 µl of competent cells and incubated in ice for 30 min, followed by 42°C for 1.5 min and again kept on ice for 5 min.900 µl of LB medium was added to the cells and kept for shaking at 200rpm for 1hr at 37°C. The cells were spread on LA plates containing carbenicillin (100mg/liter) and incubated at 37°C for overnight.

Transformed colonies were used to start overnight cultures. 5ml of LB with antibiotic were inoculated with single colonies and incubated overnight at 37°C under shaking condition at 200 rpm. Plasmid DNA was isolated using the plasmid DNA purification kit. Plasmids were checked for the insert by digesting with restriction enzymes *KpnI* and *EcoRI* and digested products were analysed on 1% agarose gel. Plasmid containing insert were sequenced.

2.2.5. Sequence based analyses:

BLAST search was performed to find the homology of cloned insert. Comparative alignment analysis of the genomic sequence and the cDNA sequence of *BjERD4* gene

was performed. Gene structure analysis was performed using popular gene finding pipeline (FGENESH at www.softberry.com). Motifs were identified using motif scan tools from annotated protein sequence [139].

2.2.6. Phylogenetic analysis:

The homologs of *B. juncea* ERD4 protein sharing more than 40% sequence identity were obtained from UniProt database using FASTA search engine. The search for ERD4 homologs using BLAST search engine was carried out also against the non-redundant protein sequences and against translated individual proteome of *C. reinhardtii*, *C. merolae*, several fungi and cyanobacterial (*Synechococcus* sp. RS9916, *Cyanothece* sp., *Nostoc punctiforme*) genomes. To detect ERD4-like proteins in animals, BLAST search against non-redundant protein sequences of animalia (taxid: 33208) kingdom was also carried out. Since complete proteome database for *Triticum aestivum* is yet not available, the search for its homolog was carried out in Ensembl [140] employing tBLASTn search engine [141]. The search of distantly related genomes or those of unrelated species was constrained for the presence of two tandem RNA-recognition motifs and a DUF221 domain detected in the closely related plant species. Multiple sequence analyses were carried out using clustalW and PROMALS3D tools [142]. The phylogenetic tree was derived from the multiple alignment using Neighbor-Joining method in MEGA4 [143].

2.2.7. Topology prediction:

Secondary structure of the plant ERD4 orthologs was predicted using PsiPred [144] and Prof (http://www.aber.ac.uk/,phiwww/prof/) suites. The web-versions of nine different topology prediction methods were used to estimate membrane topology of

ERD4 and these were: DAS [145], HMMTOP [146], MEMSAT [147], TMHMM [148], TMMod [149], TMpred [150], Conpred [151] and phobias [152]. Modeling of transmembrane topology was done using TOPO2 (http://www. sacs.ucsf.edu/TOPO-run/topoanal-adv2.pl).

2.2.8. Localization prediction using *in silico* approach:

The prediction for sub-cellular localization of the *B. juncea* ERD4 protein and its orthologs was done using web-tools: wolf PSORT [136], YLoc [137], TargetP [153] and an ambiguous targeting predictor [154]. Further a subset was analyzed consisting of 123 chloroplastic envelope proteins of *A. thaliana* chloroplast proteome for their chloroplast localization signatures [155]. These proteins were identified from the experimentally validated chloroplast envelope protein dataset; these were not showing any similarity with ribosomal proteins. Amino acid content of the complete protein, N-terminal sixteen and sixty amino acid residues of this subset of validated chloroplastic proteins and for plant ERD4 proteins were calculated.

2.2.9. Preparation of plasmid constructs

A. **Construction of ERD4-GFP transformation vector**:

An intermediate vector, pART7 harboring CaMV 35S promoter, a multiple cloning site and *ocs* terminator was used as transformation vector. The GFP coding sequence was amplified from pCAMBIA 1302 vector using gfp-F primers (GATGCTGGTACCATGGTAGATCTGACTAGTAA AG, having a *KpnI* restriction site and a start codon) and gfp-R primer (ACTACGAA GCTTTCAGCTAGCTTTGTATAG TTCATC, having a *HindIII* restriction site). The GFP PCR amplified product was then digested with *KpnI* and *HindIII* and ligated with the *KpnI* and *HindIII* digested sites of

the pART7 vector to form the intermediate vector. The full length *BjERD4* gene was amplified from the cDNA using primers ERD4F (GATCACGAATTCATGGAGTTTG CATCGTTTC with *EcoRI* restriction site) and ERD4 R2 (ACTGATGGTACCAGCAG CTGATATTGCAGC *KpnI* restriction site without stop codon). The *ERD4* PCR amplified product was digested with *EcoRI* and *KpnI*, sub cloned in frame into the N-terminal side of GFP at the *EcoRI* and *KpnI* site in the intermediate vector pART7-GFP. This resulted in another intermediate vector which was named pART7-BjERD4-GFP. The correct orientation of the inserts in the vector was confirmed by restriction digestion analysis and PCR analysis using *GFP* and *ERD4* specific primers and sequenced to ensure insert accuracy. The CaMV- BjERD4-GFP-OCS fragment was then released from pART-7-BjERD4-GFP by *NotI* digestion and subcloned into *NotI* digested pART27 plant expression vector. The resulting transformation vector pART27-BjERD4-GFP was used to transform *Agrobacterium tumefaciens* EHA105.

B. Preparation of GFP overexpression construct:

The *GFP* gene was amplified from pCAMBIA 1302 and cloned in pART7 and then subcloned in pART27 plant expression vector which was further used as positive control for subcellular localization.

2.2.10. Transformation of *Agrobacterium tumefaciens* EHA 105:

Agrobacterium tumefaciens EHA105 was streaked from a glycerol stock onto a LB plate containing appropriate antibiotics (rifampicin and chloramphenicol, 100 µg/ml and 50 µg/ml respectively). One individual colony was inoculated into 50 ml LB containing the above antibiotics and grown overnight at 30 ºC. The cells were harvested and resuspended in 1.5 ml LB. To an aliquot of 200 µL of the resuspended cells, 10 µg of

the transformation vector DNA was added and immediately frozen in liquid nitrogen. The tubes were thawed in a 37 ºC water bath for 5 minutes. One ml LB was added to each tube and incubated at 30 ºC for 2 hours and then plated on LB plates (containing rifampicin 100 µg/ml, kanamycin 50 µg/ml). Transformed colonies appeared in 2-3 days after the plates were incubated at 28 ºC.

2.2.11. Agro-infiltration in tobacco:

Agrobacterium was cultured to stationary phase (overnight grown culture) in at 26°C. Cells were pelleted down by centrifugation at 6000 rpm for 5 minutes in a microcentrifuge at room temperature. The pellet was resuspended in 1 ml of the infiltration buffer and centrifuged again twice. The bacterial suspension was diluted with infiltration buffer to adjust the inoculum density (OD600 - 0.1). Eight-week old tobacco plants were used for infiltration. Using a yellow micropipette tip, small holes were created in the epidermal layer of leaf. The nozzle of a 1 ml syringe (without needle) was pressed against the lower (abaxial) epidermis of a tobacco leaf, covering the small hole with the nozzle and holding a gloved finger to the other side of the leaf. Infiltrated area was marked. Plants were incubated under normal growing conditions for 2-3 days. A marked area was excised and examined under the florescence microscope. A Carl Zeiss Axioplan I microscope and the axiovision software (Carl Zeiss, Oberkochen, Germany) were used for visualization and documentation of tobacco infiltered leaves.

2.2.12. Protoplast isolation:

Transformed tobacco leaves were cut together into approximately 0.5 mm strips using sharp razors. Leaf pieces were laid in Petri dishes containing 5 ml of the protoplast isolation enzyme mixture (0.125% (w/v) Macerozyme R-10, 0.2% (w/v) Cellulase R-10

in 5 mM $CaCl_2$, 0.5 M sucrose, 0.1 % BSA, 2.5 M MES-HCl, pH 5.2) for 19 h in the dark at 25°C. After digestion, undigested pieces of leaves were removed and 4 ml of floating MLO6 medium (15 mM $CaCl_2$, 600 mM sucrose, 7.5 mM MES-KOH pH 6.0) was added. The protoplast suspension was then filtered through a 100 µm nylon filter and centrifuged at 110 g for 7 min in a swinging rotor. The protoplasts localized in the floating band were harvested and diluted with four volumes of autoclaved washing W5 medium (154 mM NaCl, 125 mM $CaCl_2$, 5 mM KCl, 5 mM glucose and 1.5 mM MES-KOH, pH 5.6). The cells were then pelleted (110 g for 7 min in a swinging rotor) and washed in 40 ml and then in 20 ml of "mannitol /Mg" solution (15 mM $MgCl_2$, 400 mM mannitol, 5 mM MES-KOH pH 5.6). Protoplasts were finally resuspended in Mannitol/Mg solution. Purified protoplast was treated with mitotraker dye (50nM) to see the mitochondrial localization. Protoplast was treated with 50nM mitotracker dye for 20 min and then rinsed three times with PBS buffer (137mMNaCl, 1.4mM KH_2PO_4 , 4.3mM Na_2HPO_4, 2.7mM KCl, pH-7.4)

2.2.13. Confocal microscopy analysis:

Transiently transformed epidermal peel cells and protoplasts were analyzed in TCS SP5 laser scanning confocal microscope (Leica, Germany) to detect the GFP fluorescence at 500-535 nm wave length after excitation at 488 nm. The mitotracker signals were detected at 600-630 nm laser band width excitation at 543 nm. Autofluorescence of plastids was detected at 650 to 720 nm. For the co-localization experiments, sequential scanning was done for both the channels and then merged/superimposed together to shows overlapping signals. All the images were further processed using Leica LAS AF Lite software.

2.3. Results

2.3.1. PCR amplification and cloning:

RNA free genomic DNA and total RNA were isolated from leaf samples and the concentration was determined with spectrophotometer readings. Concentration was found to be 300 ng/µl and 800 ng/µl for DNA and RNA respectively. To check the integrity of DNA and RNA, gel electrophoresis was performed which showed intact high molecular band in case of DNA and intact 18S and 28S rRNA band along with smear in case of total RNA (Fig. 2.3.1A&B). Total RNA was further used for PCR to check the presence of DNA and the result showed no amplification product when PCR was performed up to 40 cycles. cDNA synthesis was performed using gene specific reverse primer. PCR amplification was performed with genomic and cDNA and results showed the presence of 3.1 KB and 2kb band for genomic and cDNA respectively (Fig. 2.3.1C&D). These bands were eluted from the gel and cloned into TA cloning vector. After confirmation, clones (Fig. 2.3.1E &F) were sent for sequencing.

2.3.2. Sequence analysis:

The sequences were used for blast search to see the homology identification with other sequences. BLASTn result showed the 81% identity with the *Arabidopsis ERD4* (At1G30360). Comparative alignment analysis of the sequence of amplified product from *B. juncea* genomic (3291 bp) and cDNA (2172 bp) showed the presence of six exons and five introns. All introns displayed distinctive characteristics of plant introns, being relatively AT rich and having conserved 5'-GT and AG-3' splicing sites (Fig. 2.3.2A & B). The *BjERD4*cDNA encodes a protein of 723 amino acid residues with the predicted molecular weight of 80.93KDa and a theoretical PI of 9.23.

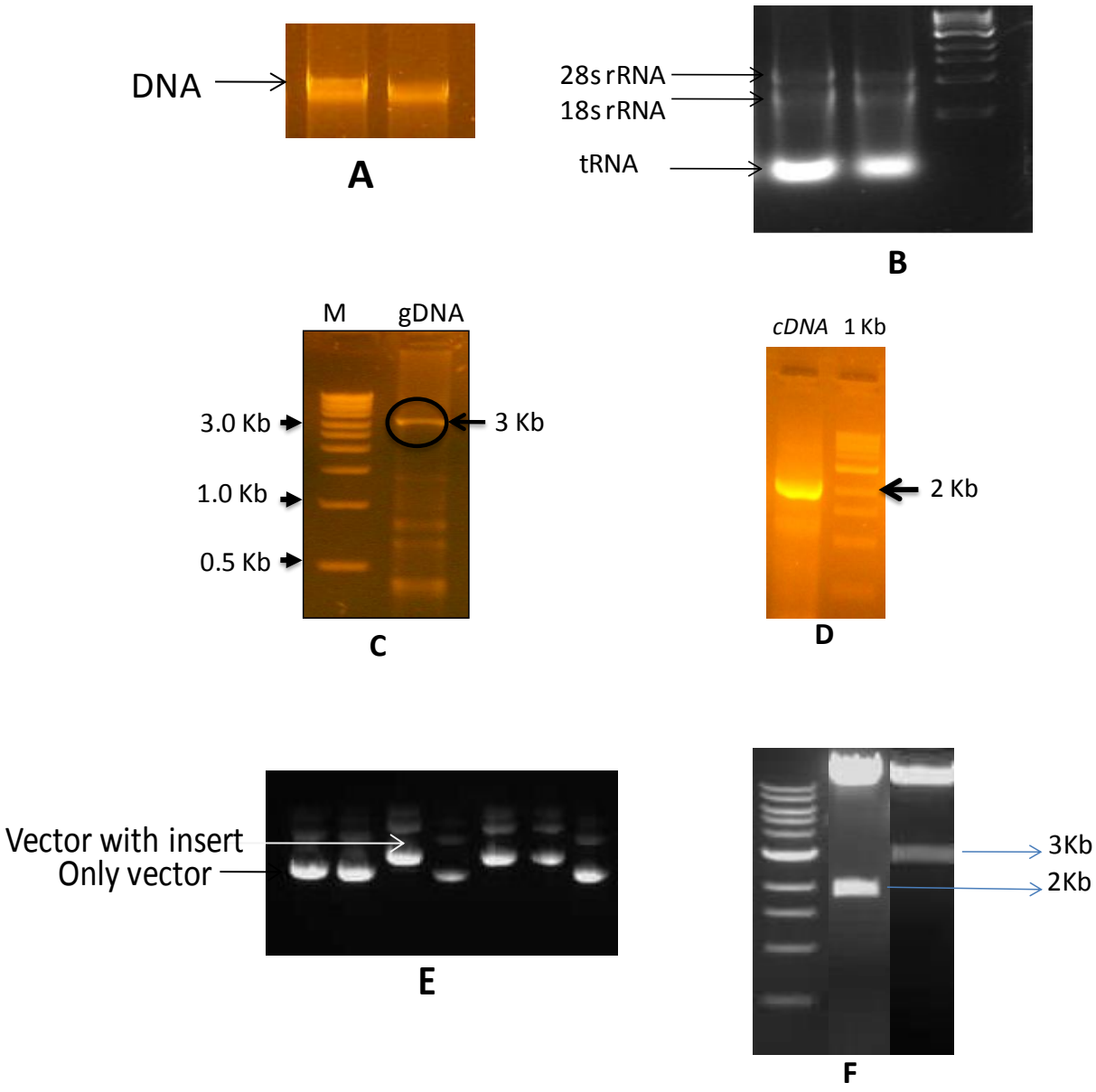


Fig.2.3.1: Amplification and cloning of *ERD4*. *B. juncea* genomic DNA (A), total RNA (B), PCR amplified product from gDNA (C) and cDNA (D), Different clones(E) and confirmation of clones with restriction digestion (F).

tagcctgtcttctctcacctgcaatttaccagaagaagcaagaacttacaaacggttactttccctttttttaaaacggcaactccttttcactttcccctcat
cttaaatcttccccagaaactttgagtaacagagagagagagaattgtagATGgagtttgcatcgtttcttgtgtccttagggacatcagcaattatcttcgtcg
ttctcatgttcctcttcacctggctctctcgtagacccggtaacgtccccgtttactacccgaatcggatcctcaaaggtatggatccatgggaaggcagctcct
tgacccgaaacccattcgcctggatccgtgaagctttcacttccaccgagcaagacgtcgttaagctatccggcgtcgatactgccgtctacttcgtcttccaga
gcactGTtcgcttcttcttctctctcttctgatctgggtttctaattttacctaataaagtttccatctttggatacaattggggagaaaaaggtttccttttg
aagatttttgtttgtatattgttggggtctgaagatcttgtttgctatgttatgtcactttgtccttatctgctttaaataaagtacctttttgagttttgggt
ctagacacttgagatcgaaattagaaccctagaccatgttttgcattgataactttggtttaaggtacatgagcgtttatgtacaccttatagcatcttcggttt
ggtttaagttattagtgatcttcggtgtatcttcttctgtatgttcttgtttgactgttgctcggaccatctatataatgagcctatgatgctggtctagccaaa
aagtttctttctttggatatgttcagtgtcttcttcgttgctcaacataacatgacaaaagtgttcctgattgttttttttgttttgttttgctttgaaacAGtt
ctggggatatttgctttatccgctctacttcttttaccgactctactcccaatagccgctacagacaacaacttagagacttcaaggagcgccactgacaccca
agcaacggaacctttagccaacttgataacctatcaatggctaacatcaccGTaagtagttttaaacacacataacttcattcagataccaccaaggtcctgatc
ctttctgttttttaacAGaaaagtagttcaaggctgtgggcgttcctaggagcggtttactgggtatccgtagtcacatatttcatgttgtggaaagcttacaag
catgtggctgcattgagagctcaagctctgatgactagcgaagaagtattaccggagcagttcgctattctcgttagggacataccttccccacctaacggggag
actcagaaagagtttgtagattcttacttcagagatatctaccctgagacattctacaggtcgcttgttgtaacagaaaacagcaagGTaatctatctatctacc
tgcttgttctgtcttatatagtttcagtctttttaaccactcgatggctcatgttctgtAGattaataagatatgggaagacctggaaggttacaagaagaagc
tcgcgcgtgcagaagcagcgtttgcagccactagcaaccgaccaacgaacaaaaccggcttgcttgggctagtcggagagcgagtagacagcatcgattattaca
cgaagctaatcaacgagtcggtagccaaactagaagcagagcagagaacggttcttgctgagaggcagcaaaccgcagcggttgtgttcttcacagacagagtta
ctgctgctttagccgctcagtctcttcactgccagatggttgataaatggacagtgaccgaagcacctgagcctcgccagctcatctgggagaatctcaagatca
agttcttcagtagaatagtcagacagtacgtgatctacttcctcgttgccataaccattctgttctacatgatccccatcgctttcgtctctgcaatcaccactc
ttgctaatctccagaaggctcttcctttccttaagccgattgtggatattgccttcataagaaccatcttggagtcttaccttcctcagattgcactcatcgtct
tcttggctatgctgcctaagttcctcatgtttctctccaaatcagaggggattccttcgcagagccatgccattagagcaacctccgggaagtacttttacttct
cggtcttgaatgtcttcatcggtgtaacccttgctgggtctttgttcgagaatttgaaggctcttgaagagaaaccaaactctttcattacccttttggctacta
gtctccctaagagtgctactttcttcttgacctacgttgctctcaaGTgagtttcccctcctccctcttcacttacataacatatacacaagcatctgttctcta
aaatcgtatgttactttcAGgttctttgttggttatggccttgagctgtctcggatcatacctttgataatattccatctgaaaaagaagtatctatgcaaaact
gaagcagaggtcaaagaagcttggtatccaggagacttaagctacgcgacaagggttcccagcgacatgctcatcctcactatcaccttctgctactccgtaatc
gctcctctaatcctcgtattcggtgttatctactttggcttaggatggctcatcctgaggaatcagGTcagactttctccctagttatgtttaagtacacgtttt
caggtgatcttgatctgtgcttttaacggtttatgacAGgctctgaaagtgtacgtaccatcatacgagagctacgggaggatgtggccgcatattcacacccgc
attctagcggcattgtttctgttccagctggttatgtttggttacttgggagtcaagatattcgtttgggccattctcttggttccactcatcttcatctccctc
atcttcggttacgtgtgccgccagaaattctacggagggttcgaacacacggctctagaggtggcttgccgtgaactgaagcagaggcctgacctcgaggaggtt
ttcagagcgtacatcccgcatagcttgagcactcacaaaggagacgatcaccagtttaaaggagctatgtctcgttatcaggactatgctgcaatatcagctgct
TAAgatttgaatctcttatctttaattatctcacagatttgttgtttgtttttttttagaagcttttattattgttgtgcgtgtgattctagggtttgtttagggt
ttttcgtctgtatgtgttccatggttgttgtaaca

A

Gene structure

B

Fig. 2.3.2: (A) Cloned product gene sequence. Initiation and termination codon has been depicted in capital red, inter sequence are in small red letters. (B) gene structure of *ERD4* (exon in blue and introns in yellow color).

2.3.3. Phylogenetic analyses:

The homologs of *B. juncea* ERD4 protein were identified in various plant lineages, for instance in bryophyta (*Physcomitrella patens*), in traceaophyta (*Selaginella moellendorffii*), and in euphylophyta (*Oryza sativa*, *Arabidopsis thaliana*). The protein was found to be conserved in all the plants for which proteome data was available (Fig. 2.3.3). Phylogenetic tree of plant ERD4 homologs showed four distinct clades and the evolution pattern of this gene followed the lineages evolution (Fig. 2.3.3). The presence of both putative RNA-binding and DUF221 domains, a characteristic of plant ERD4 proteins, was also detected in unicellular (*C. reinhardtii*) and multicellular (*V. carteri*) green algae genomes by iterative PSIBLAST search.

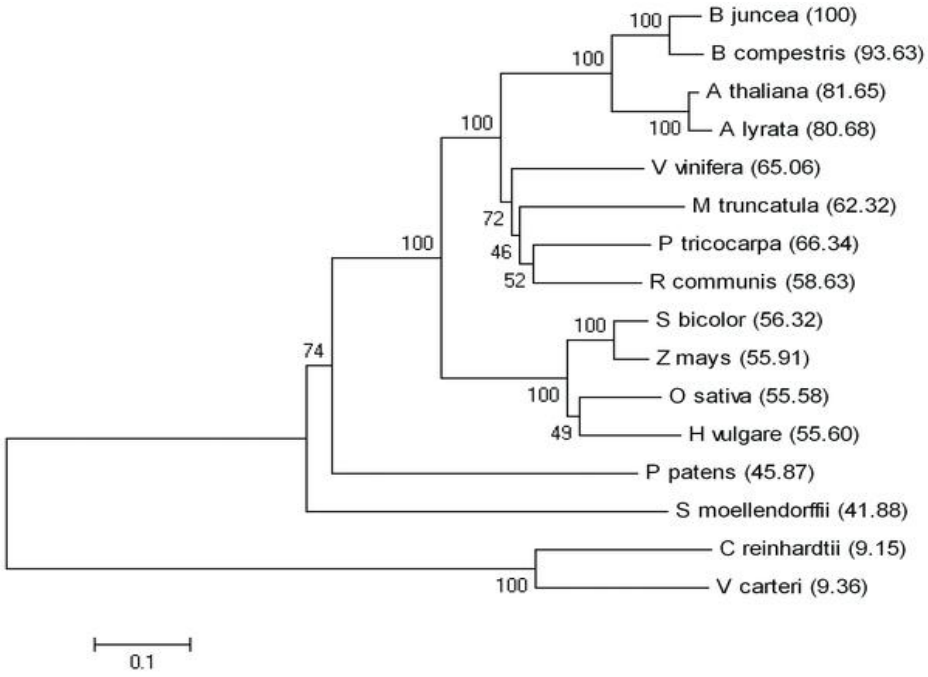


Fig. 2.3.3.Evolutionary relationship among ERD4 homologs. Evolutionary relationship was inferred using the Neighbor-Joining method in MEGA4 software. The percentage of replicate trees in which the associated taxa clustered together in the bootstrap test (100 replicates) is shown next to the branches. The tree is drawn to scale with branch lengths in the same units as those of the evolutionary distances used to infer the phylogenetic tree. The evolutionary distances are in the units of the number of amino acid substitutions per site. Also shown in brackets are the pair-wise percentage identity between *B. juncea* ERD4 and other plant proteins, including green algae.

The algal proteins, however, consisted of 1746 and 1172 residues, respectively (UniProtKB, A8HT24 and D8TSA1). However, homologs of plant ERD4, possessing both the RRM and DUF221 domains, were not detected in bacteria (including cyanobacteria) and archae. Counter intuitively, ERD4-like proteins were detected in unicellular non-photosynthetic eukaryotes like *Dictyostelium fasciculatum* (slime mould) and colonial flagellates like Choanoflagellates. These proteins showed 24.5% (52.7%) and 19% (40%) sequence identity (similarity), respectively, with *B. juncea* ERD4 protein over the complete length. In this study, proteins possessing both the RNA-binding and DUF221 domains were detected in fungi including many plant pathogens (for instance, in *Phytophthora sojae*) and in animals. A *Homo sapien* ortholog of the identified animal proteins has recently been characterized as ‘‘transmembrane protein 63A’’ (UniProt/KB, O94886; TM63A_- human). The human protein consisted of 807 amino acid residues and showed 24% (41%) sequence identity (similarity) over 608 residues with *B. juncea* ERD4 protein (Fig. S1). The motif scanning (motif_scan) and domain detection tools (Pfam, DOUTfinder and SMART) detected presence of DUF221 domain (residues 312–634) in the ERD4 sequence with very high confidence (E-value, 7e-146).

2.3.4. Transmembrane topology:

Transmembrane helices in the ERD4 sequence were identified using several web-servers although with some differences. The number of identified helices varied from 9 to 11 and the suggested starting and end- points for predicted transmembrane segments also differed (Table 2.3.1). Based on high-confidence predictions from different servers, nine transmembrane helices belonging to the sequence regions of 6–26, 90–111, 149–167,

365–385, 419–437, 457–476, 501–531, 573–593 and 638–659 were identified (Fig. 2.3.4A).

The identification of the transmembrane helices was consistent with the predicted secondary structure which suggested that the ERD4 protein is mainly helical with 64.3, 5.4 and 30.3% residues in helix, extended and coil structures, respectively. Interestingly, all the transmembrane prediction tools showed that a long polypeptide segment (residues 170–360) did not possess transmembrane helices (non-transmembrane segment). A globular domain was subsequently detected in this segment.

Table 2.3.1: Transmembrane helices in the ERD4 sequence were identified using several web-servers

Prediction server	Orientation Predicted N-terminus	Orientation Predicted C-terminus	No. of helices
DAS	--	--	11
HMMTOP	Inside	Outside	9
MEMSAT	Outside	Inside	9
TMHMM	Inside	Outside	9
TMMod	Inside	Inside	10
TMpred	Outside	Inside	9
ConPredII	Outside	Inside	11
Phobius	Inside	Inside	8

The inside or outside localization of the non-transmembrane fragment (inside or outside the chloroplast membrane) depended upon the orientation of N-terminal transmembrane helix. While MEMSAT and TMpred showed its placement inside the

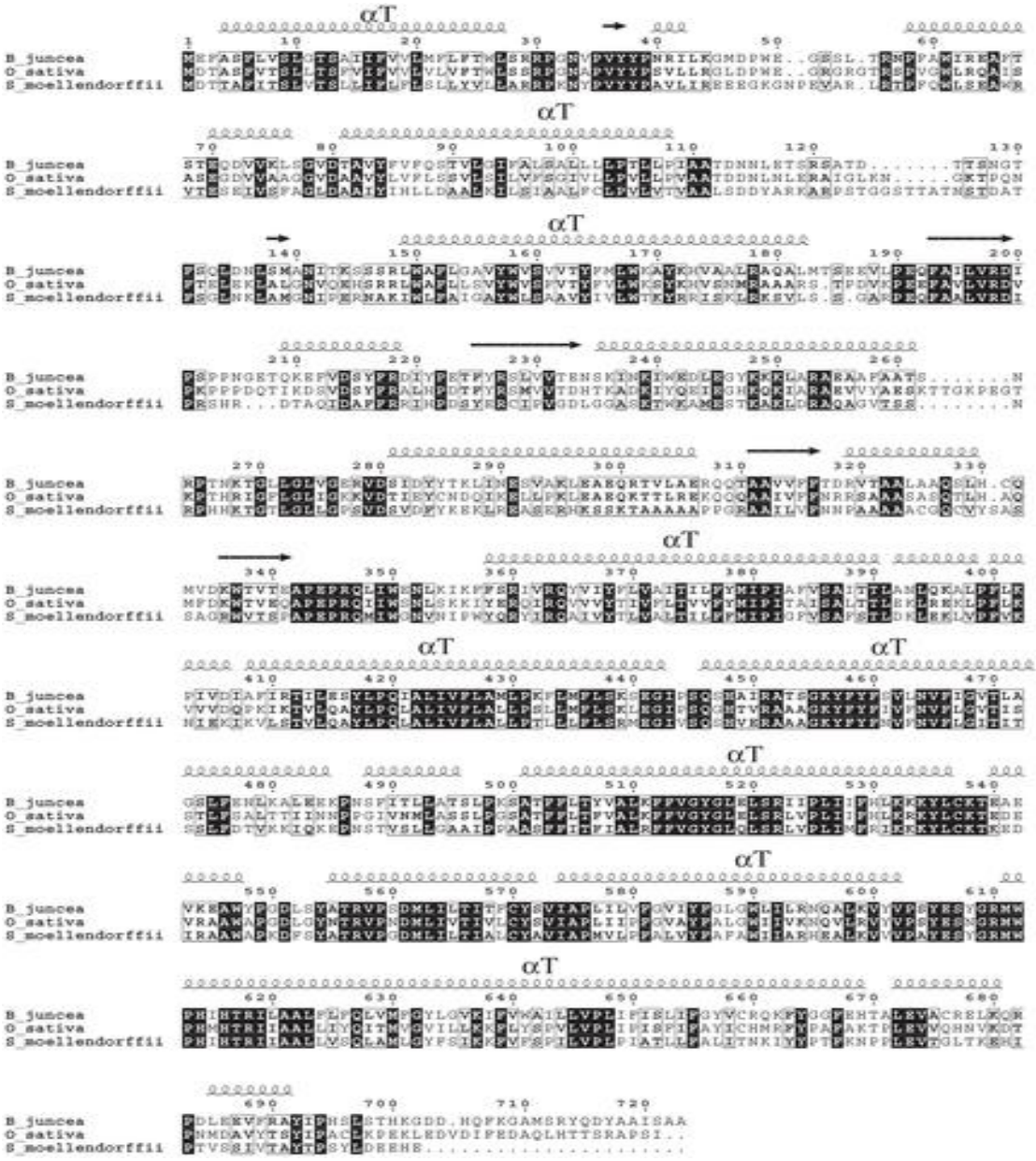


Fig. 2.3.4A. Multiple sequence alignment of plant ERD4 sequences. The alignment of all available plant ERD4 sequences was achieved using PROMALS3D and only three diverse sequences are shown here. Also shown is the consensus secondary structure predicted by PsiPred; helices are shown as coils and strands are shown as arrows. The nine transmembrane helices are marked as αT. The strictly conserved residues in all the plant ERD4 sequences are shaded, while similar residues are boxed. The residues numbering is of the full-length *B. juncea* ERD4 protein. The figure was prepared with EsPript suite.

membrane, several other tools like HMMTOP, TMHMM, TMMod predicted its presence outside the membrane. These predictions resulted in two distinct membrane topologies (Fig 2.3.4B). To resolve this ambiguity, frequency of the positively charged residues in both the possible topologies was calculated. Results showed N-terminus of ERD4 was outside the membrane as nearly 79% of the positively charged residues were observed to reside on inside loops and vice versa. The corresponding transmembrane topology model revealed presence of the non-transmembrane segment (residues 170–360) (Fig. 2.3.4B).

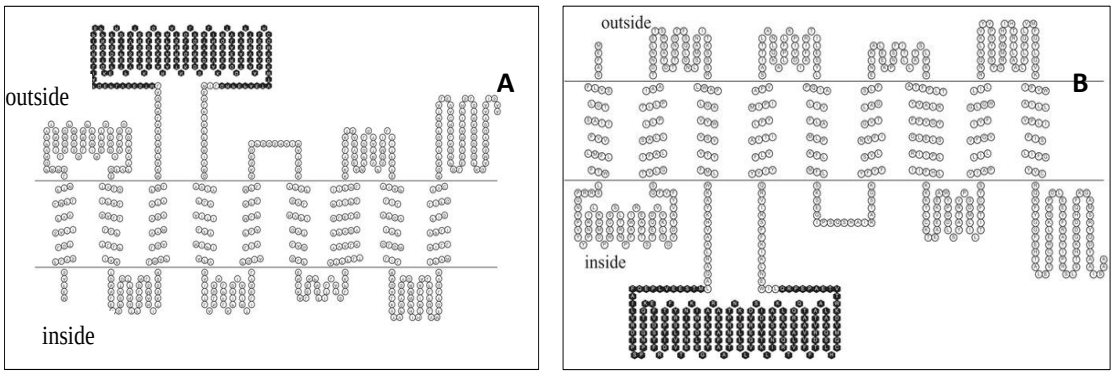


Fig. 2.3.4 B. The topology of the *B. juncea* ERD4 protein: The toplogy was drawn using TOPO2 tools. The nine transmembrane helices are shown. Also, shown (filled hexagons) is the globular domain. The globular domain was suggested to reside outside (A) and inside (B) depending on N-terminal position.

2.3.5. Prediction of ERD4 Subcellular localization using bioinformatics tools

Maximum probability of localization of this protein was predicted in plasma membrane (with score of 10) followed by chloroplast (score 2) using Wolf PSORT tool. The YLoc tool, however, suggested its presence in chloroplast with 53.9% probability and a small confidence (0.27). The TargetP server predicted this protein to be a secretory protein with high confidence (score 0.92) (Table 2.3.2). The analysis of *B. juncea* ERD4 by the ambiguous targeting predictor (ATP) suggested a score of 0.39, which weakly suggested dual targeting of the ERD4 protein. The analysis of ERD4 orthologs by the

ambiguous targeting predictor, however, suggested wide variations in the confidence score (Table 2.3.3) with a low score of 0.19 for some ERD4 proteins that clearly indicated localization of ERD4 in only one compartment (Table 2.3.3).

Table 2.3.2: Prediction of ERD4 subcellular localization using different web based tools

Prediction server	Localization (score)
Wolf PSORT	Plasma membrane (10), Chloroplast (2)
Yloc tool	Chloroplast (53.9% probability)
Ambiguous target predictor (ATP)	Dual target (Mitochondria/Chloroplast) (0.39 for BjERD4), (0.19 for other orthologs)

In order to get detailed information on the amino acid composition of presequences for chloroplast envelope targeting, chloroplastic envelope proteins of A. thaliana were analyzed and experimentally validated. An overall amino acid composition and N-terminal sequence logo plots of the 123 selected proteins (ENV dataset) from Arabidopsis proteome were analyzed [155]. The positional abundance of amino acids in sequence logos showed abundance of serine residues and underrepresentation of arginine residues in the ENV dataset. However, no clear position-specific pattern was observed in sequence logo plots. The amino acid composition analysis also showed much higher abundance of Ser, Ala and Leu residues in the N-terminal sixteen residues as compared to the full-length proteins (Fig. 2.3.5A). Also, the percentage of Arg residues in the N-terminal sixteen residues was observed to be lower than thatobserved in full-length or N-terminal sixty residues.

