In silico identification and characterization of gene/s involved in flowering and cold stress response regulation in *Cajanus cajans*

A Thesis submitted to the Orissa University of Agriculture & Technology,
in Partial fulfillment of the Requirement for the award of degree of Master of Science in Bioinformatics

By

ANIMESH PATTNAIK
ADM NO: 10BI/13

DEPARTMENT OF BIOINFORMATICS
CENTRE FOR POST GRADUATE STUDIES
ORISSA UNIVERSITY OF AGRICULTURE AND TECHNOLOGY
BHUBANESWAR-751003
2015

NATIONAL BUREAU OF PALNT GENETIC RESOURCES
DEPARTMENT OF GENOMICS RESOURCES
NEW DELHI

Dr. Dhammaprakash Pandhari Wankhede **Place:-New Delhi**
Scientist, Division of Genomics Resources **Date-**

CERTIFICATE I

This is to certify that the thesis entitled **"In silico identification and characterization of gene/s involved in flowering and cold stress response regulation in *Cajanus cajans*"** submitted in partial fulfilment of the requirements for the award of the degree of **Master of Science** in the subject **Bioinformatics** to the Orissa University of Agriculture and Technology is a faithful of bonafide and original research work carried out by **Animesh Pattnaik (Adm No. 10BI/13)** under my guidance and supervision. No part of this thesis has been submitted for any other degree or diploma.

It is further certified that the assistance and help received by him from various sources during the course of investigation has been duly acknowledged.

CHAIRMAN
ADVISORY COMMITEE

CERTIFICATE II

This is to certify that the thesis entitled **"In silico identification and characterization of gene/s involved in flowering and cold stress response regulation in *Cajanus cajans*"** submitted by **Animesh Pattnaik** to Orissa University of Agriculture and Technology, Bhubaneswar in partial fulfilment of the requirements for the degree of **Master of Science** in **Bioinformatics** has been approved/disapproved by the students advisory committee and the external examiner.

Advisory Committee

1. Dr. Dhammaprakash Pandhari Wankhede
Scientist, Division of Genomics Resources Chairman ____________________

2. Mrs. Sucharita Balabantaray
Asst. Prof.
Department of Bioinformatics Member ____________________

3. Mr. Mr. Sukanta Kumar Pradhan
Head of the Department
Department of Bioinformatics Member ____________________

External Examiner ____________________

ACKNOWLEDGEMENT

Irrespective of the magnitude of the task, it cannot be completed effectively without the help and support of other people. It also pleasant to acknowledge the help one receives from different individuals. I wish to thanks them for their time and expertise.

Firstly, it is my priviledge to share my deep sense of gratitude to Dr. Dhammaprakash Pandhari Wankhede for allowing me with this opportunity to work as a trainee in his laboratory and guiding me throughout my training period.

I would also like to thanks Dr. K C Bansal, Director, NBPGR, Prof. PGR and Dr. K. V. Bhatt for providing me the opportunity to work in the one of the best laboratories of DGR, NBPGR for my dissertation work. I would heartily thank to Dr. Soma S Marla for encouraging me during my project work. I also thank Prof. Bisht, Professor, PGR, NBPGR and Mr Joshi sir for their kind support in fulfilling the formalities in joining NBPGR for the project work.

I am thankful to the member of the advisory committee Mr Sukanta kumar Pradhan, HOD, Dept of Bioinformatics and Mrs. Sucharita Balabantaray, Asst. Prof., Department of Bioinformatics ,OUAT for their support, encouragement and allowing me to carry out this dissertation work successfully.

I convey my cordial thanks to all my faculty members Mrs. Sushma Rani Martha, Mr. Surya Narayan Rath and Sujit ku. Das for their valuable guidance.

I also immensely thank to late Gopal Dada, Budheswar Bhai, Ajaya Bhai, Sambhu Bhai, Jitendra Bhai, Ananda Dada, Showkat Ahmad Bhai for their help, ideas, encouragement and support for my project work as well as studies.

I would also like to extend my heartfelt thanks to Mr Satyaprakash, Dr Meenu Kumai, Dr Neelofar Mirza, Dr Minakshi and Mr Bharat for their help and support during entire period of project work at NBPGR.

Diction is not enough to express my deep sense of gratitude and regards for my beloved Bapa, Maa and my family for all the love, care and guidance throughout my life. My younger brother has always been a real inspiration for me in every step of my life and giving me unbound emotional support. Thanks Bhai for everything you have done for me. I feel great pleasure to express my love to all my sweetest friends Nihar, Prajana, Manisha, Debashrita, Subharaj, Sishirakanta, Sanjaya and Sameer for their strong mental support and humbly gave their valuable time for providing the relevant information.

I feel honoured to be a part of this auspicious university for providing me a healthy atmosphere in these two years. Last but not the least I express my gratitude to Thakur ji and Maa Samlei for invaluable inspiration for accomplishment blessings of such a splendid work.

ANIMESH

CONTENTS

LIST OF FIGURES

LIST OF TABLES

ABBREVIATIONS

COR – Cold Regulator

CBF – C Binding Factor

SNP – Single Nucleotide Polymorphism

ICE1 – Inducer of CBF Expression

HOS1 – High Expression of Osmotically Response gene 1

RNA – Ribonucleic Acid

DNA – Deoxy ribonucleic Acid

TAIR – Arabidopsis Information Resource

BLAST – Basic Local Alignment Search Tool

B2G – Blast2Go

EST – Expressed Sequence Tags

HMM – Hidden Markov Model

SRA – Sequence Read Archive

NGS – Next Generation Sequence

NCBI – National Center for Biotechnology Information

GOs – Gene Ontology

PI – Isoelectric Point

MW – Molecular weight

RNAseq – RNA sequence Analysis

ABSTRACT

Flowering timing is a one of the most important and very useful genetically studies for development for the crop improvement. Flowering time is a major life-history trait that contributes to fitness in annual plants. In this study, six important flowering related genes have been identified using bioinformatics approaches from Pigeonpea (*Cajanus cajan*). To achieve this, flowering related genes in Arabidopsis and other plants have been selected (FRIGIDA, FRIGIDA LIKE 1, FRIGIDA LIKE 2,FRIGIDA ESSENTIAL 1,TERMINAL FLOWERING 1, TERMINAL FLOWERING 2). Protein sequences of these genes were used as query against total proteins sequences of Pigeonpea which was used as database. Top hits from the standalone blast hits were manual curated and searched for known motifs. Identified genes were studied for their gene structure, post translation modifications, expression pattern, conservation of putative interacting partners and phylogenetic relationship with other plants. The study showed computational identification and characterization of six important flowering related genes from Pigeonpea.

Additionally, using similar approach, ICE1 (Inducer of CBF Expression 1), an important regulator of cold stress have also been identified from Pigeonpea, sorghum, foxtail millet and maize. Further, bioinformatics approaches have been used to understand ICE1 in temperate, tropical and subtropical crop plants. ICE1 sequences were obtained for 18 crop plants including temperate crops, tropical and subtropical plants (*Brassica rapa, Raphanus sativus, Arabidopsis thaliana, Oryza sativa, Glycine max, Solanum lycopersicum, Malus domestica, Solanum tuberosum, Citrus sinensis, Cicer arietinum, Cucumis sativus, Cucumis melo, Cajanus Cajan, Setaria italic, Sorghum bicolor, Triticum aestivum, Zea mays* and *Gossypium arboretum*). Multiple sequence alignments of ICE1 from these diverse plant species gives insights in sequence conservation and divergence at important sites. Further, ICE1 were studied for their gene structure and motif conservation, putative phosphorylation and sumoylation sites. This part of the work gives a brief insight into ICE1from crop plants across different climatic conditions and its possible role in differential cold stress responses.

CHAPTER I

INTRODUCTION

CHAPTER I

INTRODUCTION

Pigeon pea (*Cajanus cajan*, Family-Fabaceae) is one of the most important pulse crop all over world. Pigeon pea is domestication in South Africa in 3500 years ago now a day's cultivated all over world mostly Asia, Africa and Latin America. It is cultivated on more than five million hectares all over the world and it is rich source of protein and fiber. Pigeon pea is a most important crop for the people of Asia, Africa and America. In India it cultivated mostly in eastern and southern part. Pigeon pea is closely related to the *Cajanus cajanifolia* which is the oldest species found in the tropical region of India.

Flowering timing is a major life cycle for the development of a plant. It is most important thing for that plant, which domesticated highly like an annual crop. Depending on biotic (competition, pollinators, herbivores) and abiotic (photoperiod, temperature, nutrients) conditions, different flowering time is adopted by different species. The growing conditions of late flowering leads a longer growth period that promotes accumulation and allocation of more resources to seed production, but early flowering is resources to seed production with a short period and unexpected growth time. The pattern and time of flowering are important adaptive traits in flowering plants controlled by physiological signals, genes, gene interactions and interactions of genes with the environment. The progress in early flowering is the area of isolation and characterization of plant genes for crop improvement due to emergence of plant genomics. Availability of genome sequence of a number of plant species together with comparative genomics have helped in answering some of the fundamental aspects of plant biology including identification and analysis of genes involved in adaptive traits in crop species. One of the best examples of such evolutionary developmental studies in plant species is the identification and analysis of MADS box genes involved in flower development. Subsequently, orthologous genes have been isolated in many species providing insights into the conservation and diversification of such genes and their functions in plant development (Reyazul Rouf Mir et al. 2014).

The Floral repressive genes are the main target of selection of the early flowering of plans. As well as promoter gene are also take a main character in the leads of early flowering**.** Floral repressive genes are like FLC and its Activators, other repressive gene including to FLC. These genes are involved in gibberellin signaling pathways. Variation in the domain of this gene causes early flowering. Several loss-of-function mutation in FRIGIDA confer early flowering. Genetic approaches have found that FRIGIDA (FRI) gene as the main promoter of the early flowering. Most early-flowering contain carry a FRI allele with loss-of-function mutation. That the variation in flowering time that appears to be independent of FRI in accessions both with and without functional FRI alleles .It characterized several accessions with interesting flowering-time phenotypes, given their FRI status to determine the genetic and molecular origins of their flowering-time (Werner et al. 2005).

At present there is little information available on flowering genes from Pigeonpea and their role in flowering trait. For any such studies, it is essential to have information on flowering genes.

Cold stress is one of the most crucial stress conditions which have deleterious impact on crop growth and development and thus resulting in significant economic loss due to reduced crop productivity. In response to cold stress, plants show expression of cold-regulated (COR) genes which help to cope up cold stress. The COR genes harbour C-repeat at their promoter element and their expression is activated by C-repeat (CRT)-binding transcription factors (CBF). In Arabidopsis ICE1 (Inducer of CBF Expression 1) a MYC-type basic helix–loop–helix transcription factor is considered to be a central regulator of cold response as it is shown to regulate expression of CBF3 and downstream COR genes (Chinnusamy et al. 2003). However, in Pigeonpea there is no information on ICE1 mediated cold stress and this genes have not been cloned so far. Further it is not known so far, whether ICE1 regulation mechanism is distinct on temperate, tropical and sub-tropical crop plants.

Considering importance of abovementioned subjects objective of the present thesis were set as mentioned below.

OBJECTIVES

- In-silico identification and characterization of genes involved in flowering trait in Pigeon pea.
- In-silico identification and characterization of ICE1 from Pigeon pea.
- Comparison of ICE1 from temperate, tropical and subtropical crop plants.

CHAPTER II
REVIEW OF LITERATURE

CHAPTER II

REVIEW OF LITERATURE

Regulation in flowering traits

Pigeonpea [*Cajanus Cajan* (L.) Millsp.] is cultivated in tropical and sub-tropical areas between 300N and 300S latitude. It is an important grain legume of Asia (especially, the Indian subcontinent), Latin America and Africa. Globally, it is grown on ~5 million hectares (m ha) in about 82 countries of the world which are located in South America, Caribbean islands, southern and eastern Africa. In Asia, besides India, Nepal and Myanmar grow considerable acreages of pigeon pea. Other major contributors are Malawi, Uganda and Tanzania. Pigeon pea has a unique place in Indian farming and India accounts for about 90% of the global production. It is the second most important pulse crop next to chickpea, covering an area of around 4.42 m ha (occupying about 14.5% of area under pulses) and production of 2.86 mt (contributing to 16% of total pulse production) and productivity of about 707 kg/ha . The major states in terms of area and production are Maharashtra, Uttar Pradesh (UP), Madhya Pradesh (MP), Karnataka, Gujarat, Andhra Pradesh (AP) and Bihar.