Table 2.3.3: Prediction scores for dual organelle targeting of plant ERD4 proteins assessed by ambiguous targeting predictor (ATS)

Plant species	Accession code	Source	APS prediction score
Brassica juncea	A9LIW2	UniProtKB	0.39122
Brassica campestris	A8IXK5	UniProtKB	0.39122
Arabidopsis thaliana	Q9C8G5	UniProtKB	0.19248
Arabidopsis lyrata	D7KET4	UniProtKB	0.19248
Populus tricocarpa	B9GJG0	UniProtKB	0.39122
Sorghum bicolor	C5X9J3	UniProtKB	0.47346
Vitis vinifera	F6HLU8	UniProtKB	0.30121
Oryza sativa	Q6ZLQ0	UniProtKB	0.34804
Zea mays	B0FSL2	UniProtKB	0.47346
Medicago truncatula	AES64128	GenBank	0.20827
Ricinus communis	B9SY14	UniProtKB	0.39122
Hordeum vulgare	F2DDW1	UniProtKB	0.34804
Physcomitrella patens	A9TEC4	UniProtKB	0.41759
Selagilella moellendorffii	D8STJ2	UniProtKB	0.29168
Chlamydomomas reinhardtii	A8HT24	UniProtKB	0.49063
Volvox carteri	D8TSA1	UniProtKB	0.21542

The analysis of the N-terminal sixteen residues of the ERD4 orthologs also showed similar trend of higher abundance of potentially hydroxylated Ser/Thr residues and of hydrophobic Phe/Ile residues. The N-terminal sixteen residues also showed high

differences in the abundance of Arg and Lys residues, as compared to the N-terminal sixty and overall composition of these proteins. These positively charged residues are underrepresented in the N-terminal sixteen residues of the ERD4 orthologs (Fig. 2.3.5B). The lower abundance of Arg and Lys residues in the N-terminal sixteen residues of chloroplast proteins, compared to mitochondrial proteins, has been earlier observed by [156]. The low percentages of the positively charged Arg/Lys residues and significantly higher percentage of Ser residues in the N-terminal sixteen residues of ERD4 proteins thus corroborated experimental determination of the ERD4 protein in *A. thaliana* chloroplast envelope proteome.

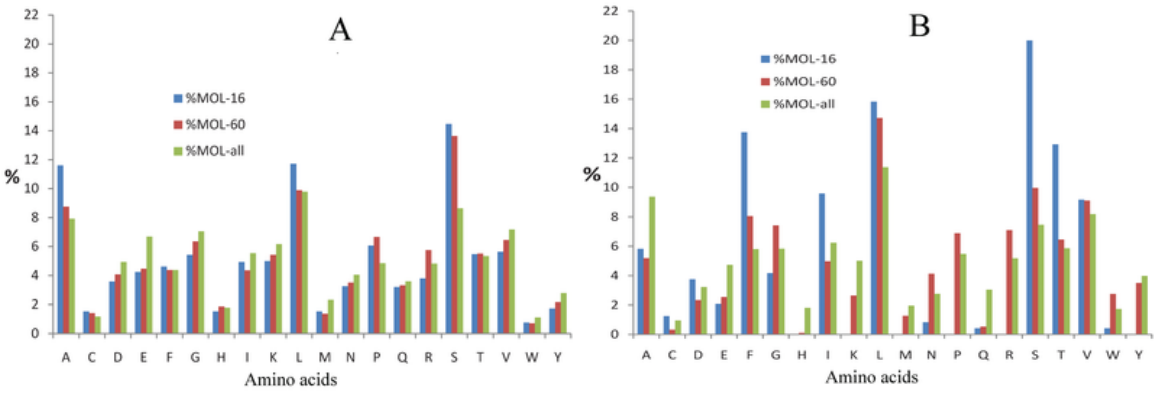


Fig. 2.3.5. Amino acid composition of presequences. Analysis of the amino acid composition of the N-terminal sixteen residues (%MOL-16), N-terminal sixty residues (%MOL-60) and full-length protein (%MOL-all) (A) analysis of the 123 chloroplast envelope proteins of *A. thaliana* (B) analysis of plant ERD4 orthologs.

2.3.6. Preparation of plasmid constructs

The GFP protein is widely used as a reporter protein in localization of proteins at subcellular level in plants. For this study, sequence for *ERD4* gene and *gfp* gene were

amplified from *Brassica juncea* cDNA and pCAMBIA vector respectively and construct was prepared by cloning in pART7 vector sequentially. Subcloning was further done in plant expression vector pART27 and construct named as pART27-ERD4-gfp (size approx. 12.2 Kb). The vector was further double digested with *EcoRI* and *KpnI* and ligated after blunting for generation of a construct named as pART27-GFP (size approx. 11.5 Kb), which was used as positive control (Fig. 2.3.6A).

Construct size and orientation were further confirmed with restriction digestion analysis. *NotI* digestion was performed with pART27, pART27-ERD4-GFP and pART27-GFP vector. Gel electrophoresis results showed the release of 3 Kb and 5 Kb fragment from pART27-ERD4-GFP and pART27-GFP vector respectively. For confirmation of in-frame cloning of *ERD4* and *gfp* genes, restriction analysis was performed with *ClaI*. Electrophoresis results showed only linearization in pART27-GFP vector and release of 2 KB fragment from pART27-ERD4-GFP, which was found to be in line with *in silico* restriction analysis data of these construct (Fig. 2.3.6B).

2.3.7. Subcellular localization using tobacco infiltration:

To determine the localization of ERD4, the tobacco leaf infiltration method was used. *Agrobacterium* carrying above mentioned construct was infiltered into *N. benthamiana* leaves. After three days of Agro-infiltration leaf discs were cut and observed under fluorescence microscope.

Protoplasts were isolated from the leaves that showed green florescence (Fig. 3.3.7A). Confocal microscopy analysis was performed with transformed protoplasts. Transient expression of GFP was detected in both the constructs i.e. only *GFP* and *ERD4* tagged with *GFP*. Florescence microscopy of transformed protoplasts revealed that *ERD4*

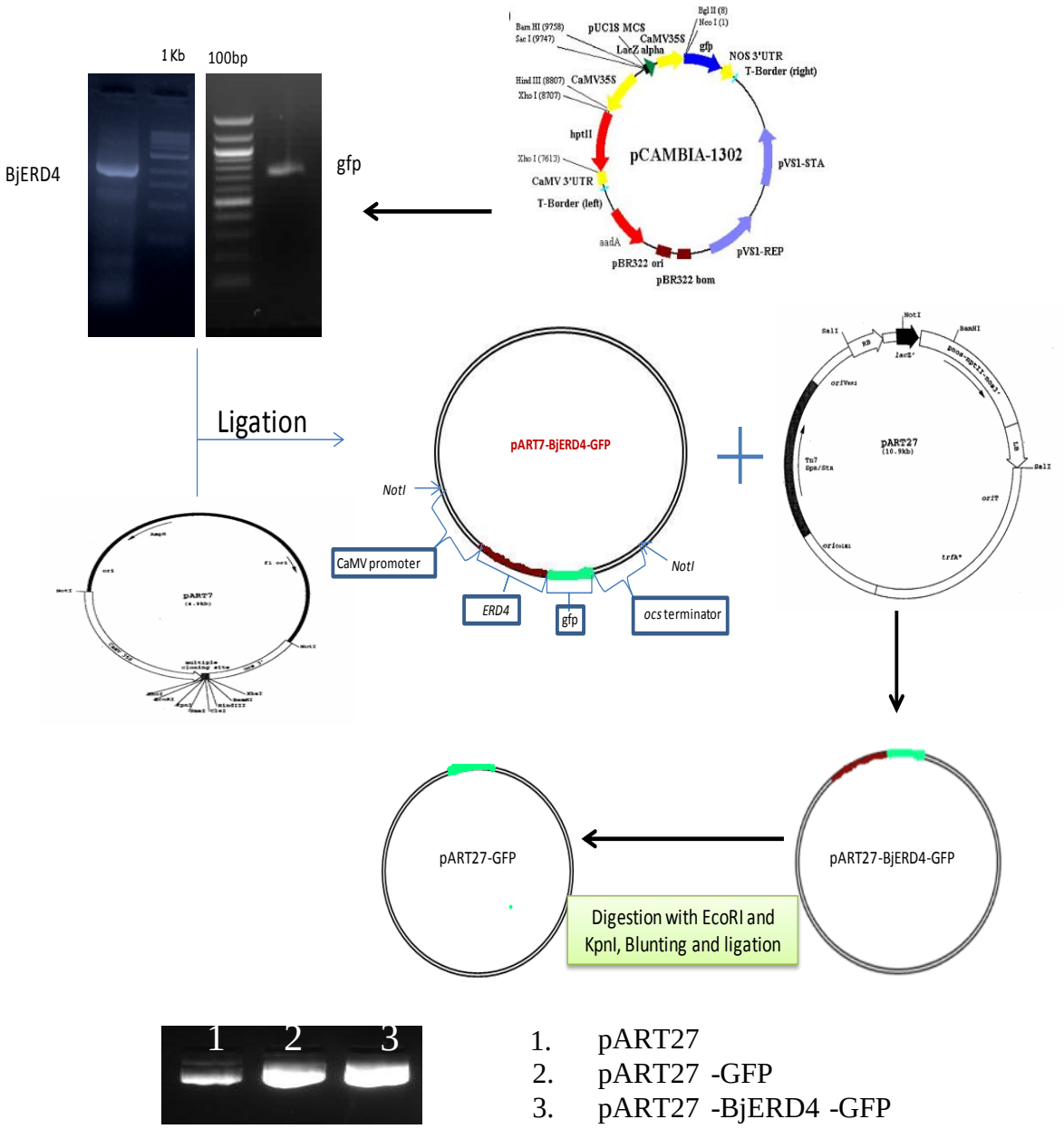


Fig. 2.3.6A: Schematic presentation of preparation of constructs for subcellular localization. *BjERD4* gene and gfp gene amplified and cloned in pART7 vector sequentially at *EcoRI* – *KpnI* and *KpnI* – *HindIII* site. Ligated product was digested with *Not1* and subcloned in pART27 plant expression vector. pART27-BjERD4-GFP was digested with *EcoRI* and *KpnI* and ligated to generate pART27-GFP vector.

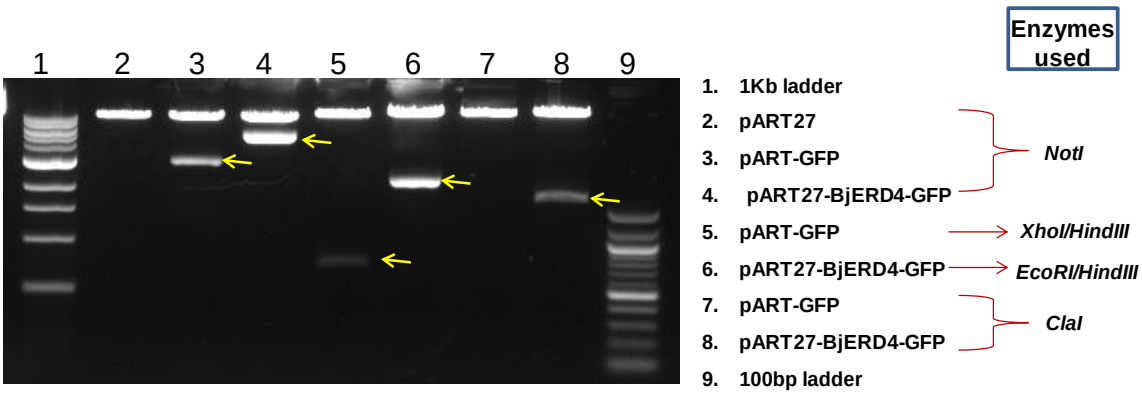


Fig. 2.3.6B: Restriction analysis of constructs for subcellular localization

tagged GFP fluorescence (Fig.3.3.7B) was localized exclusively in oval structures which perfectly matched (Fig. 2d) with the red autofluorescence of chlorophyll (Fig. 3.3.7B) demonstrating that pART27-BjERD4-GFP fusion protein is targeted to chloroplasts, while expression of only 'GFP' construct was observed throughout the protoplast. Protoplast was also treated with MitoTraker dye to see the location of mitochondria. It was observed that GFP signal of ERD4-GFP was not overlapping with MitoTraker dye fluorescence (as shown in pseudo color of blue).

2.4. Discussion

ERD4 gene is one of the important abiotic stress related genes whose enhanced expression is observed in different plant species like Arabidopsis [46], sugarcane [157] and maize [126] under different abiotic stress conditions. As a first approach to investigate the role of ERD4 in abiotic stress tolerance, molecular characterization of ERD4 was undertaken. To clone the *ERD4* gene from *B. juncea*, homology based gene cloning method was adopted. Since Arabidopsis genome showed high level of homology and synteny with Brassica genome [158], primers for amplification of ERD4 gene were designed from the gene sequence of Arabidopsis. A 3.1Kb and 2 kb gene were amplified

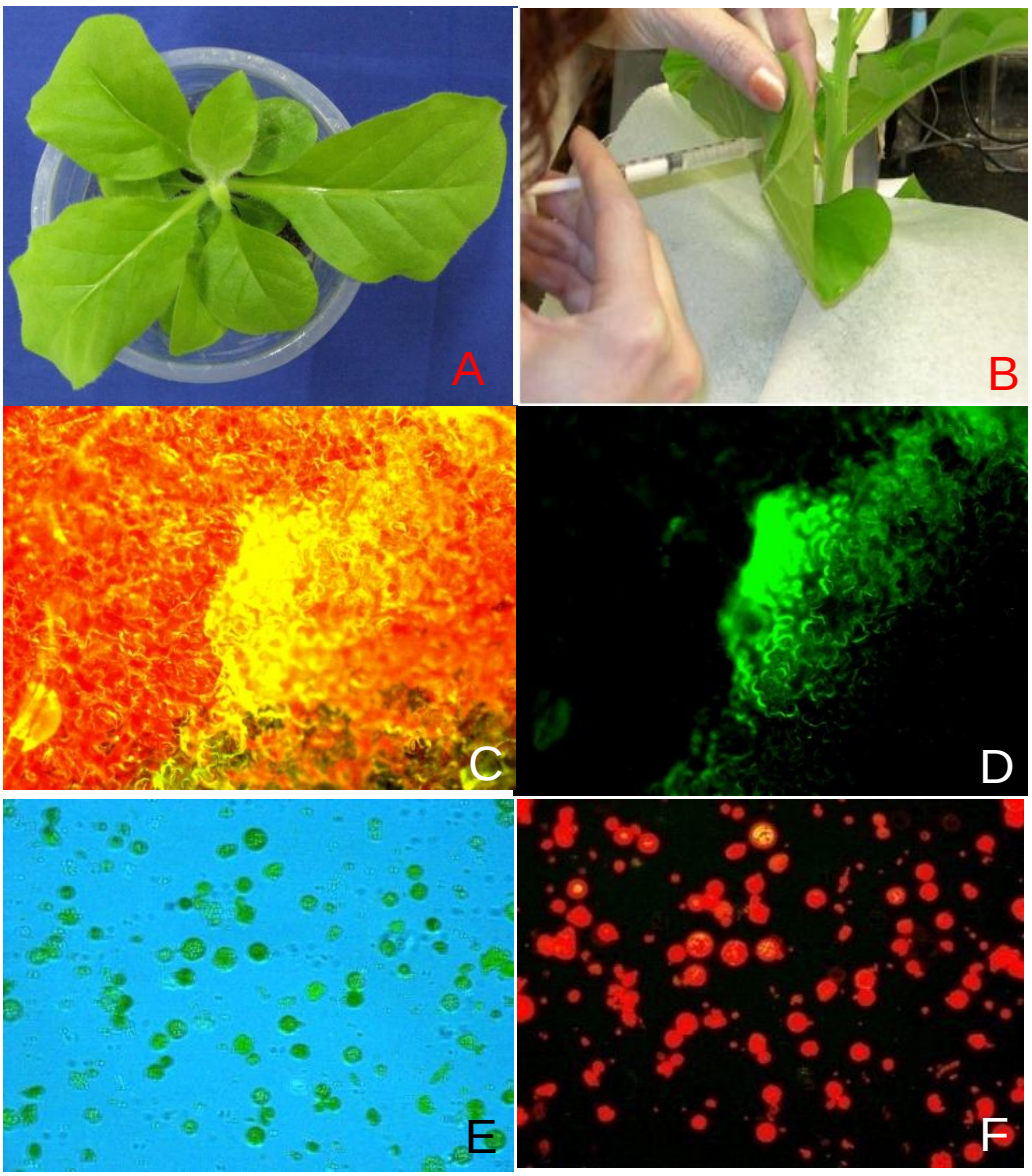


Fig. 2.3.7A: Agro-infiltration in tobacco leaves and protoplast isolation. A. Tobacco plant, B. Agro-infiltration, C. Fluorescence microscopy of transformed leaves with AF filter, D. Fluorescence microscopy of transformed leaves with GFP filter, E. Fluorescence microscopy of protoplast under normal light, F. Fluorescence microscopy of protoplast with AF filter.

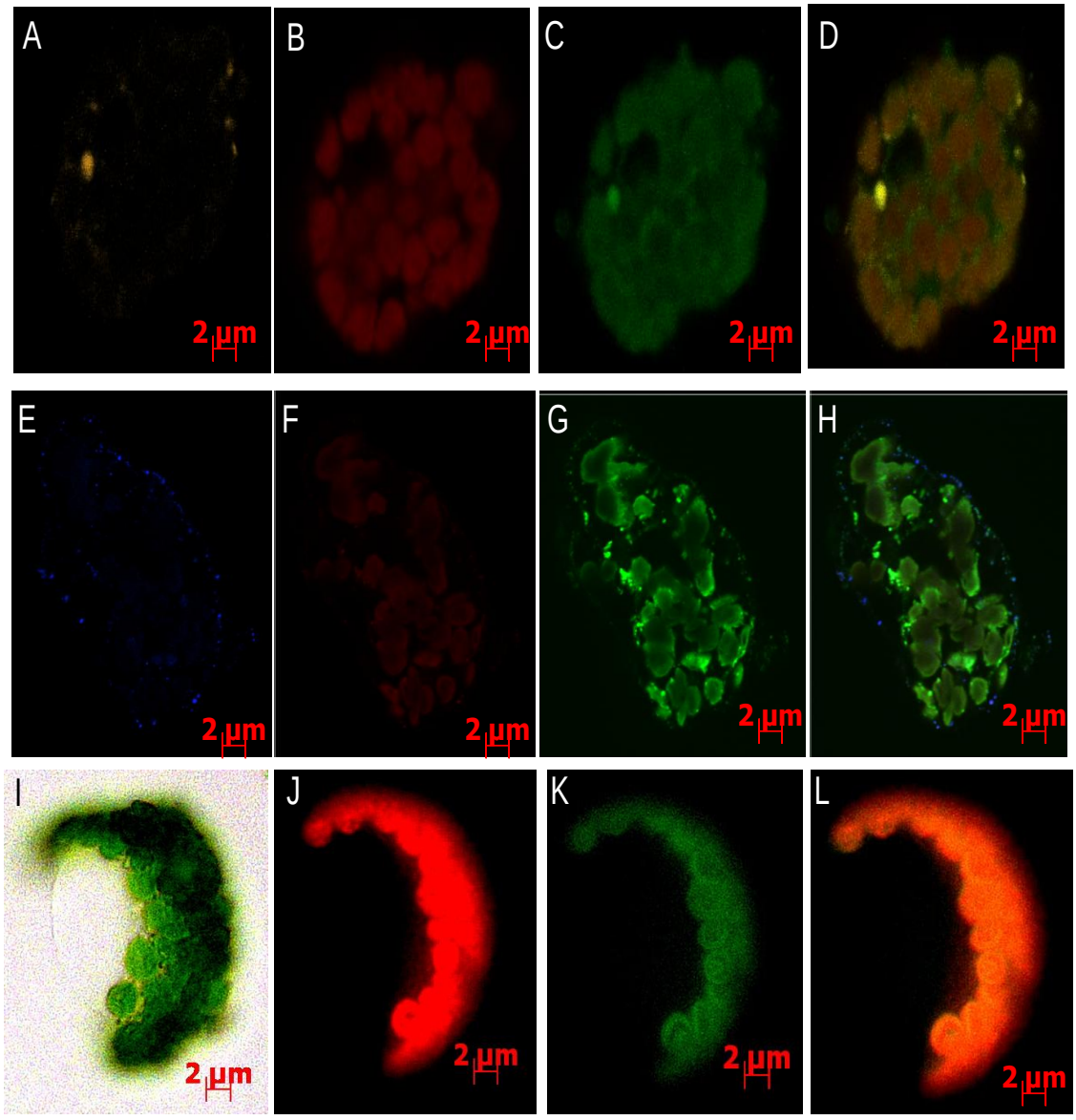


Fig. 2.3.7B: Visualization of fusion protein ERD4-GFP and GFP alone in transformed Tobacco protoplast. GFP alone with MitoTraker dye (A), Auto fluorescence (B), Green indicates fluorescence from GFP (C) and merged image of A, B and C (D). ERD4-GFP fusion protein with MitoTraker dye (E), Auto fluorescence (F, J), Green indicates fluorescence from GFP (G,K) and merged image (H,L). Protoplast under normal light (I). GFP aGreen indicates fluorescence from GFP, red indicates fluorescence from chlorophyll in chloroplasts, blue indicates mitochondria. Bars = 2µm

from gDNA and cDNA respectively which is in line of the previous report of ERD4 gene size amplified from maize [126]. Sequence of amplified product was used for blast search which showed 81% sequence identity with Arabidopsis *ERD4* gene thus the gene amplified from *B. juncea* was named as *BjERD4*.

A close homolog of *Brassica juncea* ERD4 protein was detected in all plant species indicating conservation of the protein in plantae kingdom. Phylogenetic relationship of this gene showed similar pattern of divergence as different plant lineages have evolved, emphasizing that ERD4 gene has essentially been maintained during the course of plant evolution. There are several genes, like ERD1 involved in abiotic stress tolerance which has been found to be conserved in plant species [159].

A close homolog of *Brassica juncea* ERD4 protein was detected in all plant species indicating conservation of the protein in plantae kingdom. Phylogenetic relationship of this gene showed similar pattern of divergence as different plant lineages have evolved, emphasizing that ERD4 gene has essentially been maintained during the course of plant evolution. There are several genes, like ERD1 involved in abiotic stress tolerance which has been found to be conserved in plant species [159].

Based on motif identification and topology prediction results, the presence of DUF 221 domain (312–634) and globular domain (183–347) were identified in BjERD4 protein. The DUF221 domain has been identified in all forms of eukaryotic organisms and has been observed in nearly 23 different domain architectures in combination with a variety of other functional domains like Dnaj, UBQ, VWD etc. The existence of structural domain, with a common function, in combination with variety of other domains has been known to be responsible for evolution of protein repertoire [160]. The DUF221

domain has no other known function, except for membrane integration. It is likely that biological function of the ERD4 protein is attributed mainly to the globular domain, and DUF221 helps in localization of the functional (globular) domain.

The predicted topology showed two models for this protein. Based upon positive inside rule, it was concluded that N-terminus of ERD4 was outside the membrane as nearly 79% of the positively charged residues were observed to reside on inside loops. The smaller loops reside on outside the membrane confirming also to the observation that periplasmic loops are short possibly because of difficult translocation of intermediate-length loops [161]. The corresponding transmembrane topology model revealed presence of the non-transmembrane segment (residues 170–360) inside the closed membranous structure.

ERD4 protein localization in different plant organelles has been the subject of intense discussion. Its localization in the chloroplast membrane was earlier suggested by using the *Arabidopsis* chloroplast envelope proteome analysis by Froehlich et al. [46] and Kleffmann et al. [155], while Alexandersson et al. (128) identified its location in plasma membrane of *Arabidopsis thaliana* that could have been due to organelle contamination [162]. Further, mitochondrial and plastid dual targeting of *A. thaliana* ERD4 was suggested [129]. Thus to resolve this ambiguity of ERD4 subcellular localization, *in silico* approach was first applied using different available bioinformatics tools and then the ERD4 homologues protein sequence were compared with already known that of chloroplast envelop of *Arabidopsis thaliana.*

The analysis of homologous plant ERD4 sequences was done to confirm its organelle localization on the premise that localization signatures must be strictly

conserved in all the plant ERD4 sequences. The analysis of ERD4 orthologs by the ambiguous targeting predictor suggested wide variations in the confidence score; a low score of 0.19 for a number of ERD4 orthologs (Table 2.3.1) clearly indicated its localization in only one compartment. Its presence in chloroplast membrane, however, was inferred on the basis of higher abundance of Ser/Thr and under representation of Arg/Lys residues in the N-terminal sixteen residues of ERD4 orthologs, as also observed earlier for the chloroplast proteins [156]. In this study, marked increase in percentage of hydrophobic Ala/Leu residues was also seen in the N-terminal sixteen residues for chloroplast envelope proteins of *A. thaliana*. Similar high percentage of hydrophobic Phe/Ile residues was observed in the N-terminal sixteen residues of ERD4 orthologs (Fig. 2.3.5). Taken together these data support the experimental finding of its localization in chloroplast membrane.

Localization in eukaryotic cells has proven harder to predict, for example PSORT II [163] only achieved a somewhat disappointing accuracy of 60% for ten yeast localization sites, whereas TargetP could classify plant and nonplant eukaryotic cells proteins into four sites with an accuracy of 85% and 90% respectively. Nair and Rost [123] compared TargetP [101], SubLoc [102], and NNPSL [164] on a common dataset obtaining high (99% and 93%) coverage for extracellular and mitochondrial proteins with TargetP but with low (51% and 46%) precision. For cytoplasmic and nuclear proteins, the study found SubLoc to yield coverage of 67% and 82% with precisions of 60% and 76% respectively. Based upon these, it is clearly seen about the limitation of online tools available for subcellular localization. With this background, an experimental method was sought to prove the hypothesis regarding subcellular localization.

To confirm the *in silico* results on the subcellular localization of ERD4, it was proposed to use the transient expression of ERD4 fusion protein with GFP. Fluorescent proteins have been used for localization studies in a wide variety of systems since GFP was first cloned in 1992 [165]. Fluorescent proteins are useful in labeling cells/proteins and monitoring the localization or dynamic movement of proteins without addition of exogenous substrates [125]. In the present study, tobacco leaf two system was employed to see the localization of ERD4. Transient expression of ERD4-GFP fusion protein could be identified in the tobacco system. The confocal microscopy results of transformed protoplasts confirmed the presence of ERD4 protein in chloroplast since green fluorescence was observed with the overlapping with red auto fluorescence of chloroplast but not with mitochondria.

The presence of ERD4 in the chloroplast is also consistent with predominance localization of the organelle stress response proteins in chloroplast as was observed by Taylor et al. [154]. The detection of ERD4-like protein in uni- and multicellular green algae provides further credence to the suggested chloroplastic localization of the ERD4 protein, as all plastids are derived from a single endosymbiosis and after plastid acquisition only photosynthetic eukaryotes diverged into glaucocystophytes, rhodophytes, and viridiplantae lineages [166, 167, 168]. However, ERD4-like protein was not detected in the cyanobacteria. Previous findings have also suggested that plant proteins encoded by genes of cynobacterial origin are not, as a rule, targeted to chloroplasts, whereas many non-cynobacterial proteins can be targeted to plastids [169].

CHAPTER III:

FUNCTIONAL CHARACTERIZATION OF ERD4 PROTEIN

3.1. Introduction

The functional annotation of a protein is an important challenge in the post-genomic era due to the critical roles of proteins in various biological processes. However, it is expensive and time-consuming to experimentally determine function of an unknown protein. With rapid advances in large scale genome sequencing technologies, there is an increasingly widening of gap between the number of newly found proteins and the completeness of their annotations, necessitating a faster and more effective way to annotate or interpret unknown proteins automatically.

Computational prediction of protein function is based on the idea of assigning functions to unknown proteins according to their similarity with proteins of already known functions. The most common and reliable methods are using homology mapping to transfer annotations to newly sequenced proteins. One of the ways to infer homology is detecting sequence similarity by using BLAST [170] and FAST [171]. Another way is to identify protein domains by using the databases or tools, such as Pfam [172], PRODOM [173], SCOP [174]. The identification of geometric relationships between protein structures by the use of structural alignment methods is also a powerful approach in identifying structural and functional relationships between highly divergent proteins [63].

Domains are some compactly structured components of a protein that can evolve, function, and exist independently of the rest of the protein chain. It is well established that proteins evolve partly through rearrangements of larger fragments, typically domains, and nature of these fragments determine biological function of proteins [64]. The analysis of proteins at individual domain levels can facilitate functional annotation of

uncharacterized genes and proteins [65, 175, 176]. Function of a large number of proteins of DUF families has been proposed based on the structural homology of experimentally determined structures to functionally annotated proteins [57]. The functional domains can also be identified reliably by computational analysis such as prediction of the secondary structure, transmembrane segments, and by fold-recognition [66, 67]. An atomic model of the identified domain can further be obtained from the sequence alone by identifying homologs using sequence-sequence comparison or by fold assignment using structure-sequence alignment [68, 177]. With the available computational tools, it is also possible to identify residues involved in the biological function based on the structure-structure comparison. The utility of these approaches can be extended for predicted structural models of uncharacterized proteins enabling functional annotation of related proteins. Such a strategy is particularly useful for membrane proteins as their experimental structure-function determination is a difficult task.

For validation of *in silico* identified domain structure and function, heterologous expression can be used which allows the production of plant proteins in an organism which is simpler than the natural source. This technology is widely used for large-scale purification of plant proteins from microorganisms for biochemical and biophysical analyses. *Escherichia coli* (*E. coli*) is the most widely used prokaryotic expression system for heterologous protein production [178] due to its simplicity, rapid growth rate, and relatively low cost.

RNA binding proteins (RBPs) play a fundamental role in regulating gene expression. RBPs regulate primarily by binding to specific sequence elements in nascent or mature transcripts. There are several hundreds of RBPs in plants, but the targets of

most of them are unknown. A variety of experimental methods have been developed to identify targets of an RBP [179]. These include RNA immunoprecipitation (RIP), UV cross-linking and immunoprecipitation (CLIP) and many variations of CLIP (e.g. PAR-CLIP, iCLIP). These approaches depend on immunoprecipitation of RNAs bound to a specific RBP using an antibody to that RBP. Electrophoretic mobility shift assay (EMSA), also called gel shift assay, has been used to analyze protein-nucleic acid interactions [179]. It is a simple and powerful method to analyze protein-RNA/DNA interactions.

This chapter includes the study of structure and function of ERD4 protein; initially the prediction of functional domain in this protein using 'fold recognition method' was adopted followed by the 'structure sequence alignment' method. Further, the heterologous expression of predicted functional domain in *E. coli* BL21 (DE3) and purification of recombinant protein using Ni-chelated affinity column were undertaken and its RNA binding function was demonstrated.

3.2. Materials and methods

3.2.1. Prediction of the functional domains and 3D structure:

The *Brassica juncea* ERD4 sequence was subjected to Pfam [180], DOUTfinder [181] and SMART [182] analyses for identification of similarity with the known domains and domain architecture. An independent analysis for detecting globular domains of structural-folds similar to the known protein structures was also carried out using structure prediction meta-server (http://bioinfo.pl/meta) accessing various fold-recognition and function prediction methods. The database of known protein structures (Protein Data Bank, PDB) was searched for a structure homolog to the detected globular

domain using sequence-sequence comparison search engines. In the absence of any known homologous structure, the tertiary fold of the globular domain was independently predicted using the meta-server. The collected results from fold-prediction servers were screened with 3D-jury [183]. The 3D structural model of the globular domain was constructed with Modeller [184] using sequence-to-structure alignment returned by the meta-server, and RNA-binding domains from human nucleolin (PDB code, 2KRR) and poly(a)-binding protein (PDB code, 1CVJ) as templates. The structural neighbors of the theoretical structural model of the globular domain were identified by the DALI [185] programs.

3.2.2. Identification of functional residues:

The 3D structural model of the identified globular domain was superposed onto the known structures of RNA-binding proteins which possessed RNA-recognition domains. The atomic coordinates of these were obtained from the PDB. The superposition was achieved using DALI programs and Swiss PDB viewer [186]. The amino acid residues of the ERD4 domain equivalent to the residues interacting with RNA substrates in the known RNA-binding proteins were identified as putative RNA-binding residues. The conservation of these was verified in the alignment of the amino acid sequences of the identified RRM domains of ERD4 homologs.

3.2.3. Cloning of RRM motif of ERD4 protein:

The non-transmembrane domain of ERD4 protein was predicted as RNA binding motif using bioinformatic tools. For validation, a pair of primers designed having restriction sites *BamH1* and *Nde1* and named as RRMF (5’CATATGACTAGCGAAGA AGTATTACCG3’) and RRMR (5’GAATCTCAAGATCAAGTTCTTCTAGGATCC3’)

and 528 bp gene fragment was amplified from *Brassica* cDNA. The amplified fragment was cloned in pET28-His tag vector. The plasmids were subsequently transferred to *E coli* cells. The recombinants were selected and named as pET-His-RRM. Clones were confirmed by restriction digestion with *BamHI* and *NdeI* followed by sequencing. The pET-His-RRM was then transferred to *E. coli* BL21.

3.2.4. Expression of RRM motif of ERD4 protein:

The selected clones were incubated in LB medium induced with IPTG and the cultures were sampled before and after induction. The samples were processed and cellular proteins were analyzed by SDS-PAGE. The recombinant protein of ~20 kD was observed and showing its presence in inclusion bodies as well as in soluble proteins. For enhancing the presence of this protein in soluble fraction, treatment conditions like different concentrations of IPTG, temperature and ethanol were used.

3.2.5. Purification and confirmation of RRM domain of ERD4 protein:

Purification of RRM domain of ERD4 protein was attempted using soluble fraction of cell lysate by using immobilized-metal (Ni^{2+}) affinity chromatography (IMAC). The bacterial cultures for the proteins were prepared by inoculating 50 mL LB medium with single bacterial colony and were grown for overnight. The cells grown at 37ºC in LB medium containing 50 μg ml^{-1} kanamycin (OD_{600} ~0.2) were transferred to 18ºC. the protein expression was induced at the late log phase (OD_{600} ~0.7)by the addition of 0.4 mM isopropyl β-D-1 thiogalactopyranoside (IPTG) and was incubated for overnight. The cells harvested by centrifugation at 10,000×*g* at 4 °C for 1 hr and were lysed in the lysis buffer (50 mM Tris–HCl (pH, 8.0), 15% sucrose, 2 mM dithiothreitol and 1 tablet of Complete® protease inhibitor cocktail (Roche Applied Science,

Mannheim, Germany) containing lysozyme to a final concentration of 2 mg mL^{-1}. The cell suspension was incubated for 2 h and was subsequently sonicated in pulse mode (5–10 pulses of 15 s each) by using Vibra Ultrasonic processor (model VC-250, Sonics and Materials Inc, USA). The suspension was centrifuged at 21,000×*g* at 4 °C for 30 min. The supernatant was loaded onto the pre-equilibrated nickel-nitrilotriacetic acid (Ni-NTA) column. The column was washed with buffer A (50 mM Tris–HCl (pH, 8.0), 100 mM NaCl) containing 25 mM imidazole. The bound proteins were eluted using 100–500 mM imidazole gradient over six-column volumes in buffer A, and eluted protein was dialyzed against phosphate buffer. Purified RRM domain protein was run on SDS –PAGE and 20.12 kD band was eluted and fingerprinted. Then the same band was further eluted from SDS –PAGE and antibody was generated against this domain.

3.2.6 RNA Electrophoretic Mobility Shift Assay (EMSA):

The gel electrophoretic mobility shift assay (EMSA) is used to detect protein complexes with nucleic acids. The technique is premised on the knowledge that electrophoretic mobility of protein-nucleic acid complexes is less than the corresponding free nucleic acids [187].

For RNA EMSA, purified RRM motif of ERD4 protein was incubated with *in vitro* transcribed RNA (2.5 µg) using Riboprobe *in vitro* transcription system (Promega) in reaction buffer (50mM Tris-Cl, pH 8.0, 100 mM NaCl) for 30 min at 37°C with gentle shaking. The RNA products were separated on a 1% agarose gel in TBE buffer. Bovine Serum Albumin (BSA) protein and Reverse Transcriptase (RT) protein treated with the same procedures served as a negative and positive control respectively. For small RNA

and ssDNA binding with RRM protein BSA and Translin protein were taken as negative and positive control respectively.

3.3. Results

3.3.1. Structural analysis of the globular domain:

A BLAST search with the amino-acid sequence did not reveal any close homologue in the database of known protein structures (PDB). This is not unusual as sequence comparison methods cannot reliably detect evolutionary relationship between highly divergent proteins. The structural fold of the ERD4 domain was then found by fold-recognition methods, which use sequence-structure alignment. This method allowed detection of remote homologies beyond the detection limits of other sequence comparison methods. The input for fold-recognition was *B. juncea* ERD4 sequence from which generated profile was compared to sequence profiles of proteins and domains of known structures. The search for ERD4 protein fold using fold-recognition meta- server suggested structural homology of about 165 amino acid residues (183–347) with the known RNA-binding globular proteins. Interestingly, all the best hits identified by the 3D-jury from the meta-server were RNA-binding proteins possessing two well known RNA-recognition motifs (RRM) (Table 3.3.1). The residues 183–347 of the ERD4 sequence were thus expected to adopt a globular fold with structural similarity with RNA-binding proteins.

The 3D structural models of the globular domain were constructed using the solution structure of the RBD1,2 domains from human nucleolin (PDB code, 2KRR; Jscore, 55.3) and using X-ray crystal structure of the poly(a)-binding protein in complex with polyadenylate RNA (PDB code, 1CVJ; J score, 48) as templates. Given the high

divergence between ERD4 globular domain and the RNA-recognition proteins used for constructing the theoretical models with pair-wise sequence identity of about 10 % (Table 4.3.1), the general atomic resolution of the theoretical model to be low (.3A ˚) is expected. However, all the structural neighbors of the ERD4 globular domain were found by DALI program [185] to belong to RNA-binding domain superfamily. The computationally constructed structural models for the ERD4 chloroplastic domain clearly showed the presence of two tandem RNA-recognition motifs, each having βαββαβ topology (Fig. 3.3.1.1).

Table 3.3.1: The best five structural models predicted for the ERD4 globular domain by the fold-recognition servers and their ranking by 3D-Jury method.

Model [(1)]	**3D-Jury score (JScore)**	**Scop Classification**	**Super family**	**Percentage identity/similarity with *B. juncea* ERD4 globular domain**
2krr _A	55.3	54928	RNA-binding domain	9.6/27.1
2dhs_A	54.0	54928	RNA-binding domain	12.1/36.4
1cvj_A	48.0	54928	RNA-binding domain	7.8/26.1
2g4b_A	41.0	54928	RNA-binding domain	11.5/30.9
3md3_A	39.7	54928	RNA-binding domain	10.3/32.1

[(1).] PDB designation of domain.

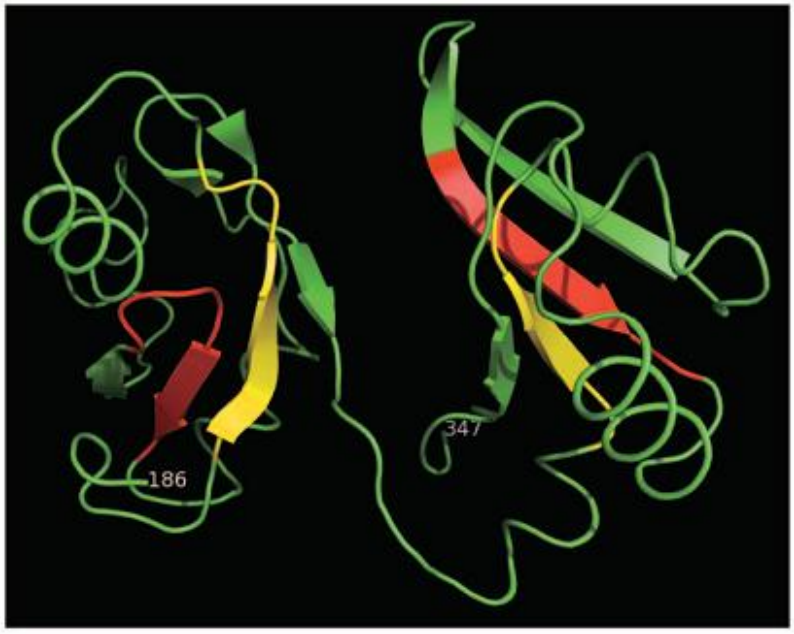


Fig. 3.3.1.1 Ribbon model of the putative RNA-binding globular domain. The ribbon model was constructed by comparative homology approaches. The fold of the domain was identified by fold-prediction meta-server. Due to low pair-wise sequence identity of nearly 10% between the query and identified template, the derived atomic coordinates for the ERD4 globular domain were expected to be of low-resolution. The two ribonucleoprotein motifs (RNP1 and RNP2) in each of the RNA-recognition domains are shown in red and yellow, respectively. The figure was prepared by PyMol (http://www.pymol.org/).

The two RRM domains were composed of amino acid residues 183–269 (RRM1) and 273–347 (RRM2) respectively, and joined by an inter domain linker peptide. The inter domain linker peptide is a typical characteristic of known RNA-binding proteins with multiple RRM domains [188]. The two RRM domains could be flexibly tethered via the linker peptide. Analogous to the well characterized RNA-binding proteins, the β-sheets of the two RNA-binding domains of ERD4 face each other and RNA substrates

could bind in the cleft. The two RNA-recognition domains of ERD4 were individually superposed onto the known RNA-binding domains of sex lethal protein (PDB code, 1B7F) and adenosine-uridine (AU)-rich binding Hu protein (PDB code, 1FXL). These proteins had similar number of amino acids as ERD4 globular domain but differed significantly from the latter (DaliLite Z-scores for ERD4/ 1B7F and ERD4/1FXL pairs were 5.8 and 5.5, respectively) and thus formed highly diverse pairs. Additionally, these structures had been refined to high precision against single crystal diffraction data and coordinates of protein-RNA complexes were available, which could hint RNA-binding mode in the ERD4 protein (Fig. 3.3.1.2).

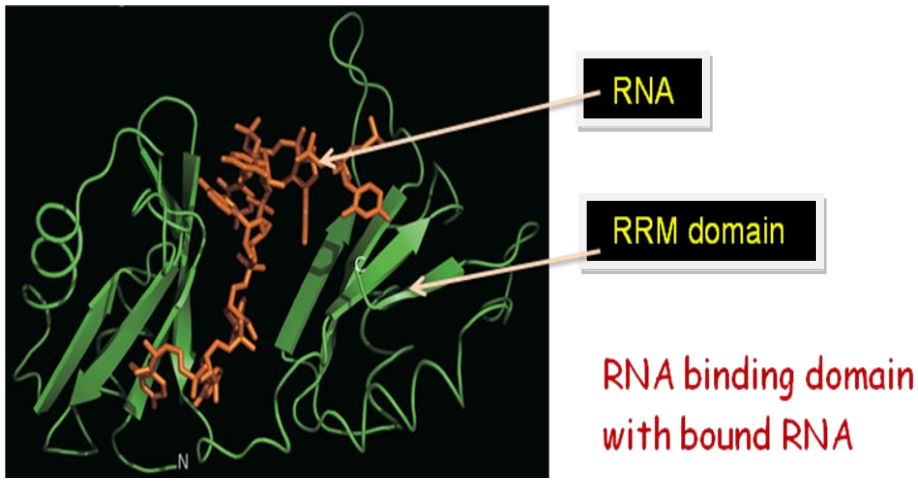


Fig. 3.3.1.2: Cartoon of HuD1, 2–cfos-11 RNA complex structure (PDB code 1FXL; 3). The RNA is shown as a stick model (orange). The N- & C- termini of the protein are marked as N and C, respectively. The two RRM domains form a cleft with the RNA bound between the b-sheets surfaces. In several RNA-binding proteins the two RRM domains are flexibly tethered via a linker peptide.

The structural alignment showed the presence of two non-canonical ribonucleoprotein sub-motifs (RNP1 and RNP2) in both the ERD4 domains (Fig.

4.3.1.3). One of the ribonucleoprotein sub-motifs (RNP2) resides on the first β-strand, while residues from third β- strand contribute towards RNP1. The putative RNP sub-motifs of RRM1 are 195-ILVRDI-200 (RNP2) and 237-INKIWEDL-244 (RNP1) and those of RRM2 are 283-DYYTKL-288 (RNP2) and 307-RQQTAAVVF-315 (RNP1). In the multiple sequence alignment of ERD4 orthologs, the RRM1 domain has conserved hydrophobic (Leu/Val) at position-2 of the RNP2 and aromatic (Trp/Tyr) at position-5 in RNP1 (Fig. 3.3.1.3). Also, Tyr/His and Ala are conserved in RNP2 position-2 and RNP1 position-5, respectively, in the RRM2 domain. A positively charged amino acid residue (Arg/Lys) was also found in most of the plant ERD4 proteins at RNP1 position-1 of RRM2. In addition to the α-strands, the loops β1/α1 (connecting b1 and a1 elements), β2/β3 and α2/β4 have also been observed in RNA-binding proteins to interact with nucleic acid substrates [188]. Most of these residues are conserved in ERD4 orthologs (Fig. 3.3.1.3). Interestingly *B. juncea* Pro-201, residing on the loop β1/α1, is strictly conserved in all the plant ERD4 proteins. This position is occupied by Pro/Ser residues in majority of RNA-binding domains identified in NCBI conserved domains database CD00590 [189].

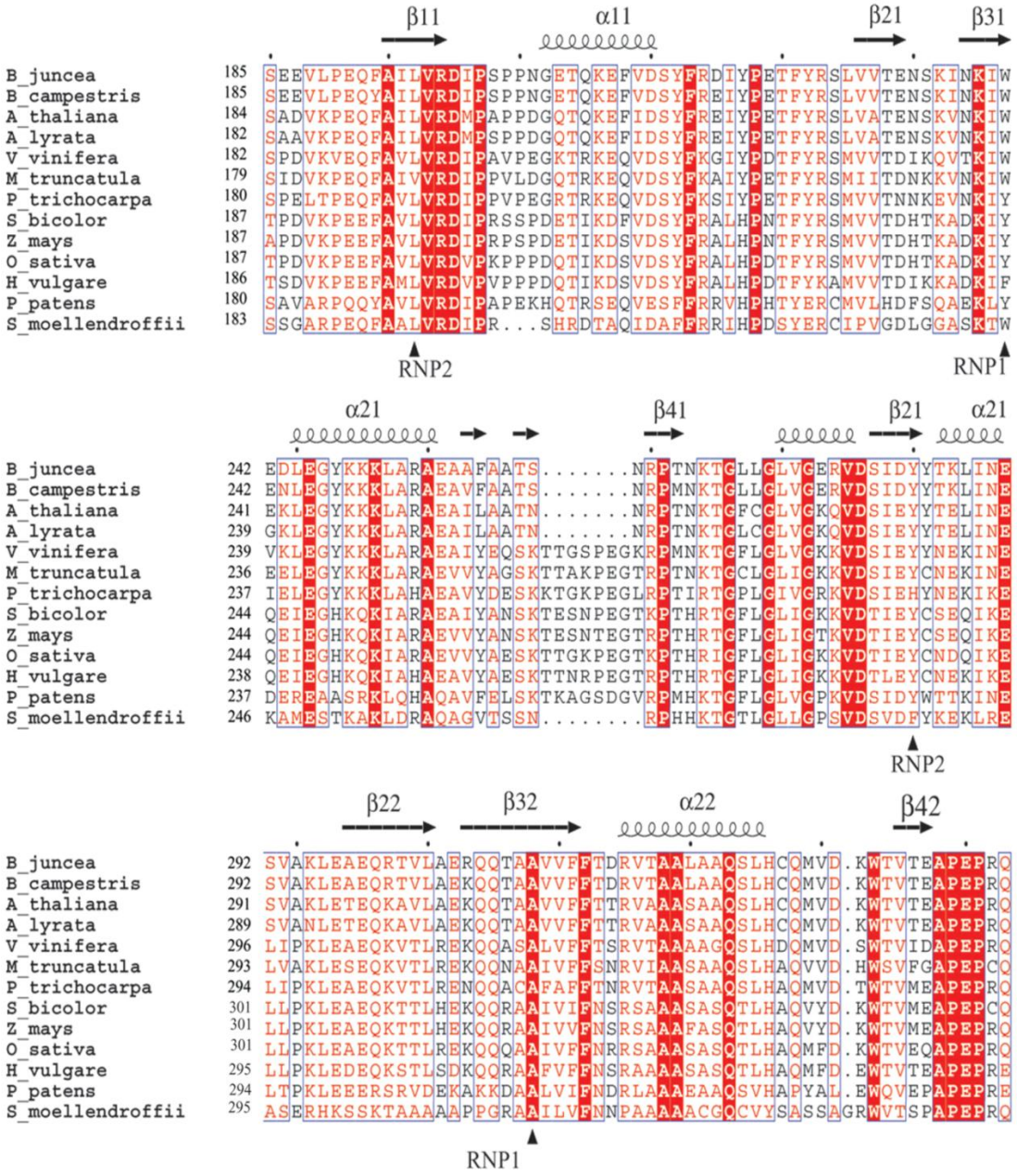


Fig. 3.3.1.3. **Multiple sequence alignment of the ERD4 globular domain.** The alignment was generated by ClustalW. The two RNA-recognition domains are composed of amino acid residues 183–269 (RRM1) and 273–347 (RRM2), respectively. The two ribonucleoprotein motifs of each RRM domain are marked as RNP1 and RNP2. The suggested RNA-interacting residues are marked with filled triangle. The secondary structure elements of each RRM domain in the theoretical structural model are also shown. The strictly conserved residues in all the plant ERD4 sequences are shaded, while similar residues are boxed. The residues numbering is of the full-length ERD4 proteins.

Phylogenetic relationship of RRM domain of ERD4 protein and other known RNA binding protein domain:

Protein sequence of structurally characterized RRM domain of RNA binding proteins were retrieved from PDB database and a phylogenetic tree was constructed by NJ method. Maximum similarity was found with the human poly pyramidine track binding protein which is regulator of alternate splicing (Fig 3.3.1.4)

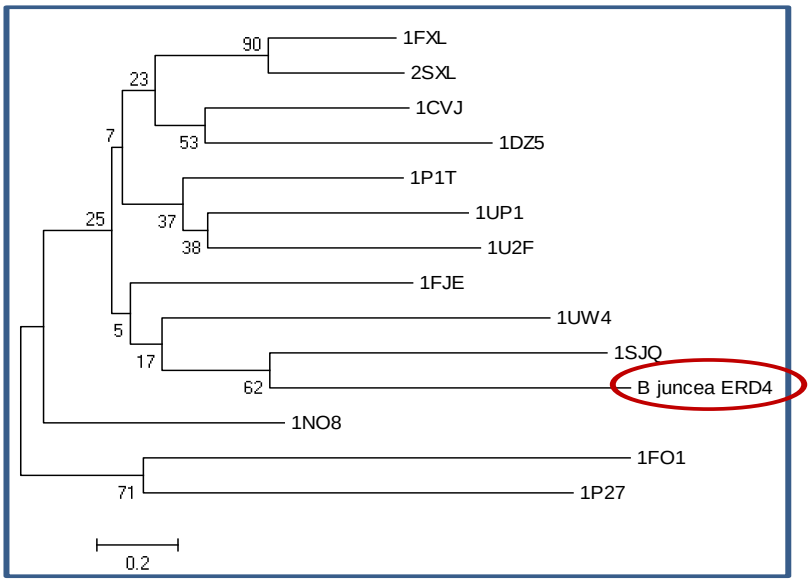


Fig. 3.3.1.4: Phylogenetic relationship of RRM domain of ERD4 protein and other known RNA binding protein domain

3.3.2. Functional validation of Globular domain (RRM) of ERD4

3.3.2.1. Cloning and expression of RRM motif of ERD4 protein:

Non-transmembrane domain of ERD4 protein was predicted as RNA binding motif using bioinformatic tools. For validation, a pair of primers was designed for the amplification only globular domain. These primers had restriction site of s *BamH1* and

Nde1, and 528bp gene fragment was amplified from *Brassica* cDNA. The amplified fragment was cloned in TA cloning vector and later it was cloned in pET28-His tag vector. The plasmids were subsequently transferred to *E coli* cells. The recombinants were selected on kanamycin selection medium. Colony PCR and further sequencing of the PCR product was performed for the confirmation of the clones (Fig. 3.3.2.1). The confirmed clones were named as pET-His-RRM. The pET-His-RRM was then transferred to *E.coli* BL21(codon plus strain) for protein expression.

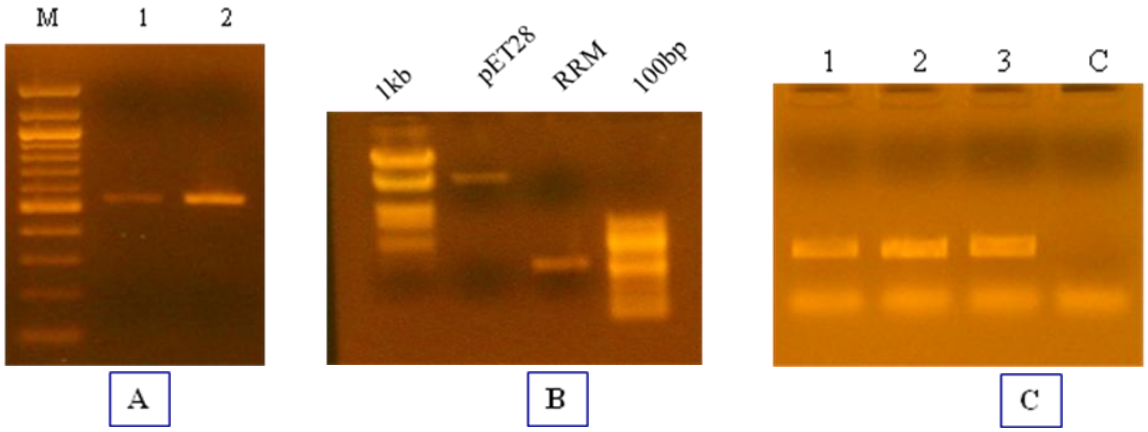


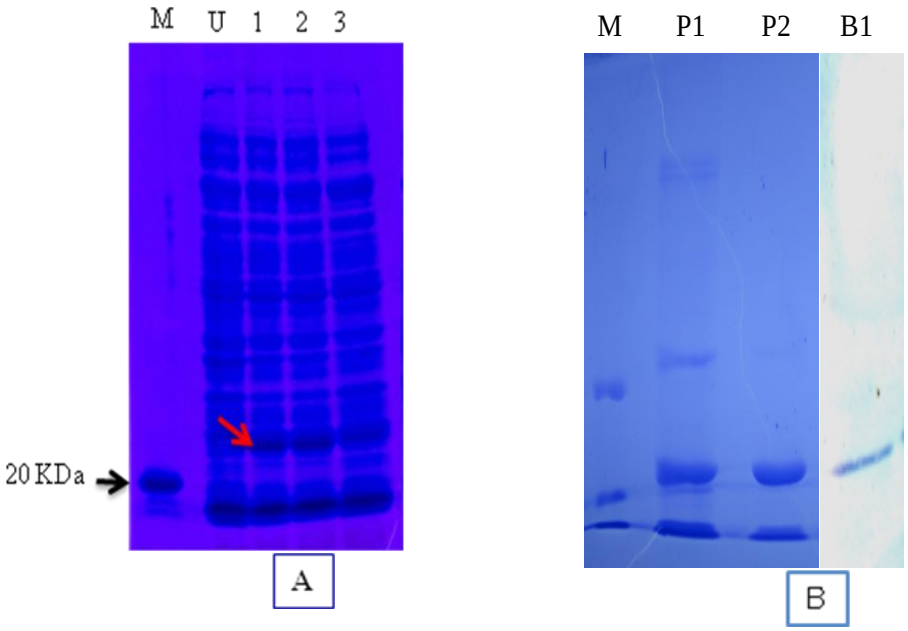
Fig. 3.3.2.1: Construction of expression cassette for RRM domain. PCR amplification of RRM domain using *B. juncea* cDNA. M: 100 bp marker; 1: cDNA from root; 2: cDNA from shoot **(**A). Size and concentration check of pET28 and RRM domain after double digestion with *NdeI* and *BamH1* (B). Colony PCR with RRM primers of pET-RRM clone. Lane 1-3: colony; C: control (C).

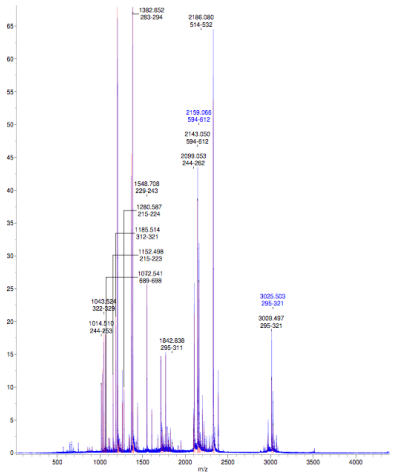
3.3.2.2. Expression, purification and confirmation of RRM domain of ERD4 protein:

The confirmed clone was incubated in LB medium induced with IPTG and the cultures were sampled before and after induction. The samples were processed and cellular proteins were analyzed by SDS-PAGE. The recombinant protein of approx. size

of 20KD was found to be overexpressed as compared to the uninduced sample. The overexpressed band was observed in inclusion bodies as well as in soluble proteins (Fig. 3.3.2.2A). For enhancing the presence of this protein in soluble fraction different treatment conditions like different concentration of IPTG, different temperature and ethanol was used. 0.4mM IPTG, 18^{o}C and 3% ethanol was found to be optimum condition for this protein expression in soluble fraction. Purification of RRM domain of ERD4 protein was attempted using soluble fraction of cell lysate by using immobilized-metal (Ni^{2+}) affinity chromatography (IMAC). Purification of induced protein by column was done successfully as specific size band for RRM-His were observed in SDS–PAGE. Eluted protein was dialyzed against phosphate buffer. After dialysis once again protein was purified with Ni- column. After confirmation, antibody was generated against this domain. Eluted protein was probed with antibody which sowed specific binding with RRM domain (Fig. 3.3.2.2B).

Purified RRM domain protein was run on SDS –PAGE and 20 KD band was eluted and send for fingerprinting. Fingerprinting result for the band shows top score of 72 for gi|161019615, ERD4 protein (*Brassica juncea*) (Fig. 3.3.2.2C).





Match sequence	Score	Database	Accesskey
ERD4 protein (*B. juncea*)	72.400000	NCBInr	gi|161019615
Hypothetical protein SORBIDRAFT_02g005590 (*S. bicolor*)	59.500000	NCBInr	gi|242043250
Hypothetical protein Ycf1 (*Elleanthus caricoides*)	55.800000	NCBInr	gi|187372626

C

Fig. 3.3.2.2: Expression, purification and confirmation of RRM domain of ERD4 protein. SDS-PAGE analysis of RRM domain protein expression. M-Trypsin inhibitor, U-Uninduced lysate, 1- hole cell lysate, 2- Pellet, 3-Supernatant (A).Purified RRM domain protein. M-Low range marker, P1-Ni column purified, P2- after dialysis second time Ni column purified sample, B1-western blot of purified sample with antibody generated against RRM domain of ERD4 protein (B). Confirmation of expressed band with MALDI spectra (C).

3.3.2.3. RRM domain of ERD4 protein binds RNA

For confirmation of RNA binding property of ERD4 protein EMSA was performed. *In vitro* RNA synthesis was performed using Riboprobe *in vitro* Transcription systems. RNA transcript was generated using the T7 RNA polymerase and the pGEM express positive control as template. Two transcripts, of size 1,065 and 2,346 bp long were synthesized (3.3.2.3A). Further the Electrophoresis mobility shift assay (EMSA) was performed with synthesized probe and result showed that ERD4 –RRM protein domain can bind RNA efficiently and could show shift in RNA position. In contrast, comparable amounts of bovine serum albumin and Reverse transcriptase enzyme showed nil and maximum shift with RNA probe respectively. But ERD4 –RRM protein domain could not show any shift when incubated with ssDNA and small RNA (Fig. 3.3.2.3B, C &D).

3.4. Discussion

The structural analysis of a protein is known to reveal the distant evolutionary links that could provide the first hypothesis about biological function of the uncharacterized domains [57]. RNA‐binding proteins as well as RNA‐protein complexes have been investigated in a variety of living organisms, including microorganisms, animals and plants. This has led to the discovery of several conserved protein motifs, such as RNA‐recognition motifs (RRMs), glycine‐rich domains, arginine‐rich domains, SR‐repeats, RD‐repeats and zinc finger motifs [75]. The RRM is the most widely found and best characterized RNA‐binding motif. Much less is known about plant RNA-binding proteins, and in only a few cases has their function been studied. The

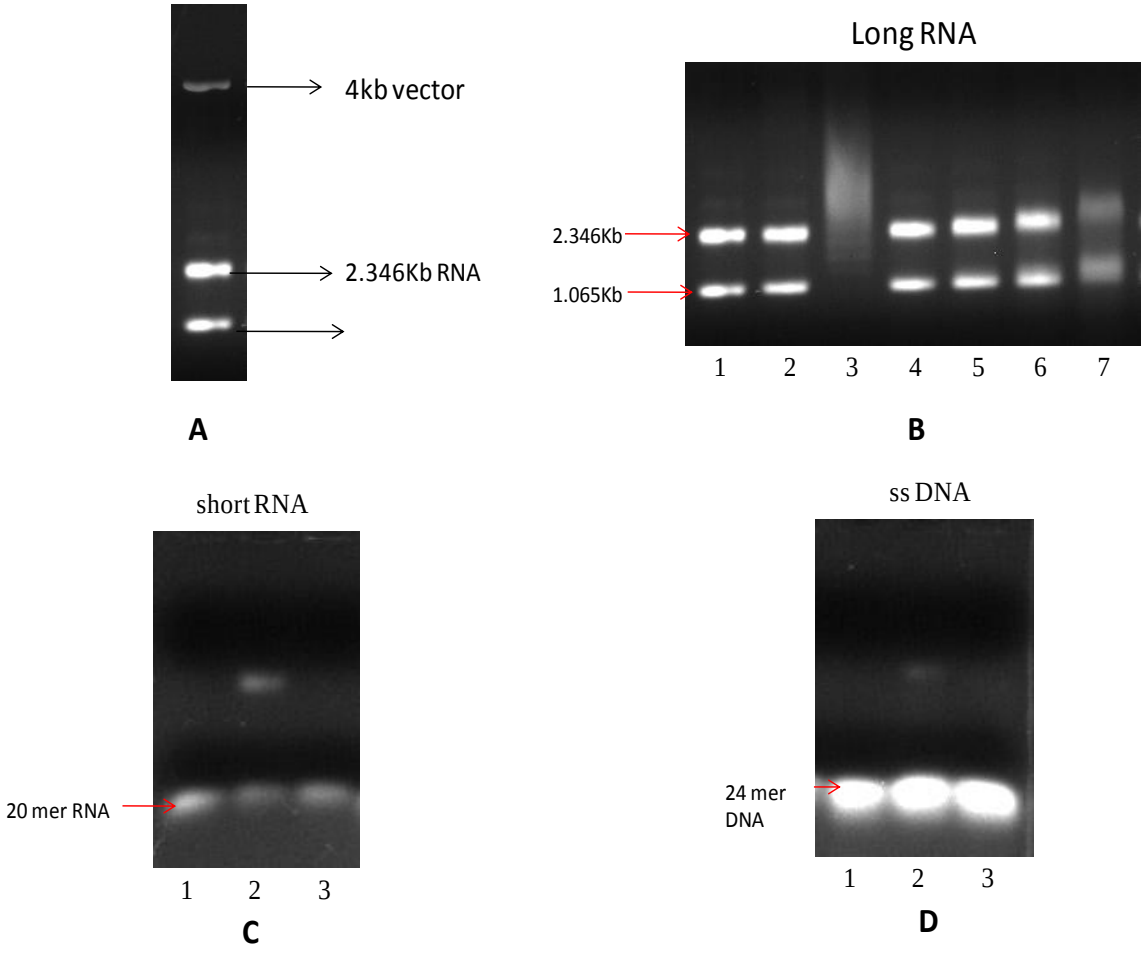


Fig. 3.3.2.3: EMSA showing RNA and RRM domain interaction. A. *In vitro* RNA synthesis, B. Interaction of protein with long ssRNA, 1. Only RNA, 2. RNA+BSA, 3. RNA+reverse transcriptase, 4-7 increasing concentration of RRM domain protein +RNA. C. Interaction of protein with short ssRNA,1. RNA+ BSA, 2. RNA+ Translin, 3. RNA+ RRm domain protein. D. Interaction of protein with short ssDNA,1. DNA+ BSA, 2. DNA+ Translin, 3. DNA+ RRM domain protein.

particular arrangement of the domains serves to define different protein families. For example, chloroplast RNA binding proteins comprise an acidic region at the N-terminus and two repeats of the RRM [71]. A characteristic type of nuclear-encoded, RNA-binding proteins is present in the chloroplasts of higher plants, with a unique structure formed by an acidic region at the N-terminus and two repeats of the RRM motif. These types of protein appear to be involved in post-transcriptional regulation of chloroplast gene expression [190].

In the present study, the tertiary structure of the ERD4 chloroplastic globular domain was predicted by fold-prediction algorithms that suggested presence of two RNA-recognition motifs in its sequence. Each of the RRM was predicted to adopt βαββαβ topology (Fig. 3.3.1.1). The fold of the ERD4 globular domain was found to be shared only by RNA- binding domains, as observed in the search for structural neighbors with DALI programs. Structural and sequence comparison with the known RNA-binding proteins showed the presence of RNP1 and RNP2 ribonucleoprotein sub-motifs in both the identified RNA-recognition motifs of ERD4. The four RNP's in two RRM domains reside on the α-strands creating a RNA binding cleft (Fig. 3.3.1.1). A hydrophobic and an aromatic amino acid residue at 2^{nd} and 5^{th} positions of RNP2 and RNP1, respectively, were conserved in RNA-binding proteins and ERD4 homologs (Fig. 3.3.1.3). These residues stack against the two bases of substrate RNA in the known RNA-binding proteins. The 1^{st} position of RNP1 in RRM2 of ERD4 was also found to be conserved as positively charged amino acid that could neutralize the negatively charged phosphodiester group [191]. In most of the RRM-RNA complex structures only one to three of these contacts are observed with two stacking interactions involving RNP2

position-2 and RNP1 position-5 observed most frequently [192]. The orthologs of TM63A_human protein identified by BLAST search due to its sequence similarity with plant ERD4 proteins, however, do not show strict conservation in the residues corresponding to the proposed RNA-binding domain of ERD4. In contrast to RNA-binding ability, polypeptides that recognize protein substrates, and not RNA, have only one RRM domain. The combination of two or more RNA-recognition motifs, as observed in ERD4 sequences, often results in dramatically increased RNA-binding affinity [72, 188].

For validating hypothesis regarding RNA binding activity of ERD4, RRM domain was cloned and expressed in *E. coli*. Further this domain was purified and interaction with RNA was checked. RRM domain of ERD4 protein binding with large ssRNA as the concentration of protein increases but significant shift was not observed. Whereas the protein interaction with small ssRNA and ssDNA binding was not visible. This suggests sequence specific binding of this domain, but the exact sequence for the binding site has to be further studied.

Chloroplasts not only carry out photosynthesis but also represent sites for important processes like starch, fatty acid and amino acid metabolism. These processes are essential for plant growth and development, thus chloroplast gene expression must be tightly regulated. The chloroplast genome contains only 87 potential protein-coding genes [193]. Much of the regulation of chloroplast processes is exerted by nuclear gene products, which frequently participate in chloroplast gene expression. The RNA binding domain carrying RNP signature sequences is a highly abundant domain in eukaryotes. This domain has been found in a variety of heterogeneous nuclear ribonucleoproteins

(hnRNPs), proteins implicated in regulation of alternative splicing, and protein components of small nuclear ribonucleoproteins (snRNPs), and is involved in post-transcriptional gene expression processes including mRNA and rRNA processing, RNA export, and RNA stability. The domain binds a variable number of nucleotides, ranging from two to eight. It is, however, known that despite using the same β-sheet surface to bind RNA, each protein achieves sequence-specificity slightly differently [188]. The conservation of two tandem RNA-recognition motifs and the substrate binding residues suggests that globular domain of ERD4 protein may be RNA-binding competent. The ERD4 protein can participate in mRNA metabolism such as sequestering and protecting mRNAs during conditions of limiting transcription. In plants, the RNA-binding proteins may modulate ABA signaling through the alteration of mRNA processing events such as splicing, processing, nuclear export, transcript stability and RNA degradation [194]. Also, induction of ERD4 could influence the membrane fluidity as its DUF221 domain is expected to be integrated in the chloroplast membrane. It hence assumes significance to study functionally important residues and domains that are critical for ERD4 activity in response to various environmental conditions. It is also suggested from the analysis that ERD4 proteins may be characterized by the presence of both RRM and DUF221 domains and not by DUF221 domain alone as is the current practice in putative annotations in the sequence databases.

CHAPTER IV:

EXPRESSION ANALYSIS AND FUNCTIONAL VALIDATION OF *BjERD4*

4.1. Introduction

The interaction between a plant's genome and its environment determines it's growth and development. Expression profiling has become an important tool to investigate responses of an organism to environmental changes at the transcriptional level [96, 195]. Sometimes these transcriptional changes are successful adaptations leading to tolerance while at other times, the gene expression is merely a response to stress and the plant fails to adapt to the new environment and is considered sensitive to that condition. Expression profiling can define both tolerant and sensitive genotype and is useful tool for studying regulatory genetic circuitry, which has an application in biotechnological approaches to improve stress tolerance. Beyond transcript profiling, genomics also facilitates the functional analysis of genes. As signaling cascades and metabolic pathways are elucidated in model systems and crop plants, key regulatory genes can be targeted for silencing or over-expression to study the role of these pathways in plant responses to stress.

The ultimate goal of any genomics studies is to identify the biological function of every gene in the genome. Molecular and genomic analyses using model plants facilitated the resolution of complex networks and led to the discovery of additional mechanism(s) of stress tolerance [196]. By employing molecular biology tools and genetic approaches, several abiotic stress-inducible genes were isolated and their functions have been precisely characterized in transgenic plants [84]. The functions of some of the genes have been identified directly by the appropriate assay, or have been inferred by homology to gene of known function in the other organisms [197]. Loss of function has been very informative about the role of some of these genes. Transformation of dsRNA can trigger specific RNA degradation, in a process known as RNA-interference [198]. This process

facilitates targeted post-transcriptional gene silencing (PTGS). Using hairpin RNA (hpRNA) constructs containing sense/antisense arms ranging from 98 to 853 nt can give efficient silencing in a wide range of plant species and inclusion of an intron in these construct has a consistently enhancing effect.

Further several genes that are involved in signaling and regulatory pathways or genes that encode proteins conferring stress tolerance or enzymes present in pathways leading to the synthesis of functional and structural metabolites have been transferred into crop plants to improve their tolerance against specific stress conditions. Over expression studies using a variety of genes associated with stress tolerance pathways has been employed to generate transgenic plants. Overexpression of the cold-induced plasma membrane protein gene (MpRCI) of plantain (*Musa paradisiaca*) in tobacco resulted in increased tolerance to low temperature [199]. A NAC-type transcription factor (OsNAC5) of rice, when overexpressed in transgenic rice, caused increased tolerance to salinity [200].