Pigeon pea has several unique features, by which it occupies a predominant place as a food and in several cropping systems all over India. This crop is used in more diverse ways than any other pulse crop. Besides being a major source of protein, it has several other economic uses in the form of feed, fuel, forage, green manure etc. Pigeon pea has the potential to produce a very high biomass and fix atmospheric nitrogen up to 40 kg/ha in a cropping season (Rao *et al.*, 1983). The deep root system allows optimum moisture and nutrient utilization. It also has a mechanism to extract phosphorus from the soil. Because of its perennial, hardy nature and profuse foliage, it is also used as a hedge and for wind-break.

The productivity of pigeon pea has not increased over the years despite its importance in diverse cropping systems, low input requirement, drought tolerance, and availability of a large spectrum of variability. The large gap between the potential yield (2,500 kg/ha) and yields obtained on farmers' fields (866.2 kg/ha in Asia, 736.2 kg/ha in Africa) (Mula and Saxena, 2010) has been attributed to several agronomic and genetic constraints like poor crop management, cultivation in marginal land (with ~4.4% coverage under irrigation), photoperiod and low temperature sensitivity of most of the available cultivars, lack of physiologically efficient genotypes and of tolerance to certain biotic and abiotic stresses, difficulties in quality seed production due to frequent out-crossing in certain conditions, the narrow genetic base of cultivars so far released and inadequate crop protection technologies, particularly against pests (Singh *et al*., 2006).

It is a multipurpose plant as it is extensively eaten as a dal. It is rich in proteins. In India its leaves are used for rearing silkworms; green pods are used as a vegetable; husk, green leaves and tops are used as fodder and also as green manure. Amongst its many medicinal uses, *C. Cajan* is indicated in the relief of pain in traditional Chinese medicine and as a sedative (Ambasta SP. 2004). In recent years it has also been explored for the treatment of ischemic necrosis of the caput femoris, aphtha, bedsore and wound healing (Ahsan R et al 2009). Chemical investigations have revealed the presence of two globulins, cajanin and concajanin. It has been used widely for many years for treating diabetes, sores, skin irritations, hepatitis, measles, jaundice, dysentery and many other illnesses; for expelling bladder stones and stabilizing menstrual period (Dilipkumar Pal et al 2011).

Flowering time: a keys stone in plants adaptation

Flowering timing is a major life cycle for the development of a plant. It is most important thing for that plant, which domesticated highly like an annual crop. Which have to synchronize their reproduction with favourable environmental conditions. Depending on biotic (competition, pollinators, herbivores) and a biotic (photoperiod, temperature, nutrients) conditions, different flowering time is adopted by different species. The growing conditions of late flowering leads a longer growth period that promotes accumulation and allocation of more resources to seed production, but early flowering is resources to seed production with a short period and unexpected growth time. The pattern and time of flowering are important adaptive traits in flowering plants controlled by physiological signals, genes, gene interactions and interactions of genes with the environment. The trade-off between resource accumulation and stress avoidance is also of primary importance for crop yield and quality, and the identification of molecular variation associated with flowering time is a key step when selecting varieties adapted to different latitudes and cropping seasons. Studying on the genetic regulation of flowering time in the model plant Arabidopsis thaliana has increased recently. Genes that regulate flowering time have been identified mainly by analysing Arabidopsis laboratory mutants with an altered flowering phenotype, many of which are early flowering. The laboratory-generated variation, adaptive and natural variation results from the long process of natural selection, screening mutations according to their global phenotypic effect. The genetic determinism of naturally occurring early flowering in Arabidopsis using the following approach: (i) examination of the genetic network controlling flowering time, which should enable likely targets of selection for early flowering to be identified; (ii) the study of their associated negative pleiotropic effects should reduce the number of these potential targets; (iii) the number and distribution of the genetic effects that are eventually fixed under selection can be predicted using recent genetic theories of adaptation. Deciphering the genetic determinism of quantitative variation is of interest not only to evolutionary biologists studying the genetics of adaptation in wild species, but also to crop breeders because it could provide useful guidelines for quantitative trait loci (QTL) studies and identification of target genes for selection as well as Marker Assisted Selection and crop species development. (Roux et al TIPS 2006) The progress in early flowering is the area of isolation and characterization of plant genes for crop improvement due to emergence of plant genomics.

Genetic dissection of flowering timing in *Cajanus cajan*

Determinacy is an ergonomically important trait selected during pigeon pea domestication. In the present study, seven genes related to determinacy/flowering pattern in pigeon pea were isolated through a comparative genomics approach. Single nucleotide polymorphism (SNP) analysis of these candidate genes on 142 pigeon pea lines found a strong association of SNPs with the determinacy trait for three of the genes. Subsequently, QTL analysis highlighted one gene, *CcTFL1*, as a likely candidate for determinacy in pigeon pea since it explained 45–96 % of phenotypic variation for determinacy, 45 % for flowering time and 77 % for plant height. Comparative genomics analysis of *CcTFL1* with the soybean (*Glycine max*) and common bean (*Phaseolus vulgaris*) genomes at the micro-syntenic level further enhanced our confidence in *CcTFL1* as a likely candidate gene. These findings have been validated by expression analysis that showed down regulation of *CcTFL1* in a determinate line in comparison to an indeterminate line. Gene-based markers developed in the present study will allow faster manipulation of the determinacy trait in future breeding programs of pigeon pea and will also help in the development of markers for these traits in other related legume species (Reyazul Rouf Mir et. al. 2014**).**

Elements of FRIGIDA in early flowering

Floral repressive genes are like FLC and its Activators, other repressive gene including to FLC.This gene are contains gibberellins signalling path ways. Variation in the domain of this gene causes early flowering. This point is supported by the recent studies on Arabidopsis (J. D. Werner et al.).Several loss-of-function mutation in FRIGIDA confer early flowering. Genetic approaches have found that FRIGIDA (FRI) gene as the main promoter of the early flowering. Most early-flowering contain carry a FRI allele with loss-of-function mutation. That the variation in flowering time that appears to be independent of FRI in accessions both with and without functional FRI alleles. It characterized several accessions with interesting flowering-time phenotypes, given their FRI status to determine the genetic and molecular origins of their flowering-time (Werner et al. 2005).

The molecular genetics of flowering time also fail to provide a clear explanation in flowering time have not been observed. The flowering time gene FRIGIDA (FRI) is considered to be the major determinant of flowering time variation in natural accessions of Arabidopsis. Functional FRI alleles lead to an accumulation of Flowering Locus C (FLC) mRNA, which in turn inhibits flowering. Unless functional FRI prevents the plants from flowering until winter has passed. The loss of functional *FRI* alleles is promoting the early function of flowering in the absence of vernalization.THE loss-of-function mutations in *FRI* are associated with earlier flowering time under controlled conditions in 13 of 18 ecotypes screened to date (Simpson, G. G. & Dean, C. 2002).An excess of no synonymous polymorphisms have been described in the first exon of *FRI*, which has been interpreted as evidence of natural selection for early flowering time in this plant(Le Corre et al 2002) .

The regulatory mechanisms by which FLC expression is controlled are of great interest given its central role in controlling flowering time. The Recent screening for mutants that in winter-annual, FRI-containing lines early flowering have led to the identification of genes that are additional positive regulators of FLC expression. Some of these are required for the elevated FLC expression in both autonomous-pathway mutants and FRI-containing lines (e.g. PHOTOPERIOD-INDEPENDENTEARLY FLOWERING 1 (PIE1), VERNALIZATION INDEPENDENCE 3 (VIP3) and VIP4}, whereas others are specifically required only for Remediated FLC activation(Zhang, et al. 2003), Noh, Y.S. and Amasino, R.M. (2003), Michaels, S.D. et al. (2004), Yuehui He et al).

Given these results, one would expect a latitudinal divergence in *FRI* functionality, with early-flowering ecotypes containing non functional *FRI* restricted to either Southern latitudes or climates with mild winters as a result of natural selection. However, to date there is no evidence of a relationship between winter temperatures at the site of origin of accessions and flowering time as measured under controlled conditions of a predominance of ecotypes with non-functional *FRI* alleles in Southern.(John R. Stinchcombe et al 2004)

The report on latitudinal cline in days in *Arabidopsis* ecotypes allowed to over winter under natural conditions. These results demonstrate a latitudinal cline in a life history trait closely related to flowering time in *A. thaliana*. However, this cline is detected only in ecotypes with *FRI* alleles lacking deletions that would disrupt protein function. Surprisingly, in ecotypes from Southern latitudes, such putatively functional *FRI* alleles are associated with accelerated flowering relative to ecotypes with non-functional *FRI* under the winter conditions of our experiment. These results suggest that the ecological function of the vernalization requirement conferred by *FRI* may differ across latitudes.

Studies of natural variation in the winter annual habit of many accessions of Arabidopsis is conferred by two genes, FRIGIDA (FRI) and FLOWERING LOCUS C (FLC), whose activities impose a vernalization requirement. To better understand the mechanism underlying the winter-annual habit, a genetic screen was performed to identify mutants that suppress the late- flowering behaviour of a non-vernalized winter-annual strain. A locus, FRIGIDA-ESSENTIAL 1 (FES1), which like FRI is specifically required for the up regulation of FLC expression. FES1 is predicted to encode a protein with a CCCH zinc finger, but the predicted sequence does not otherwise share significant similarity with other known proteins. Fes1 is a complete suppressor of FRI-Gene flowering, but has little effect on the late-flowering phenotype of autonomous pathway mutants. Thus, FES1 activity is required for the FRI-mediated winter-annual habit, but not for the similar phenotype, also resulting from autonomous-pathway mutations. Analysis between FES1, FRI and another specific suppressor of FRI-containing lines, FRIGIDA-LIKE 1 (FRL1), indicates that these genes do not function in a linear pathway, but instead act cooperatively to promote the expression of FLC. (Robert J. Schmitz et al 2005)

Elements of terminal flowering likes gene in early flowering

Genes in the *TERMINAL FLOWER1*(*TFL1*) family are important key regulatory genes involved in the control of flowering time and floral architecture in several different plant species**.** For some fundamental aspects of plant biology, the genes involved have been identified through a molecular genetics approach using the model species Arabidopsis. From this basic information, comparative studies between species can begin, in particular to understand the genetic and molecular mechanisms responsible for the large diversity in plant morphology and to identify the genes involved in adaptive evolution (Cronk, 2001). In plants, the best example for find the evolutionary developmental studies is the identification and analysis of MADS box genes involved in flower development in several plant species, including gymnosperms (reviewed by Ma and De Pamphilis, 2000; Fabrice Foucher et al). Isolation of putative orthologs in different species and studies of RNA and/or protein expression patterns provide insights into the conservation and diversification of gene function in plant development (Hofer and Ellis, 2002). Different *TFL1* homolog's control two distinct aspects of plant development in pea, whereas a single gene, *TFL1*, performs both functions in Arabidopsis (Fabrice Foucher, 2003).

The used a molecular approach to study homolog's of the snapdragon *CENTRORADIALIS*/Arabidopsis *TERMINAL FLOWER1* (*CEN/TFL1*) genes in pea. Orthologs of *CEN* have been found in different species: *TFL1*in Arabidopsis (Bradley et al., 1997), *SELF PRUNING* (*SP*) in tomato (Pnueli et al., 1998), *CET* in tobacco (Amaya et al., 1999), and *LpTFL1* in *Lolium perenne* (Jensen et al., 2001). These studies showing that flowering time/flowering Pattern/determinacy has been selected long ago by breeders in combination with photoperiod insensitivity to obtain varieties with shorter flowering period, earlier maturation and ease of mechanized harvest

(Repinski et al. 2012). Genetic mechanism responsible for these traits has been uncovered in model plant *Arabidopsis* (*Arabidopsis thaliana*), pea (*Pisumsativum*), soybean (*Glycine max*), common bean (*Phaseolusvulgaris*) etc. (Foucher et al. 2003; Hecht et al. 2005; Kwak et al. 2008; Liu et al. 2010; Repinski et al. 2012). In some cases it was proved that determinacy is controlled by a single gene, whereas in other studies more than one gene was found responsible for the transition of different growth habits (Tina et al. 2010). In pea, it was shown that the determinate mutant (*det*) is caused by mutations in a homologue of the *Arabidopsis TFL1* gene (Foucher et al. 2003). In soybean, the gene responsible for determinacy "*GmTfl1*" was isolated and found to complement the functions of *TFL1* in *Arabidopsis* (Liu et al. 2010; Tian et al. 2010). Similarly, in common bean, it was proved that gene "*PvTFL1y*" co-segregated with the determinacy locus "*fin*" (Kwak et al. 2008) and later the same was validated and found as a functional homolog of *ArabidopsisTFL1* gene (Repinski et al. 2012). In pigeon pea, both indeterminate (IDT) and determinate (DT) type flowering pattern exist (Mir et al. 2012). (Reyazul Rouf Mir, 2014)

This early-flowering phenotype was not observed in snapdragon. TFL1 may play a role in inflorescence meristem identity as well as in floral initiation control as a repressor of flowering. It was proposed that these two distinct roles are in fact one, with*TFL1*controllingthe length of both the vegetative and reproductive phases (Ratcliffe et al., 1998; Fabrice Foucher et al 2003). CEN and TFL1 are similar to a family of mammalian phosphate dylethanol amine binding proteins (PEBPs) also known asRaf-1 kinas inhibitor proteins. Crystallography analysis shows that CEN may be involved in interaction with a kinase (Banfieldand Brady, 2000). In tomato, SP was shown to interact with multiple proteins and was proposed to encode a modular protein with the potential to interact with a variety of signalling pathways (Pnueli et al., 2001). Expression analysis has shows that genes closely related to*TFL1*are expressed mainly in the shoot apical meristem in the region below the terminal meristem.*CEN*is induced during floral initiation (Bradley et al., 1996), whereas*TFL1*expression also is found during the vegetative phase of flowering in Arabidopsis (Bradley et al., 1997). Analysis of mutants and sequencing of the entire Arabidopsis genome have revealed that the*TFL1*genes belong to a small family (atleast six genes) with functional divergence (Mimida et al., 2001).One of them, *FT*, has a*TFL1*-antagonistic role by promoting flowering in Arabidopsis (Kardailsky et al., 1999; Kobayashi et al., 1999). Studies of*TFL1*homologs in other species may help us better understand the function and the evolution of the*TFL1* family in flowering plants.