In this chapter, study related to changes in transcript expression of *ERD4* gene under salt, PEG and other abiotic stress in *Brassica juncea* cv. Pusa Bold are presented. The role of *Brassica juncea* ERD4 protein in abiotic stress tolerance has been elucidated using knockdown and overexpression approaches.

4.2. Materials and methods

4.2.1. Plant materials, growth conditions and stress treatments:

Brassica juncea cv. Pusa Bold was used in all the experiments. Seeds were soaked in distilled water for 0.5 h, and then germinated in plastic Petri dishes containing filter paper saturated with distilled water in darkness at 22°C for 2 days. Seedlings were then transferred to hydroponic containers containing continuously aerated 1/2 Murashige

and Skoog (MS) liquid solution (pH 5.8, without agar and sugar). The 1/2 MS liquid solution was changed once every 3 days.

Basal expression of *ERD4* during different stages (germinating seed, root, shoot, young leaves, mature leaves and pod) of *Brassica juncea* "Pusa Bold" was checked with semi quantitative PCR. For stress treatments (salinity and osmotic) three-week-old seedlings were transferred to new 1/2 MS liquid solution (pH 5.8, without agar and sugar) under a continuous time course (0, 0.5, 1, 2, 4, 8 and 16 h). For salt and osmotic treatments, seedlings were exposed to 1/2 MS solution (pH 5.8) containing 200 mM NaCl and 20 % (w/v) polyethylene glycol (PEG). For cold and heat treatments, seedlings were exposed to the 4°C and 37°C conditions in 1/2 MS solution (pH 5.8) for one hr. For mannitol, ABA and salicylic acid treatments, seedlings were exposed for 1hr to 1/2 MS solution (pH 5.8) containing 200mM mannitol, 100 µM ABA and 100 µM salicylic acid. All seedlings were placed under the same growth conditions, except for the different treatment factors, and exposed to 1/2 MS solution at 25°C as controls. The root and shoot samples were harvested in three biological replicates for RNA preparation.

4.2.2. Isolation of RNA from plants:

100 mg leaf and root tissue of treated samples were ground into a fine powder in liquid nitrogen. The powder was transferred to tubes containing 2 ml Trizol and 400 µl of chloroform, vortexed vigorously, and then spun down at 13000 rpm for 20 minutes at 4 ^{o}C. The supernatant was transferred to a fresh tube and an equal amount of ice cold isopropyl alcohol was added. The contents were mixed by gentle inversion and kept at room temperature for 10 minutes to aid RNA precipitation. The samples were spun at

13000 rpm for 20 minutes at 4 ºC to collect the RNA pellet. The pellet was washed with ice cold 70% ethanol, dried in air and dissolved in 25µlof DEPC water.

For checking the RNA integrity, the samples were loaded on 2% agarose gel with 1 kb ladder. Two intact bands representing 28S and 18S RNA was visualized. The concentration of RNA was determined by measuring the absorbance at 260nm in a spectrophotometer by diluting 5µl of RNA sample with 995µl of DEPC autoclaved water. The ratio of 260nm and 280nm (A_{260}/A_{280}) was recorded as an estimate of RNA purity with respect to contaminants that absorb in the U.V spectra such as protein. Pure RNA has as A_{260}/A_{280} ratio of 1.8- 1.9. RNA concentration (µg/ml) was calculated using formula: 40 x Dilution factor x OD_{260}.

4.2.3. cDNA Synthesis:

First strand cDNA was synthesized from 2.5µg total RNA. The oligo (dT) primer was used for cDNA synthesis so that the same cDNA pool could be used for internal control and the target gene transcript expression analysis. To minimize the potential effects of the efficiency of synthesis during the reverse transcription reaction, three separate cDNA syntheses were performed and pooled for each RNA preparation. The cDNAs were then stored at −20°C until used for real-time PCR. In order to check DNA contamination, −RT control reaction was kept. The forward (BEEF: 5′TCCGTGAAGCTTTCACTTCC3′) and reverse (BEER: 5′GTTGGCTAAAGGTTCCG TTG 3′) primers for *ERD4* and, forward (*AraAct*F: 5′GGCTCCTCTTAACCCAAAGG 3′) and reverse (*AraAct*R: 5′ CAGTAAGGTCACGTCCAGCA3′) primers for *β actin* were designed using Primer3 software [201]. Real-time quantitative RT-PCR was carried out. The *Brassica juncea* β-actin gene was amplified in parallel with the target gene

ERD4, for gene expression normalization and providing relative quantification. Detection of real-time RT-PCR products was done using a SYBR Green master Mix kit. The quantity of cDNA used as a template for PCR was 2.0µl (the equivalent of 250ng of total RNA).

4.2.4. Real time PCR for transcript quantification:

The 2µl of diluted cDNA from each sample was taken for the real time RT PCR quantification of *ERD4* gene. *Arabidopsis thaliana actin* gene was also run in parallel with the target gene which allowed the gene expression normalization. The detection of real time PCR product was done using the Sybr Green 2X Master Mix kit (Sigma 048K6272). the quantity of cDNA used as a template for PCR was 2µl and the PCR cycling condition was comprised of an initial cycle at 94^0C for 5min, followed by at 94^0C for 30 sec; 55^0C for 30 sec and 72^0C for 20 sec. For each sample, reactions were set up in triplicates to ensure the reproducibility of the results. At the end of each PCR run, melting curve was generated and analyzed with the dissociation curve software. The melt curve obtained depends on the GC/ AT ratio and the overall length of the amplicon. This analysis allowed products to be distinguished from one another and also to be identified primer dimmers and other erroneous dsDNA. The exact quantification was done using the software REST-MCS.

4.2.5A. Preparation of siRNA construct for the *ERD4* gene:

A 442 nt piece of *ERD4* gene coding region was amplified from *Brassica juncea* cDNA using a primer that added *Xho*I and *Kpn*I site on the ends of one product (sense forward: CTCGAGAGGCTCTTGAGACGAAACCA, reverse: GGTACCACAGAAAC AACGCCGCTAGT) and *Cla*I and *BamH*I site on the ends of the other product

(antisense, forward: GGATCCAGGCTCTTGAGACGAAACA, reverse: ATCGATACA GAAACAACGCCGCTAGT). These two amplified products were cloned separately in TA cloning vector. After restriction digestion these two products were further directionally cloned into Hannibal (EHSA) andtransformed in *E. coli* XL1 blue strain. Double digestion with *Xho*I and *Kpn*I and *BamH*I and *Cla*I confirmed the presence of sense and antisense arm. EHSA clone was digested with *Not*I and upper band of *Not*I-digested product of EHSA clone was then ligated with plant expression vector pART27 (digested with *Not*I). Plant expression vector with expression cassette was transformed into *E. coli* XL1 blue strain. Blue white screening method was used for selection and 10 colonies were double digested with XhoI and BamHI for confirmation of expression cassette. The siRNA construct was then transformed in *Agrobacterium Eha105*. Presence of siRNA construct was confirmed by PCR using primers for *nptII* gene and CaMv promoter region.

4.2.5B. Construction of overexpression cassette of *ERD4* gene:

For construction of overexpression cassette of *ERD4* gene, pART27-ERD4-GFP construct prepared for the subcellular localization study was used (details given in section 2.2.9A). pART27-ERD4-GFP construct was digested with *KpnI* and *HindIII* restriction enzyme and *GFP* gene was released. The digested construct was blunted and ligation was performed. The resulted product named as pART27-ERD4 consisting of CaMV-*ERD4*-OCS expression cassette in pART27 plant expression vector was used for overexpression study.

4.2.6. Transformation of *Agrobacterium tumefaciens EHA105*:

Agrobacterium tumefaciens Eha105 competent cells were prepared. To an aliquot of 100 µL of the resuspended cells, 2 µg of the transformation vector DNA was added. Electroporation was performed using multiporator (Eppendrof make). One ml LB was added to each tube and incubated at 28 ºC for 2 hours and then plated on LB plates (rifampicin 100 µg/ml, Spectinomycin 100 µg/ml). Transformed colonies appeared in 2-3 days after the plates were incubated at 28 ºC.

4.2.7. *Agrobacterium* mediated Arabidopsis transformation of *ERD4* siRNA and overexpression constructs:

The floral dip method developed by Clough and Bent was adapted [202]. Wild type Columbia was used to transform with knockdown transformation vectors. Plants were sown in soil that was sterilized by presoaking it in fertilizer for 10 minutes. The seeds were stratified for 5 days and then moved under constant light at room temperature. Once the first two true leaves were formed, they were pruned so that the number of plants per pot was ~25. The first bolts were cut to encourage the formation of secondary bolts. At this stage plants that did not produce the first bolts were removed. When most plants had produced inflorescences, they were once again pruned such that none of the plants possessed fruits or flowers that might have been pollinated. The plants were then bound at the base with elastic bands to keep the soil from dropping into the transformation solution. The transformation solution was made of 5 % sucrose solution and resuspended Agrobacterium cells. It was subjected to gentle stirring until the detergent was mixed completely. The plants were dipped into the transformation solution by inverting the pot into it and swirled gently for exactly 10 seconds. The dripping sucrose solution was

carefully drained off from the plants. The plants were placed on slanting trays, covered with plastic wrap and kept under low light for 24 hours. The wrap was slit to allow slow entry of air for the next 24 hours. After a total of 48 hours, the plants were returned to their erect positions and allowed to continue to grow until the maximum amount of seeds were harvested.

4.2.8. Screening of transgenic plants:

The bulk harvested T1seeds were surface sterilized by vortexing them in 1 ml of 70% alcohol twice for 10 minutes each. Then seeds were thoroughly washed five times with autoclaved distilled water by vortexing them for 10 minutes each time. The sterilized T1 seeds were spread evenly on MS plates with kanamycin (100 µg/ml). The seeds were stratified at 4 ºC for 7 days, moved to room temperature under low light to induce germination. Resistant plants that grew on kanamycin-MS plates were considered as T1 transformants. They were transplanted into sterilized soil, allowed to self pollinate and individually harvested to collect T2 seeds. The T2 seeds were surface sterilized as described above and plated on kanamycin-MS plates. The resistant vs self pollinate and individually harvested to collect T2 seeds. The T2 seeds were surface sterilized as described above and plated on kanamycin-MS plates. The resistant Vs sensitive segregation ratio of T2 seeds on kanamycin-MS plates was used to identify single (3:1). The resistant T2 plants were transplanted into sterile soil, allowed to self pollinate, and individually harvested to collect T3 seeds. The T3 seeds that gave a 100% kanamycin resistance when grown on kanamycin-MS plantes were considered homozygous for single insertion of the gene-of-interest. These lines were used for further analysis.

4.2.9. Characterization of *BjERD4* knockdown and overexpressed lines

4.2.9.1. Molecular characterization of knockdown lines:

Based upon kanamycin selection, six knockdown and overexopressed lines were selected separately. Genomic DNA was isolated from all these lines. PCR was performed using primer for *nptII* gene (forward- TGTTCCGGCTGTCAGCGCAG, reverse- GATCCTCGCCGTCGGGCATG) and the PCR cycling condition was comprised of an initial cycle at 94^0C for 5min, followed by at 94^0C for 1 min; 55^0C for 1 min and 72^0C for 30 sec. To see the change in transcript level of *ERD4* gene in knock down and overexpressed lines, quantitative real time PCR method was used. Total RNA isolation, cDNA preparation and real time PCR was performed as explained under expression analysis studies (sections 5.2.2 and 5.2.3).

4.2.9.2. Determination of germination:

Wild-type and erd4 mutant seeds sown on 1/2 MS were cold-stratified at 4^0C for 2 days in the dark and then incubated at 22^0C under the long-day condition for 5 days. Each day, the number of germinated seeds with protruding radicals was counted (Oh et al., 2006).

4.2.9.3. Assessment of Knockdown and overexpressed lines under stress treatment:

Wild-type, *Bjerd4* and overexpressed lines seeds sown on 1/2 MS were cold-stratified at 4^0C for 2 days in the dark and then incubated at 22^0C under the long-day condition for 5 days. Ten days old knockdown and wild type seedlings were transferred to different treatment conditions (Control-MS, Salinity- MS+100mM NaCl and Drought- 10% PEG or 150mM Mannitol) for 7 days. Performance of knockdown and

overexpressed lines were compared with wild type in terms of chlorophyll content, lipid peroxidation, GSH content and ROS production using DAB and NBTdye.

A. Determination of chlorophyll content:

Total chlorophyll content was determined spectrophotometrically according to the method described by Arnon [203]. The 300 mg leaves were grinded into powder with liquid nitrogen and then were transferred to a 15 ml Falcon tube. 5 ml of 80% acetone was added to the tube and mixed thoroughly. Centrifugation was performed at 4°C for 15 min (3,000 rpm). Supernatant was transferred to a new centrifuge tube and the absorbance of chlorophyll was measure using spectrophotometer. The chlorophyll concentrations are calculated as follows (80% acetone as a blank control was used as blank). Ca+b (mg/g) = $[8.02\times A_{663}+20.20\times A_{645}]\times V/1000\times W$. Where V = volume of the extract (ml); W = Weight of fresh leaves (g).

B. Estimation of lipid peroxidation (LP):

The LP level in plant tissues was determined by measuring the malondialdehyde (MDA) content via the 2-thiobarbituric acid (TBA) reaction [204]. Leaf tissue (100 mg) was homogenized in 1 ml of 10 mM sodium phosphate buffer (pH 7.4) and centrifuged at 4000 g for 5 min at room temperature. A 200 µl aliquot of the supernatant was added to a reaction mixture containing 100 µl of 8.1% (w/v) SDS, 750 µl of 20% (w/v) acetic acid (pH 3.5), 750 µl of 0.8% (w/v) aqueous TBA, and 200 µl of Milli-Q water. An identical reaction mixture in which 200 µl of supernatant was substituted by an equal volume of buffer was simultaneously set up as a blank. Both reaction mixtures were then incubated at 98 ^{o}C for 1 h. After cooling to room temperature the mixtures were centrifuged for 5 min. Absorbance at 535 nm was measured and corrected for non-specific absorbance at

600 nm. The level of LP was expressed as µmol of MDA formed derived from the difference in absorbance at 535 nm and 600 nm using an extinction coefficient of 156 mM^{-1} cm^{-1}.

C. Measurement of glutathione content:

For estimation of reduced (GSH) plant material (500 mg) was frozen in liquid nitrogen and homogenized in 0.1 M phosphate- EDTA buffer (pH 8.0) containing 25% meta-phosphoric acid. The homogenate was centrifuged at 20,000 g for 20 min at 4^{o}C. GSH content was determined fluorometrically in the supernatant after 15 min incubation with o-phthaldialdehyde (OPT) [205]. Fluorescence intensity was recorded at 420 nm after excitation at 350 nm on a fluorescence spectrophotometer.

D. Qualitative Assay of H_2O_2:

Detection of H_2O_2 was performed by infiltrating leaves with a solution of 1 mg mL^{-1} DAB in MES buffer (pH 6.5) as described by Thordal-Christensen et al. [206]. H_2O_2 was visualized as a reddish-brown coloration. Prior to imaging, chlorophyll was removed from leaves with 70% (v/v) ethanol.

E. Qualitative assay for superoxide radicals:

For visualization of superoxide generation as a result of abiotic stress, 2-week-old seedlings treated with mannitol (200mM), NaCl (150mM), for 7 d, along with unstressed plants were employed for NBT staining. Stained samples were transferred to 80% ethanol and incubated at 70°C for 10min to remove the chlorophyll. The NBT staining method described by Wang et al. [207] was used for superoxide detection.

4.3. Results

4.3.1. RNA isolation and spatial expression analysis:

The quality of the isolated total RNA was assessed in terms of spectrophotometric absorbance (*A*260/*A*280). These values ranged 1.9 to 2.0, indicating good purification. Further, the total RNA samples were separated on formaldehyde denatured agarose gel. The gel electrophoresis revealed integrity of the extracted RNA and suitability for further transcript expression analysis (Fig 4.3.1A).

Basal expression of *ERD4* was checked in different stages of plant growth of *Brassica juncea* using semi quantitative method. ERD4 transcript expression was detected in germinating seeds, root, young and mature leaves, shoot, and in pod. Maximum basal expression was observed in young leaves (Fig. 4.3.1B).

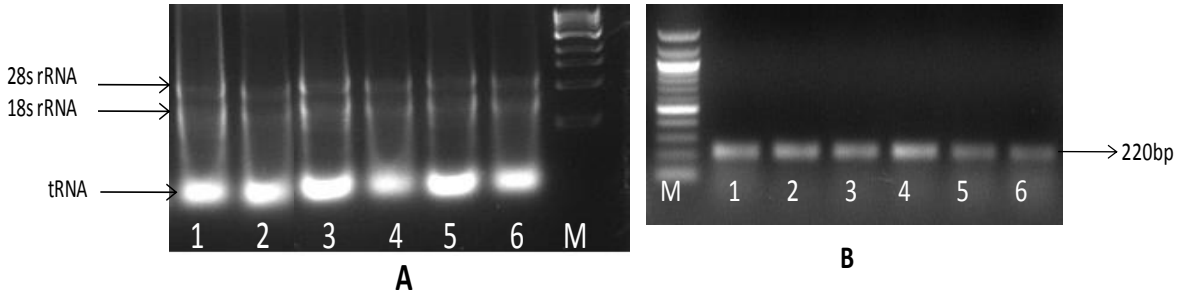


Fig. 4.3.1.A & B: RNA isolation from different plant parts (A) and *ERD4* gene expression using semiquantitative PCR (B). Gel was run for the PCR product after 28 cycles. M. Marker, 1.Germinating seed, 2.Root, 3.Shoot, 4.Young leaves, 5. Mature leaves and, 6. Pod.

4.3.2. Temporal expression analysis of *ERD4* gene under different abiotic stress conditions:

To investigate the effect of different stress conditions on the expression of the *BrassicaERD4* gene, real time PCR was performed using total RNA isolated from stressed and non stressed leaves and roots of *Brassica juncea* cv. Pusa bold. In case of salinity treatment, significant increase in *BjERD4* gene expression was detected as early as within 0.5 hr of treatment (2.165 fold) which reached maximum after 4 hr of treatment (6 fold) in roots, while in shoots maximum expression was observed only after 2 hr of treatment (2.7 fold induction). With increasing time of treatment, down expression of this gene was observed (Fig.45.3.2A).

In case of PEG treatment, early induction of *BjERD4* gene was observed in root (2.75 fold) after 0.5 hr of treatment, which reached maximum after 1hr of treatment (4.491 fold). In shoots also, significant fold increase was observed within 0.5 hr of treatment whereas maximum induction was seen after 4 hr of treatment (1.4 fold) (Fig 5.3.2B). Significant increase in *BjERD4* gene expression was also detected in root and shoot within 1hr when treated with mannitol (100mM), ABA (100µM), SA(100µM), cold (4^0C) and heat (37^0C) (Fig. 4.3.2C).

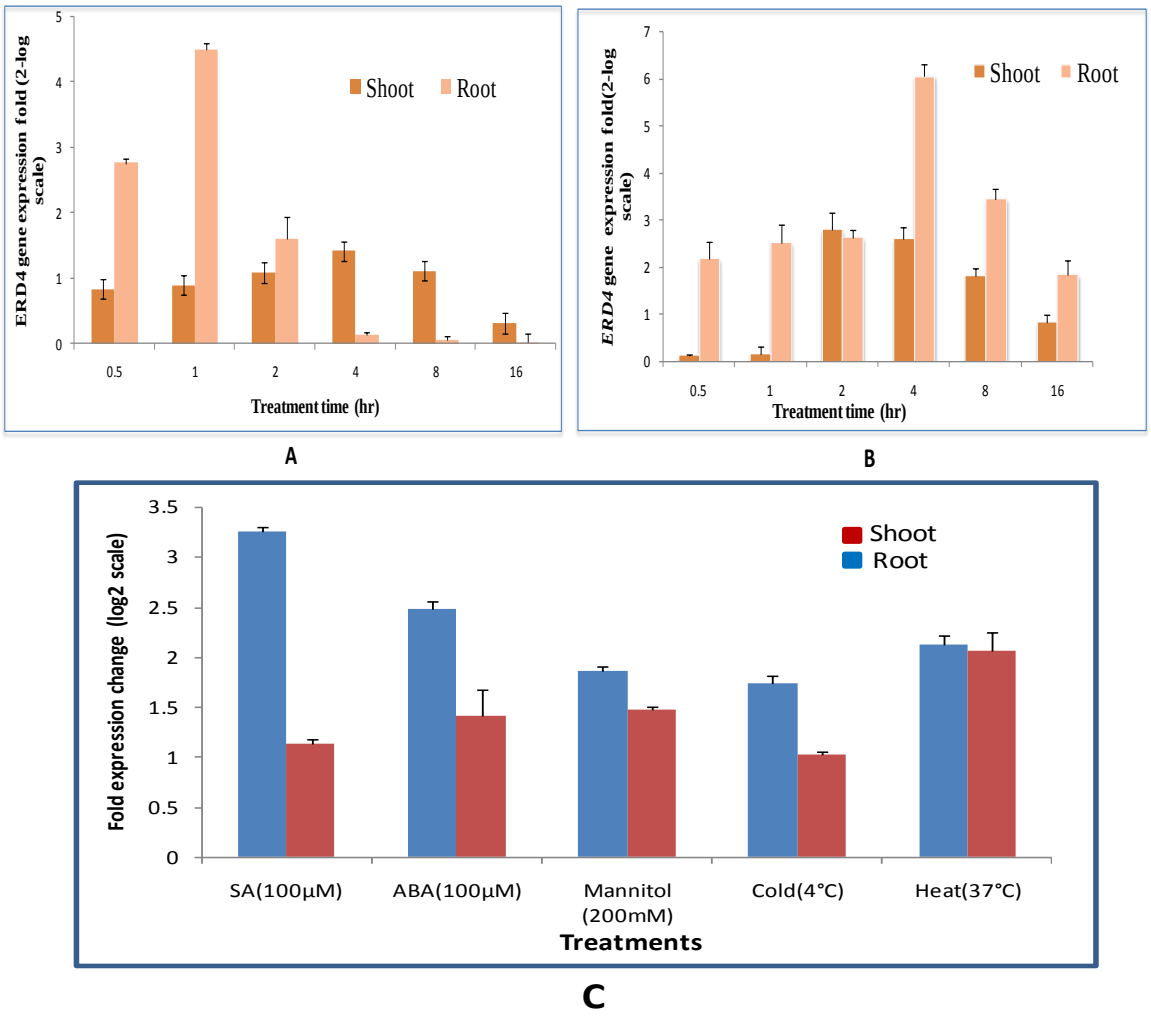


Fig4.3.2: Expression analysis of *ERD4* gene under abiotic stress conditions. Change in *ERD4* transcript under drought (A), salinity (B) at different time periods (temporal expression) and other abiotic stress after 1hr (C).

4.3.3. Preparation of Knock-down and overexpression constructs, transformation and screening of *erd4* and *OERD4* lines:

To understand the functional role of *BjERD4*, knockdown and overexpressed Arabidopsis lines were generated. For knockdown construct, a 442 bp long fragment was

cloned in sense and antisense orientation in Hannibal vector under CaMV promoter. This was further subcloned in pART27 plant expression vector (Fig. 4.3.3.A, B & C). For overexpression construct full length ERD4 gene was cloned under CaMV promoter and subcloned in pART27 plant expression vector (4.3.3.D).

Knockdown and overexpression constructs were transformed in Arabidopsis using vaccum infiltration method and screening was performed on Kanamycin antibiotic for transformed lines. Six independent transgenic lines were selected based on the survival on kanamycin. All the transformed lines (T4 generation) were selected on kanamycin to attain homozygosity. The homozygous nature of these lines was also confirmed by the segregation ratio of kanamycin in all the generations (Fig45.3.3E).

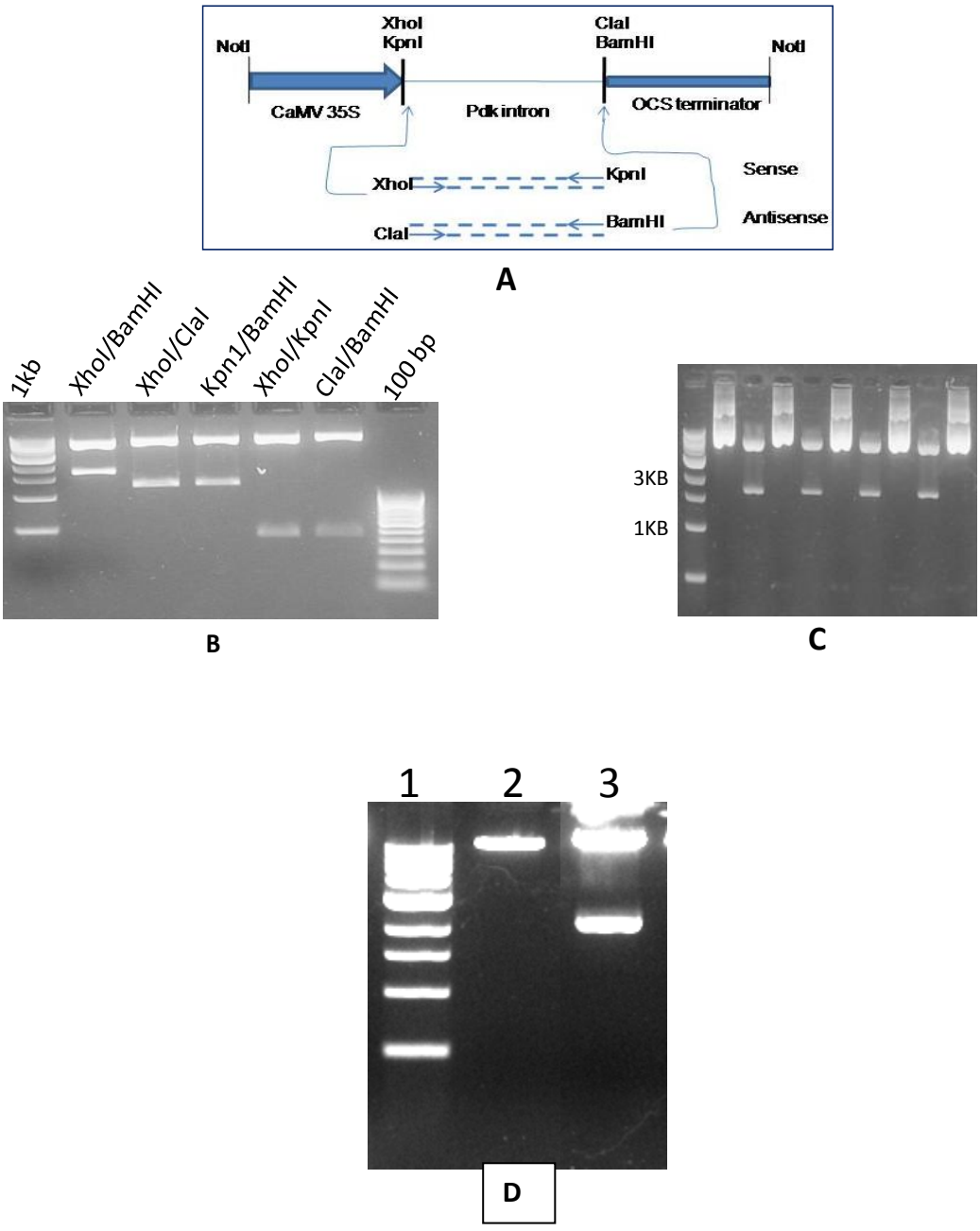


Fig. 4.3.3.A, B, C & D: Preparation of knockdown and overexpression construct. (A) Schematic representation of the construct used for the knockdown of *ERD4* in Arabidopsis. Confirmation of knockdown clone in Haniibal (B) and in pART27 vector with NotI releases 1.6 Kb expression cartridges, (C) knockdown in Arabidopsis by restriction digestion. (D) Confirmation of overexpression cassette 1.1Kb ladder, 2. pART27 (*NotI* digestion) & 3. pART27-BjERD4 (*EcoRI/KpnI double digestion)*.

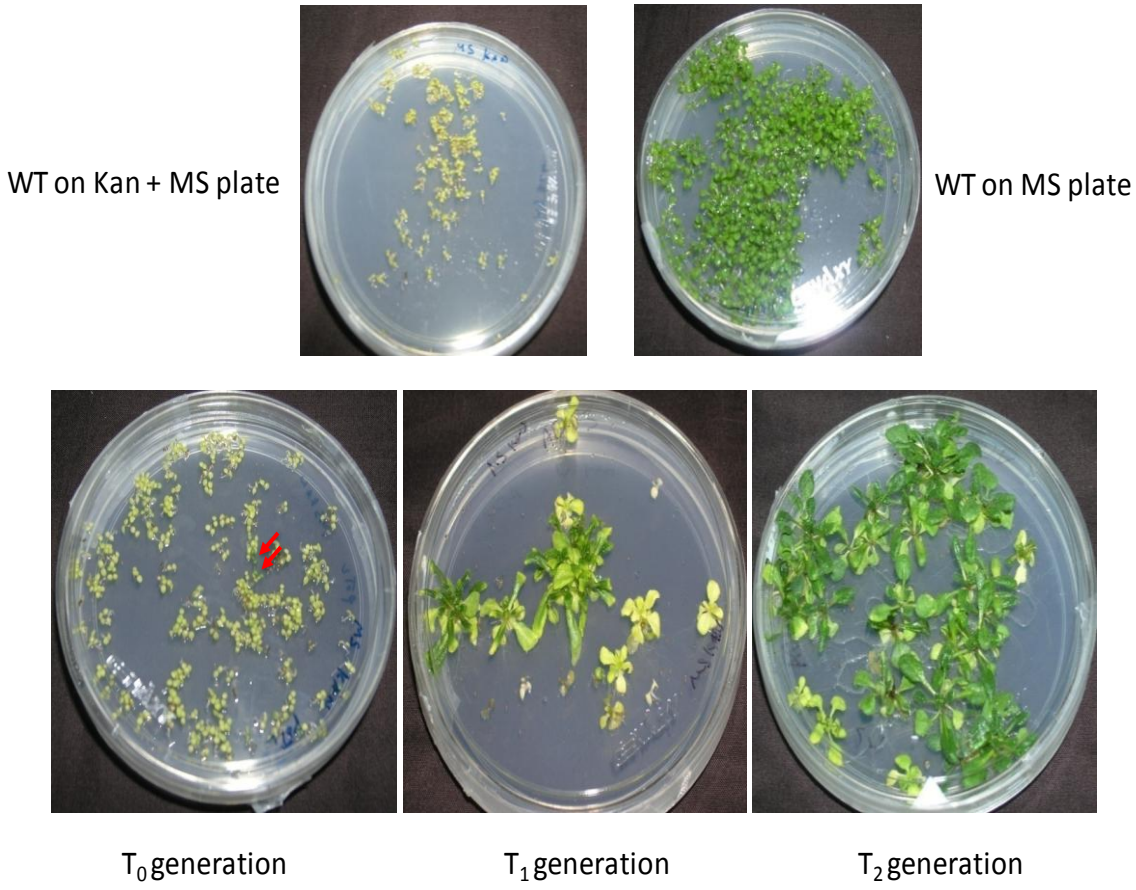


Fig 4.3.3.E: Selection of transformed lines of Arabidopsis. Transgenic lines were selected on kanamycin (100 mg/lit) on MS plate based screening.

4.3.4. Molecular characterization of Knockdown lines:

Six Arabidopsis *ERD4* knockdown and overexpressed lines were confirmed by PCR using nptII primer which showed amplification of *nptII* gene in transformed lines as in positive control, whereas no amplification product was observed in the Wild type (Fig. 4.3.4 A &C). The Real time PCR analysis was done for all these lines using primers designed to amplify specific gene sequence cloned in RNAi construct. Significant reduction of transcript in two lines was observed. Transcript reduction was up to 3.6 fold (Fig4.3.4B). In case of overexpressed lines significant increase in transcript was observed in all lines but maximum fold increase in transcript was up to 4 times. (Fig 4.3.4D)

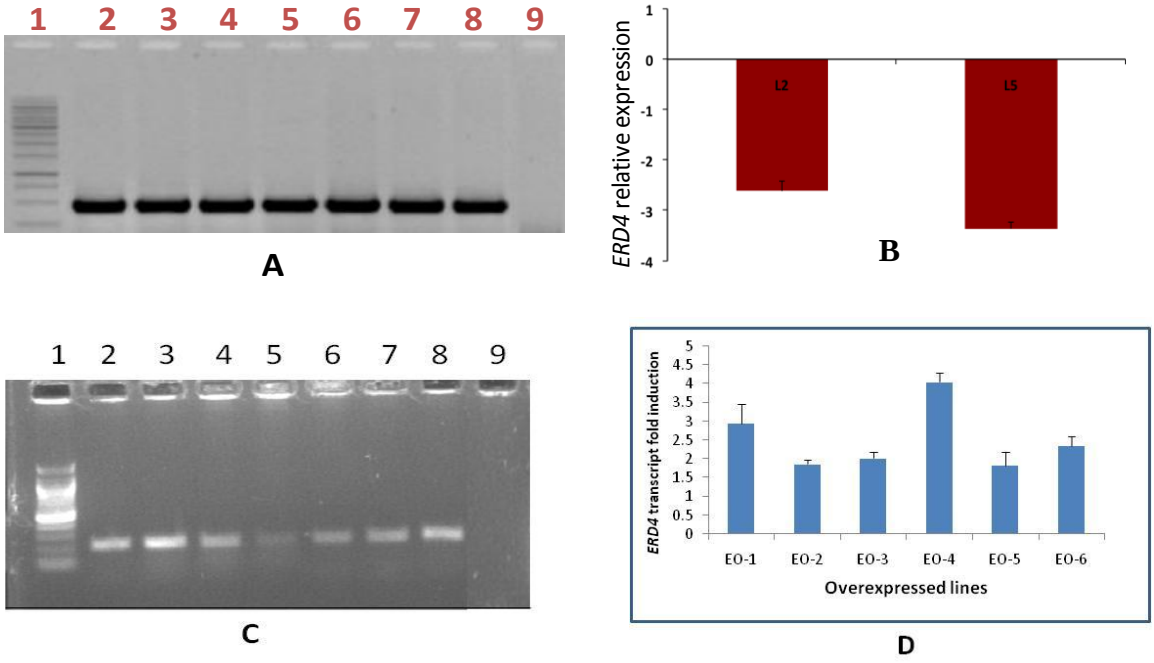


Fig. 4.3.4 A&B: Molecular characterization of transformed lines. PCR amplification with *nptII* primers of Knockdown (A) & overexpressed (C) lines . Lane 1: 1 kb ladder; Lanes 2 to 7: knockdown lines 1- 6, Lane 8: Plasmid and Lane 9. Wild type. Real time PCR with *ERD4* knockdown (B) & overexpressed (D) lines.

4.3.5. Phenotypic analysis of *ERD4* RNAi lines:

Germination was scored following seed imbibition and subsequent development of roots and shoots was monitored. Germination time was delayed by two days in most of the *ERD4*RNAi seeds. When roots and shoots were examined, the RNAi transgenic Arabidopsis seedlings exhibited marked dwarf phenotype compared to the wild type seedlings (4.3.5A&B)..

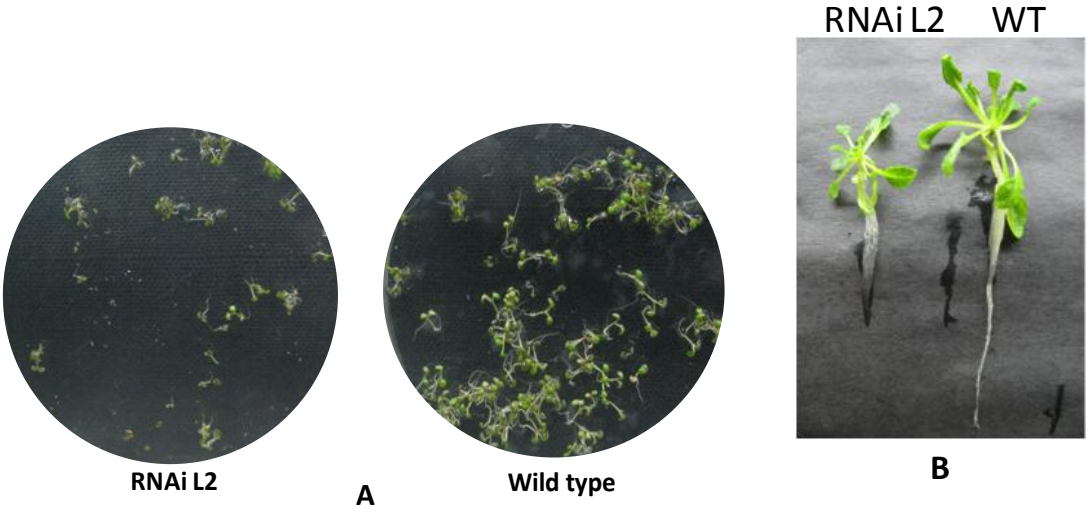


Fig. 4.3.5A&B: Germination and phenotypic difference of *ERD4* knockdown and wild type Arabidopsis

4.3.6. Performance of Knockdown lines under salinity and PEG treatment:

Ten days old Wt, RNAiL-2 and RNAiL-5 seedlings were transferred to different treatment conditions (Control-MS, Salinity- MS+100mM NaCl and Drought- 10% PEG/ 150mM mannitol) for 7 days.

Root growth: Root growth was restricted under drought and salinity stress. But reduction was more in the knockdown lines as compared to wild type plant (Fig. 4.3.6A).

Chlorophyll content: In *ERD4*RNAi plants, chlorophyll content and the carotenoid pigment per fresh wt was significantly decreased compared to control, while the Chl a/Chl b ratio was slightly higher in knockdown plants compared to WT. Salinity and drought stresses showed significant decrease in total chlorophyll content and the carotenoid pigment in both wild type and knockdown lines (Fig 4.3.4B).

MDA assay: As a marker of oxidative damage under abiotic stress, lipid peroxidation was measured by the TBARS assays. On the 7^{th} day of exposure to salt and PEG, lipid peroxidation showed drastic increase of 225% and 145% in the shoot of knockdown seedlings compared to the WT seedlings (Fig4.3.4C).

Redox state of ERD4 RNAi lines: High cellular GSH is crucial for the redox state of the cell which determines the survival of cells under any adverse condition and thought to act as redox sensor. When the concentration of GSH was measured in WT and *ERD4* knockdown Arabidopsis plants, it was found that sodium chloride treatment did not perturb the GSH content in WT plants whereas, significant decrease (53%) was observed in knockdown lines (Fig 4.3.4D).

Qualitative Assay of H_2O_2: To test whether down expression of *ERD4* gene leads to ROS production under stress conditions, wild type and knockdown lines leaves were stained with diaminobenzidine (DAB) for the detection of H_2O_2. As compared to wild type, intense staining was observed in knockdown lines compared to control implying that H_2O_2 level was highin knockdown lines under salt and drought stress (Fig 4.3.4 E).

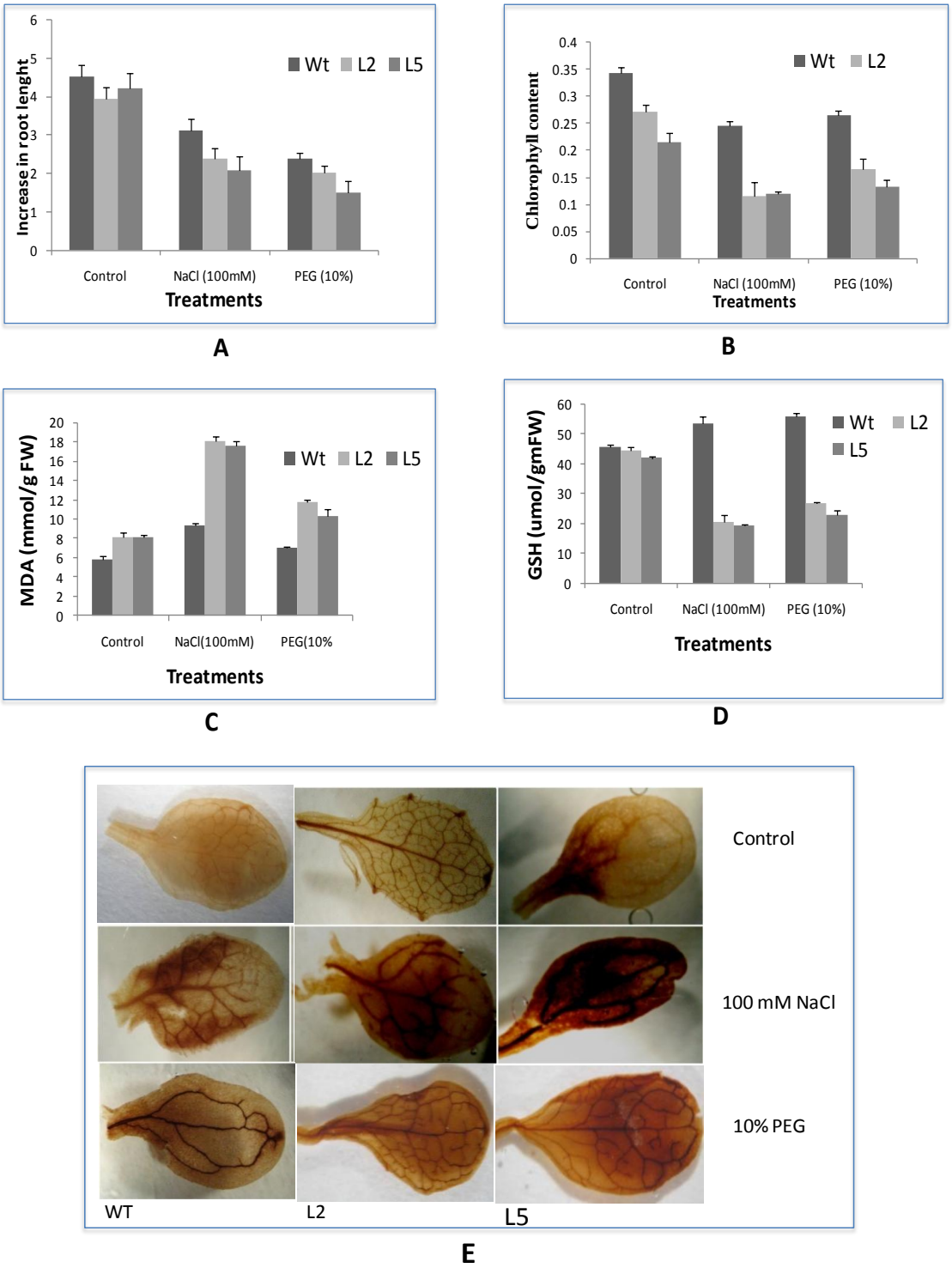


Fig 4.3.6: Performance of knockdown lines under salinity and drought. Change in root length (A), Chlorophyll content (B), MDA content, (GSH content) and Qualitative Assay of H_2O_2 (E).

4.3.7. Phenotypic analysis of *ERD4* overexpressed lines:

Overexpressed lines showed significant change in plant morphology i.e enhanced plant size with more number of branches and increased leaf and pod size as compared to wild type when grown in MS medium (Fig 4.3.7).

Fig. 4.3.7: Morphological changes in overexpressed lines

4.3.8. Assessing the performance of overexpressed lines under salinity and mannitol treatment:

Ten days old Wt, EO-1 and EO-4 seedlings were transferred to different treatment conditions (Control-MS, Salinity- MS+150 mM NaCl and Drought- 150mM mannitol) for 7 days.

Chlorophyll content: In *ERD4* overexpressed plants, chlorophyll content per fresh wt was significantly increased compared to wild type under control condition. Salinity and

drought stresses showed significant decrease in total chlorophyll content in both wild type and overexpressed lines (Fig 4.3.8A).

MDA assay: Lipid peroxidation was measured by the TBARS assays. On the 7^{th} day of exposure to salt and mannitol, MDA content showed drastic increase of 110% in the WT seedlings whereas the increse in MDA content observed in overexpressed lines was not significant as compared to wild type. (Fig. 4..3.8 B).

Qualitative Assay of hydrogen peroxide and superoxide radical: To test whether down expression of *ERD4* gene leads to ROS production under stress conditions, wild type and overexpressed lines seedlings were stained with diaminobenzidine (DAB) for the detection of H_2O_2 and NBT for superoxide radical. As compared to wild type, intense staining was observed in wild type compared to overexpressed lines implying that H_2O_2 and super oxide radical level were high in wild type under salt and drought stress (Fig. 4.3.8 C& D).

4.4. Discussion:

To study the gene function and role in plant development, *ERD4* gene was amplified from *Brassica juncea*. The organ specific expression of *BjERD4* in *Brassica juncea* was detected by semiquantitative PCR. This expression pattern in germinating seeds, root, young and mature leaves, shoots, and in pod indicated that BjERD4 may also function in the normal programme of the plant growth and development. The early induction of gene expression was observed as early as 0.5 hr under drought and salinity treatments in *Brassica juncea*. Significant increase in *ERD4* transcript under other stress factors like mannitol, heat and cold was also observed. The early induction of *ERD4* at transcript level under these abiotic conditions was also observed in Arabidopsis [208],

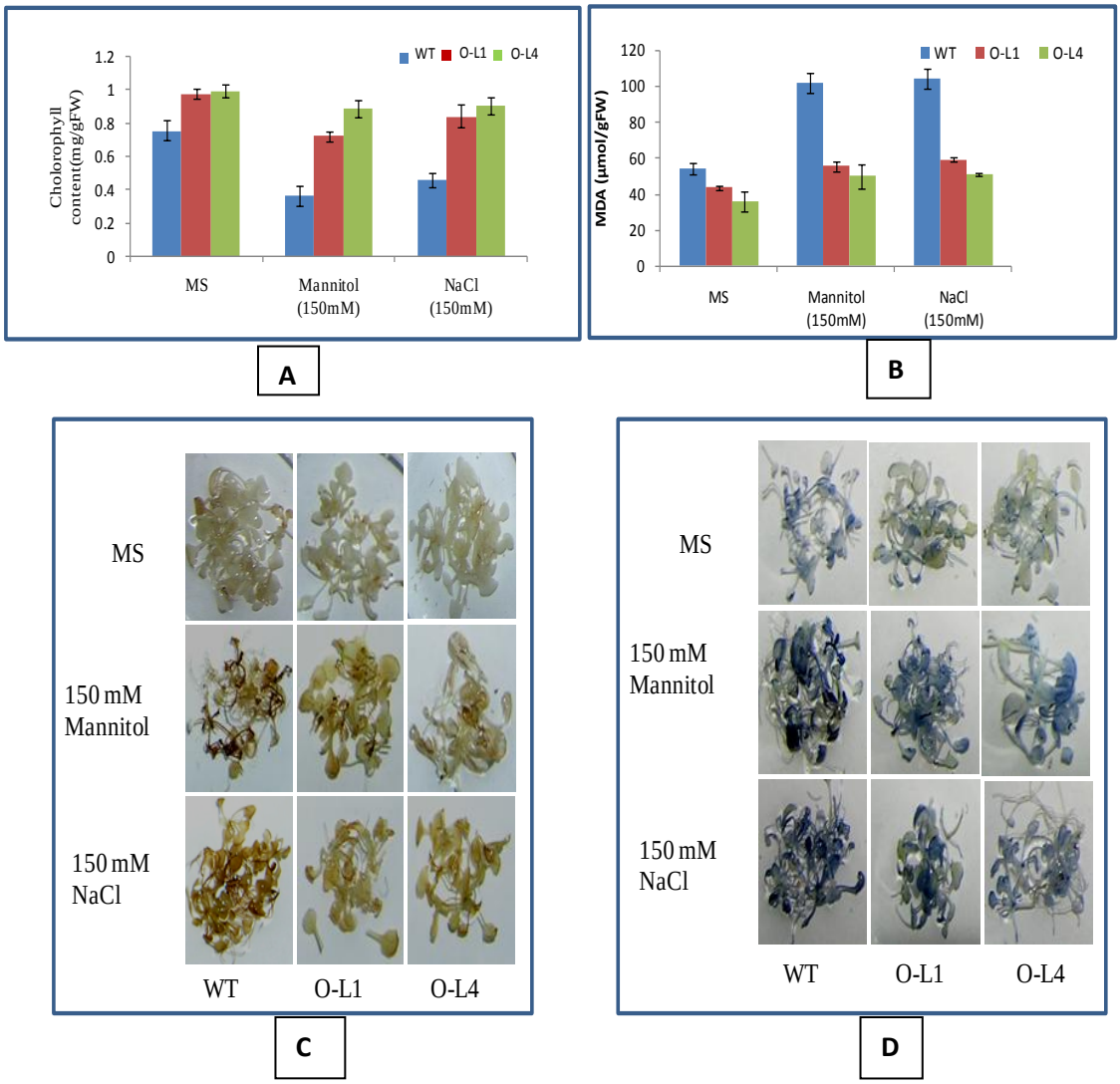


Fig. 4.3.8: **Performance of overexpressed lines under salinity and drought.** Change in (A) Chlorophyll content, (B) MDA content, and Qualitative Assay of H_2O_2 (C) & superoxide radical (D).

sugarcane [127] and maize [126]. Different subsets of ERD family members have been shown to be up-regulated or downregulated by various other environmental stimuli, such as cold [209], light [210], excessive arsenate [211] or a transient increase in cytoplasmic Ca^{2+} [212].

Rapid adaptation to changing environmental conditions is essential for plant survival and for the development of tolerance to both abiotic and biotic stresses. Such tolerances can be achieved by distinct metabolic and physiological adjustments, which are mediated by a number of plant hormones, and are often specific to a certain type of stress [52]. As a central regulator of plants' adaptation to environmental stress, ABA plays a crucial role in the regulation of transpirational water loss [31, 213]. The early responsive to dehydration (ERD) gene is one of the key negative regulators of ABA responses in plants. Changes to the abundance of ERD transcript abundance modulate ABA responsiveness in Arabidopsis. Song et al. demonstrated that the expression of ERD3, ERD4, and ERD7 responded rapidly to water deficit [214]. This implies that ERD genes are rapid drought responsive genes, as are the ERD genes in Arabidopsis. The ERD gene family has at least 21 members. The ERD3, ERD4, and ERD7 showed different expression patterns, indicating that various members of the ERD gene family may have separate functions in the water stress response.

In Arabidopsis *ERD4* gene induction was reported before the accumulation of ABA, but in *B. juncea* we noted a significant increase in *ERD4* transcript when exogenous ABA and SA treatment were given. Similarly the ZmERD4 was also induced in the presence of ABA [126]. This suggests that *ERD4* gene expression is also modulated by hormones like ABA which is mainly a stress hormone [21]. Thus based

upon the expression analysis, it can be concluded that *BjERD4* expression is constitutive as well as inducible.

To further confirm the role of *BjERD4*, *ERD4* knockdown and overexpression lines were generated in Arabidopsis. In knockdown lines, seed germination was delayed compared to wild type and at a later stage (at 10 days) the developmental difference was also visible with dwarf plant type in case of knock down mutant. But in case of overexpressed lines the enhanced growth was observed in terms of leaf size, pod size, and branching under control condition itself. This suggests that this gene has pleiotropic effect and may be playing important role in plant development.

When knockdown mutant lines were tested for their performance under salinity and drought conditions, knockdown lines showed significant difference with the wild type plants (Fig 5.3.6). Significant increase in MDA content and ROS production was observed whereas chlorophyll content and GSH content were decreased. On the other hand, in case of overexpressed lines, increase in MDA content and ROS generation was not significant as compared to wild type plants. was It has been shown in Chapter III that the protein was localized in chloroplast (plastids) which is one of the major sites for reactive oxygen species production. Further, chloroplast is also the major site for the production of antioxidant compounds and enzyme like GSH, ascorbate, SOD etc [215]. In higher plants, pigments mainly accumulate in the thylakoid membrane of chloroplasts, where they function to harvest light and protect the photosynthetic apparatus from oxidative damage by quenching the triplet excited state of Chl (3Chl) and reactive singlet oxygen (^{1}O2) and dissipating excess energy [216, 217]. There is a possibility that decrease in production of GSH, other antioxidants and enzymes may be because of

structural changes in chloroplasts in knockdown lines. This ultimately may lead to poor defense against ROS and hence plants become more susceptible as compared to wild type. Chloroplast localized RNA interference (RNAi) study of *Ostrxm* rice plants has also shown developmental defects, including semi-dwarfism, pale-green leaves, abnormal chloroplast structure, and reduced carotenoid and chlorophyll content. OstrxmRNAi plants showed remarkably decreased Fv/Fm values under high irradiance conditions (1,000 μmol m^{-2} s^{-1}) with delayed recovery [218]. Lack of ERD10 protein accumulation (Early responsive to dehydration 10) in late stage of seed maturation resulted in the reduction of germination of erd10 mutant seeds. In addition ERD10 mutant also showed reduced tolerance to drought stress [219]. In contrast, overexpression of *BjERD4* enhanced tolerance in transgenic plants to both drought and salt stress as it was also reported earlier for *ZmERD4* [126].

ERD15 from Arabidopsis is functionally characterized as a common regulator of the abscisic acid (ABA) response and the salicylic acid (SA)-dependent defense pathway [52]. The overexpression of ERD15 reduced the sensitivity to ABA, as the transgenic plants were less tolerant to drought and were impaired in increasing their freezing tolerance in response to ABA. In contrast, the loss of the ERD15 function caused a hypersensitivity to ABA, and the silenced plants displayed enhanced tolerance to both drought and freezing [52]. A NAC-type transcription factor (OsNAC5) of rice, when overexpressed in transgenic rice, caused increased tolerance to salinity [200]. Overexpression of OsLEA3-1 in rice resulted in enhanced tolerance to drought under field conditions [220]. Similarly, expression of DREB1A of Arabidopsis, hybrid-proline-rich protein (CcHyPRP) genes of pigeonpea (*Cajanus cajan* L.), and TaSnRK2.8 and

TaSnRK2.4 of wheat in Arabidopsis conferred tolerance to drought, salinity, and extreme temperatures [65, 221, 222].

Post-transcriptional metabolism of RNA involves both housekeeping and regulatory mechanisms. These processes require the interaction of RNA-binding proteins with specific RNA sequences [71, 75]. But at the primary sequence level RNA binding proteins are poorly conserved, making it difficult to detect any domain in ERD4 primary sequences, while the structural analysis is known to reveal the distant evolutionary links that could provide the first hypothesis about biological function of the uncharacterized domains [57]. Hence by using fold-prediction algorithms it was possible to identify the presence of two RNA-recognition motifs in its sequence [223]. For validation of this hypothesis the RNA EMSA was performed. Results suggested that ERD4-RRM domain binds to only RNA but not the ssDNA and binding is probably sequence specific.Independent estimations predict the existence of about 60 chloroplast RNA-binding proteins (cpRBPs) in *A. thaliana*, based on the computer-assisted analysis of putative chloroplast-targeting signals. Different plastid RBPs possess differentially regulated RNA-binding activities [224]. Some of these exhibit sequence-specific and, thus, gene-specific binding affinities, while others represent more general RBPs that might establish a protein scaffold for chloroplast transcripts enhancing RNA stability and/or RNA-folding. Previous studies have supported translational roles of 46- and 47-kD RNA-binding chloroplastic membrane protein [225]. RB47 is associated with the second class of low density chloroplastic membranes, has been proposed to activate the translation of the chloroplast *psbA* mRNA [226]. Several previous reports suggest a role of the inner envelope membrane in chloroplast gene expression and thylakoid biogenesis

[225]. Extensions of the inner envelope membrane and membrane vesicles in the stroma have been observed by electron microscopy in the chloroplasts of *C. reinhardtii* [227], tobacco, pea, soybean and spinach [228], and in the chromoplasts of red pepper fruits [229]. It has also demonstrated the presence of a homologue of the *Escherichia coli* ribosome releasing factor associated with the chloroplast envelope in spinach [225]. Thus, these evidences suggest that possibly translation of chloroplast mRNAs occurs at the chloroplast inner envelope membrane. Probably ERD4 protein could be one of the components involved in the protein biosynthesis or may participate in post-transcriptional modification of unidentified target RNAs or might hold a pool of mRNAs at the chloroplast envelope as a reserve to sustain protein synthesis during stress conditions of limiting transcription. Further studies are required to identify the target RNAs of ERD4 and to define how ERD4 regulates post-transcriptional modification during stress conditions.

CHAPTER V: CONCLUSIONS AND FUTURE DIRECTIONS

Conclusions:

Plants are often exposed to various biotic and abiotic stresses in their natural habitats. Abiotic stresses limit plant growth and productivity and in crop plant lead to significant yield reduction. Thus the study has been taken up to get insight into the molecular characterization of one of the important abiotic stress tolerance genes, *Early Responsive to Dehydration (ERD4)*. Investigations related to its structure, function, localization and role in abiotic stress tolerance of *BjERD4* gene have been made. *ERD4* gene was cloned from *Brassica juncea* and sequence comparison showed the presence of five introns in this gene. Motif scan results showed the presence of a transmembrane DUF221 domain and Topology prediction suggested the presence of one more motif which was globular and posited inside the membranous structure with at least nine transmembrane helices.

Subcellular localization of any protein influences protein function by controlling access to different molecular interactions. Studies on localization of ERD4 protein were performed using bioinformatics as well as GFP tagging approaches. *In silico* studies established that the protein is localized in chloroplast membrane and confocal microscopy of transformed protoplasts with *GFP* tagged *ERD4* coding sequence confirmed the chloroplastic (plastids) localization of ERD4.

It is well established that proteins evolve partly through rearrangements of larger fragments, typically domains, and nature of these fragments determine biological function of proteins. The analysis of proteins at individual domain levels can facilitate functional annotation of uncharacterized genes and proteins. Thus the globular domain was characterized for its structure and function. By fold-prediction algorithms, the

presence of two RNA-recognition motifs was detected in ERD4 protein sequence along with the presence of two RNP sites for RNA binding. These RNA binding sites were found to be conserved when multiple sequence analysis was performed with orthologs of ERD4 protein. For validation of these RRM domains in vitro RNA was synthesized and interaction with RRM domain was studied with these RNA using RNA EMSA. This showed the specific interaction of RRM domain with only RNA molecule.

The ERD gene encoded proteins show a great structural and functional diversity, with a particular class of proteins acting as connectors of different stress response pathways. The induction of this gene under abiotic stress conditions was studied using transcript expression analysis. Results showed the early expression of this gene under drought, salinity and other abiotic stress treatments. Basal expression of this gene was also studied in different plant organ and result showed high basal expression indicated that *BjERD4* may function in the normal programme of the plant growth and development.

To further study the role of *BjERD4*, knockdown and overexpression lines were generated in Arabidopsis. The performance of knockdown mutant lines under salinity and drought conditions showed significant increase in MDA content and ROS production in knockdown lines where as chlorophyll content was found to be decreased as compared to wild type. Overexpressed lines showed significant change in plant morphology i.e. enhanced plant size with more number of branches and increased leaf and pod size as compared to wild type.

Future directions:

Elucidation of the stress responses of crop plants assumes great relevance in view of the challenges posed by abiotic stresses to crop productivity and the need to develop stress tolerant crop varieties. Towards this, full complement of stress responses will have to be analyzed by relating data on single stresses to data on multiple stress responses to provide clues about signaling "cross talk" between different stress factors.

In this study, the role of *ERD4* in abiotic stress tolerance has been validated using over expression strategy. Further transgenic experiments using *ERD4* gene as a candidate gene can be done in other crop plants so as to achieve tolerance to different abiotic stresses. The Overexpressed *ERD4* lines can be analyzed to see the modulations in plant hormones which could have contributed to superior plant growth and pod characteristics. The OE lines can be further analyzed in understanding the role of *ERD4* in other abiotic and biotic stresses tolerance and in molecular cross talk. Further studies involving microarray experiments can be conducted to fully establish different roles of *BjERD4* gene in other stress responses and plant development. Chloroplast structural changes can be studied to offer more insight in the role of ERD4 protein in plastid membrane structure. Since Arabidopsis database mining has showed nine more ERD4 like proteins, investigations can be taken up to decipher their role in relation to abiotic stress tolerance.

In this study, ERD4 protein function has been shown to have chloroplastic RNA binding activity. Further studies are warranted to identify RNAs which will bind to ERD4 protein. To dissect such RNA-protein interaction networks ('RNA interactome'), it is necessary to identify the RNAs with which each RBP interacts and to determine how those interactions influence RNA fate and downstream processes. Study of related RBPs

using next generation sequencing approaches such as RNA-seq will vastly enable to analyse RBPs and knockout mutants of RBPs.

In recent years, the area of research involving *ERD* genes has generated renewed interest in understanding the molecular mechanism of stress tolerance. Towards this end, the present study has, for the first time, shown that the *ERD4* gene and its protein have role in tolerance to different abiotic stress factors, and RNA binding activity respectively.

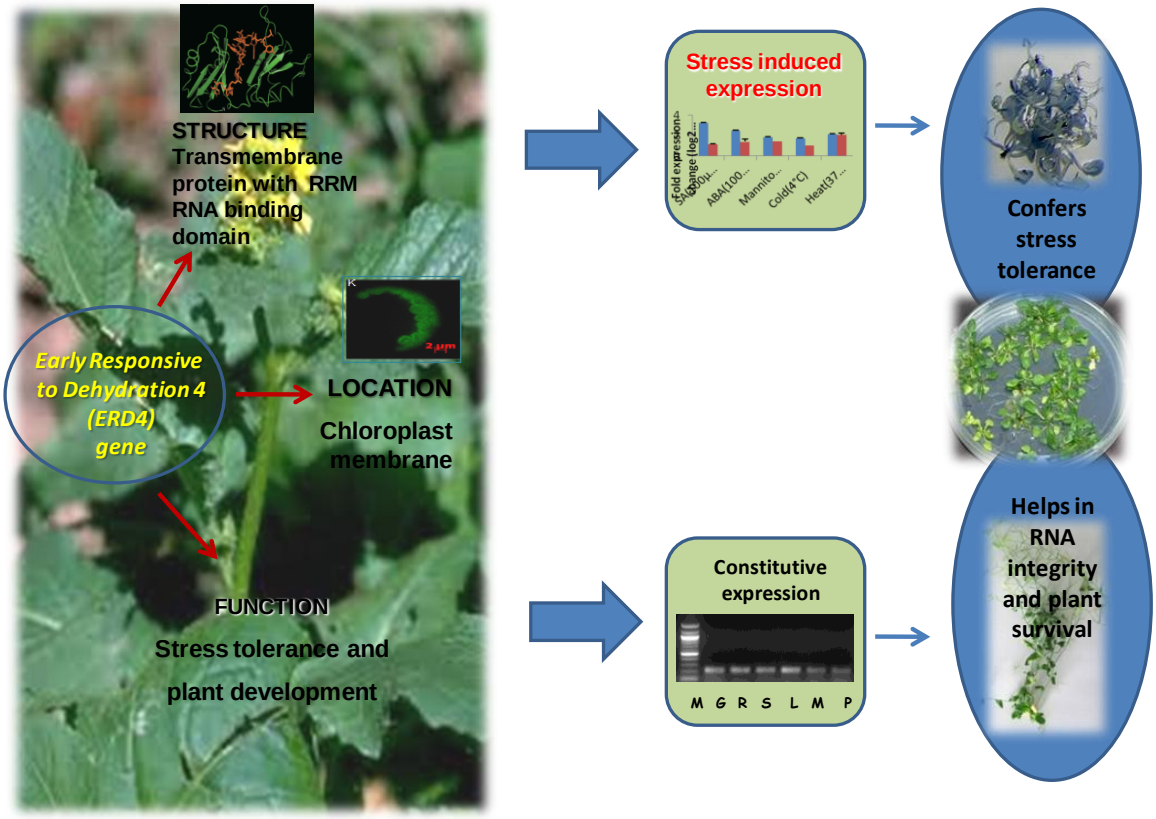


Fig. 5.1. Diagrammatic representation of the role of ERD4 gene in plant stress tolerance and development

REFERENCES

1. Pardo JM (2010) Biotechnology of water and salinity stress tolerance. Current Opinion in Biotechnology 21(2): 185-196.
2. Xiong L, Lee BH, Ishitani M, Lee H, Zhang C, Zhu JK (2001) FIERY1 encoding an inositol polyphosphate 1 phosphatase is a negative regulator of abscisic acid and stress signaling Arabidopsis. Genes Development 15: 1971-1984.
3. Peleg Z, Blumwald E (2011) Hormone balance and abiotic stress tolerance in crop plants. Curr. Opin. Plant Biol. 14(3):290-295.
4. Kosová K, Vítámvás P, Prášil IT, Renaut J (2011) Plant proteome changes under abiotic stress- Contribution of proteomics studies to understanding plant stress response. Journal of Proteomics 74: 1301–1322.
5. Taiz L, Zeiger E (2006) Plant Physiology. 4th ed. Sinauer Associates, Inc. Publishers, Massachusetts.
6. Nonami H (1998) Plant water relations and control of cell elongation at low water potentials. Journal of Plant Research 111(3): 373-382.
7. Kaya MD, Okcu G, Atak M, Cikili Y, Kolsarici O (2006) Seed treatments to overcome salt and drought stress during germination in sunflower (*Helianthus annuus* L.). Eur. J. Agron. 24: 291-295.
8. Hussain M, Malik MA, Farooq M, Ashraf MY, Cheema MA (2008) Improving drought tolerance by exogenous application of glycine betaine and salicylic acid in sunflower. J. Agron. Crop Sci. 194: 193-199.
9. Garg K (2003) Nutrient uptake and management under drought: Nutrient-moisture interaction. Curr. Agric. 27: 1-8.
10. McWilliams D (2003) Drought strategies for cotton, cooperative extension service circular 582, College of Agriculture and Home Economics. New Mexico State University, USA.

11. Wahid A, Rasul E (2005) Photosynthesis in leaf, stem, flower and fruit. In: Pessarakli M (ed), Handbook of Photosynthesis, 2nd ed, CRC Press, Florida, pp 479–497.
12. Munné-Bosch S, Peñuelas J (2003a) Photo- and antioxidative protection, and a role for salicylic acid during drought and recovery in field-grown Phillyrea angustifolia plants. Planta 217: 758–766.
13. Apel K, Hirt H (2004) Reactive oxygen species: metabolism, oxidative stress, and signal transduction. Annu. Rev. Plant Biol. 55: 373-399.
14. Sairam RK, Srivastava GC, Agarwal S, Meena RC (2005) Differences in antioxidant activity in response to salinity stress in tolerant and susceptible wheat genotypes. Biol. Plant. 49: 85-91.
15. Mantri N, Patade V, Penna S, Ford R, Pang ECK (2012) Abiotic stress responses in plants- present and future. In: Ahmad P, Prasad MNV (eds) Abiotic stress responses in plants: Metabolism to productivity. Springer, Science + Business Media NY, USA, pp 1-19.
16. Farooq M, Wahid A, Kobayashi N, Fujita D, Basra SMA (2009) Plant drought stress: effects, mechanisms and management. Agron. Sustain. Dev. 29: 185–212.
17. Morgan PW (1990) Effects of abiotic stresses on plant hormone systems, in: Stress Responses in plants: adaptation and acclimation mechanisms, Wiley-Liss, Inc., pp 113–146.
18. Subbarao GV, Nam NH, Chauhan YS, Johansen C. (2000) Osmotic adjustment, water relations and carbohydrate remobilization in pigeonpea under water deficits, J. Plant Physiol. 157: 651–659.
19. Gong H, Zhu X, Chen K, Wang S, Zhang C (2005) Silicon alleviates oxidative damage of wheat plants in pots under drought. Plant Science, 169: 313-321.

20. Shinozaki K, Yamaguchi-Shinozaki K, Seki M (2003) Regulatory network of gene expression in the drought and cold stress responses. Curr. Opin. Plant Biol. 6: 410–417.
21. Tuteja N (2007) Abscisic Acid and Abiotic Stress Signaling. Plant Signal Behav. 2(3): 135–138.
22. Mahajan S, Tuteja N (2005) Cold, salinity and drought stresses: an overview. Arch Biochem Biophys. 444: 139–158.
23. Munns R, Tester M (2008) Mechanisms of salinity tolerance. Annual Review of Plant Biology 59: 651–681.
24. Hasegawa PM, Bressan RA, Zhu JK, Bohnert HJ (2000) Plant cellular and molecular responses to high salinity. Annual Review of Plant Physiology and Plant Molecular Biology 51: 463-499.
25. Munns R (2002) Comparative physiology of salt and water stress. Plant, Cell and Environment: 25(2): 239-250.
26. Munns R, Termaat A (1986) Whole-Plant Responses to Salinity. Functional Plant Biology 13(1): 143-160.
27. Bensen RJ, Boyer JS, Mullet JE (1988) Water deficit-induced changes in abscisic acid, growth, polysomes, and translatable RNA in soybean hypocotyls. Plant Physiol. 88: 289-294.
28. He T, Cramer GR (1996) Abscisic acid concentrations are correlated with leaf area reductions in two salt-stressed rapid cycling Brassica species. Plant Soil 179: 25-33.
29. Platten DJ, Cotsaftis O, Berthomieu P, Bohnert H, Davenport RJ, Fairbairn DJ, Horie T, Leigh RA, Lin HX, Luan S, Maser P, Pantoja O, Rodriguez-Navarro A, Schachtman DP, Schroeder JI, Sentenac H, Uozumi N, Very AA, Zhu JK, Dennis

ES, Tester M (2006) Nomenclature for HKT transporters, key determinants of plant salinity tolerance. Trends Plant Sci. 11: 372–374.

30. Mahajan S, Sopoy SK, Tuteja N (2006) CBL-CIPK paradigm: role in calcium and stress signaling in plants. Proc Indian Natl Sci Acad USA 72: 63-78.
31. Zhu JK (2002) Salt and drought stress signal transduction in plants. Annu. Rev. Plant Biol. 53: 247–273.
32. Atlin GN, Lafitte (2002) Developing and testing rice varieties for water saving systems in the tropics. In: Bouman BAM, Hengsdijk H, Hardy B, Bindraban PS, Tuong TP, Ladha JK (eds) Water wise rice production. Proceedings of the International Workshop on Water wise Rice Production, Los Banos, Philippines, pp 275-283.
33. Babu RC, Nguyen BD, Chamarerk V, Shanmugasundaram P, Chezhian P, Jeyaprakash P, Ganesh SK, Palchamy A, Sadasivam S, Sarkarung S, Wade LJ, Nguyen HT (2003) Genetic analysis of drought resistance in rice by molecular markers: Association between secondary traits and field performance. Crop Sci. 43: 1457-1469.
34. Cattivelli L, Rizza F, Badeck FW, Mazzucotelli E, Mastrangelo AM, Francia E, Mare C, Tondelli A, Stanca AM (2008) Drought tolerance improvement in crop plants: An integrated view from breeding to genomics. Field Crop Res. 105: 1-14.
35. Zhu Y, Mang H, Sun Q, Qian J, Hipps A, Hua J (2012) Gene discovery using mutagen-induced polymorphisms and deep sequencing: application to plant disease resistance Genetics 192: 139–146.
36. Qureshi MI, Qadir S, Zolla L (2007) Proteomics-based dissection of stress-responsive pathways in plants. Journal of Plant Physiology 164: 1239-1260.