COLD STRESS GENE RESPONSE IN PLANT

Plants are constantly exposed to a variety of environmental stresses. Freezing or extremely low temperature constitutes is adversely affecting factor for plant growth, development and crop productivity. Most temperate plants acquire freezing tolerance by a process called cold acclimation. Here in the cold stress we study in the recent progress in transcriptional, post-transcriptional and post-translational regulation of gene expression that is critical for cold acclimation. Cold stress, which includes chilling (<20 8C) and/or freezing (<0 8C) temperatures, adversely affects the growth and development of plants and significantly constraints in plants and agricultural productivity. Cold stress prevents the expression of full genetic potential of plants. Its direct inhibition of metabolic reactions and indirectly, through cold-induced osmotic (chilling-induced inhibition of water uptake and freezing-induced cellular dehydration), oxidative and other stresses. Cold regulation is a process by which plants acquire freezing tolerance upon prior exposure to low non-freezing temperatures.

Most temperate plants can cold-acclimate and acquire tolerance to extracellular ice formation in their vegetative tissues. Winter-habit plants (winter wheat, barley, oat, rye, oilseed rape, etc) have a vernalization requirement, which prevents premature transition to the reproductive phase before the threat of freezing stress during winter has passed. Many important crops, such as rice, maize, soybean, cotton and tomato, are chilling sensitive and incapable of cold acclimation; moreover, they cannot tolerate ice formation in their tissues. (Thomashow, M.F. 1999; Chinnusamy et al. 2003)

Studies on acquired freezing tolerance in Arabidopsis have contributed substantially towards the understanding of cold acclimation mechanisms. Cold acclimation involves the remodelling of cell and tissue structures and the reprogramming of metabolism and gene expression.

Cold stress signalling

Cellular membranes are fluid structures, and cold temperatures can reduce their fluidity, causing increased rigidity. Plant cells can sense cold stress through low temperature-induced changes in membrane fluidity, protein and nucleic acid conformation and metabolite concentration. Using a pharmacological approach, plasma membrane rigidification has been shown previously to induce COR (COLD RESPONSIVE) genes and result in cold acclimation in alfalfa and Brassica napus (Orvar, B.L. et al. 2000; Chinnusamy et al. 2003).

Secondary signals, such as abscisic acid (ABA) and reactive oxygen species (ROS) can also induce Ca2+ signatures that impact cold signaling. Arabidopsis mutants defective in the activation of the molybdenum cofactor of abscisic aldehyde oxidase, namely aba3/freezing sensitive 1 (frs1), also known as los5 (low expression of osmotically responsive genes 5), exhibit hypersensitivity to freezing stress. (Xiong, L. et al. 2001; Chinnusamy et al. 2003)

The Arabidopsis fro1 (frostbite1) mutant, which constitutively accumulates high levels of ROS, exhibits impaired expression of COR genes and hypersensitivity to chilling and freezing. FRO1 encodes the Fe–S subunit of complex I (NADH dehydrogenase) of the respiratory electron transfer chain in mitochondria, and its disruption leads to high levels of ROS generation. (Lee, B-H. et al. 2002; Chinnusamy et al. 2003)

Low temperature affects water and nutrient uptake, membrane fluidity and protein and nucleic acid conformation, and it drastically influences cellular metabolism either directly by reducing the rates of biochemical reactions or indirectly through gene expression reprogramming. Metabolic profiling revealed that cold acclimation increases _75% of the 434 metabolites detected in Arabidopsis plants (Lee, B-H. et al. 2002; Chinnusamy et al. 2003), although metabolite profiles do not appear to correlate with cold acclimation capacity of Arabidopsis. (Hannah, M.A. et al. 2006)

Transcriptional regulation

Cold acclimation temperatures induce profound changes in the plant transcriptome. In Arabidopsis, cold-regulated genes have been estimated to constitute _4% to 20% of the genome.

ICE1–CBF transcriptional cascade

Cold stress induces the expression of APETALA2/ETHYLENE RESPONSE FACTOR family transcription factors, that is, CBFs (C-repeat binding factors, also known as dehydration-responsive element-binding protein 1s or DREB1s), which can bind to cis-elements in the promoters of COR genes and activate their expression. CBFs regulate the expression of genes involved in phosphoinositide metabolism, transcription, osmolyte biosynthesis, ROS detoxification, membrane transport, hormone metabolism and signaling and many others with known cellular protective functions (S. and Thomashow, M.F. 2002; K. et al. 2004; Yamaguchi-Shinozaki, K. and Shinozaki, K. 2006; Chinnusamy et al. 2003). CBF homologs have been cloned from both cold-tolerant (wheat, barley and Brassica napus) and cold-sensitive (rice, maize, tomato and cherry) crops. Transgenic expression of Arabidopsis CBFs in different plant species was able to enhance chilling/freezing tolerance, and, conversely, the ectopic expression of CBFs from other plant species could enhance the freezing tolerance of transgenic Arabidopsis. Microarray analysis of transgenic Arabidopsis plants shows the CBFs expression of downstream cold-responsive transcription factor genes RAP2.1 and RAP2.7, which might control subregulons of the CBF regulon (S. and Thomashow, M.F. 2002). CBF regulons from freezing -tolerant and -sensitive plant species can differ, as evident from microarray analysis of transgenic tomato and Arabidopsis plants over expressing LeCBF1 and AtCBF3, respectively (Zhang, X. et al. 2004). However, the molecular basis of constitutive freezing tolerance is poorly understood. Transcriptome and metabolome analyses in Arabidopsis accessions differing in constitutive freezing tolerance suggest that the CBF pathway might also have an active role in constitutive freezing tolerance (Hannah, M.A. et al. 2006).

In Arabidopsis, ICE1 (INDUCER OF CBF EXPRESSION1), a MYC-type basic helix–loop–helix transcription factor, can bind to MYC recognition elements in the CBF3 promoter and is important for the expression of CBF3 during cold acclimation.

Constitutive over expression of ICE1 enhanced the expression of CBF3, CBF2 and COR genes during cold acclimation, and increased freezing tolerance of the transgenic Arabidopsis. ICE1 is constitutively expressed and localized in the nucleus, but it induces expression of CBFs only under cold stress. This suggests that cold stress-induced posttranslational modification is necessary for ICE1 to activate downstream genes in plants (Chinnusamy, V. et al. 2003). Transcriptome analysis that expression of _40% of cold-regulated genes, and in particular 46% of cold-regulated transcription factor genes is impaired in the dominant ice1 mutant. The cold induction of genes involved in calcium signalling, lipid signalling or encoding receptor-like protein kinases are also affected by the ice1 mutation (Lee, B-H. et al. 2005). Bioinformatics analysis of microarray data on the cold-responsive transcriptome of wild type and mutants or of transgenic Arabidopsis plants over expressing specific transcription factors led to the prediction of a cold-acclimation transcriptional network. In this network, ICE1 is predicted to be a transcriptional inducer of CBFs (CBF1–CBF3), ZAT12, NAC072 and the constitutively expressed transcription factor HOS9 in Arabidopsis (Benedict, C. et al. 2006).

It is likely that ICE1 and related proteins also play a critical role in the regulation of the expression of genes important in the chilling tolerance of Arabidopsis. The ice1 mutation renders Arabidopsis plants chilling sensitive and it affects the basal transcript levels of 204 of the 939 cold-regulated genes under non-stress conditions (Lee, B-H. et al. 2005). Basal expression of these genes could be important for chilling tolerance of Arabidopsis, as altered expression of these genes in ice1 is correlated with chilling sensitivity (Chinnusamy et al. 2003).

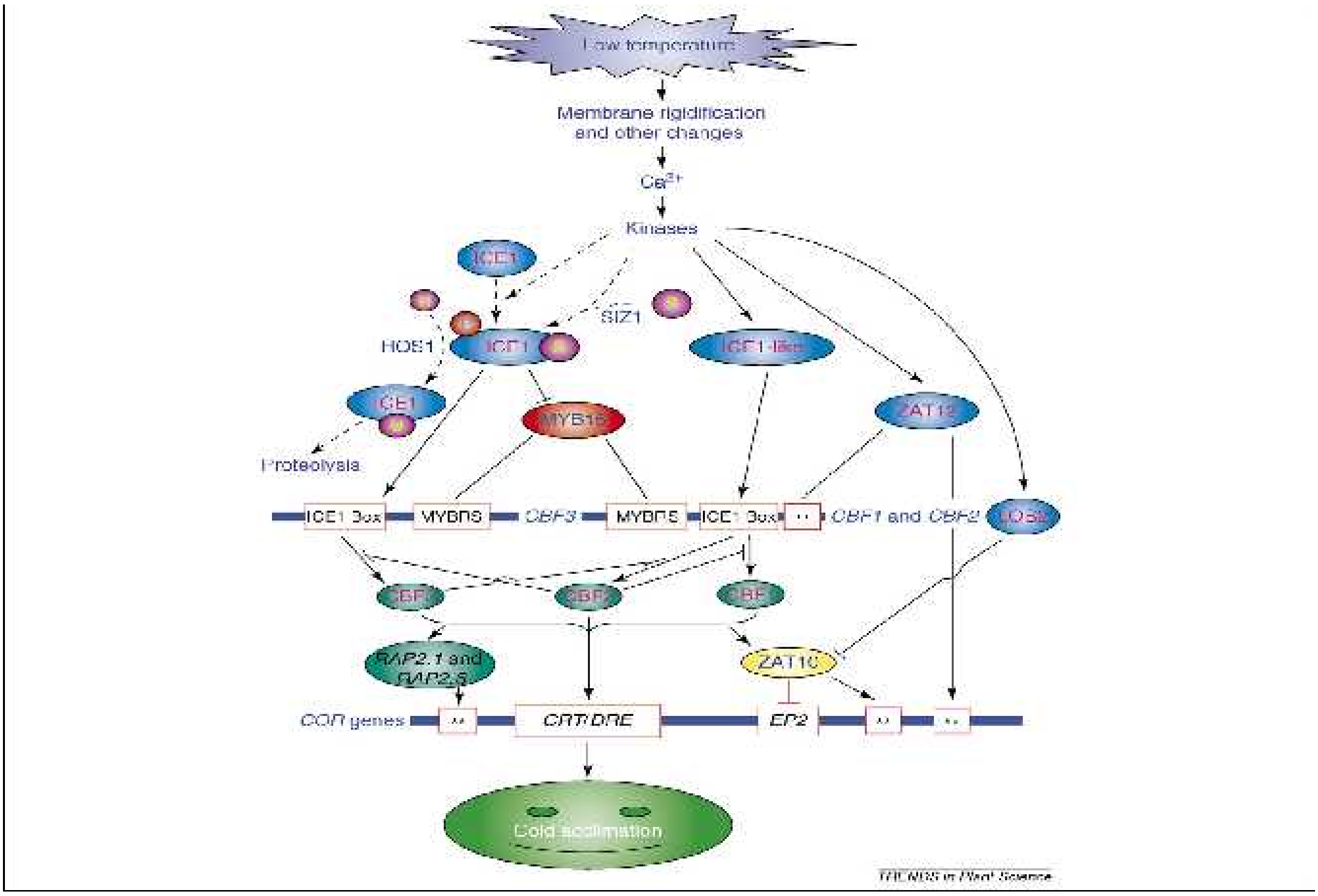


Figure 2.1 Regulation of low temperature response in plants (source: Trends in Plant Sciences)

Post-transcriptional regulation

In addition to transcriptional regulation, gene expression is regulated post-transcriptionally at pre-mRNA processing, mRNA stability, and export from nucleus and translation steps. Recent studies revealed that post-transcriptional regulation plays critical roles during cold acclimation.