37. Cao W, Epstein C, Liu H, DeLoughery C, Ge N, Lin J, Diao R, Cao H, Long F, Zhang X, Chen Y, Wright PS, Busch S, Wenck M, Wong K, Saltzman AG, Tang Z, Liu L, Zilberstein A (2004) Comparing gene discovery from Affymetrix GeneChip microarrays and Clontech PCR-select cDNA subtraction: a case study. BMC Genomics 27; 5(1): 26.
38. Seki M, Ishida J, Nakajima M, Enju A, Iida K, Satou M, Fujita M, Narusaka Y, Narusaka M, Sakurai T, Akiyama K, Oono Y, Kamei A, Umezawa T, Mizukado S, Maruyama K, Yamaguchi-Shinozaki K, Shinozaki K (2005) Genomic analysis of stress response. In: Plant Abiotic Stress. Jenks MA and Hasegawa PM (eds). Blackwell Publ. UK. pp 248-265.
39. Shinozaki K, Shinozaki KY (2007) Gene networks involved in drought stress response and tolerance. J. Exp. Bot. 58 (2): 221-227.
40. Harb A, Krishnan A, Ambavaram MMR, Pereira A (2010) Molecular and physiological analysis of drought stress in Arabidopsis reveals early responses leading to acclimation in plant growth. Plant Physiology 154 :1254-1271
41. Dubouzet JG, Sakuma Y, Ito Y, Kasuga M, Dubouzet EG, Miura S, Seki M, Shinozaki K, Yamaguchi-Shinozaki K (2003) OsDREB genes in rice, Oryza sativa L., encode transcription activators that function in drought, high-salt and cold-responsive gene expression. Plant J. 33: 751-763.
42. Christmann A, Moes D, Himmelbach A, Yang Y, Tang Y, Grill E (2006) Integration of abscisic acid signalling into plant responses.Plant Biol (Stuttg) 8(3):314-25.
43. Hirayama T, Shinozaki K (2007) Perception and transduction of abscisic acid signals: keys to the function of the versatile plant hormone ABA. Trends of Plant Science, 12-343.

44. Kiyosue T, Yamaguchi-Shinozaki K, Shinozaki K (1994) Characterization of two cDNAs (ERD10 and ERD14) corresponding to genes that respond rapidly to dehydration stress in *Arabidopsis thaliana*. Plant Cell Physiology 35-225.
45. González-Martínez SC, Ersoz E, Brown GR, Wheeler NC and Neale DB (2006) DNA sequence variation and selection of tag single-nucleotide polymorphisms at candidate genes for drought-stress response *Pinus taeda* L. Genetics 172(3): 1915-1926.
46. Froehlich JE, Wilkerson CG, Ray WK, Mc Andrew RS, Osteryoung KW, Gage DA, Phinney BS (2003) Proteomic study of the Arabidopsis thaliana chloroplastic envelope membrane utilizing alternatives to traditional two-dimensional electrophoresis. Journal of Proteome Research 2(4): 413-425.
47. Nakashima K, Satoh R, Kiyosue T, Yamaguchi-Shinozaki K, Shinozaki K (1998) A gene encoding proline dehydrogenase is not only induced by proline and hypoosmolarity, but is also developmentally regulated in the reproductive organs of Arabidopsis. Plant Physiology 118: 1233.
48. Kiyosue T, Abe H, Yamaguchi-Shinozaki K, Shinozaki K (1998) ERD6, a cDNA clone for an early dehydration-induced gene of Arabidopsis, encodes a putative sugar transporter. Biochimica Biophysica Acta, 1370-187.
49. Kimura M, Yamamoto YY, Seki M, Sakurai T, Sato M, Abe T, Yoshida S, Manabe K, Shinozaki K, Matsui M (2003) Identification of Arabidopsis genes regulated by high light-stress using cDNA microarray. Photochemical and Photobiology 77(2): 226- 330.
50. Kiyosue T, Yamaguchi-Shinozaki K, Shinozaki K (1993) Characterization of two cDNAs (ERD11 and ERD13) for dehydration inducible genes that encode

putative glutathione S-transferases *Arabidopsis thaliana* L. FEBS Letters 335: 189.

51. Kiyosue T, Yamaguchi-Shinozaki K, Shinozaki K (1994) Cloning of cDNAs for genes that are early-responsive to dehydration stress (ERDs) in Arabidopsis thaliana L.: identification of three ERDs as HSP cognate genes. Plant Molecular Biology 25: 791.
52. Kariola T, Brader G, Helenius E, Li J, Heino P, Palva ET (2006) EARLY RESPONSIVE to DEHYDRATION 15, a negative regulator of abscisic acid responses in Arabidopsis. Plant Physiology 142: 1559.
53. Gutierrez RA, Green PJ, Keegstra K, Ohlrogge JB (2004. a) Phylogenetic profiling of the Arabidopsis thaliana proteome: What proteins distinguish plants from other organisms? Genome Biol 5: R53
54. Zhang XL, Zhang ZJ, Chen J, Chen Q, Wang XC, Huang RF (2005) Expressing TERF1 in tobacco enhances drought tolerance and abscisic acid sensitivity during seedling development. Planta 222: 494.
55. Ouyang B, Yang T, Li H, Zhang L, Zhang Y, Zhang J, Fei Z, Ye Z (2007) Identification of early salt stress response genes in tomato root by suppression subtractive hybridization and microarray analysis. J. Exp. Bot. 58(3): 507-520.
56. Wang XF, Auler AS, Edwards RL, Cheng H, Cristalli PS, Smart PL, Richards DA, Shen CC (2004) Wet periods in northeastern Brazil over the past 210 kyr linked to distant climate anomalies. Nature 432: 740–743.
57. Jaroszewski L, Li Z, Krishna SS, Bakolitsa C, Wooley J, et al. (2009) Exploration of uncharted regions of the protein universe. PLoS Biol. 7: e1000205.

58. Finn RD, Tate J, Mistry J, Coggill PC, Sammut SJ, Hotz HR, Ceric G, Forslund K, Eddy SR, Sonnhammer EL, Bateman A (2008) Nucleic Acids Res. 36(Database issue): D281-8.

59. Orengo CA, Thornton JM (2005) Protein families and their evolution-a structural perspective. Annu. Rev. Biochem. 74:867-900.

60. Marsden RL, Lee D, Maibaum M, Yeats C, Orengo CA (2006) Comprehensive genome analysis of 203 genomes provides structural genomics with new insights into protein family space. Nucleic Acids Res. 34: 1066-1080.

61. Gollery M, Harper J, Cushman J, Mittler T, Girke T, Zhu JK, Bailey-Serres J, Mittler R (2006) What makes species unique? The contribution of proteins with obscure features. Genome Biol. 7: 757.

62. Gollery M, Harper J, Cushman J, Mittler T, Mittler R (2007) POFs: what we don't know can hurt us. Trends Plant Sci. 12: 492–496.

63. Kolodny R, Petrey D, Honig B (2006) Protein structure comparison: implications for the nature of 'fold space', and structure and function prediction. Curr. Opin. Struct. Biol. 16: 393–398.

64. Vogel C, Bashton M, Kerrison ND, Chothia C, Teichmann SA (2004) Structure, function and evolution of multidomain proteins. Curr. Opin. Struct. Biol. 14: 208–216.

65. Zhang QC, Petrey D, Norel R, Honig BH (2010) Protein interface conservation across structure space. Proc. Natl. Acad. Sci. 107: 10896–10901.

66. Bryson K, McGuffin LJ, Marsden RL, Ward JJ, Sodhi JS, Jones DT (2005) Protein structure prediction servers at University College London. Nucleic Acids Res 33: W36-38.

67. Rost B (2003) Prediction in 1D: secondary structure, membrane helices, and accessibility. Methods Biochem. Anal. 44: 559–587.

68. Petrey D, Honig B (2005) Protein structure prediction: inroads to biology. Mol. Cell 20: 811–819.

69. Luhua S, Ciftci-Yilmaz S, Harper J, Cushman J, Mittler R (2008) Enhanced tolerance to oxidative stress in transgenic Arabidopsis plants expressing proteins of unknown function. Plant Physiol. 148(1): 280-292.

70. Lorkovic ZJ, Barta A (2002) Genomic analysis: RNA recognition motif (RRM) and K homology (KH) domain RNA-binding proteins from the flowering plant Arabidopsis thaliana. Nucleic Acids Res. 30: 623–635.

71. Alba MM, Pages M (1998) Plant proteins containing the RNA recognition motif. Trends Plant Sci. 3: 15–21.

72. Maris C, Dominguez C, Allain FHT (2005) The RNA recognition motif, a plastic RNA-binding platform to regulate post-transcriptional gene expression. FEBS J 272: 2118–2131.

73. Nagai K, Oubridge C, Ito N, Avis J, Evans P (1995) The RNP domain: a sequence specific RNA binding domain involved in processing and transport of RNA. Trends Biochem. Sci. 20: 235–240.

74. Ambrosone A, Costa A, Leone A, Grillo S (2012) Beyond transcription: RNA-binding proteins as emerging regulators of plant response to environmental constraints. Plant Sci. 182: 12-18.

75. Kang H, Kwak KJ (2012) RNA‐Binding Proteins in Plant Response to Abiotic Stresses. RNA Binding Proteins, edited by Zdravko J. Lorković. Publ. Landes Bioscience. pp 125-136.

76. Herschlag D (1995) RNA chaperones and the RNA folding problem. J. Biol. Chem. 270: 20871‑20874.

77. Phadtare S, Inouye M, Severinov K. (2002) The nucleic acid melting activity of Escherichia coli CspE is critical for transcription antitermination and cold acclimation of cells. J. Biol. Chem. 277:7239-7245.

78. Kim JS, Park SJ, Kwak KJ, Kim YO, Kim JY, Song J, Jang B, Jung CH, Kang H (2007b) Cold shock domain proteins and glycine rich RNA-binding proteins from *Arabidopsis thaliana* can promote the cold adaptation process in E. coli. Nucleic Acids Research, 35: 506–516.

79. Kwak KJ, Kim YO, Kang H (2005) Characterization of transgenic Arabidopsis plants overexpressing GR-RBP4 under high salinity, dehydration, or cold stress. Journal of Experimental Botany 56: 3007–3016.

80. Kim JS, Jung HJ, Lee HJ, Kim KA, Goh C-H, Woo Y, Oh SH, Han YS, Kang H (2008) Glycine-rich RNA-binding protein7 affects abiotic stress responses by regulating stomata opening and closing in *Arabidopsis thaliana*. The Plant Journal 55: 455–466.

81. Kim JY, Kim WY, Kwak KJ, Oh SH, Han YS, Kang H (2010) Glycine-rich RNA-binding proteins are functionally conserved in *Arabidopsis thaliana* and *Oryza sativa* during cold adaptation process. Journal of Experimental Botany 61(9): 2317–2325.

82. Kim YO, Kim JS, Kang H (2005) Cold-inducible zinc finger-containing glycine-rich RNA-binding protein contributes to the enhancement of freezing tolerance in *Arabidopsis thaliana*. The Plant Journal 42: 890–900.

83. Park SJ, Kwak KJ, Jung HJ, Lee HJ, Kang H (2010) The C‐terminal zinc finger domain of Arabidopsis cold shock domain proteins is important for RNA chaperone activity during cold adaptation. Phytochem 71: 543‐547.

84. Park SJ, Kwak KJ, Oh TR, Kim YO, Kang H (2009) Cold shock domain proteins affect seed germination and growth of *Arabidopsis thaliana* under abiotic stress conditions. Plant and Cell Physiology 50: 869–878.

85. Kim MH, Sasaki K, Imai R (2009) Cold shock domain protein 3 regulates freezing tolerance in *Arabidopsis thaliana*. J. Biol. Chem. 284: 23454–23460.

86. Chaikam V, Karlson D (2008) Functional characterization of two cold shock domain proteins from *Oryza sativa*. Plant Cell Environ. 31: 995‐1006.

87. Kant P, Kant S, Gordon M, Shaked R, Barak S (2007) Stress response suppressor1 and stress response suppressor2, two DEAD-box RNA helicases that attenuate Arabidopsis responses to multiple abiotic stresses. Plant Physiol. 145: 814–830.

88. Reed CF (1976) Information summaries on 1000 economic plants. Typescripts submitted to the USDA.

89. Hemingway JS (1976) Mustards: Brassica spp. and *Sinapis alba* (Cruciferae). In: Evolution of Crop Plants. Simmonds NW, Longman (ed), London, pp 56-59.

90. Wright PR, Morgan JM, Jessop RS (1997) Turgor maintenance by osmoregulation in *Brassica napus* and *B. juncea* under field conditions. Annals of Botany 80: 313–319.

91. Ashraf M, McNeilly T (2004) Salinity tolerance in brassica oilseeds. Critical Reviews in Plant Sci. 23(2): 157-174.

92. Oram RN, Kirk JTO, Veness PE, Hurlstone CJ, Edlington JP, Halsall DM (2005) Breeding Indian mustard (*Brassica juncea* (L.) Czern.) for cold-pressed, edible

oil production-a review. Australian Journal of Agricultural Research 56: 581–596.

93. Ashraf M (2001) Relationship between growth and gas exchange characteristics in some salt tolerant amphidiploid Brassica species to sodium chloride. Environ. Exp. Bot. 30: 475-487.
94. Valdez-Aguilar LA, Reed DW (2010) Growth and nutrition of young bean plants under high alkalinity as affected by mixtures of ammonium, potassium, and sodium. Journal of Plant Nutrition 33: 1472–1488.
95. Yang C, Wang P, Li C, Shi D, Wang D (2008) Comparison of effects of salt and alkali stresses on the growth and photosynthesis of wheat. Photosynthetica 46: 107–114.
96. Baginsky S, Hennig L, Zimmermann P, Gruissem W (2010) Gene Expression Analysis, Proteomics, and Network Discovery. Plant Physiology 152: 402-410
97. O'Neill CM, Bancroft I (2000) Comparative physical mapping of segments of the genome of Brassica oleracea var alboglabra that are homoeologous to sequenced regions of the chromosomes 4 and 5 of *Arabidopsis thaliana*. Plant J. 23: 233–243.
98. Rana D, van den Boogaart T, O'Neill CM, Hynes L, Bent E, Macpherson L, Park JY, Lim YP, Bancroft I (2004) Conservation of the microstructure of genome segments in *Brassica napus* and its diploid relatives. Plant J. 40: 725–733.
99. Yang TJ, Kim JS, Kwon SJ, Lim KB, Choi BS, Kim JA, Jin M, Park JY, Lim MH, Kim HI (2006) Sequence level analysis of the diploidization process in the triplicated FLOWERING LOCUS C region of *Brassica rapa*. Plant Cell NU 1339-1347.

100. Cong L, Chai T, Zhang Y (2008) Characterization of the novel gene BjDREB1B encoding a DRE-binding transcription factor from *Brassica juncea* L. Biochemical and Biophysical Research communication 371: 702-706.

101. Jami SK, Clark GB, Turlapati SA, Handley CA, Roux SJ, Kirti PB (2008) Ectopic expression of an annexin from *Brassica juncea* confers tolerance to abiotic and biotic stress treatments in transgenic tobacco. Plant Physiol Biochem. 46; 1019-1030.

102. Yusuf MA, Kumar D, Rajwanshi R, Strasser RJ, Tsimilli-Michael M, Govindjee SNB (2010) Overexpression of gamma-tocopherol methyl transferase gene in transgenic Brassica juncea plants alleviates abiotic stress: physiological and chlorophyll a fluorescence measurements. Biochim Biophys Acta 1797: 1428–1438.

103. Mi-Jung B, Kim YS, Kim IS, Choe YH, Lee EJ, Kim YH, Park HM, Yoon HS (2013) Transgenic rice overexpressing the *Brassica juncea* gamma-glutamyl cysteine synthetase gene enhances tolerance to abiotic stress and improves grain yield under paddy field conditions. Molecular Breeding 31(4): pp 931.

104. Verma SS, Chinnusamy V, Asif M, Bansal KC (2013) Synergistic effect of C2H2 type Zinc-finger protein with LEA promoter to enhance abiotic stress tolerance in *Brassica juncea*. Plant Omics 6(3): 208-214.

105. Veena,Vanga S, Reddy, Sopory SK (2002) Glyoxalase I from *Brassica juncea*: molecular cloning, regulation and its over-expression confer tolerance in transgenic tobacco under stress. The Plant Journal 17(4): 385–395.

106. Mohd AY, Deepak K, Ravi R, Reto JS, Merope TM, Govindjee, Neera BS (2010) Overexpression of γ-tocopherol methyl transferase gene in transgenic

Brassica juncea plants alleviates abiotic stress: Physiological and chlorophyll a fluorescence measurements. Biochimica et Biophysica Acta, 1797: 1428–1438.

107. Jami SK, Clark GB, Swathi Anuradha T, Handley C, Roux SJ, Kirti PB (2008) Ectopic expression of an annexin from *Brassica juncea* confers tolerance to abiotic and biotic stress treatments in transgenic tobacco. Plant Physiol Biochem 46:1019–1030

108. Farinati S, DalCorso G, Varotto S, Furin A (2010) The *Brassica juncea* BjCdR15, an ortholog of Arabidopsis TGA3, is a regulator of cadmium uptake, transport and accumulation in shoots and confers cadmium tolerance in transgenic plants. New Phytologist 185(4): 964–978.

109. Bogs J, Bourbouloux A, Cagnac O, Wachter A, Rausch T, Delrot S (2003) Functional characterization and expression analysis of a glutathione transporter, BjGT1, from *Brassica juncea*: evidence for regulation by heavy metal exposure. Plant Cell Environ. 26: 1703-1711.

110. Hanson MR, Köhler RH (2001) GFP imaging: methodology and application to investigate cellular compartmentation in plants. J. Exp. Bot. 52 (356): 529-539.

111. Emanuelsson O (2002) Predicting protein subcellular localization from amino acid sequence information. Briefings in Bioinformatics 3(4): 361–376.

112. Hua S, Sun Z (2001) Support vector machine approach for protein subcellular localization prediction. Bioinformatics 17(8):721-728.

113. Wang J, Sung WK, Krishnan A, Li KB (2005) Protein subcellular localization prediction for Gram-negative bacteria using amino acid subalphabets and a combination of multiple support vector machines. BMC Bioinformatics 6:174.

114. Yu CS, Lin CJ, Hwang JK (2004) Predicting subcellular localization of proteins for Gram-negative bacteria by support vector machines based on n-peptide compositions. Protein Sci. 13(5):1402-1406.

115. Nakai K, Kanehisa M (1991) Expert system for predicting protein localization sites in gram-negative bacteria. Proteins 11(2): 95-110.

116. Chou KC, Shen HB (2006) Predicting eukaryotic protein subcellular location by fusing optimized evidence-theoretic K-nearest neighbor classifiers. Journal of Proteome research 5: 1888-1897.

117. Su CY, Lo A, Chiu HS, Sung TY, Hsu WL (2006) Protein subcellular localization prediction based on compartment-specific biological features. Proceedings of IEEE Computational Systems Bioinformatics Conference (CSB'06): 14–18 August 2006; Stanford, California pp 325-330.

118. Lee K, Kim DW, Na D, Lee KH, Lee D (2006) PLPD: reliable protein localization prediction from imbalanced and overlapped datasets. Nucleic Acids Res. 34(17):4655-4666.

119. Lu Z, Szafron D, Greiner R, Lu P, Wishart DS, Poulin B, Anvik J, Macdonell C, Eisner R (2004) Predicting subcellular localization of proteins using machine-learned classifiers. Bioinformatics 20: 547–556.

120. Gardy JL, Laird MR, Chen F, Rey S, Walsh CJ, Ester M, Brinkman FS (2005) PSORTb v.2.0: expanded prediction of bacterial protein subcellular localization and insights gained from comparative proteome analysis. Bioinformatics 21(5):617-623.

121. Bhasin M, Garg A, Raghava GP (2005) PSL pred: prediction of subcellular localization of bacterial proteins. Bioinformatics 21(10): 2522-2524.

122. Gardy JL, Brinkman FS (2006) Methods for predicting bacterial protein subcellular localization. Nat. Rev. Microbiol 4(10):741-751.

123. Nair R, Rost B (2002) Sequence conserved for subcellular localization. Protein Sci. 11(12):2836-2847.

124. Yu CS, Chen YC, Lu CH, Hwang JK. (2006) Prediction of protein subcellular localization, Proteins, 64, 643-651.

125. Chalfie M, Tu Y, Euskirchen G, Ward WW, Prasher DC (1994) Green fluorescent protein as a marker for gene expression. Science 263:802–805.

126. Liu Y, Li H, Shi Y, Song Y, Wang T, (2009) A maize early responsive to dehydration gene, ZmERD4, provides enhanced drought and salt tolerance in Arabidopsis. Plant Molecular Biology Reporter 27: 542–548.

127. McQualter RB, Dookun-Saumtally A (2007) Expression profiling of abiotic-stress inducible genes in sugarcane. XXVI Congress, International Society of Sugar Cane Technologists. pp 878-888.

128. Alexandersson E, Saalbach G, Larsson C, Kjellbom P (2004) Arabidopsis plasma membrane proteomics identifies components of transport, signal transduction and membrane trafficking. Plant Cell Physiol. 45: 1543-1556.

129. Mitschke J, Fuss J, Blum T, Höglund A, Reski R (2009) Prediction of dual protein targeting to plant organelles. New Phytol. 183: 224–235.

130. Kumar LS (1999) DNA markers in plant improvement: An overview. Biotechnology Advances 17: 143–182.

131. Lang M, Orgogozo V (2011) Identification of homologous gene sequences by PCR with degenerate primers. Methods Mol. Biol. 772: 245-56.

132. Doyle JJ, Gaut BS (2000) Evolution of genes and taxa: a primer. Plant Mol Biol. 42(1): 1-23.

133. Soltis ED, Soltis PS (2000) Contributions of plant molecular systematics to studies of molecular evolution. Plant Mol. Biol. 42(1):45-75.

134. Claros MG, Vincens P (1996) Computational method to predict mitochondrially imported proteins and their targeting sequences. Eur. J. Biochem. 241: 779–786.

135. Bannai H, Tamada Y, Maruyama O, Nakai K, Miyano S (2002) Extensive feature detection of N-terminal protein sorting signals. Bioinformatics18: 298–305.

136. Horton P, Park KJ, Obayashi T, Fujita N, Harada H, Adams-Collier CJ, Nakai K (2007) WoLF PSORT: protein localization predictor. Nucleic Acids Res. 35: W585–587.

137. Briesemeister S, Rahnenfuhrer J, Kohlbacher O (2010) YLoc – an interpretable web server for predicting subcellular localization. Nucleic Acids Res. 38: W497–502.

138. Sambrook J, Russell DW (2002) Molecular cloning: a laboratory manual. Cold Spring Harbor, NY: Cold Spring Harbor Laboratory Press.

139. Hulo N, Bairoch A, Bulliard V, Cerutti L, Cuche BA (2008) The 20 years of PROSITE. Nucleic Acids Res 36: D245–249.

140. Flicek P, Amode MR, Barrell D, Beal K, Brent S (2011) Ensembl 2012. Nucleic Acids Res. 40: D84–90.

141. Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ (1990) Basic local alignment search tool. J Mol Biol. 215: 403–410.

142. Pei J, Kim BH, Grishin NV (2008) PROMALS3D: a tool for multiple protein sequence and structure alignments. Nucleic Acids Res 36: 2295–2300.

143. Tamura K, Dudley J, Nei M, Kumar S (2007) MEGA4: Molecular Evolutionary Genetics Analysis (MEGA) software version 4.0. Mol Biol Evol 24: 1596–1599.

144. McGuffin LJ, Bryson K, Jones DT (2000) The PSIPRED protein structure prediction server. Bioinformatics 16: 404–405.

145. Cserzo M, Wallin E, Simon I, von Heijne G, Elofsson A (1997) Prediction of transmembrane alpha-helices in procariotic membrane proteins: the Dense Alignment Surface method. Protein Eng 10: 673–676.

146. Tusna´dy GE, Simon I (2001) The HMMTOP transmembrane topology prediction server. Bioinformatics 17: 849–850.

147. Jones DT, Taylor WR, Thornton JM (1994) A model recognition approach to the prediction of all-helical membrane protein structure and topology. Biochemistry 33: 3038–3049.

148. Sonnhammer EL, von Heijne G, Krogh A (1998) A hidden Markov model for predicting transmembrane helices in protein sequences. Proc Int Conf Intell Syst Mol Biol 6: 175–182.

149. Kahsay RY, Gao G, Liao L (2005) An improved hidden Markov model for transmembrane protein detection and topology prediction and its applications to complete genomes. Bioinformatics 21: 1853–1858.

150. Hofmann K, Stoffel W (1993) TMbase — A database of membrane spanning proteins segments. Biol Chem Hoppe–Seyler 374: 166–170.

151. Arai M, Mitsuke H, Ikeda M, Xia JX, Kikuchi T (2004) ConPred II: a consensus prediction method for obtaining transmembrane topology models with high reliability. Nucleic Acids Res 32: W390–393.

152. Ka¨ ll L, Krogh A, Sonnhammer ELL (2004) A combined transmembrane topology and signal peptide prediction method. J Mol Biol 338: 1027–1036.

153. Emanuelsson O, Brunak S, von Heijne G, Nielsen H (2007) Locating proteins in the cell using TargetP, SignalP, and related tools. Nature Protocols 2: 953–971.

154. Taylor NL, Tan YF, Jacoby RP and Millar AH (2009) Abiotic environmental stress induced changes in the *Arabidopsis thaliana* chloroplast, mitochondria and peroxisome proteomes. J. Proteomics 72: 367–378.

155. Kleffmann T, Russenberger D, von Zychlinski A, Christopher W, Sjolander K, Gruissem W, Baginsky S (2004) The *Arabidopsis thaliana* chloroplast proteome reveals pathway abundance and novel protein functions. Curr. Biol. 14: 354–362.

156. Bhushan S, Kuhn C, Berglund AK, Roth C, Glaser E (2006) The role of the N-terminal domain of chloroplast targeting peptides in organellar protein import and miss-sorting. FEBS Lett. 580: 3966–3972.

157. Camargo SR, Canc¸ado GMA, Ulian EC, Menossi M (2007) Identification of genes responsive to the application of ethanol on sugarcane leaves. Plant Cell Rep 26: 2119–2128.

158. Cheng F, Wu J, Fang L, Wang X (2012) Syntenic gene analysis between *Brassica rapa* and other Brassicaceae species. Front Plant Sci. 3: 198.

159. Nakashima K, Ito Y, Yamaguchi-Shinozaki K (2009) Transcriptional regulatory networks in response to abiotic stresses in Arabidopsis and grasses. Plant Physiol. 149: 88–95

160. Chothia C, Gough J, Vogel C, Teichmann S (2003) Evolution of the protein repertoire. Science 300: 1701–1703.

161. Andersson H, von Heijne G (1993) Position-specific Asp-Lys pairing can affect signal sequence function and membrane protein topology. J Biol Chem 268: 21389–21393.

162. Ferro M, Salvi D, Brugiere S, Miras S, Kowalski S, Louwagie M, Garin J, Joyard J, Rolland N (2003) Proteomics of the chloroplast envelope membranes from *Arabidopsis thaliana*. Mol. Cell Proteomics 2: 325–345.

163. Horton P, Nakai K (1997) Better prediction of protein cellular localization sites with the k nearest neighbors classifier. In: Gaasterland T, Karp P, Karplus K, Ouzounis C, Sander C, Valencia A (eds), Proceeding of the Fifth International Conference on Intelligent Systems for Molecular Biology, Halkidiki, Greece, AAAI Press, pp 147–152.
164. Reinhardt A, Hubbard T (1998) Using neural networks for prediction of the subcellular location of proteins. Nucleic Acids Research 26(9): 2230–2236.
165. Prasher DC, Eckenrode VK, Ward WW, Prendergast FG, Cor MJ (1992) Primary structure of the *Aequorea victoria* a green fluorescent protein. Gene 111: 229–233.
166. Martin W, Hermann RG (1998) Gene transfer from organelles to the nucleus: how much, what happens and why? Plant Physiol. 118: 9–17.
167. Cavalier-Smith T (2000) Membrane heredity and early chloroplast evolution. Trends Plant Sci. 5: 174–182.
168. Lopez-Juez E (2007) Plastid biogenesis, between light and shadows. J. Exp. Bot. 58: 11–26.
169. Martin W, Rujan T, Richly E, Hansen A, Cornelsen S, Lins T, Leister D, Stoebe B, Hasegawa M, David Penny (2002) Evolutionary analysis of Arabidopsis, cyanobacterial, and chloroplast genomes reveals plastid phylogeny and thousands of cyanobacterial genes in the nucleus. Proc. Natl. Acad. Sci. USA 99: 12246–12251.
170. Altschul SF, Madden TL, Schäffer AA, Zhang J, Zhang Z, Miller W, Lipman DJ (1997) Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. Nucleic Acids Res. 25(17): 3389-3402.

171. Pearson WR (1990) Rapid and sensitive sequence comparison with FASTP and FASTA. Methods Enzymol. 183:63-98.

172. Bateman A, Coin L, Durbin R, Finn RD, Hollich V, Griffiths Jones S, Khanna A, Marshall M, Moxon S, Sonnhammer ELL (2004) The Pfam protein families database. Nucleic Acids Res. 32(1): D138-D141.

173. Corpet F, Servant F, Gouzy J, Kahn D (2000) ProDom and ProDom-CG: tools for protein domain analysis and whole genome comparisons. Nucleic Acids Res. 28(1): 267-269.

174. Andreeva A, Howorth D, Brenner SE, Hubbard TJP, Chothia C, Murzin AG (2004) SCOP database in 2004: refinements integrate structure and sequence family data. Nucleic Acids Res. 32 (1): D226-D229.

175. Moore AD, Bjorklund AK, Ekman D, Bornberg-Bauer E, Elofsson A (2008) Arrangements in the modular evolution of proteins. Trends Biochem Sci. 33: 444–451.

176. Petrey D, Fischer M, Honig B (2009) Structural relationships among proteins with different global topologies and their implications for function annotation strategies. Proc Natl Acad Sci USA 106: 17377–17382. 177.

177. Godzik A (2003) Fold recognition methods. Methods Biochem. Anal. 44: 525–546.

178. Frommer WB, Ninnemann O (1995) Heterologous expression of genes in bacterial, fungal, animal, plant cells. Annual Review of Plant Physiology and Plant Molecular Biology 46: 419–444.

179. Palusa SG, Wilusz J (2013) Approaches for the identification and characterization of RNA-protein interactions. Biophysical approaches to

translational control of gene expression, Biophysics for the Life Sciences, JD Dinman (eds) Springer 199-212.

180. Finn RD, Mistry J, Tate J, Coggill P, Heger A, Pollington JE, Gavin OL, Gunasekaran P, Ceric G, Forslund K, Holm L, Sonnhammer EL, Eddy SR, Bateman A (2010) The Pfam protein families database. Nucleic Acids Res. 38: D211–222.

181. Novatchkova M, Schneider G, Fritz R, Eisenhaber F, Schleiffer A (2006) DOUTfinder-identification of distant domain outliers using sub-significant sequence similarity. Nucleic Acids Res. 34: 214–218.

182. Letunic I, Doerks T, Bork P (2011) SMART 7: recent updates to the protein domain annotation resource. Nucleic Acids Res. 40: D302–305.

183. Ginalski K, Elofsson A, Fischer D, Rychlewski L (2003) 3D-Jury: a simple approach to improve protein structure predictions. Bioinformatics 19:1015–1018.

184. Marti-Renom MA, Stuart A, Fiser A, Sanchez R, Melo F, Sali A (2000) Comparative protein structure modeling of genes and genomes. Annu. Rev. Biophys. Biomol. Struct. 29: 291–325.

185. Holm L, Rosenstrom P (2010) Dali server: conservation mapping in 3D. Nucleic Acids Res. 38: W545–549.

186. Guex N, Peitsch MC (1997) SWISS-MODEL and the Swiss-PdbViewer: an environment for comparative protein modeling. Electrophoresis 18: 2714–2723.

187. Hellman LM, Fried MG (2007) Electrophoretic mobility shift assay (EMSA) for detecting protein-nucleic acid interactions. Nat Protoc. 2(8):1849-61.

188. Clery A, Blatter M, Allain FH (2008) RNA recognition motifs: boring? Not quite. Curr. Opin. Struct. Biol. 18: 290–298.

189. Marchler-Bauer A, Lu S, Anderson JB, Chitsaz F, Derbyshire MK, DeWeese-Scott C, Fong J H., Geer LY, Geer RC, Gonzales NR, Gwadz M, Hurwitz DI, Jackson JD, Ke Z, Lanczycki CJ, Lu F, Marchler GH, Mullokandov M, Omelchenko MV, Robertson CL, Song JS, Thanki N, Yamashita RA, Zhang D, Zhang N, Zheng C, Bryant SH (2011) DD: a Conserved Domain Database for the functional annotation of proteins. Nucleic Acids Res 39: D225–229.

190. Sugita M, Sugiura M (1996) Regulation of gene expression in chloroplasts of higher plants. Plant Mol. Biol. 32: 315–326.

191. Birney E, Kumar S, Krainer AR (1993) Analysis of the RNA-recognition motif and RS and RGG domains: conservation in metazoan pre-mRNA splicing factors. Nucleic Acids Res 21: 5803–5816.

192. Ding J, Hayashi MK, Zhang Y, Manche L, Krainer AR, Xu RM (1999) Crystal structure of the two-RRM domain of hnRNP A1 (UP1) complexed with single-stranded telomeric DNA. Genes Dev. 1: 1102–1115.

193. Sato S, Nakamura Y, Kaneko T, Asamizu E, Tabata S (1999) Complete structure of the chloroplast genome of *Arabidopsis thaliana*. DNA Res 6:283–290.

194. Kuhn JM, Schroeder JI (2003) Impacts of altered RNA metabolism on abscisic acid signaling. Curr. Opin. Plant Biol. 6: 463–469.

195. Hazen SP, Wu Y, Kreps JA (2003) Gene expression profiling of plant responses to abiotic stress. Funct. Integr. Genomics 3(3):105-11.

196. Umezava T, Fujita M, Fujita Y, Yamaguchi-Shinozaki K, Shinozaki K(2006) Engineering drought tolerance in plants: discovering and tailoring genes to unlock the future. Curr Opin Biotechnol 17: 113-122

197. Martienssen R (1998) Functional genomics: probing plant gene function and expression with transposons. Proc. Natl Acad. Sci. USA, 95:2021 – 2026.

198. Fire A, Xu S, Montgomery MK, Kostas SA, Driver SE, Mello CC (1998) Potent and specific genetic interference by double stranded RNA in *Caenorhabditis elegans*. Nature, 391: 806 –811.

199. Feng DR, Liu B, Li WY, He YM, Qi KB, Wang BH, Wang FJ (2009) Over-expression of a cold-induced plasma membrane protein gene (MpRCI) from plantain enhances low temperature-resistance in transgenic tobacco. Environmental and Experimental Botany 65: 395–402.