RNA processing and export from the nucleus Pre-mRNA is a crucial nuclear process for the of functional mRNAs of intron-containing genes, and this process is coupled with nuclear export of mRNAs. Two early COR genes (a ribokinase and a C3H2C3 RING-finger protein) were shown to be regulated by intron retention in their mature mRNAs under cold stress (Mastrangelo, A.M. et al. 2005). By using the PRD29A::LUC genetic screen, Lee et al. (2006) identified STABILIZED1 (STA1), a nuclear pre-mRNA splicing factor, as a regulator of pre-mRNA splicing that is of particular importance to cold tolerance in Arabidopsis. Pre-mRNA and is hypersensitive to chilling, ABA and salt stresses. These results showed that a fully functional STA1 is required for splicing and turnover of specific transcripts, and, under cold stress, there is an increased demand for this factor (Lee et al. 2006). In Arabidopsis, cold and heat stresses regulate the alternatively by pre-mRNAs of many SR genes, which might produce different isoforms of SR proteins with altered splicing functions under stress conditions.

Small RNAs have potentially a large share in plant stress responses

Small non-coding RNAs of _21 to 24 nucleotides in length, namely microRNAs (miRNAs) and short interfering RNAs (siRNAs), are ubiquitous repressors of gene expression in animals and plants. miRNAs and siRNAs regulate gene expression by directing the cleavage or translational repression of complementary target mRNAs or by inducing transcriptional silencing of target genes. Microarray analysis indicated that _17% of coldupregulated genes encode transcription factors, whereas only 7% of cold-down regulated genes encode transcriptional regulators (Lee, B-H. et al. 2005

Post-translational regulation

Controlled proteolysis of transcriptional regulators has an important role in shaping the cold-responsive transcriptome in plants, as evidenced from studies on HOS1 (HIGH EXPRESSION OF OSMOTICALLY RESPONSIVE GENE 1). HOS1 encodes a RING finger ubiquitin E3 ligase, the nuclear localization of which is also enhanced by cold stress (Lee, H. et al. 2001). Ubiquitin E3 ligases are known to confer substrate specificity for regulated proteolysis by the ubiquitination / 26S proteasome pathway. It was proposed that HOS1 might target upstreamsignaling components or transcriptional regulators of CBFs for proteolysis to regulate negatively the expression CBFs (Lee, H. et al. 2001). ICE1 was identified as a target of HOS1 (Dong, C.H. et al. 2006). HOS1 physically interacts with ICE1 and mediates the ubiquitination of ICE1 both in vitro and in vivo. Polyubiquitination and proteolysis of ICE1 after 12 h of cold stress was found only in wild type but not in hos1 mutants. Transgenic Arabidopsis plants over expressing HOS1 show a substantial reduction in GFP–ICE1 protein levels, and a reduction in transcript levels of CBFs and CBF regulon genes and hypersensitivity to freezing stress. These results demonstrate thatHOS1 ubiquitinates ICE1 to regulate negatively the expression of ICE1 target genes and are thus critical for the de-sensitization of plant cells to cold stress.

Sumoylation is a post-translational protein modification where SUMO (small ubiquitin-related modifier) proteins are conjugated to protein substrates in a process dependent on SUMO E3 ligases, whereas desumoylation is the removal of SUMO proteins from their target proteins by SUMO proteases. Sumoylation might protect target from proteasomal degradation because sumoylation prevents ubiquitination. Sumoylation/desumoylation of proteins has been shown to have a pivotal role in plant responses to abiotic and biotic stress responses and in ABA and salicylic acid signalling. Recently, a role for sumoylation in cold acclimation was found through studies on the Arabidopsis SUMO E3 ligase, SIZ1 (SAP and Miz1). SIZ1 is required for the accumulation of SUMO conjugates during cold stress, and the siz1 null mutant is hypersensitive to chilling and freezing stresses. The siz1 mutation significantly reduces cold-induction of CBFs and its target COR genes (COR15A, COR47 and KIN1), but it enhances the cold induction of AtMYB15, a negative regulator of CBFs. In contrast to HOS1, which promotes the proteolysis of ICE1, SIZ1 mediates SUMO conjugation to K393 of ICE1 during cold acclimation, and this reduces polyubiquitination of ICE1. Transgenic Arabidopsis plants over expressing ICE1 but not ICE1 (K393R) exhibit an enhanced cold induction of CBFs and increased freezing tolerance. Further, similar to ice1 mutant plants, ICE1(K393R) over expressing transgenic plants exhibit a moderate increase in MYB15 expression under cold stress, and display a hypersensitivity to freezing stress. These results suggest that SIZ1-mediated sumoylation might facilitate ICE1 stability and activity, which is necessary for CBF expression and MYB15 repression to fine-tune the transcription of COR genes during cold acclimation (Ulrich, H.D. 2005; Miura, K. et al. 2007; Chinnusamy et al. 2003).

CHAPTER III
MATERIALS AND METHODS

CHAPTER III

MATERIALS AND METHODS

DNA ISOLATION

- 3 g of fresh leaf sample was ground to a fine powder using liquid nitrogen, mortar and pestle
- The ground leaf powder was transferred to 15ml of pre-warmed (60°C) isolation buffer in a capped polypropylene tube and clump was suspended using spatula.
- It was incubated for 30-60 minutes at 60°C in a water bath and mixed every 10 minutes.
- 1 Volume of chloroform iso-amyl alcohol was added; the tubes were capped and extracted for 10 minutes on a rotatory shaker.
- It was centrifuged for 10 minutes (5,000g, room temperature) and the aqueous phase was re-extracted with fresh chloroform isoamylalcohol and centrifuged again.
- The final solution was transferred to a glass centrifuge tube using large bore pipette. Heat treated RNase A was added to a final concentration of 100µg/ml mixed and incubated at room temperature for 30 minutes.
- 0.6 Volume of ice-cold isoproponal was added and mixed gently but thoroughly by inverting the tube several times. The amorphous precipitate was collected by centrifugation (5000g, 10 minutes at 4°C). Remove the residual isopropanol well.
- Add 500 ul of 70% ethanol and centrifuge for 5 minutes at 7K RPM, remove the residual ethanol well and allow the pellet to dry in, keep the tubes open for 10-20 minutes.
- Add 100 ul TE buffer and allow it dissolve at room temperature for 10 minutes and then on ice. Store DNA in 4 °C overnight. Aliquot DNA as working and stock sample. Store in -20 for long term.

CHEMICAL AND RAGENTS PREPARETION

- **CTAB 10%**: 40 GM in 350 ml water. Dissolve well and make up the volume to 400ML.
- **EDTA: 0.5M EDTA pH 8.0 for 400ml.**

74.44g EDTA (mw 372.24) to 250 ml distilled H20. Adjust pH to 8.0 and the EDTA will start to dissolve once dissolved adjust volume to 400ml. **Autoclave and store at 4°C**. (Wear lab coat, latex gloves, dust respirator and safety spectacles.)

- **NaCl 5M solution 200ML**: (molecular wt for NaCl; it is 58.44.)

  Add 58.44 GM in 160 ML water. Make up the volume to 200ML
  Autoclave and store at room temperature.
- **Tris-HCL, 1M pH 8.0, 400ML** (MW: 121.1)

  48.44 g Tris to 250ml distilled water. Adjust pH to 8 with Concentrated HCl. Make up to 400ml with distilled water. Autoclave and store at 4°C.
- **Tris-HCL, 1M pH 7.5, 400ML** (MW: 121.1)

  48.44 g Tris to 250ml distilled water. Adjust pH to 7.5 with Concentrated HCl. Make up to 400ml with distilled water. Autoclave and store at 4°C. Wear lab coat, latex gloves.

Extraction buffer

2% [W/v] CTAB, 1.4M NaCl,20mM EDTA, 100mM Tris HCl pH 8.0,
0.2% β-mercaptoethanol (added just before use)

Reagent	Conc	Concentration required	100ML	500ML
CTAB	10%	2%	20ML	100ML
EDTA	500mM	20 mM	20ML	100ML
NaCl	5M	1.4 M	28ML	140
Tris HCL pH 8	1M	100mM	10ML	50
β-mercaptoethanol		0.2%	200μl	1ML
WATER			21.8ML	109

- **Chloroform Isoamyl Alcohol, 200ML (24:1):** 192 ML Chloroform and 8 ML Isoamyl alcohol

- **RNAse solution: 10mg/ML**

Reagent	Conc required	FOR 10 ML
RNAse A	10mg/ML	100MG
Tris HCl 1M pH7.5	10 mM	100μL
NaCl 5M pH7.5	15 mM	30μL
WATER		TO 10 ML
BOIL FOR 15 MINUTES, LET IT COLL ALIQUATE IN 1O VIAL AND KEEP IT IN -20°c FREEZER		

- **Washing solution:** 70% ethanol

- **TE buffer:**

Reagent	Conc needed	10 ML
Tris HCl 1 M, pH 8	10mM	100µl
EDTA 0.5M	1mM	20 µl
		Make up 10 ML

Identification of possible genes involved in flowering trait in Pigeonpea

- Sequences of the genes protein and nucleotide which are regulate in the flowering traits are retrieved from the TAIR databases. (https://www.arabidopsis.org). It is The Arabidopsis Information Resource (TAIR) maintains database where we get the genetic and molecular biology data for the model higher plant *Arabidopsis thaliana*. Data available from TAIR includes the complete genome sequence along with gene structure, gene product information, gene expression, DNA and seed stocks, genome maps, genetic and physical markers, publications, and information about the Arabidopsis research community.
- After retrieving all gene protein sequences, protein BLAST was performed against database sequence and selected the top six sequences from the result.Here we used the protein query agenised the protein data base for similarity search (http://blast.ncbi.nlm.nih.gov/Blast.cgi).
- Stand alone blast was done using query sequence against all the genes from Pigeonpea genome (*Cajanus Cajan*) which was formatted as database. Local blast also knows as Standalone blast or Offline Blast. This blast describes overall quality of alignment and higher score and higher similarity. Sequences with more than 30% were selected and from the selected sequences top scoring genes were retrieved.

Annotation of identified genes

- Top scoring genes were checked for presence of related features of the respective gene family. The selected genes were the used for gene ontology study using Blast2go tool. Blast2Go is a universal gene ontology annotation, visualization and analysis tool for functional genomics research (nhttp://blast2go.com/webstart/blast2go1000.jlp). Blast2GO (B2G) is a comprehensive bioinformatics tool for the functional annotation and analysis of gene or protein sequences. The tool was originally developed to provide a user-friendly interface for Gene Ontology The gene ontology consortium annotation. Recent improvements have considerably increased the annotation functionality of the tool and currently Enzyme code (EC), KEGG Maps and InterPro motifs are also supported. Additionally the application over's a wide array of graphical and analytical tools for annotation manipulation and data mining.

Searching for ESTs of identified flowering related genes

- Total available Pigeonpea ETSs (Expressed Sequence Tags) were downloaded from ESTdb, NCBI. This EST database is collection of short single read transcript sequence provide a resources provides a resources to evaluate gene expression.Standalone blast was performed against total Pigeonpea ESTs (as database) using identified flowering related genes as query. BLAST hits with significant and top identity were considered as ESTs of the genes in question.

Multiple sequence alignment

- Multiple sequence alignments of the flowering traits gene from the all diverse plants was performed • using ClustalW2. ClustalW2 is a general purpose multiple sequence alignment program for DNA or proteins. It attempts to calculate the best match for the selected sequences, and lines them up so that the identities, similarities and differences can be seen. (http://www.ebi.ac.uk/Tools/msa/clustalw2/help/)

Identification of flowering related genes from Pigeonpea using HMM based approach

- Results from clustalw multiple sequence alignment used to construct for HMM profiling. HMMER-3.0 is the software which is used for searching sequence databases for homologs of protein sequences. It also knows as profiling HMM (Hidden markov model). Multiple sequence alignment and used to construct HMM profile. Using HMMER-3.0 the local protein database was searched on the basis of the HMM profile. HMMER and BLAST hits were compared and parsed.