200. Takasaki H, Maruyama K, Kidokoro S, Ito Y, Fujita Y, Shinozaki K, Yamaguchi-Shinozaki K, Nakashima K (2010) The abiotic stress-responsive NAC-type transcription factor OsNAC5 regulates stress-inducible genes and stress tolerance in rice. Molecular Genetics and Genomics 284:173–183.

201. Rozen S, Skaletsky H (2000) Primer3 on the WWW for general users and for biologist programmers. Methods Mol Biol 132: 365-386.

202. Clough SJ, Bent AF (1998) Floral dip: a simplified method for *Agrobacterium*-mediated transformation of *Arabidopsis thaliana*. Plant J 16: 735-43.

203. Arnon DI (1949) Copper enzymes in isolated chloroplasts, polyphenoxidase in beta vulgaris. Plant Physiology 24: 1-15.

204. Hodgson RA, Raison JK (1991) Superoxide production by thylakoids during chilling and its implication in the susceptibility of plants to chilling-induced photoinhibition. Planta 183: 222–228.

205. Hissin PJ, Hilf R (1976) A fluorometric method for determination of oxidized and reduced glutathione in tissues. Anal Biochem 74: 214-226.

206. Thordal-Christensen H, Zhang Z, Wei Y, Collinge DB (1997) Subcellular localization of H_2O_2 in plants, H_2O_2 accumulation in papillae and hypersensitive response during barley-powdery mildew interaction. Plant J. 11:1187–1194.

207. Wang C, Jing R, Mao X, Chang X, Li A (2010) TaABC1, a member of the activity of bc1 complex protein kinase family from common wheat, confers enhanced tolerance to abiotic stresses in Arabidopsis. Journal of Experimental Botany 63:1299–1311.

208. Shinozaki K, Yamaguchi-Shinozaki K (1997) Gene expression and signal transduction in water-stress response. Plant Physiol 115: 327-334.

209. Yamaguchi-Shinozaki K, Shinozaki K (2005) Organization of cis-acting regulatory elements in osmotic- and cold-stress-responsive promoters. Trends Plant Sci, 10:88–94.

210. Soitamo AJ, Piippo M, Allahverdiyeva Y, Battchikova, Aro EM (2008) Light has a specific role in modulating Arabidopsis gene expression at low temperature. BMC Plant Biology 8: 13

211. Abercrombie JM, Halfhill MD, Ranjan P, Rao MR, Saxton AM et al. (2008) Transcriptional responses of *Arabidopsis thaliana* plants to As (V) stress. BMC Plant Biology 87: 1471–2229.

212. Kaplan B, Davydov O, Knight H, Galon Y, Knight MR, Fluhr R, Fromm H. (2006) Rapid transcriptome changes induced by cytosolic Ca^{2+} transient reveal ABRE- related sequences as Ca^{2+} responsive cis elements in Arabidopsis. Plant Cell 18: 2733-2748.

213. Shinozaki K, Shinozaki KY (2007) Gene networks involved in drought stress response and tolerance. J Exp Bot 58(2): 221-227.

214. Song Y, Wang Z, Bo W, Ren Y, Zhang Z, Zhang D (2012) Transcriptional profiling by cDNA-AFLP analysis showed differential transcript abundance in response to water stress in *Populus hopeiensis*. BMC Genomics, 13: 286.

215. Gill SS, Tuteja N (2010) Reactive oxygen species and antioxidant machinery in abiotic stress tolerance in crop plants. Plant Physiol Biochem. 48(12):909-30.

216. Demmig-Adams B, Adams WW (1996) The role of xanthophyll cycle carotenoids in the protection of photosynthesis. Trends Plant Sci 1: 21–26

217. Tracewell CA, Vrettos JS, Bautista JA, Frank HA, Brudvig GW (2001) Carotenoid photooxidation in photosystem II. Arch Biochem Biophys 385: 61–69.

218. Chi YH, Moon JC, Park JH, Kim HS, Zulfugarov IS, Fanata WI, Jang HH, Lee JR, Lee YM, Kim ST, Chung YY, Lim CO, Kim JY, Yun DJ, Lee CH, Lee KO, Lee SY (2008) Remarkably decreased Fv/Fm values under high irradiance conditions (1,000 μmol m^{-2} s^{-1}) with delayed recovery. Plant Physiology 148: 808–817.

219. Kim SY, Nam KH (2010) Physiological roles of ERD10 in abiotic stresses and seed germination of Arabidopsis. Plant cell reports 29 (2): 203-209.

220. Xiao B, Huang Y, Tang N, Xiong L (2007) Over-expression of a LEA gene in rice improves drought resistance under the field conditions. Theoretical and Applied Genetics 115: 35–46.

221. Mao X, Zhang H, Tian S, Chang X, Jing R. (2009) TaSnRK2.4, an SNF1-type serine-threonine protein kinase of wheat (*Triticum aestivum* L.) confers enhanced multi-stress tolerance in Arabidopsis. Journal of Experimental Botany 61: 683–696.

222. Priyanka B, Sekhar K, Reddy VD, Rao KV (2010) Expression of pigeonpea hybrid-proline-rich protein encoding gene (CcHyPRP) in yeast and Arabidopsis affords multiple abiotic stress tolerance. Plant Biotechnology Journal 8:76–87.

223. Rai A, Suprasanna P, D'Souza SF, Kumar V (2012) Membrane topology and predicted RNA-binding function of the 'Early Responsive to Dehydration (ERD4)' plant protein. Plos One 7(3):e32658.

224. Nickelsen J (2003) Chloroplast RNA-binding proteins. Curr. Genet. 43: 392–399.

225. Zerges W, Rochaix JD (1998) Low Density Membranes Are Associated with RNA-binding Proteins and Thylakoids in the Chloroplast of *Chlamydomonas reinhardtii*. J Cell Biol. 140(1): 101–110.

226. Danon A, Mayfield SP (1991) Light regulated translational activators: identification of chloroplast gene specific mRNA binding proteins. Eur Mol Biol Organ 10:3993–4001.

227. Hoober, JK, CO Boyd, Paavola L (1991) Origin of thylakoid membranes in *Chlamydomonas reinhardtii y-1*at 38°C. Plant Physiol 96:1321–1328.

228. Morré DJ, Selldén GC, Sundqvist, Sandelius AS (1991) Stromal low temperature compartment derived from the inner membrane of the chloroplast envelope. Plant Physiol.97:1558–1564.

229. Hugueney P, Bouvier F, Badillo A, D'Harlingue A, Kuntz M, Camara B (1995) Identification of a plastid protein involved in vesicle fusion and/or membrane protein translocation. Proc Natl Acad Sci USA. 92: 5630–5634.

LIST OF PUBLICATIONS ARISING FROM THE THESIS

Journal

Published: One

Membrane topology and predicted RNA-binding function of the 'Early Responsive to Dehydration (ERD4)' plant protein. **Rai A., Suprasanna P., D'Souza SF, Kumar V., *Plos One* 2012,** 7(3):e32658.

Under Communication: One

Novel chloroplast targeted RNA binding protein ERD4 improves salinity and drought tolerance.

Conference

1. Molecular characterization of abiotic stress responsive *Brassica juncea BjERD4* gene **Archana N. Rai** and P. Suprasanna. National Conference of Plant Physiology; National Research Centre on Groundnut, Junagadh; 11-13, 2013**.**

Others:

1. Gene Bank Submissions:

- Brassica juncea ERD4 protein (ERD4) gene, complete cds gi|161006797|gb|EU126607.2|[161006797]
- Brassica juncea early responsive to dehydration 4 protein-like mRNA, partial sequence. gi|183376723|gb|EU596450.1|[183376723].

Archana N. Rai

OPEN ACCESS Freely available online

PLoS one

Membrane Topology and Predicted RNA-Binding Function of the 'Early Responsive to Dehydration (ERD4)' Plant Protein

Archana Rai[1], Penna Suprasanna[1], Stanislaus F. D'Souza[1]*, Vinay Kumar[2]*

1 Nuclear Agricultural & Biotechnology Division, Bhabha Atomic Research Centre, Mumbai, India, **2** High Pressure & Synchrotron Radiation Physics Division, Bhabha Atomic Research Centre, Mumbai, India

Abstract

Functional annotation of uncharacterized genes is the main focus of computational methods in the post genomic era. These tools search for similarity between proteins on the premise that those sharing sequence or structural motifs usually perform related functions, and are thus particularly useful for membrane proteins. Early responsive to dehydration (*ERD*) genes are rapidly induced in response to dehydration stress in a variety of plant species. In the present work we characterized function of *Brassica juncea ERD4* gene using computational approaches. The ERD4 protein of unknown function possesses ubiquitous DUF221 domain (residues 312–634) and is conserved in all plant species. We suggest that the protein is localized in chloroplast membrane with at least nine transmembrane helices. We detected a globular domain of 165 amino acid residues (183–347) in plant ERD4 proteins and expect this to be posited inside the chloroplast. The structural-functional annotation of the globular domain was arrived at using fold recognition methods, which suggested in its sequence presence of two tandem RNA-recognition motif (RRM) domains each folded into βαββαβ topology. The structure based sequence alignment with the known RNA-binding proteins revealed conservation of two non-canonical ribonucleoprotein sub-motifs in both the putative RNA-recognition domains of the ERD4 protein. The function of highly conserved ERD4 protein may thus be associated with its RNA-binding ability during the stress response. This is the first functional annotation of ERD4 family of proteins that can be useful in designing experiments to unravel crucial aspects of stress tolerance mechanism.

Citation: Rai A, Suprasanna P, D'Souza SF, Kumar V (2012) Membrane Topology and Predicted RNA-Binding Function of the 'Early Responsive to Dehydration (ERD4)' Plant Protein. PLoS ONE 7(3): e32658. doi:10.1371/journal.pone.0032658

Editor: Vasilis J. Promponas, University of Cyprus, Cyprus

Received August 26, 2011; **Accepted** February 2, 2012; **Published** March 14, 2012

Funding: Department of Atomic Energy, India, supported the research. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Competing Interests: The authors have declared that no competing interests exist.

* E-mail: sfdsouza@barc.gov.in (SFD); vinay@barc.gov.in (VK)

Introduction

Dehydration is one of the most common environmental stresses that soil plants are exposed to affecting their growth and development through alternation in metabolism and gene expression [1]. Plants induce a large number of genes under water stress, which can be divided into two categories based on the time of induction: responsive to dehydration and early responsive to dehydration [2,3]. However, the exact function of many stress tolerance associated gene products is still unknown and the encoded proteins have been grouped as hypothetical domains of uncharacterized functions (DUF).

Early responsive to dehydration (*ERD*) genes are rapidly induced to respond to dehydration and various other abiotic stresses. A total of sixteen complementary DNAs for early response to dehydration genes have been isolated from 1 hour dehydrated *Arabidopsis thaliana* which included the *ERD4* gene [4]. The *ERD4* encoded protein (ERD4) has been validated as gene product in *A. thaliana* [2,4–5], in *Zea Mays* [6], and in *Saccharum officinarum* [7]. However, due to lack of information of its structure and function, ERD4 has been classified as belonging to DUF221 protein family (Pfam, PF02714) found in a family of hypothetical transmembrane proteins, none of which have any known function. Also, the organelle localization of the ERD4 protein has been debated in plasma, mitochondria and chloroplast membranes.

The identification of geometric relationships between protein structures, by the use of structural alignment methods, offers a powerful approach in identifying structural and functional relationships between highly divergent proteins [8]. It is well established that proteins evolve partly through rearrangements of larger fragments, typically domains, and nature of these fragments determine biological function of proteins [9]. The analysis of proteins at individual domain levels can facilitate functional annotation of uncharacterized genes and proteins [10–12]. Recently, function of a large number of proteins of DUF families has been proposed based on the structural homology of experimentally determined structures to functionally annotated proteins [13]. The functional domains can also be identified reliably by computational analysis such as prediction of the secondary structure, transmembrane segments, and by fold-recognition [14,15]. An atomic model of the identified domain can further be obtained from the sequence alone by identifying homologs using sequence-sequence comparison or by fold assignment using structure-sequence alignment [16,17]. With the available computational tools, it is also possible to identify residues involved in the biological function based on the structure-structure

comparison. The utility of these approaches can be extended for predicted structural models of uncharacterized proteins enabling functional annotation of related proteins. Such a strategy is particularly useful for membrane proteins as their experimental structure-function determination is a difficult task.

We investigated the function of the *Brassica juncea* ERD4 protein using a combination of advanced sequence profile searches and structure prediction bioinformatics approaches like fold recognition and comparative modeling. We found a globular domain in ERD4 sequence. The globular domain resides inside the chloroplast and belongs to RNA-binding protein superfamily. The domain has two RNA-recognition motifs, typical of RNA-binding proteins. Also, conservation of the RNA-binding residues was observed by structure comparison methods.We suggest that ERD4 has a role in post transcriptional gene regulation. The bioinformatics analyses presented here offers the first hypothesis about the function of the ERD4 family of proteins.

Results

Sequence and phylogenetic analyses

The 3291 bp long nucleotide sequence of *B. juncea ERD4* gene structure study suggests that this gene codes for mRNA of length 2172 (6 exons and 5 introns) which encodes 723 amino acids long protein (UniProtKB, A9LIW2). The homologs of *B. juncea* ERD4 protein were identified in various plant lineages, for instance in bryophyta (*Physcomitrella patens*), in traceaophyta (*Selaginella moellendorffii*), in euphylophyta (*O. sativa*, *A. thaliana*). The protein was found to be conserved in all the plants for which proteome data was available (Fig. 1). Phylogenetic tree of plant ERD4 homologs showed four distinct clades and the evolution pattern of this gene followed the lineages evolution (Fig. 1). The presence of both putative RNA-binding and DUF221 domains, a characteristic of plant ERD4 proteins, was also detected in unicellular (*C. reinhardtii*) and multicellular (*V. carteri*) green algae genomes by iterative PSI-BLAST search. The algal proteins, however, consists of 1746 and 1172 residues, respectively (UniProtKB, A8HT24 and D8TSA1). However, homolog of plant ERD4, possessing both the RRM and DUF221 domains, were not detected in bacteria (including cyanobacteria) and archae. Counter intuitively, ERD4-like proteins were detected in unicellular non-photosynthetic eukaryotes like *Dictyostelium fasciculatum* (slime mould) and colonial flagellates like *Choanoflagellates*. These proteins showed 24.5% (52.7%) and 19% (40%) sequence identity (similarity), respectively, with *B. juncea* ERD4 protein over the complete length. We also detected proteins possessing both the RNA-binding and DUF221 domains in fungi including many plant pathogens (for instance, in *Phytophthora sojae*) and in animals. A *Homo sapien* ortholog of the identified animal proteins has recently been characterized as ''transmembrane protein 63A'' (UniProt/KB, O94886; TM63A_-human). The human protein consists of 807 amino acid residues and shows 24% (41%) sequence identity (similarity) over 608 residues with *B. juncea* ERD4 protein (Fig. S1).

The motif scanning (motif_scan) and domain detection tools (Pfam, DOUTfinder and SMART) detected presence of DUF221 domain (residues 312–634) in the ERD4 sequence with very high confidence (E-value, 7e-146). The DUF221 domain is found in a family of hypothetical transmembrane proteins none of which have any known function. This domain has been identified in all forms of eukaryotic organisms and has been observed in different domain architectures in combination with a variety of other functional domains like PIWI, phosphate metabolism protein etc. The DOUTfinder also identified potential similarity with eukaryotic RNA-recognition motif with 10% false-positive rate.

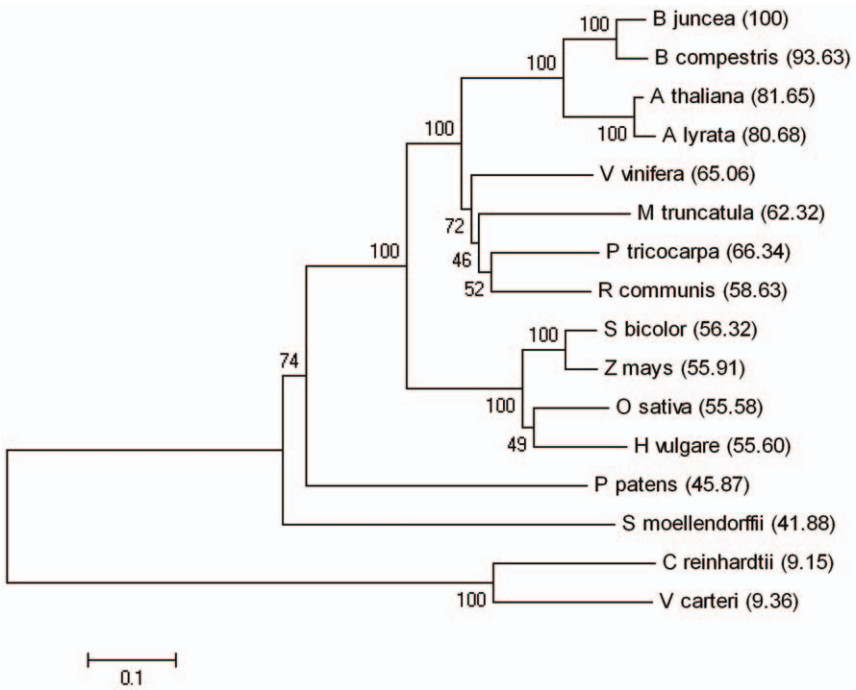


Figure 1. Evolutionary relationship among ERD4 homologs. Evolutionary relationship was inferred using the Neighbor-Joining method in MEGA4 software. The percentage of replicate trees in which the associated taxa clustered together in the bootstrap test (100 replicates) is shown next to the branches. The tree is drawn to scale with branch lengths in the same units as those of the evolutionary distances used to infer the phylogenetic tree. The evolutionary distances are in the units of the number of amino acid substitutions per site. Also shown in brackets are the pair-wise percentage identity between *B. juncea* ERD4 and other plant proteins, including green algae.
doi:10.1371/journal.pone.0032658.g001

The biological relevance of this was, however, not clear owing to highly distant sequence similarity as suggested by poor D-score of 163 [18].

Transmembrane topology and localization

Transmembrane helices in the ERD4 sequence were identified using several web-servers *albeit* with some differences. The number of identified helices varied from 9 to 11 and the suggested starting- and end- points for predicted transmembrane segments also differed. Based on high-confidence predictions from different servers, nine transmembrane helices belonging to the sequence regions of 6–26, 90–111, 149–167, 365–385, 419–437, 457–476, 501–531, 573–593 and 638–659 were identified (Fig. 2). The identification of the transmembrane helices was consistent with the predicted secondary structure which suggested that the ERD4 protein is mainly helical with 64.3, 5.4 and 30.3% residues in helix, extended and coil structures, respectively. Interestingly, all the transmembrane prediction tools showed that a long polypeptide segment (residues 170–360) did not possess transmembrane helices (non-transmembrane segment). A globular domain was subsequently detected in this segment.

Maximum probability of localization of this protein was predicted in plasma membrane (with score of 10) followed by chloroplast (score 2) using Wolf PSORT tool. The YLoc tool, however, suggested its presence in chloroplast with 53.9% probability and a small confidence (0.27). The TargetP server predicted this protein to be a secretory protein with high confidence (score 0.92). The analysis of *B. juncea* ERD4 by the ambiguous targeting predictor (ATP) suggested a score of 0.39, which weakly suggested dual targeting of the ERD4 protein. The analysis of ERD4 orthologs by the ambiguous targeting predictor, however, suggested wide variations in the confidence score (Table 1) with a low score of 0.19 for some ERD4 proteins that clearly indicated localization of ERD4 in only one compartment. Although the used predictors failed to identify unambiguously the chloroplastic localization of the ERD4 protein, its localization in chloroplast membrane has been shown experimentally in Arabidopsis [19].

It has been earlier shown that N-terminal sixty residues contain signal sequence for chloroplastic localization, sixteen of which could be used to discriminate between mitochondrial and chloroplastic localization [20]. In order to get detailed information on the amino acid composition of presequences for chloroplast envelope targeting, we analyzed experimentally validated chloroplastic envelope proteins of *A. thaliana*. An overall amino acid composition and N-terminal sequence logo plots of the 123 selected proteins (ENV dataset) from Arabidopsis proteome [19] were analyzed. The positional abundance of amino acids in sequence logos showed abundance of Ser residues and underrepresentation of Arg residues in the ENV dataset. However, no clear position-specific pattern was observed in sequence logo plots. Similar trends have earlier been observed for the total chloroplast proteins, including stroma proteins [20,21]. The amino acid composition analysis also showed much higher abundance of Ser, Ala and Leu residues in the N-terminal sixteen residues as compared to the full-length proteins (Fig. 3A). Also, the percentage of Arg residues in the N-terminal sixteen residues was observed to be lower than that observed in full-length or N-terminal sixty residues. The analysis of the N-terminal sixteen residues of the ERD4 orthologs also showed similar trends; higher abundance of potentially hydroxylated Ser/Thr residues and of hydrophobic Phe/Ile residues. The N-terminal sixteen residues also showed high differences in the abundance of Arg and Lys residues, as compared to the N-terminal sixty and overall composition of these proteins. These positively charged residues are underrepresented in the N-terminal sixteen residues of the ERD4 orthologs (Fig. 3B). The lower abundance of Arg and Lys residues in the N-terminal sixteen residues of chloroplast proteins, compared to mitochondrial proteins, has been earlier observed by Bhushan et al. [20]. The low percentages of the positively charged Arg/Lys residues and significantly higher percentage of Ser residues in the N-terminal sixteen residues of ERD4 proteins thus corroborated experimental determination of the ERD4 protein in *A. thaliana* chloroplast envelope proteome.

The inside or outside localization of the non-transmembrane fragment (inside or outside the chloroplast membrane) depended upon the orientation of N-terminal transmembrane helix. While MEMSAT and TMpred showed its placement inside the membrane, several other tools like HMMTOP, TMHMM, TMMod predicted its presence outside the membrane. These predictions resulted in two distinct membrane topologies and the ambiguity was resolved using frequency of the positively charged residues in both the possible topologies. It was concluded that N-terminus of ERD4 was outside the membrane as nearly 79% of the positively charged residues were observed to reside on inside loops. The corresponding transmembrane topology model revealed presence of the non-transmembrane segment (residues 170–360) inside the chloroplast (Fig. 4). The predicted secondary structure showed nearly 47% residues in helix, 12.6% residues in β-strand and 40.4% residues in the coil structure, respectively, in this segment.

Structural analysis of the globular domain

A BLAST search with the amino-acid sequence did not reveal any close homologue in the database of known protein structures (PDB). This is not unusual as sequence comparison methods cannot reliably detect evolutionary relationship between highly divergent proteins. The structural fold of the ERD4 domain was then found by fold-recognition methods, which use sequence-structure alignment. This method allows detection of remote homologies beyond the detection limits of other sequence comparison methods. The input for fold-recognition was *B. juncea* ERD4 sequence from which generated profile was compared to sequence profiles of proteins and domains of known structures. The search for ERD4 protein fold using fold-recognition meta-server suggested structural homology of about 165 amino acid residues (183–347) with the known RNA-binding globular proteins. Interestingly, all the best hits identified by the 3D-jury from the meta-server were RNA-binding proteins possessing two well known RNA-recognition motifs (RRM) (Table 2). The residues 183–347 of the ERD4 sequence were thus expected to adopt a globular fold with structural similarity with RNA-binding proteins

The 3D structural models of the globular domain were constructed using the solution structure of the RBD1,2 domains from human nucleolin (PDB code, 2KRR; Jscore, 55.3) and using X-ray crystal structure of the poly(a)-binding protein in complex with polyadenylate RNA (PDB code, 1CVJ; Jscore, 48) as templates. Given the high divergence between ERD4 globular domain and the RNA-recognition proteins used for constructing the theoretical models with pair-wise sequence identity of about 10% (Table 2), we would expect the general atomic resolution of the theoretical model to be low (>3 Å). However, all the structural neighbors of the ERD4 globular domain were found by DALI program [22] to belong to RNA-binding domain superfamily. The computationally constructed structural models for the ERD4 chloroplastic domain clearly showed the presence of two tandem RNA-recognition motifs, each having βαββαβ topology (Fig. 5).

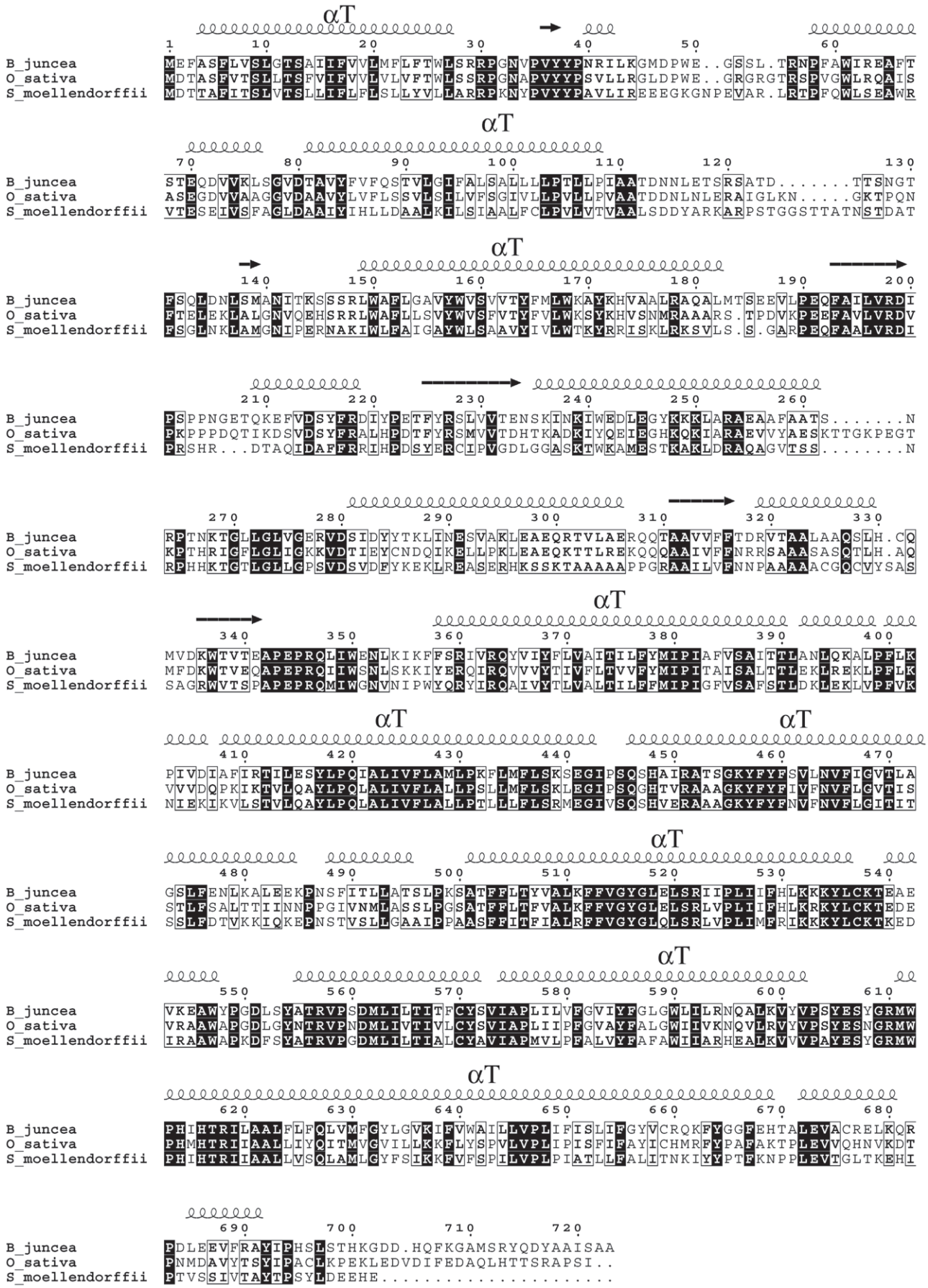

αT
αT
αT
αT
αT
αT
αT
B_juncea
O_sativa
S_moellendorffii

Figure 2. Multiple sequence alignment of plant ERD4 sequences. The alignment of all available plant ERD4 sequences was achieved using PROMALS3D [42] and only three diverse sequences are shown here. Also shown is the consensus secondary structure predicted by PsiPred; helices are shown as coils and strands are shown as arrows. The nine transmembrane helices are marked as αT. The strictly conserved residues in all the plant ERD4 sequences are shaded, while similar residues are boxed. The residues numbering is of the full-length *B. juncea* ERD4 protein. The figure was prepared with EsPript suite [64].
doi:10.1371/journal.pone.0032658.g002

The two RRM domains are composed of amino acid residues 183–269 (RRM1) and 273–347 (RRM2) respectively, and are joined by an interdomain linker peptide. The interdomain linker peptide is a typical characteristic of known RNA-binding proteins with multiple RRM domains [23]. The two RRM domains could be flexibly tethered via the linker peptide. Analogous to the well characterized RNA-binding proteins, the β-sheets of the two RNA-binding domains of ERD4 face each other and RNA substrates could bind in the cleft.

The two RNA-recognition domains of ERD4 were individually superposed onto the known RNA-binding domains of sex lethal protein (PDB code, 1B7F) and adenosine-uridine (AU)-rich binding Hu protein (PDB code, 1FXL). These proteins had similar number of amino acids as ERD4 globular domain but differed significantly from the latter (DaliLite Z-scores for ERD4/1B7F and ERD4/1FXL pairs were 5.8 and 5.5, respectively) and thus formed highly diverse pairs. Additionally, these structures had been refined to high precision against single crystal diffraction data and coordinates of protein-RNA complexes were available, which could hint RNA-binding mode in the ERD4 protein (Fig. S2). The structural alignment showed the presence of two non-canonical ribonucleoprotein sub-motifs (RNP1 and RNP2) in both the ERD4 domains (Fig. 6). One of the ribonucleoprotein sub-motifs (RNP2) resides on the first β-strand, while residues from third β-strand contribute towards RNP1. The putative RNP sub-motifs of RRM1 are 195-ILVRDI-200 (RNP2) and 237-INKIWEDL-244 (RNP1) and those of RRM2 are 283-DYYTKL-288 (RNP2) and 307-RQQTAAVVF-315 (RNP1). In the multiple sequence alignment of ERD4 orthologs, the RRM1 domain has conserved hydrophobic (Leu/Val) at position-2 of the RNP2 and aromatic (Trp/Tyr) at position-5 in RNP1 (Fig. 6). Also, Tyr/His and Ala are conserved in RNP2 position-2 and RNP1 position-5, respectively, in the RRM2 domain. A positively charged amino acid residue (Arg/Lys) was also found in most of the plant ERD4 proteins at RNP1 position-1 of RRM2. In addition to the â-strands, the loops β1/α1 (connecting β1 and α1 elements), β2/β3 and α2/β4 have also been observed in RNA-binding proteins to interact with nucleic acid substrates [23]. Most of these residues are conserved in ERD4 orthologs (Fig. 6). Interestingly *B. juncea* Pro-201, residing on the loop β1/α1, is strictly conserved in all the plant ERD4 proteins. This position is occupied by Pro/Ser residues in majority of RNA-binding domains identified in NCBI conserved domains database CD00590 [24].

Table 1. Prediction scores for dual organelle targeting of plant ERD4 proteins assessed by ambiguous targeting predictor (APS).

Plant species	Accession code	Source	APS prediction score
Brassica juncea	A9LIW2	UniProtKB	0.39122
Brassica campestris	A8IXK5	UniProtKB	0.39122
Arabidopsis thaliana	Q9C8G5	UniProtKB	0.19248
Arabidopsis lyrata	D7KET4	UniProtKB	0.19248
Populus tricocarpa	B9GJG0	UniProtKB	0.39122
Sorghum bicolor	C5X9J3	UniProtKB	0.47346
Vitis vinifera	F6HLU8	UniProtKB	0.30121
Oryza sativa	Q6ZLQ0	UniProtKB	0.34804
Zea mays	B0FSL2	UniProtKB	0.47346
Medicago truncatula	AES64128	GenBank	0.20827
Ricinus communis	B9SY14	UniProtKB	0.39122
Hordeum vulgare	F2DDW1	UniProtKB	0.34804
Physcomitrella patens	A9TEC4	UniProtKB	0.41759
Selagilella moellendorffii	D8STJ2	UniProtKB	0.29168
Chlamydomomas reinhardtii	A8HT24	UniProtKB	0.49063
Volvox carteri	D8TSA1	UniProtKB	0.21542

doi:10.1371/journal.pone.0032658.t001

Discussion

A close homolog of *Brassica juncea* ERD4 protein was detected in all plant species indicating conservation of the protein in plantae kingdom. Phylogenetic relationship of this gene showed similar pattern of divergence as different plant lineages have evolved, emphasizing that *ERD4* gene has been essentially maintained during the course of plant evolution (Fig. 1).

A consensus assignment using high confidence prediction scores suggested that ERD4 is a transmembrane protein with at least nine transmembrane helices in the ERD4 sequence (Fig. 2). Its localization in different plant organelle has been subject of intense discussion recently. Its localization in the chloroplast membrane was earlier suggested from the Arabidopsis chloroplast envelope proteome analysis [5,19], while Alexandersson et al. [25] identified its location in plasma membrane of *Arabidopsis thaliana* that could have been due to organelle contamination [26]. Further, mitochondrial and plastid dual targeting of *A. thaliana* ERD4 was suggested [27]. The analysis of homologous plant ERD4 sequences was used here for confirming its organelle localization on the premise that localization signatures must be strictly conserved in all the plant ERD4 sequences. The analysis of ERD4 orthologs by the ambiguous targeting predictor suggested wide variations in the confidence score; a low score of 0.19 for a number of ERD4 orthologs (Table 1) clearly indicated its localization in only one compartment. Its presence in chloroplast membrane, however, was inferred on the basis of higher abundance of Ser/Thr and underrepresentation of Arg/Lys residues in the N-terminal sixteen residues of ERD4 orthologs, as also observed earlier for the chloroplast proteins [20]. We also found marked increase in percentage of hydrophobic Ala/Leu residues in the N-terminal sixteen residues for chloroplast envelope proteins of *A. thaliana*. Similar high percentage of hydrophobic Phe/Ile residues was observed in the N-terminal sixteen residues of ERD4 orthologs (Fig. 3B). Taken together these data support the experimental finding of its localization in chloroplast membrane. The presence of ERD4 in the chloroplast is also consistent with predominance localization of the organelle stress response proteins in chloroplast as noted recently by Taylor et al. [28]. The detection of ERD4-like protein in uni- and multicellular green algae provides further credence to our suggested chloroplastic localization of the ERD4 protein, as all plastids derive from a

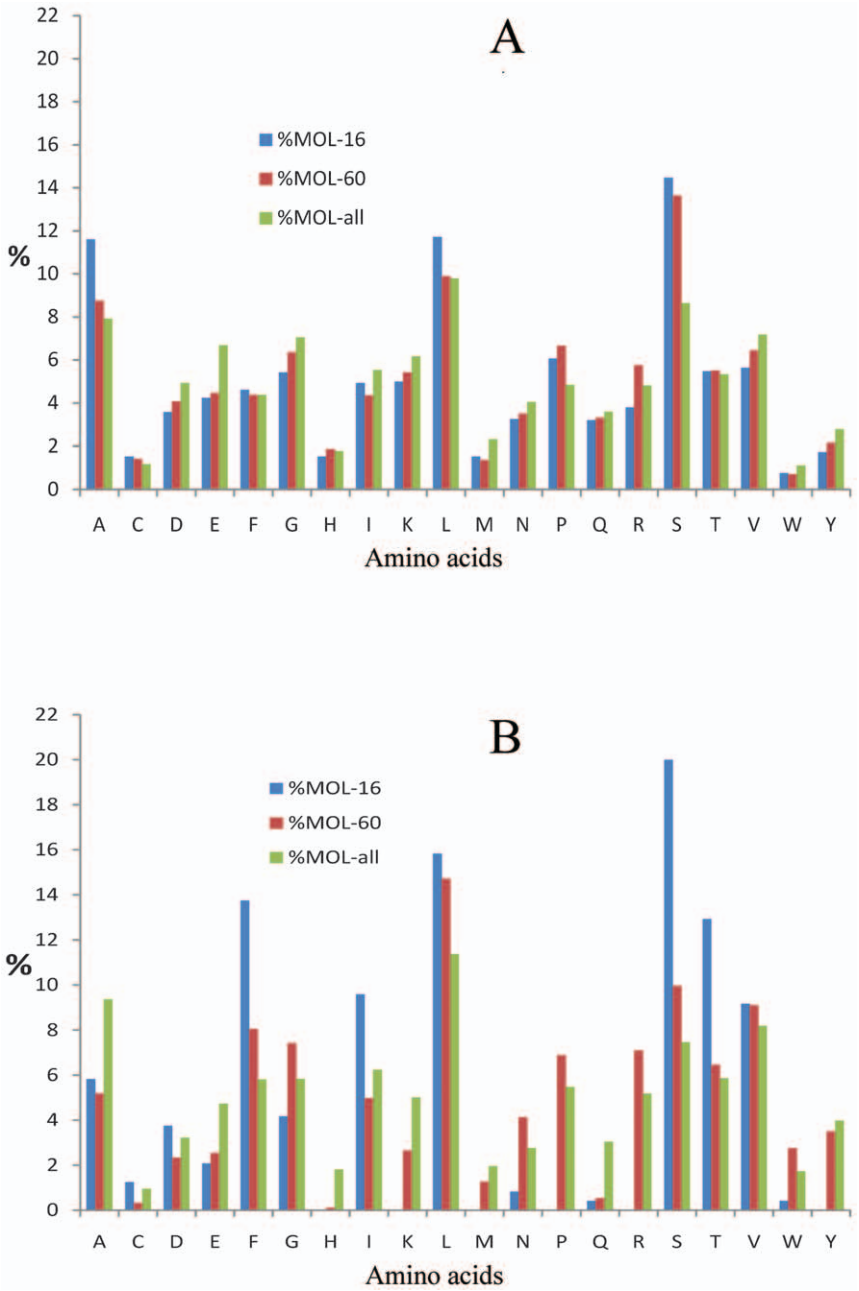


Figure 3. Amino acid composition of presequences. Analysis of the amino acid composition of the N-terminal sixteen residues (%MOL-16), N-terminal sixty residues (%MOL-60) and full-length proteins (%MOL-all) (A) analysis of the 123 chloroplast envelope proteins of *A. thaliana* (B) analysis of plant ERD4 orthologs.
doi:10.1371/journal.pone.0032658.g003

single endosymbiosis and after plastid acquisition only photosynthetic eukaryotes diverged into glaucocystophytes, rhodophytes, and viridiplantae lineages [29–31]. However, ERD4-like protein was not detected in cyanobacteria. Previous findings have also reported that plant proteins encoded by genes of cynobacterial origin are not, as a rule, targeted to chloroplast, whereas many non-cynobacterial proteins can be targeted to plastids [32].

A transmembrane DUF221 domain (312–634) and a globular domain (183–347) were identified in the Brassica ERD4 sequence. The DUF221 domain has been identified in all forms of eukaryotic organisms and has been observed in nearly 23 different domain architectures in combination with a variety of other functional domains like Dnaj, UBQ, VWD etc. The existence of structural domain, with a common function, in combination with variety of other domains has been known to be responsible for evolution of protein repertoire [33]. The DUF221 domain has no other known function, except for membrane integration. It is likely that biological function of the ERD4 protein is attributed mainly to the globular domain, and DUF221 helps in localization of the functional (globular) domain. The deduced topology, based on the positive-inside rule, reveals that the globular domain resides inside the chloroplast (Fig. 4). The smaller loops reside on outside the membrane confirming also to the observation that periplasmic loops are short possibly because of difficult translocation of intermediate-length loops [34].

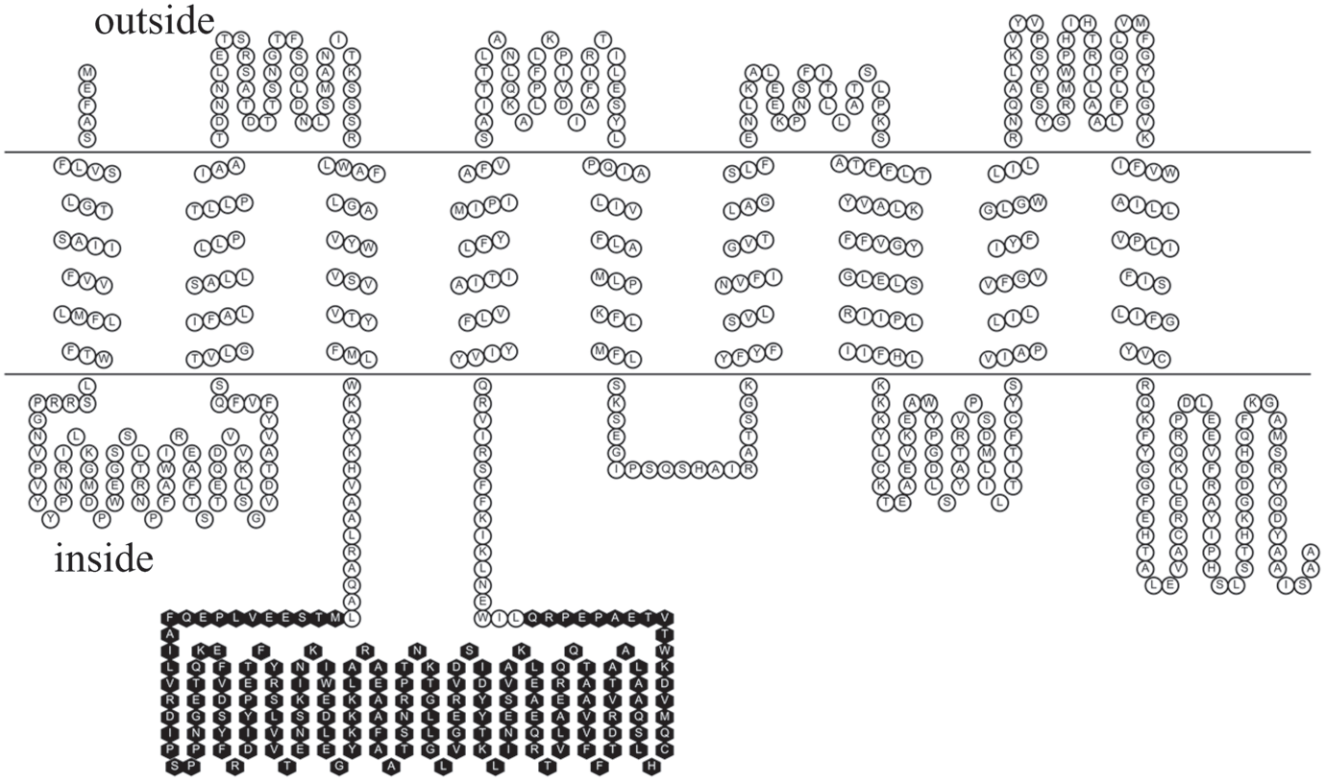


Figure 4. The topology of the *B. juncea* ERD4 protein. The toplogy was drawn using TOPO2 tools. The nine transmembrane helices are shown. Also, shown (filled hexagons) is the globular domain containing RNA-recognition domains. The globular domain is suggested to reside inside the chloroplast.
doi:10.1371/journal.pone.0032658.g004

The structural analysis is known to reveal the distant evolutionary links that could provide the first hypothesis about biological function of the uncharacterized domains [13]. The tertiary structure of the ERD4 chloroplastic globular domain was predicted by fold-prediction algorithms that suggested presence of two RNA-recognition motifsin its sequence. Each of the RRM was predicted to adopt βαββαβ topology (Fig. 5,6). The fold of the ERD4 globular domain was found to be shared only by RNA-binding domains, as observed in the search for structural neighbors with DALI programs. Structural and sequence comparison with the known RNA-binding proteins showed the presence of RNP1 and RNP2 ribonucleoprotein sub-motifs in both the identified RNA-recognition motifs of ERD4. The four RNP's in two RRM domains reside on the â-strands creating a RNA binding cleft (Fig. 5). A hydrophobic and an aromatic amino acid residue at 2nd and 5th positions of RNP2 and RNP1, respectively, were conserved in RNA-binding proteins and ERD4 homologs (Fig. 6). These residues stack against the two bases of substrate RNA in the known RNA-binding proteins. The 1st position of RNP1 in RRM2 of ERD4 was also found to be conserved as positively charged amino acid that could neutralize the negatively charged phosphodiester group [35]. In most of the RRM-RNA complex structures only one to three of these contacts are observed with two stacking interactions involving RNP2 position-2 and RNP1 position-5 observed most frequently [36]. The orthologs of TM63A_human protein identified by BLAST

Table 2. The best five structural models predicted for the ERD4 globular domain by the fold-recognition servers and their ranking by 3D-Jury method.

Model (1)	3D-Jury score (JScore)	Scop [63]		Percentage identity/similarity with *B. juncea* ERD4 globular domain
		Classification	Superfamily	
2krr _A	55.3	54928	RNA-binding domain	9.6/27.1
2dhs_A	54.0	54928	RNA-binding domain	12.1/36.4
1cvj_A	48.0	54928	RNA-binding domain	7.8/26.1
2g4b_A	41.0	54928	RNA-binding domain	11.5/30.9
3md3_A	39.7	54928	RNA-binding domain	10.3/32.1

(1) PDB identifier code.
doi:10.1371/journal.pone.0032658.t002

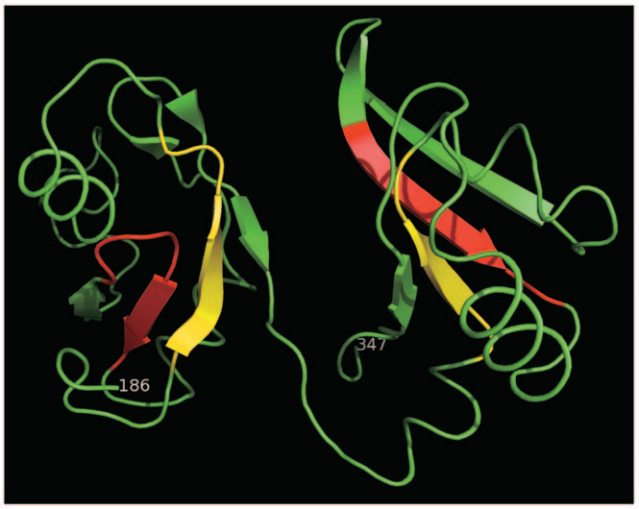


Figure 5. Ribbon model of the putative RNA-binding globular domain. The ribbon model was constructed by comparative homology approaches. The fold of the domain was identified by fold-prediction meta-server. Due to low pair-wise sequence identity of nearly 10% between the query and identified template, the derived atomic coordinates for the ERD4 globular domain were expected to be of low-resolution. The two ribonucleoprotein motifs (RNP1 and RNP2) in each of the RNA-recognition domains are shown in red and yellow, respectively. The figure was prepared by PyMol (http://www.pymol.org/).
doi:10.1371/journal.pone.0032658.g005

search due to its sequence similarity with plant ERD4 proteins, however, do not show strict conservation in the residues corresponding to the proposed RNA-binding domain of ERD4 (Fig. S1). In contrast to RNA-binding ability, polypeptides that recognize protein substrates, and not RNA, have only one RRM domain. The combination of two or more RNA-recognition motifs, as observed in ERD4 sequences, often results in dramatically increased RNA-binding affinity [23,37].

The RNA binding domain carrying RNP signature sequences is a highly abundant domain in eukaryotes. This domain has been found in a variety of heterogeneous nuclear ribonucleoproteins (hnRNPs), proteins implicated in regulation of alternative splicing, and protein components of small nuclear ribonucleoproteins (snRNPs), and is involved in post-transcriptional gene expression processes including mRNA and rRNA processing, RNA export, and RNA stability. The domain binds a variable number of nucleotides, ranging from two to eight. It is, however, known that despite using the same β-sheet surface to bind RNA, each protein achieves sequence-specificity slightly differently [23].The conservation of two tandem RNA-recognition motifs and the substrate binding residues suggests that globular domain of ERD4 protein may be RNA-binding competent.The ERD4 protein can particpate in mRNA metabolism such as sequestering and protecting mRNAs during conditions of limiting transcription. In plants, the RNA-binding proteins may modulate ABA signaling through the alteration of mRNA processing events such as splicing, processing, nuclear export, transcript stability and RNA degradation [38]. Also, induction of ERD4 could influence the membrane fluidity as its DUF221 domain is expected to be integrated in the chloroplast membrane. It hence assumes significance to study functionally important residues and domains that are critical for ERD4 activity in response to various environmental conditions. We also suggest from the analysis that ERD4 proteins may be characterized by the presence of both RRM and DUF221 domains and not by DUF221 domain alone as is the current practice in putative annotations in the sequence databases.

Conclusion

The ERD4 protein is a transmembrane protein whose role has been identified in abiotic stress amelioration in plants. Based on sequence analysis, we expect its location in chloroplast membrane. A globular chloroplastic domain was detected in its sequence that is suggested to possess two tandem RNA-recognition motifs. Detection of RNA binding residues in the globular domain further suggests that the biological function of ERD4 may be associated with its RNA-binding ability. Understanding of structure-function of *ERD4* gene product may help in understanding plant stress response and in enhancing plant tolerance to environmental stresses.

Materials and Methods

Sequence based analyses

The *Brassica juncea ERD4* gene sequence was obtained from the Genbank (accession number: EU126607). Gene structure study was performed using popular gene finding pipeline (FGENESH at www.softberry.com). The homologs of *B. juncea* ERD4 protein sharing better than 40% sequence identity were obtained from UniProt database using FASTA search engine. The search for ERD4 homologs using BLAST search engine was carried out also against the non-redundant protein sequences and against translated individual proteome of *C. reinhardtii*, *C. merolae*, several fungi and cyanobacterial (*Synechococcus* sp. RS9916, *Cyanothece* sp., *Nostoc punctiforme*) genomes. To detect ERD4-like proteins in animals, BLAST search against non-redundant protein sequences of animalia (taxid:33208) kingdom was also carried out. Since complete proteome database for *T. aestivum* is yet not available, the search for its homolog was carried out in Ensembl [39] employing tBLASTn [40] search engine.The search of distantly related genomes or those of unrelated species was constrained for the presence of two tandem RNA-recognition motifs and a DUF221 domain detected in the closely related plant species (for discussion on RRM see Results). Multiple sequence analyses were carried out using clustalW and PROMALS3D tools [41,42]. The phylogenetic tree was derived from that multiple alignment using Neighbor-Joining method in MEGA4 [43]. Motifs were identified using motif scan tools [44].

Localization and Topology prediction

The prediction for sub-cellular localization of the *B. juncea* ERD4 protein and its orthologs was done using wolf PSORT [45], YLoc [46], TargetP [47],and ambiguous targeting predictor [28] web-tools. Further a subset consisting of 123 chloroplastic envelope proteins of *A. thaliana* chloroplast proteome [19] was analyzed for chloroplast localization signatures. These proteins were identified from the experimentally validated chloroplast envelope protein dataset, those not showing similarity with ribosomal proteins. Amino acid contents were calculated from the complete protein sequence, and for N-terminal sixteen and sixty amino acid residues of this subset of validated chloroplastic proteins and for plant ERD4 proteins.

Secondary structure of the plant ERD4 orthologs were predicted using PsiPred [48] and Prof (http://www.aber.ac.uk/~phiwww/prof/) suites. The web-versions of nine different topology prediction methods were used to estimate membrane topology of ERD4 and these were: DAS [49], HMMTOP [50], MEMSAT [51], TMHMM [52], TMMod [53], TMpred [54], Toppred [55], Conpred [56] and phobias [57]. Modeling of

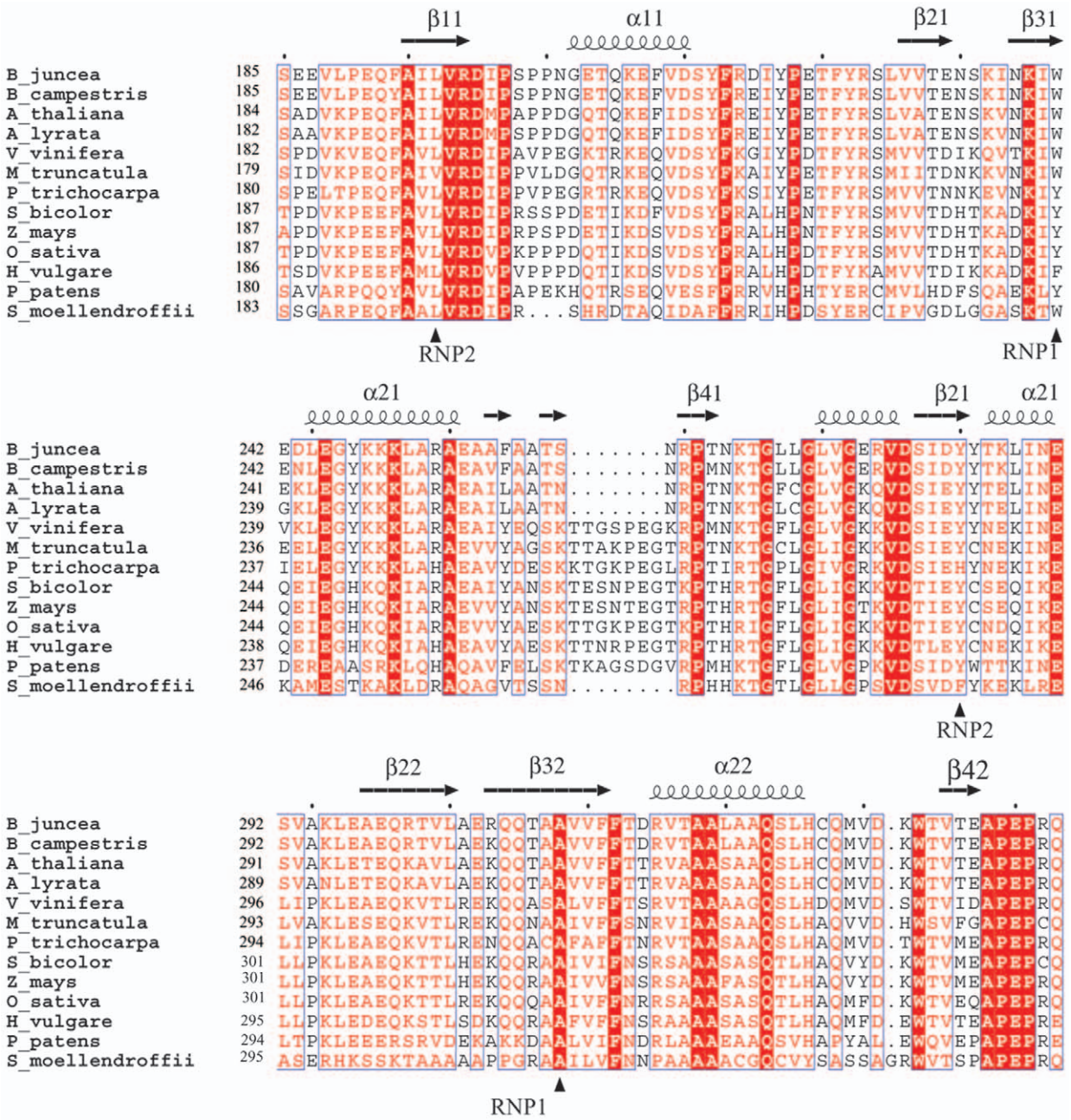


Figure 6. Multiple sequence alignment of the ERD4 globular domain. The alignment was generated by ClustalW. The two RNA-recognition domains are composed of amino acid residues 183–269 (RRM1) and 273–347 (RRM2), respectively. The two ribonucleoprotein motifs of each RRM domain are marked as RNP1 and RNP2. The suggested RNA-interacting residues are marked with filled triangle (▲). The secondary structure elements of each RRM domain in the theoretical structural model are also shown. The strictly conserved residues in all the plant ERD4 sequences are shaded, while similar residues are boxed. The residues numbering is of the full-length ERD4 proteins.
doi:10.1371/journal.pone.0032658.g006

transmembrane topology was done using TOPO2 (http://www.sacs.ucsf.edu/TOPO-run/topoanal-adv2.pl).

Prediction of the functional domains and 3D structure

The *B. juncea* ERD4 sequence was subjected to Pfam [58], DOUTfinder [18] and SMART [59] analysis for identification of the known domains and domain architecture. An independent analysis for detecting globular domains of structural-folds similar to the known protein structures was also carried out using structure prediction meta-server (http://bioinfo.pl/meta) accessing various fold-recognition and function prediction methods. A globular domain in ERD4 sequence was detected by the fold-prediction meta-server. The database of known protein structures (Protein Data Bank, PDB) was searched for a structure homologus to the detected globular domain using sequence-sequence comparison search engines. In the absence of any known homologus structure, the tertiary fold of the globular domain was independently predicted using the meta-server. The collected results from fold-prediction servers were screened with 3D-jury [60]. The 3D structural model of the globular domain was

constructed with Modeller [61] using sequence-to-structure alignment returned by the meta-server, and RNA-binding domains from human nucleolin (PDB code, 2KRR) and poly(a)-binding protein (PDB code, 1CVJ) as templates. The structural neighbors of the theoretical structural model of the globular domain were identified by the DALI [22] programs.

Identification of functional residues

The 3D structural model of the identified globular domain was superposed onto the known structures of RNA-binding proteins which possessed RNA-recognition domains. The atomic coordinates of these were obtained from the PDB. The superposition was achieved using DALI programs and Swiss PDBViewer [62]. The amino acid residues of the ERD4 domain, equivalent to the residues interacting with RNA substrates in the known RNA-binding proteins, were identified as putative RNA-binding residues. The conservation of these was verified in the alignment of the amino acid sequences of the identified RRM domains of ERD4 homologs.

Supporting Information

Figure S1 Multiple sequence alignment of plant ERD4 and proteins of animalia (taxid:33208) kingdom identified by BLAST. The alignment of plant ERD4 sequences [*B. juncea* (UniProtKB, A9LIW2) and *A. thaliana* (UniProtKB, Q9C8G5)] and diverse animal sequences [*H. sapiens* (UniprotKB, O94886), *X. laevis* (UniProtKB, Q5PQ13) and *N. vectensis* (UniProt KB, A7S3E8)] was achieved using PROMALS3D [1]. The strictly conserved residues are shaded, while similar residues are boxed. The proposed RNA-binding domain of *B. juncea* ERD4 is marked as RBD. A number of insertion/deletions and poor amino acid conservation in the corresponding domains of animal sequences do not suggest close evolutionary relationship between plant and animal proteins. The figure was prepared with EsPript suite [2].
(TIF)

Figure S2 Cartoon of RNA-binding domain with bound RNA. Cartoon of HuD1,2–cfos-11 RNA complex structure [PDB code 1FXL; 3]. The RNA is shown as a stick model (orange). The N- & C- termini of the protein are marked as N and C, respectively. The two RRM domains form a cleft with the RNA bound between the β-sheets surfaces. In several RNA-binding proteins the two RRM domains are flexibly tethered via a linker peptide.
(TIF)

Author Contributions

Conceived and designed the experiments: AR VK. Performed the experiments: AR VK. Analyzed the data: AR VK. Contributed reagents/materials/analysis tools: PS SFD. Wrote the paper: AR VK.

References

1. Manssfield TA, Atkinson CJ (1990) Stomatal behavior in water stressed plant. *In*: stress responses in plants: Adaptation and acclimation mechanism Alscher RG, Cumming JR, eds. New York, Chichester, Brisbane, Toronto, Singapore: Wiley-Liss. pp 241–264.
2. Shinozaki K, Yamaguchi-Shinozaki K (1997) Gene Expression and Signal Transduction in Water-Stress Response. Plant Physiol 115: 327–334.
3. Shinozaki K, Yamaguchi-Shinozaki K (2007) Gene networks involved in drought stress response and tolerance. J Exp Bot 58: 221–227.
4. Kiyosue T, Yamaguchi-Shinozaki K, Shinozaki K (1994) Cloning of cDNAs for genes that are early-responsive to dehydration stress (ERDs) in *Arabidopsis thaliana* L.: identification of three ERDs as HSP cognate genes. Plant Mol Biol 25: 791–798.
5. Froehlich JE, Wilkerson W, Ray WK, McAndrew RS, Osteryoung KW, et al. (2003) Proteomic study of the Arabidopsis thaliana chloroplastic envelope membrane utilizing alternatives to traditional two-dimensional electrophoresis. J Proteome Res 2: 413–425.
6. Liu Y, Li H, Shi Y, Song Y, Wang T, et al. (2009) A maize early responsive to dehydration gene, ZmERD4, provides enhanced drought and salt tolerance in Arabidopsis. Plant Molecular Biology Reporter 27: 542–548.
7. Camargo SR, Cançado GMA, Ulian EC, Menossi M (2007) Identification of genes responsive to the application of ethanol on sugarcane leaves. Plant Cell Rep 26: 2119–2128.
8. Kolodny R, Petrey D, Honig B (2006) Protein structure comparison: implications for the nature of 'fold space', and structure and function prediction. Curr Opin Struct Biol 16: 393–398.
9. Vogel C, Bashton M, Kerrison ND, Chothia C, Teichmann SA (2004) Structure, function and evolution of multidomain proteins. Curr Opin Struct Biol 14: 208–216.
10. Moore AD, Björklund AK, Ekman D, Bornberg-Bauer E, Elofsson A (2008) Arrangements in the modular evolution of proteins. Trends Biochem Sci 33: 444–451.
11. Petrey D, Fischer M, Honig B (2009) Structural relationships among proteins with different global topologies and their implications for function annotation strategies. Proc Natl Acad Sci USA 106: 17377–17382.
12. Zhang QC, Petrey D, Norel R, Honig BH (2010) Protein interface conservation across structure space. Proc Natl Acad Sci USA 107: 10896–10901.
13. Jaroszewski L, Li Z, Krishna SS, Bakolitsa C, Wooley J, et al. (2009) Exploration of uncharted regions of the protein universe. PLoS Biol 7: e1000205.
14. Bryson K, McGuffin LJ, Marsden RL, Ward JJ, Sodhi JS, et al. (2005) Protein structure prediction servers at University College London. Nucleic Acids Res 33: W36–38.
15. Rost B (2003) Prediction in 1D: secondary structure, membrane helices, and accessibility. Methods Biochem Anal 44: 559–587.
16. Godzik A (2003) Fold recognition methods. Methods Biochem Anal 44: 525–546.
17. Petrey D, Honig B (2005) Protein structure prediction: inroads to biology. Mol Cell 20: 811–819.
18. Novatchkova M, Schneider G, Fritz R, Eisenhaber F, Schleiffer A (2006) DOUTfinder—identification of distant domain outliers using sub-significant sequence similarity. Nucleic Acids Res 34: 214–218.
19. Kleffmann T, Russenberger D, von Zychlinski A, Christopher W, Sjölander K, et al. (2004) The Arabidopsis thaliana chloroplast proteome reveals pathway abundance and novel protein functions. Curr Biol 14: 354–362.
20. Bhushan S, Kuhn C, Berglund AK, Roth C, Glaser E (2006) The role of the N-terminal domain of chloroplast targeting peptides in organellar protein import and miss-sorting. FEBS Lett 580: 3966–3972.
21. Zybailov B, Rutschow H, Friso G, Rudella A, Emanuelsson O, et al. (2008) Sorting signals, N-terminal modifications and abundance of the chloroplast proteome. PLoS One 3: e1994.
22. Holm L, Rosenström P (2010) Dali server: conservation mapping in 3D. Nucleic Acids Res 38: W545–549.
23. Cléry A, Blatter M, Allain FH (2008) RNA recognition motifs: boring? Not quite. Curr Opin Struct Biol 18: 290–298.
24. Marchler-Bauer A, Lu S, Anderson JB, Chitsaz F, Derbyshire MK, et al. (2011) CDD: a Conserved Domain Database for the functional annotation of proteins. Nucleic Acids Res 39: D225–229.
25. Alexandersson E, Saalbach G, Larsson C, Kjellbom P (2004) Arabidopsis plasma membrane proteomics identifies components of transport, signal transduction and membrane trafficking. Plant Cell Physiol 45: 1543–1556.
26. Ferro M, Salvi D, Brugiere S, Miras S, Kowalski S, et al. (2003) Proteomics of the chloroplast envelope membranes from *Arabidopsis thaliana*. Mol Cell Proteomics 2: 325–345.
27. Mitschke J, Fuss J, Blum T, Höglund A, Reski R, et al. (2009) Prediction of dual protein targeting to plant organelles. New Phytol 183: 224–235.
28. Taylor NL, Tan YF, Jacoby RP, Millar AH (2009) Abiotic environmental stress induced changes in the Arabidopsis thaliana chloroplast, mitochondria and peroxisome proteomes. J Proteomics 72: 367–378.
29. Martin W, Hermann RG (1998) Gene transfer from organelles to the nucleus: how much, what happens and why? Plant Physiol 118: 9–17.
30. Cavalier-Smith T (2000) Membrane heredity and early chloroplast evolution. Trends Plant Sci 5: 174–182.
31. Lo' pez-Juez E (2007) Plastid biogenesis, between light and shadows. J Exp Bot 58: 11–26.
32. Martin W, Rujan T, Richly E, Hansen A, Cornelsen S, et al. (2002) Evolutionary analysis of Arabidopsis, cyanobacterial, and chloroplast genomes reveals plastid phylogeny and thousands of cyanobacterial genes in the nucleus. Proc Natl Acad Sci USA 99: 12246–12251.
33. Chothia C, Gough J, Vogel C, Teichmann S (2003) Evolution of the protein repertoire. Science 300: 1701–1703.
34. Andersson H, von Heijne G (1993) Position-specific Asp-Lys pairing can affect signal sequence function and membrane protein topology. J Biol Chem 268: 21389–21393.
35. Birney E, Kumar S, Krainer AR (1993) Analysis of the RNA-recognition motif and RS and RGG domains: conservation in metazoan pre-mRNA splicing factors. Nucleic Acids Res 21: 5803–5816.

36. Ding J, Hayashi MK, Zhang Y, Manche L, Krainer AR, et al. (1999) Crystal structure of the two-RRM domain of hnRNP A1 (UP1) complexed with single-stranded telomeric DNA. Genes Dev 1: 1102–1115.
37. Maris C, Dominguez C, Allain FH (2005) The RNA recognition motif, a plastic RNA-binding platform to regulate post-transcriptional gene expression. FEBS J 272: 2118–2131.
38. Kuhn JM, Schroeder JI (2003) Impacts of altered RNA metabolism on abscisic acid signaling. Curr Opin Plant Biol 6: 463–469.
39. Flicek P, Amode MR, Barrell D, Beal K, Brent S, et al. (2011) Ensembl 2012. Nucleic Acids Res 40: D84–90.
40. Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ (1990) Basic local alignment search tool. J Mol Biol 215: 403–410.
41. Thompson JD, Higgins DG, Gibson TJ (1994) CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position specific gap penalties and weight matrix choice. Nucleic Acids Res 22: 4673–4680.
42. Pei J, Kim BH, Grishin NV (2008) PROMALS3D: a tool for multiple protein sequence and structure alignments. Nucleic Acids Res 36: 2295–2300.
43. Tamura K, Dudley J, Nei M, Kumar S (2007) MEGA4: Molecular Evolutionary Genetics Analysis (MEGA) software version 4.0. Mol Biol Evol 24: 1596–1599.
44. Hulo N, Bairoch A, Bulliard V, Cerutti L, Cuche BA, et al. (2008) The 20 years of PROSITE. Nucleic Acids Res 36: D245–249.
45. Horton P, Park KJ, Obayashi T, Fujita N, Harada H, et al. (2007) WoLF PSORT: protein localization predictor. Nucleic Acids Res 35: W585–587.
46. Briesemeister S, RahnenführerJörg, Kohlbacher O (2010) YLoc - an interpretable web server for predicting subcellular localization. Nucleic Acids Res 38: W497–502.
47. Emanuelsson O, Brunak S, von Heijne G, Nielsen H (2007) Locating proteins in the cell using TargetP, SignalP, and related tools. Nature Protocols 2: 953–971.
48. McGuffin LJ, Bryson K, Jones DT (2000) The PSIPRED protein structure prediction server. Bioinformatics 16: 404–405.
49. Cserzo M, Wallin E, Simon I, von Heijne G, Elofsson A (1997) Prediction of transmembrane alpha-helices in procariotic membrane proteins: the Dense Alignment Surface method. Protein Eng 10: 673–676.
50. Tusnády GE, Simon I (2001) The HMMTOP transmembrane topology prediction server. Bioinformatics 17: 849–850.
51. Jones DT, Taylor WR, Thornton JM (1994) A model recognition approach to the prediction of all-helical membrane protein structure and topology. Biochemistry 33: 3038–3049.
52. Sonnhammer EL, von Heijne G, Krogh A (1998) A hidden Markov model for predicting transmembrane helices in protein sequences. Proc Int Conf Intell Syst Mol Biol 6: 175–182.
53. Kahsay RY, Gao G, Liao L (2005) An improved hidden Markov model for transmembrane protein detection and topology prediction and its applications to complete genomes. Bioinformatics 21: 1853–1858.
54. Hofmann K, Stoffel W (1993) TMbase — A database of membrane spanning proteins segments. Biol Chem Hoppe–Seyler 374: 166–170.
55. Claros MG, von Heijne G (1994) TopPred II: An improved software for membrane protein structure predictions. Comput Appl Biosci 10: 685–686.
56. Arai M, Mitsuke H, Ikeda M, Xia JX, Kikuchi T, et al. (2004) ConPred II: a consensus prediction method for obtaining transmembrane topology models with high reliability. Nucleic Acids Res 32: W390–393.
57. Käll L, Krogh A, Sonnhammer ELL (2004) A combined transmembrane topology and signal peptide prediction method. J Mol Biol 338: 1027–1036.
58. Finn RD, Mistry J, Tate J, Coggill P, Heger A, et al. (2010) The Pfam protein families database. Nucleic Acids Res 38: D211–222.
59. Letunic I, Doerks T, Bork P (2011) SMART 7: recent updates to the protein domain annotation resource. Nucleic Acids Res 40: D302–305.
60. Ginalski K, Elofsson A, Fischer D, Rychlewski L (2003) 3D-Jury: a simple approach to improve protein structure predictions. Bioinformatics 19: 1015–1018.
61. Marti-Renom MA, Stuart A, Fiser A, Sánchez R, Melo F, et al. (2000) Comparative protein structure modeling of genes and genomes. Annu Rev Biophys Biomol Struct 29: 291–325.
62. Guex N, Peitsch MC (1997) SWISS-MODEL and the Swiss-PdbViewer: an environment for comparative protein modeling. Electrophoresis 18: 2714–2723.
63. Andreeva A, Howorth D, Chandonia JM, Brenner SE, Hubbard TJP, et al. (2008) Data growth and its impact on the SCOP database: new developments. Nucleic Acids Res 36: D419–425.
64. Gouet P, Robert X, Courcelle E (2003) ESPript/ENDscript: extracting and rendering sequence and 3D information from atomic structures of proteins. Nucleic Acids Res 31: 3320–3323.