Gene expression analysis using SRA data and CLC genomics workbench

- Pigeonpea SRA data was downloaded from SRA database of NCBI (http://www.ncbi.nlm.nih.gov/sra/). Sequence Read Archive is store data for next generation sequence platform.SRA makes biological sequence data which is help in research community and allow the new comparison in data sets. The SRA data was used in CLC workbench for saw the expression of the genes using the RNA seq Analysis. CLC Genomics Workbench solves this problem and will enable everyone to rapidly analyze and visualize the huge amounts of data generated by NGS machines. CLC workbench used for multipurpose area of NGS like genomics, Transcriptomics Epigenomics, Classical sequence analysis tools. RNA seq analysis gives expression like RNA-seq, Small RNA, tag based expression based profiling and single colour microarray gene expression data. The interactivity of the multiple available views allows easy navigation and overview of data and analysis results. The complete integration of the expression analysis in the workbench enables the user to carry out downstream analysis of genes of interest with the comprehensive set of sequence analysis tools provided, immediately and without the hassle of switching between software's.

In-silico identification of ICE1

- To understand the ICE1 in template (*Brassica Rapa*, *Raphanus sativus, Arabidopsis thaliana, Solanum lycopersicum, Solanum tuberosum, Malus domestica, Triticum aestivum, Zea mays, Fragaria vesca*), tropical *(Oryza sativa, Citrus sinensis, Gossypium arboretum, Setaria italic, Cajanus cajan, Sorghum bicolor*, *Ricinuscommunis*, *Theobroma cacao*) and subtropical (*Glycine max, Cucumis sativus, Cucumis melo)* crop plants, sequence were retrieving from the gene bank from BLAST search using the Arabidopsis ICE1. For sorghum, Maize, Foxtail millet and pigeon pea search was performed by stand alone blast or using HMM profile against total protein sequence of the individual plant. In case of Pigeon pea total contigs, from whole genome sequence project were used to for gene prediction program Augustus. Augustus is universally used for gene prediction (Coding and Protein). The available ICE1 sequences from diverse plants were aligned using multiple sequence alignment.

- For further study and understand the relation between the diverse plants sequences from template (*Brassica Rapa, Raphanus sativus, Arabidopsis thaliana*, *Solanum lycopersicum, Solanum tuberosum, Malus domestica, Triticum aestivum*, *Zea mays*, *Fragaria vesca*), tropical *(Oryza sativa*, *Citrus sinensis, Gossypium arboretum, Setaria italic, Cajanus cajan, Sorghum bicolor, Ricinus communis, Theobroma cacao*) and subtropical (*Glycine max, Cucumis sativus, Cucumis melo)* crop plants phylogeny tree was construct.

  Phylogeny tree construct to analysing phylogenetic relationships between molecular sequences (http://phylogeny.lirmm.fr/phylo_cgi/simple_phylogeny.cgi).

In–silico characterization of flowering related/ICE1 gene/s

- Genes were studies for sub cellular localization, molecular weight (Dalton), motif conservation, sumoylation sites and introns localization. Sub cellular localization give the gives some major component like mitochondria, golgibuddy, cytoplasm, nucleus, ribosome, nuclear, extra-cellular space etc. CELLO: - sub cellular localization prediction tool (http://cello.life.nctu.edu.tw/) were used.

- Conserved motif also known as conserved domain, which gives the protein sequence structure that can evolve, function and exit independently of the rest of the chain. NCBI conserved Domain search tool used for search the conserved domain of this genes.(http://www.ncbi.nlm.nih.gov/Structure/cdd/wrpsb.cgi)
- Small Ubiquitin like modifier protein are a small family which covalently attached and detached from other proteins in the cells to mortify their function, which is known as Sumoylation site. By the help of sumoylation site prediction tool, the ubiquitin like modifier search in these genes (http://sumosp.biocuckoo.org/online.php).
- Fancy gene tool v1.4, which is used for to search the introns in the genes. This tool visualized like intron, exon, markers and domain (http://bio.ieo.eu/fancygene/).

- Again, the genes are used to search for the start codon using the ORF finder. ORF (Open Reading Frame) is a graphical analysis tool which is using to search for standard and alternative genetic codons (http://www.ncbi.nlm.nih.gov/gorf/orfig.cgi). An open reading frame is a continuous stretch of DNA beginning with start codon, usually methionine (ATG) and ending with stop codon (TAA, TAG, TGA).
- To analysis the better conservation and functional partner in other organism again the gene are used in STRING data base. STRING data base is to predicted protein interaction. This interactions include direct (physical) and indirect (functional) associations. It is also known as protein-protein interaction data base. (http://string-db.org/newstring_cgi/show_input_page.pl) STRING quantitatively integrates interaction data from these sources for a large number of organisms, and transfers information between these organisms where applicable. The database currently covers 9'643'763 proteins from 2'031 organisms.

CHAPTER IV

RESULTS

CHAPTER IV

RESULTS

4.1 In silico identification and characterization of genes involved in flowering trait regulations

Flowering time trait is one of the most important trait in plant development. In optimal condition conditions late flowering leads to longer vegetative growth to accumulate resources for seed production. However, in less favorable and unpredictable environmental conditions plants have evolved early flowering. In many of the crop plants early flowering is desirable traits. Several genes have been identified and cloned in plants especially in *Arabidopsis thaliana* and rice. However, in pulses, there is very few information regarding genes involved in flowering regulation. With the advent of sequencing technologies and availability of complete genome sequence of crop plants, it has become possible to mine genes of interest using bioinformatics approaches. In present work attempts have been made in identification of genes important regulators of flowering traits from Pigeonpea (*Cajanus cajan*).

In silico identification genes involved in flowering trait regulations

In order to identify genes involved in flowering trait regulations, six important regulators namely Frigida, Frigida Like1, Frigida Like2, Frigida Essential1, Terminal Flowering1, Terminal Flowering2 have been selected for present study. The above mentioned genes have already been known to play important role in flowering regulation in different plants (Roux et al. 2006). Protein sequences of above mentioned genes from Arabidopsis and other related plants have been retrieved from gene bank, NCBI. Two different approaches have been followed, 1. HMM build and HMM profile search and, 2. Standalone BLAST search using protein sequences of the respective genes from different plant. In both cases, total protein sequences from *C. cajan* were used as database. Results from both HMM profile search and stand alone blast results have been parsed by manually. Self blast was performed to remove the redundancy. Figure 1 depicts the work flow followed for the identification of flowering genes from Pigeonpea. Following this approach, the top hits with more than 30% identity and with known conserved domains for respective proteins have been considered to be the genuine orthologs of flowering genes in Pigeonpea. The identified Pigeonpea (*Cajanus cajan*) genes were thus named *CcFrigida*, *CcFrigida*

Like1, CcFrigida Like2, CcFrigida Essential1, CcTerminal Flowering1, and *CcTerminal Flowering2*. Table 4.1 shows identity of flowering related genes from Pigeonpea to that of Arabidopsis. Among selected genes of Pigeonpea, TERMINAL FLOWERING 1 protein showed maximum identity (74%) and similarity (88%), whereas CcFrigida essential 1 showed minimum identity (31%) and similarity (43%) with their respective orthologs in Arabidopsis.

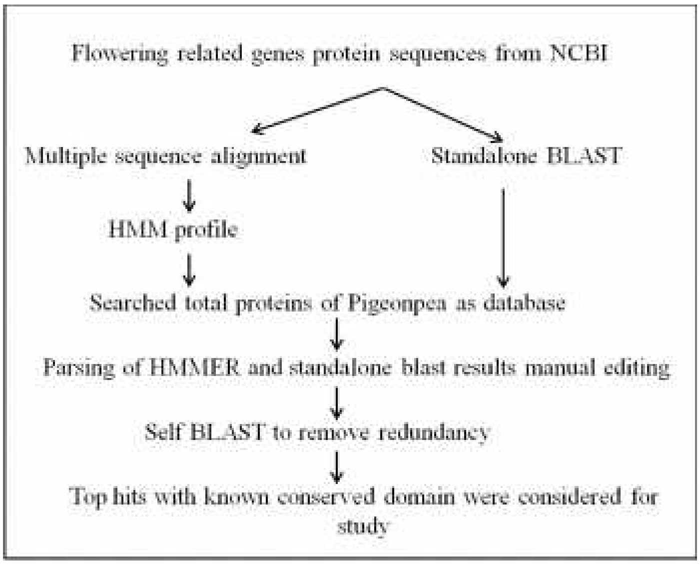


Figure 1: Work flow followed for identification of flowering genes from Pigeonpea

GENES NAME	IDENTITY	SIMILARITY	E-VALUE
CcFRIGIDA	39%	61%	3e-96
CcFRIGIDA LIKE 1	31%	55%	9e-34
CcFRIGIDA LIKE 2	31%	52%	3e-39
CcFRIGIDA ESSENTIAL 1	31%	43%	4e-24
CcTERMINAL FLOWERING 1	74%	88%	7e-96
CcTERMINAL FLOWERING 2	37%	51%	4e-57

Table 4.1. Identity/similarity of putative flowering genes from Pigeonpea with *A. thaliana*

In order to see whether, identified protein had conserved functional annotation, Blast2Go analysis was performed and results showed that all the six proteins showed similar ontology to that of respective Arabidopsis genes with respect to functional, biological process as we as cellular functions. Results of the blast2Go are shown in table 4.2.

GENE	**Seq. Description**	**min. e Value**	**mean Similarity**	**GOs**	**Enzyme Codes**	**InterProScan**
FRIGIDA	protein frigida	0	67.9%	P:regulation of multicellular organismal development	NIL	Coil (COILS); Coil (COILS); Coil (COILS); IPR012474 (PFAM); PTHR31791 (PANTHER); PTHR31791:SF3 (PANTHER)
FRIGIDA LIKE-1	frigida-like protein 3	0	88.05%	F:molecular_function; P:biological_process	NIL	Coil (COILS); Coil (COILS); IPR012474 (PFAM); PTHR31791 (PANTHER); PTHR31791:SF4 (PANTHER)
FRIDIGA LIKE-2	frigida-like protein 3	0	88.05%	F:molecular_function; P:biological_process	NIL	Coil (COILS); Coil (COILS); IPR012474 (PFAM); PTHR31791:SF4 (PANTHER); PTHR31791 (PANTHER)
FRIGIDA ESSENTIAL 1	zinc finger c-x8-c-x5-c-x3-h type family isoform 1	0	63.65%	F:binding	NIL	PTHR15242 (PANTHER)
TERMINAL FLOWERING LOCUS 1	dt1(determinant stem)	4.45E-120	96.05%	F:transcription cofactor activity; P:negative regulation of flower development; P:photoperiodism flowering; C:nucleus; C:plasma membrane	NIL	IPR000953 (SMART); IPR007630 (PFAM); IPR011991 (G3DSA:1.10.10.GENE3D); (SUPERFAMILY); IPR016197 (SUPERFAMILY)
TERMINAL FLOWERING LOCUS 2	rna polymerase sigma factor sigb	0	84.5%	P:response to far red light; P:carotenoid biosynthetic process; F:sequence-specific DNA binding transcription factor activity; P:rRNA processing; P:thylakoid membrane organization	EC:2.7.7.6	IPR000943 (PRINTS); IPR008251 (SMART); IPR000953 (SMART); IPR007630 (PFAM); IPR011991 (G3DSA:1.10.10.GENE3D); (SUPERFAMILY); IPR016197 (SUPERFAMILY)

Table 4.2: Functional annotations of Pigeonpea flowering genes as shown by Blast2Go

A few important features of the identified Pigeonpea flowering genes have also been studied such as molecular weight, PI and putative sub-cellular localization. Frigida, Frigida Like 1, Frigida Essential 1 and Terminal Flowering Locus 2 showed nuclear localization, whereas Frigida Like 2 showed both nuclear as well as cytoplasmic localization and Terminal Flowering Locus 1 showed mitochondrial localization (Table 4.3). The identified protein showed molecular weight ranging from 61 to 99kDa and also possessed conserved domains observed in their orthologs in Arabidopsis.

GENE	**PI/Mw(Dalton)**	**Sub-cellular location(score)**	**Conserved domains**
Frigida	6.19/67897.36	Nuclear (2.273)	Frigida-like protein
Frigida Like 1	6.05/61116.42	Nuclear (2.461)	Frigida-like protein Bin/Amphiphysin/Rvs (BAR) OmOutermrotepinembranepH YqaJ-like viral recombinase
Frigida Like 2	6.05/61116.42	Nuclear (2.352) Cytoplasmic (1.742)	Figida-like protein Bin/Amphiphysin/Rvs (BAR) Outer membrane protein (OmpH-like) YqaJ-like viral recombinase
Frigida Essential 1	7.09/70620.73	Nuclear (4.752)	NIL
Terminal Flowering Locus 1	9.10/19606.52	Mitochondrial (1.305)	PhosphatidylEthanolamineBinding Protein (PEBP)
Terminal Flowering Locus 2	8.052/99023.55	Nuclear (4.282)	Chromatin organization modifier (chromo Sigma70 region (SR) Sigma-70 region Chromo ShadowDomain

Table 4.3: Details of molecular weight, PI, sub-cellular localization and conserved domain of flowering related genes from pigeon pea

Conservation of important domains in the identified Pigeonpea flowering genes was also studied by comparing with their counterpart in Arabidopsis thaliana by sequence alignment using clustalw (Figure 2A-F). CcFrigida essential 1 shows more variations at N-terminal end and relatively higher conservation at C-terminal end of the protein (Figure 2A).

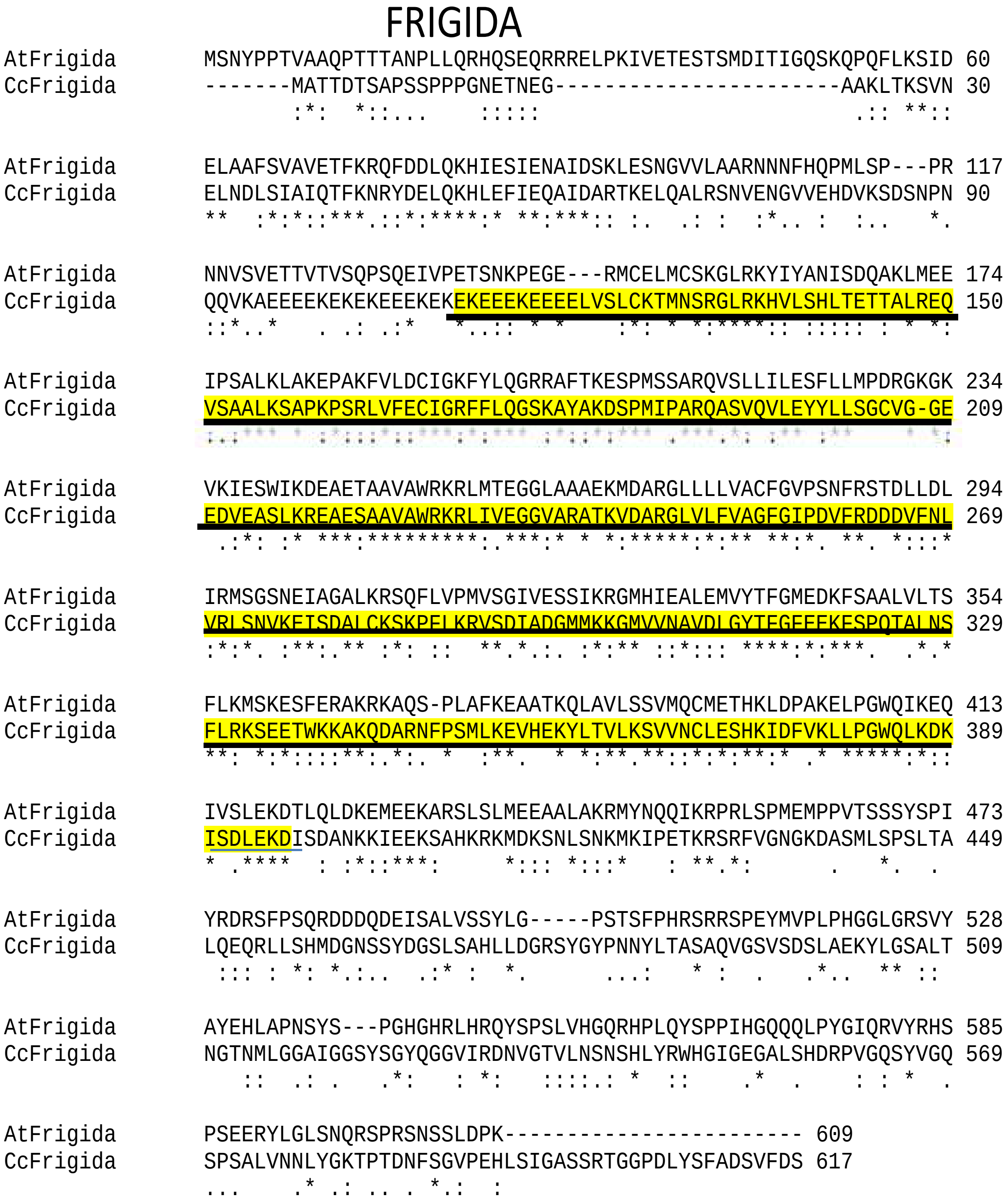
GENE	**QUERY**	**Accession no.**	**E value**	**Identity**
FRIGIDA	NA	NA	NA	NA
FRIGIDA ESSENTIAL 1	3507	GW357230.1 GW357231.1	7E-39 0.0000003	95% 94%
FRIGIDA LIKE_1	22942	GW347796.1 GW347795.1	0 e-160	96% 95%
FRIGIDA LIKE_2	22942	GW347796.1 GW347795.1	0 e-160	96% 95%
TERMINAL FLOWERING LOCUS 1	N A	NA	NA	NA
TERMINAL FLOWERING LOCUS 2	NA	NA	NA	NA

Table 4.4: Identity of putative flowering genes from Pigeonpea with EST Data

In Table no. 4.4 showing the identity of flowering related genes from Pigeonpea to that EST database. Among selected genes of Pigeonpea, TERMINAL FLOWERING 1 and TERMINAL FLOWERING 2 proteins showing no identity with the EST database. Whereas FRIGIDA ESSENTIAL 1,FRIGIDA LIKE 1 and FRIGIDA LIKE 2 showing above 90% identity with EST Database.

Figure 2A: Sequence alignment of Pigeonpea Frigida essential 1 (CcFrigida essential 1) with Arabidopsis Frigida essential 1 (AtFrigida essential 1)

Amino acid sequence alignment of Pigeonpea FRIGIDA to that of Arabidopsis FRIGIDA showed relatively high sequence conservation FRIGIDA like protein domain (Figure 2B). In similar fashion CcFIRIGIDA like 1 shows sequence conservation compared to AtFIRIGIDA like 1 largely at conserved FIRIGIDA like domain (Figure 2C) which is spanning most of the length of the protein. Similar sequence conservation at conserved FIRIGIDA like



FRIGIDA

```
AtFrigida       MSNYPPTVAAQPTTTANPLLQRHQSEQRRRELPKIVETESTSMDITIGQSKQPQFLKSID 60
CcFrigida       -------MATTDTSAPSSPPPGNETNEG-------------------------AAKLTKSVN 30
                       :*:  *::...    :::::                         .:: **::

AtFrigida       ELAAFSVAVETFKRQFDDLQKHIESIENAIDSKLESNGVVLAARNNNFHQPMLSP---PR 117
CcFrigida       ELNDLSIAIQTFKNRYDELQKHLEFIEQAIDARTKELQALRSNVENGVVEHDVKSDSNPN 90
                **  :*:*::***.::*:****:* **:***:: :.  .: :  :*.. :  :..   *.

AtFrigida       NNVSVETTVTVSQPSQEIVPETSNKPEGE---RMCELMCSKGLRKYIYANISDQAKLMEE 174
CcFrigida       QQVKAEEEEKEKEKEEEKEKEKEEEKEEEELVSLCKTMNSRGLRKHVLSHLTETTALREQ 150
                ::*..*   . .: .:*   *..::  * *    :*: * *:****:: ::::: : * *:

AtFrigida       IPSALKLAKEPAKFVLDCIGKFYLQGRRAFTKESPMSSARQVSLLILESFLLMPDRGKGK 234
CcFrigida       VSAALKSAPKPSRLVFECIGRFFLQGSKAYAKDSPMIPARQASVQVLEYYLLSGCVG-GE 209
                [illegible]

AtFrigida       VKIESWIKDEAETAAVAWRKRLMTEGGLAAAEKMDARGLLLLVACFGVPSNFRSTDLLDL 294
CcFrigida       EDVEASLKREAESAAVAWRKRLIVEGGVARATKVDARGLVLFVAGFGIPDVFRDDDVFNL 269
                 .:*: :* ***:*********..***:* * *:*****:*:** **:*. **. *:::*

AtFrigida       IRMSGSNEIAGALKRSQFLVPMVSGIVESSIKRGMHIEALEMVYTFGMEDKFSAALVLTS 354
CcFrigida       VRLSNVKEISDALCKSKPELKRVSDIADGMMKKGMVVNAVDLGYTEGEEEKESPQTALNS 329
                :*:*. :**:.** :*: ::  **.*.:. :*:** ::*::: ****:*:***.  .*.*

AtFrigida       FLKMSKESFERAKRKAQS-PLAFKEAATKQLAVLSSVMQCMETHKLDPAKELPGWQIKEQ 413
CcFrigida       FLRKSEETWKKAKQDARNFPSMLKEVHEKYLTVLKSVVNCLESHKIDFVKLLPGWQLKDK 389
                **: *:*::::**:.*:. *  :**.  * *:**.**::*:*:**:* .* *****:*::

AtFrigida       IVSLEKDTLQLDKEMEEKARSLSLMEEAALAKRMYNQQIKRPRLSPMEMPPVTSSSYSPI 473
CcFrigida       ISDLEKDISDANKKIEEKSAHKRKMDKSNLSNKMKIPETKRSRFVGNGKDASMLSPSLTA 449
                * .****  : :*::***:     *::: *:::*   : **.*:      .   *.  .

AtFrigida       YRDRSFPSQRDDDQDEISALVSSYLG-----PSTSFPHRSRRSPEYMVPLPHGGLGRSVY 528
CcFrigida       LQEQRLLSHMDGNSSYDGSLSAHLLDGRSYGYPNNYLTASAQVGSVSDSLAEKYLGSALT 509
                 ::: : *: *.:..  .:* :  *.      ...:   * :  .   .*..  ** ::

AtFrigida       AYEHLAPNSYS---PGHGHRLHRQYSPSLVHGQRHPLQYSPPIHGQQQLPYGIQRVYRHS 585
CcFrigida       NGTNMLGGAIGGSYSGYQGGVIRDNVGTVLNSNSHLYRWHGIGEGALSHDRPVGQSYVGQ 569
                   ::  .: .   .*:   : *:   ::::.: *  ::    .*  .    :  : *  .

AtFrigida       PSEERYLGLSNQRSPRSNSSLDPK------------------------ 609
CcFrigida       SPSALVNNLYGKTPTDNFSGVPEHLSIGASSRTGGPDLYSFADSVFDS 617
                ...    .* .: .. . *.:  :
```

Figure 2B: Sequence alignment of Frigida from *C. cajan* and *A. thaliana*.

domain have also been observed in CcFIRIGIDA like 2, however insertion of 81 amino acids long stretch have also been found in CcFIRIGIDA like 2 (Figure 2D). Similarly in Pigeon pea Terminal flowering 1, additional 54 amino acid long stretch has been observed at

```
AtFrigida       MTASETIATAINQIDEKKEKLKKAFDDLQAHRSLLSPSFSLSWSEIDSHFSSLQSSLASR 60
CcFrigida       MEDTDSVATLIDSTTSKIQQLQKAFAELESYR---AVSLNLKWKELEEHFHGLEKSLKRR 57
                *  ::::** *:.  .* ::*:*** :*:::*   : *:.*.*.*::.** .*:.**  *

AtFrigida       FRLLHS------------------------------------------------------ 66
CcFrigida       FHELEDQEKEFENKTMKARELLEKREADVYAKEQASLQSLQDKRDAAVFAIVNAREKHRK 117
                *: *..

AtFrigida       -------------TSPLEHDSYRIDASDAGKSSSSEEVSEQP-------VVEPELRALCE 106
CcFrigida       VSSRELAIVSNGGKGTLGVEEKLVDTVSVAAESNVEEVKLSPENGNVDLISYPELVKLCK 177
                             ...*  :.  :*: ... .*. ***. .*       :  ***  **:

AtFrigida       KIDGIGLIKYLIRIWDDETPLNQEVSAAIRYSPDTASMVLDAIEGSNYTPSSS------G 160
CcFrigida       EMDAAGLHKFISDNRKNLAAVREDIPNALRAAPNAACLVLDSLEGFYCTEVSNQDVKKDA 237
                ::*. ** *::    .: :.:.:::. *:* :*::*.:***::**   *  *.      .

AtFrigida       RSFDVRRVFVLLMEVLIEINAN-------ITVDTRNRAKKLAYHWKSKVG--------VK 205
CcFrigida       NLLGLRRTCIMLMECLCDFLSNSACVTNSISEDIKDRAKAVAEAWKPRLDALDMDASNGN 297
                . :.:**. ::*** * :: :*       *: * ::*** :*  **.::.         :

AtFrigida       PFEALVFLHLVAAFELGSEFDTEELSDYVFMIAKYKQATLVCNKIGVDRKRVGKLIKTLL 265
CcFrigida       SLEAHAFLQLVASFGIASGFDEEELSRLIPMVSRRRQTADLCRFLGLSEKMPG-VIEVLV 356
                .:** .**:***:* :.* ** ****  : *::: :*:: :*. :*:..*  * :*:.*:

AtFrigida       DSGKPILAVKFMYECGMTDEFEPIPVLKSYIKDCREAALRVCVEDNYSLKSQNEASDKEV 325
CcFrigida       NNGRQIDAVNLAFAFDLTEQFSPVSLLKSYLKDARKASSPVKSVN-SSPTMQIEVNEREL 415
                :.*: * **:: :  .:*::*.*:.:****:**.*:*:  *   :  * . * *.:.:*:

AtFrigida       SALKPLIKIIKDQNLESEFTQEKVEERVEELEKNKALRKR--NTTNPPKQEPQQKGKKRT 383
CcFrigida       VALKAVIKCIEEHKLDEKYPLDPLQKRLVQLEKAKADKKREAEATKPQPKRPRANGAGYV 475
                 ***.:** *::::*:.::. : :::*: :*** ** :**  ::*:*  :.*: :*   .

AtFrigida       RDCKNGSQVPVPSQQLLSRPEALLMPEHSHHGLQLNPYGLMTSAFSGVVVNPLTGLFGSG 443
CcFrigida       PRVTN-----IPSDKTCYARVADRYPQYVYDRSYMYPG------PTDNHCPPLMSTATYN 524
                   .*     :**::      *   *:: :.   : *        :.    ** .    .

AtFrigida       ATPQSLYYAQQTGYVLPPQYHPPYYSQ 470
CcFrigida       FSPSHGNYFGNG-----YQYQATYLH- 545
                 :*.   *  :       **:..*
```

Figure 2C: Sequence alignment of Frigida like 1 from *C. cajan* and *A. thaliana*.

N-terminal end of the protein (Figure 2E). In case of Cc Terminal flowering 2, presence of additional N terminal extension compared to AtTerminal flowering 2 was observed, however, C-terminal end showed relatively higher sequence conservation (Figure 2F).

FRIGIDA LIKE 2

```
AtFrigida       MTAAESIAASINQIDEKKQKLKKAFDDLQAHRSLLSPSFNLSWSEIDSHFSSLQSSLFNR 60
CcFrigida       MEDTDSVATLIDSTTSKIQQLQKAFAELESYR---AVSLNLKWKELEEHFHGLEKSLKRR 57
                *  ::*:*: *:.  .* *:*:*** :*:::*   : *:**.*.*::.** .*:.** .*

AtFrigida       LQSAVTS----------------------------------------------------- 67
CcFrigida       FHELEDQEKEFENKTMKARELLEKREADVYAKEQASLQSLQDKRDAAVFAIVNAREKHRK 117
                ::.   .

AtFrigida       --------------------------------SNSGNIETPTAVTTETPVLWPELRKFCE 95
CcFrigida       VSSRELAIVSNGGKGTLGVEEKLVDTVSVAAESNVEEVKLSPENGNVDLISYPELVKLCK 177
                                                **  ::: ..   .   : :*** *:*:

AtFrigida       KNDGKGLGNYMIENSRKRLSINEELPNAIRCSENPAALVLDAIEGSYHCSSPSS--SSSA 153
CcFrigida       EMDAAGLHKFISDNRKNLAAVREDIPNALRAAPNAACLVLDSLEGFYCTEVSNQDVKKDA 237
                : *. ** ::: :* ::  ::.*::***:*.: *.*.****::** *  . ...  ...*

AtFrigida       RAIDVKRIFVLLLEALIEINAN-------LTNDLRERARTIAYDWKP--------NIGNK 198
CcFrigida       NLLGLRRTCIMLMECLCDFLSNSACVTNSISEDIKDRAKAVAEAWKPRLDALDMDASNGN 297
                . :.::*  ::*:*.* :: :*       :::*:::**:::*  ***          ..:

AtFrigida       PSEALGFLHLVAAFELGSLFSTEEICDYIFLISKYKQATTICKKIGLDRNRIGVLVQKFL 258
CcFrigida       SLEAHAFLQLVASFGIASGFDEEELSRLIPMVSRRRQTADLCRFLGLS-EKMPGVIEVLV 356
                . ** .**:***:* :.* *. **:.  * ::*: :*:: :*: :**. :::  ::: ::

AtFrigida       DTGRLLVAIRFIYENEMVGEFEPVSILKTSLKNSREAAKRVCAEGNYSLKVQNEATDKEL 318
CcFrigida       NNGRQIDAVNLAFAFDLTEQFSPVSLLKSYLKDARKASSPVKSVN-SSPTMQIEVNEREL 415
                :.** : *:.: :  ::. :*.***:**: **::*:*:. * : .  * .:* *..::**

AtFrigida       SALRAVIKVVKEKNIESEFMEEKLEECVKELEDQKAQRKRATKFNSPANPQQPQEQKVDN 378
CcFrigida       VALKAVIKCIEEHKLDEKYPLDPLQKRLVQLEKAKADKKREAEATKPQP----------- 464
                 **:**** ::*::::.::  : *:: : :**. **::** :: ..*

AtFrigida       KRPRVANGSSMEYNLTIPPLRPLQQPPLLPTPSQILQVNPYGLLSSILPGVAVPYGNPRA 438
CcFrigida       KRPRANGAGYVPRVTNIPSDKTCYARVADRYPQYVYDR------SYMYPGPTDNHCPPLM 518
                ****. ... :    .**. :.         *. : :       * : ** :  :  *

AtFrigida       LFGSVPAPASRPVFYVQQTGYGMPPPQYRPPYYPQ 473
CcFrigida       STATYNFSPSHGNYFGNG-------YQYQATYLH- 545
                  .:   ..*:  :: :         **:..*
```

Figure 2D: Sequence alignment of Frigida like 2 from *C. cajan* and *A. thaliana*.

TERMINAL FLOWERING 1

```
AtTerminal      ------------------------------------------------------------
CcTerminal      MEDTDSVATLIDSTTSKIQQLQKAFAELESYRAVSLNLKWKELEEHFHGLEKSLKRRFHE 60


AtTerminal      ------------------------------------------------------MENMGT 6
CcTerminal      LEDQEKEFENKTMKARELLEKREADVYAKEQASLQSLQDKRDAAVFAIVNAREKHRKVSS 120
                                                       .::.:

AtTerminal      RVIEPLIMG-----RVVGDVLDFFTPTTKMNVSYNKKQVSNG-HELFPSSVSSKPRVEIH 60
CcTerminal      RELAIVSNGGKGTLGVEEKLVDTVSVAAESNVEEVKLSPENGNVDLISYPELVKLCKEMD 180
                * :  :  *      *  .::* .: ::: **.  * . .**  :*:. .   *   *:.

AtTerminal      GGDLRSFFT-----LVMIDPDVPGPS--DP------------------------------ 83
CcTerminal      AAGLHKFISDNRKNLAAVREDIPNALRAAPNAACLVLDSLEGFYCTEVSNQDVKKDANLL 240
                ...*:.*::     *. :  *:*..    *

AtTerminal      -------FLKEHLHWIVTNIPGTT---------------------------DATFG---- 105
CcTerminal      GLRRTCIMLMECLCDFLSNSACVTNSISEDIKDRAKAVAEAWKPRLDALDMDASNGNSLE 300
                       :* * *  :::* . .*                           **: *

AtTerminal      -----KEVVSYELPR-----------PSIGIHR-----FVFVLFRQKQRRVIFPNIPSRD 144
CcTerminal      AHAFLQLVASFGIASGFDEEELSRLIPMVSRRRQTADLCRFLGLSEKMPGVIEVLVNNGR 360
                     : *.*: :.            * :. :*       *: : :*   **   : .

AtTerminal      HFNTRKFAVEYDLGLPVAAVFFNAQRETAARKR--------------------------- 177
CcTerminal      QIDAVNLAFAFDLTEQFSPVSLLKSYLKDARKASSPVKSVNSSPTMQIEVNERELVALKA 420
                :::: ::*. :**   .:.* :  .  . ***

AtTerminal      ------------------------------------------------------------
CcTerminal      VIKCIEEHKLDEKYPLDPLQKRLVQLEKAKADKKREAEATKPQPKRPRANGAGYVPRVTN 480


AtTerminal      ------------------------------------------------------------
CcTerminal      IPSDKTCYARVADRYPQYVYDRSYMYPGPTDNHCPPLMSTATYNFSPSHGNYFGNGYQYQ 540


AtTerminal      -----
CcTerminal      ATYLH 545
```

Figure 2E: Sequence alignment of Terminal flowering 1 from *C. cajan* and *A. thaliana*.

TERMINAL FLOWERING 2

```
AtTerminal      ------------------------------------------------------------
CcTerminal      MSCLLPHFNCHLDTFRPLPYAHHPSKTRPNLCLQPQCVLSATSPPSTLLNLENKLALEAP 60

AtTerminal      ------------------------------------------------------------
CcTerminal      SHPNTPWPYITAPSQDNFKSTLSAESLLTNEEAVIAAAASEALALAKAAVKVAKDAALLV 120

AtTerminal      ------------------------------------------------------------
CcTerminal      KKKPPAEAEYRSHVSSKSDDLLLKWFQHMEAEDGVAGGSIDAGAEIMEGVDINPSEEESD 180

AtTerminal      ------------------------------------------------------------
CcTerminal      VEPSLEELECLQEELSDSIAVRSRRQIERKAKRVRAAEKATTNITSIKSGSSSRRKRVSM 240

AtTerminal      ------------------------------------------------------------
CcTerminal      QEVDYSDPLRYLRTTTSASRLLTPSEEIKLSAGIQDLLKLEKLQEDLTERFGGQPTFAQW 300

AtTerminal      ---------------------------------------------------MKGASGAVK 9
CcTerminal      AAVAGVDQKTLRKRLNYGICCKDKMIKSNIRLVISIAKNYQGSGMNLQDLVQEGCRGLVK 360
                                                                    :*. * **

AtTerminal      ---------------------------------------------------KKPQVLNEA 18
CcTerminal      GAEKEDGTKGEKESTYAHWWIKQAVRKSLSDQSRTIRLPEHMVEATYRVKEARKQLYSEN 420
                                                                    : *: .*

AtTerminal      G---EAETAVETVGESRK-----------------------------ISGD--------- 37
CcTerminal      GRQPDDEEVAEATGLSMKRLNAVLLTPKAPRSLEQKIGINQNLKPSEVIADPDAETTEEQ 480
                *   : * ..*:.* * *                             : .*

AtTerminal      --------------------------GGFGSDDGGGGGGGGSGESILREIGDDRPTEDGD 71
CcTerminal      LLKQFMKKDLEEALDSLNPRERQVVRWRFGMDDGRTKTLQEIGEMLGLEIAMKKRSTTSS 540
                                            ** ***        ** :  **. .: :  ..

AtTerminal      EEEEEDEDEDDGGD------------------EEDEEGEGEGGQEERPKLDEGFYEIEAI 113
CcTerminal      VWEVNANAIENGVDDSQLQLQPSSSAVAELPKEKQEETLVPPNPNPNPNLADGFYEIETI 600
                  * : :  ::* *                  *::**     . : .*:* :******:*

AtTerminal      RRKRVRKGKVQYLIKWRGWPETANTWEPLENLQSIADVIDAFEGSLKPGKPGRKRKRKYA 173
CcTerminal      RRKRVRKGQLQYLIKWRGWPETANTWEPLQNLQSVPDLIDAFEESLKSGK-HRKRKRKHV 659
                ********::*******************:****:.*:***** ***.**  ******:.

AtTerminal      GPHSQMKKKQRLTSTSHDATEKSDSSTSLNNSSLPDIPDPLDLSGSSLLNRDVEAKNAYV 233
CcTerminal      VHHTQLKKR----------EERSTTAYSLR-----HFPTPLS------------------ 686
                  *:*:**:           *:* :: **.     .:* **.

AtTerminal      SNQVEANSGSVGMARQVRLIDNEKEYDPTLNELRGPVNN----SNGAGCSQGGGIGSEGD 289
CcTerminal      ADQGLTNADPTTLAKPLLTNADENEYDPKLSELKATANTGVEVDNLAIHEQQPRVSTANG 746
                ::*  :*:... :*: :    :*:****.*.**:...*.    .* *   *   :.: ..

AtTerminal      NVRPNGLLKVYPKELDKNSRFIGAKRRKSGSVKRFKQDGSTSNNHTAPTDQNLTPDLTTL 349
CcTerminal      HVD--GQPQVDYVEPTQSGRCRGAKRRKSGSVKRFKRETDPCKLVDAQNGVAMP------ 798
                :*   *  :*   *  :..*  **************:: ...:   * ..  :.

AtTerminal      DSFGRIARMGNEYPGVMENCNLSQKTKIEELDITKILKPMSFTASVSDNVQEVLVTFLAL 409
CcTerminal      ---AGAACMGYNSPMKMGDAKTACN-------IVKIIKPIGYSASLSDNMQDVLVTFMAM 848
                   .  * ** : *  * :.: : :       *.**:**:.::**:***:*:*****:*:

AtTerminal      RSDGKEALVDNRFLKAHNPHLLIEFYEQHLKYNRTP-- 445
CcTerminal      RSDGTEVMVDNRYLKAYNPLLLINFYELHLRYSPTSCD 886
                ****.*.:****:***:** ***:*** **:*. *.
```

Figure 2F: Sequence alignment of Terminal flowering 2 from *C. cajan* and *A.thaliana*

Prediction of SUMOylation of flowering related proteins

SUMOylation is one of the most important post-translational modifications which plays a pivotal role in several cellular processes, including nuclear-cytosolic transport, transcriptional regulation, various biotic and abiotic stress response generation, and protein stability. Small Ubiquitin-like Modifier (SUMO) proteins are small proteins which modify the function of the target protein either by attaching or detaching to/from it.

GENE	POSITION	PEPTIDE	SCORE	CUTOFF
CcFrigida	94 217 374	SNPNQQVKAEEEEKE EDVEASLKREAESAA VNCLESHKIDFVKLL	46.88 28.021 38.054	16 16 36.625
CcFrigida Like 1	234 354 - 358 415 - 419 429	EVSNQDVKKDANLLG KMPGVIEVLVNNGRQIDAV IEVNERELVALKAVIKCIE IKCIEEHKLDEKYPL	42.742 61.262 61.369 42.466	36.625 59.29 59.29 36.625
CcFrigida Like 2	234 354 - 358 415 - 419 429	EVSNQDVKKDANLLG KMPGVIEVLVNNGRQIDAV IEVNERELVALKAVIKCIE IKCIEEHKLDEKYPL	42.742 61.262 61.369 42.466	36.625 59.29 59.29 36.625
CcFrigida Essential 1	29 36 189	SFEMLTLKKEELHLK KKEELHLKSEISCNL AHQKRELKVEEGVRE	26.033 24.476 27.478	16 16 16
CcTerminal Flowering Locus 1	53 86 - 90	STVNTIPKVEIDGGD DPYLREHLHWIVTDIPGTT	21.518 62.164	16 59.29
CcTerminal Flowering Locus 2	279 410 780 821 - 825	AGIQDLLKLEKLQED VEATYRVKEARKQLY SGSVKRFKRETDPCK DAKTACNIVKIIKPIGYSA	26.5 37.749 18.092 64.369	16 36.625 16 59.29

Table 4.5 SUMOylation sites in Pigeonpea flowering gene protein sequences

The identified Pigeonpea flowering genes encoding proteins studied for possible presence of possible SUMOylation sites. As shown in table 4.4, presences of signature SUMOylation sites have been observed for the six identified proteins. Two signature SUMOylation sites were observed in CcTerminal flowering 1 whereas, three sites were observed in CcFrigida and Frigida Essential 1. Four signature SUMOylation sites were observed in CcFrigida like 1, CcFrigida like 2 and CcTerminal flowering 2. All the predicted SUMOylation sites have shown significantly high score than the cutoff values (Table 4.4).

Gene structure of Pigeonpea flowering related genes

In order to further understand flowering related genes in Pigeonpea, their gene structure have been studied. To achieve this, results for six selected genes from Augustus gene perdition of whole genome sequence of Pigeonpea have been considered as input. As shown in Figure 3, all the six selected genes showed presence of introns in their structure. Among six studied genes, CcFrigida Essential 1 showed presence of maximum number of introns (five introns), whereas CcFrigida, CcFrigida like 1, CcFrigida like 2 showed lowest two introns in their genomic structures. CcTerminal flowering 1 and CcTerminal flowering 2 showed three and twelve number of introns, respectively.

Phylogenetic analysis of Pigeonpea flowering related genes

To understand the evolutionary relationship of Pigeonpea flowering related genes, phylogenetic analysis performed using respective orthologs from other plant species including Arabidopsis, Maize (*Zea mays*), Brassica (Brassica rapa), Sesame (*Sesamum indicum*), chick pea (*Cicer arietinum*), soybean (*Glycine max*), and rice (*Oryza sativa*). Amino acid sequences of the respective genes were considered for phylogenetic analysis (Figure 4). As shown in Figure 4A, CcFrigida Essential 1 was placed close to soybean and chickpea Frigida Essential 1. Similarly, CcFrigida was placed along with respective orthologs from soybean and chickpea (Figure 4B). In case of Frigida like 1, CcFrigida like 1 and CaFrigida like 1 were grouped together where as their soybean orthologs was distantly placed (Figure 4C). However, Frigida like 2, CcFrigida like 2 was close to maize and rice orthologs than that of soybean and chickpea (Figure 4D). Similarly, CcTerminal flowering 1 was also found in the same clad than that of rice and maize (Figure 4E). However, CcTerminal flowering 2 was found close to their orthologs in chickpea and soybean (Figure 4F).

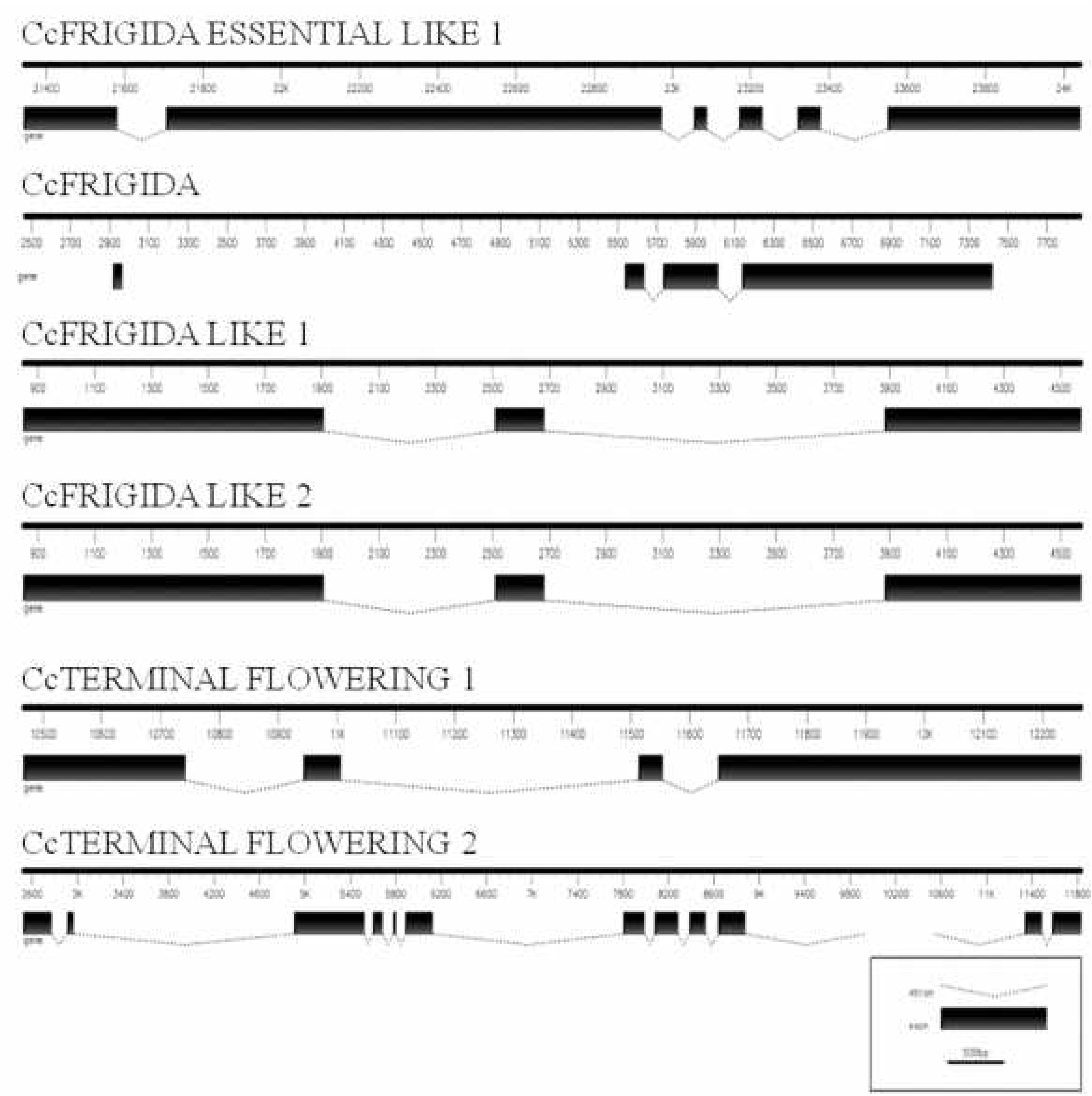


Figure 3: Genomic structure of Pigeonpea flowering genes

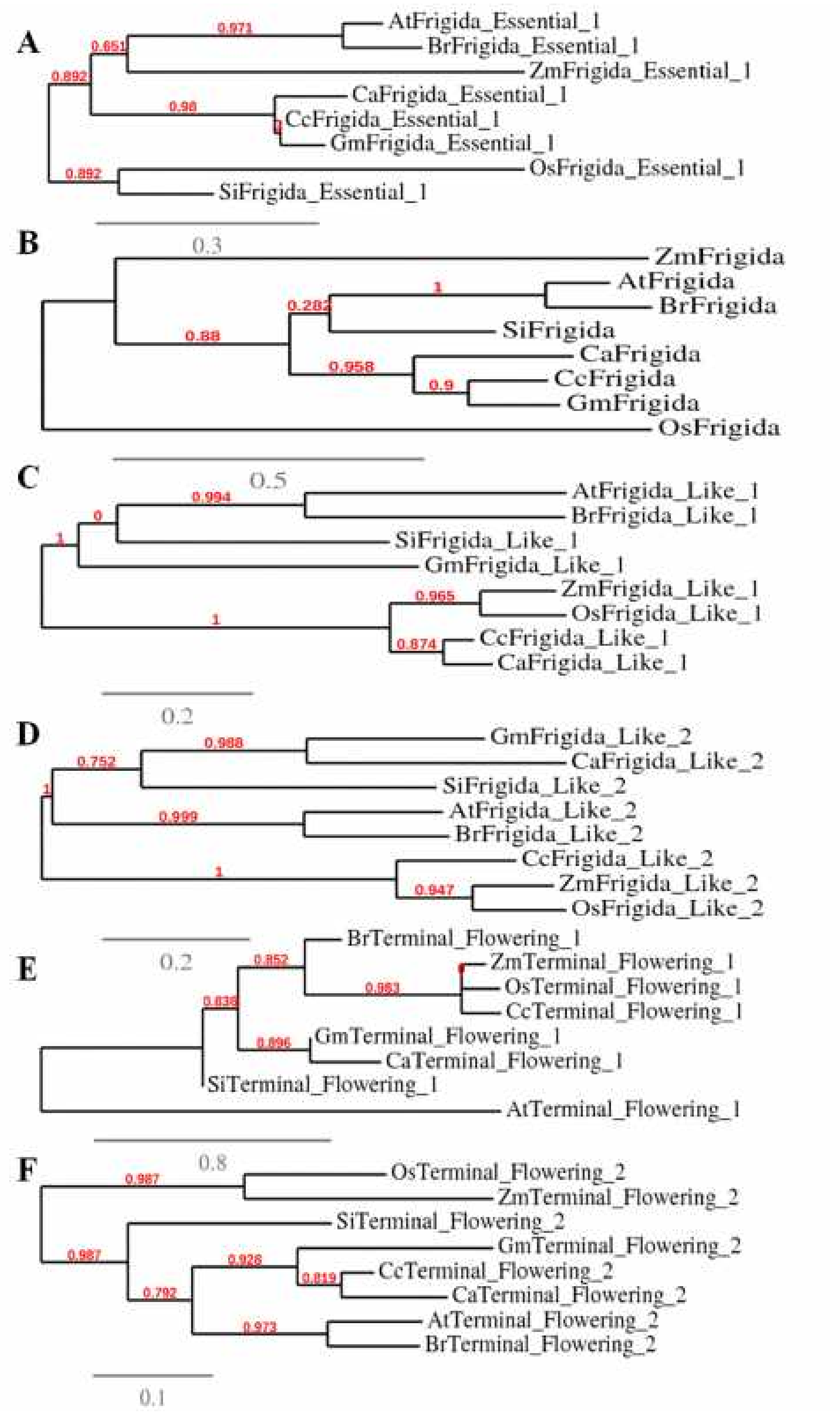


Figure 4: Phylogenetic analysis of Pigeonpea flowering genes with other plant species.

Identifying possible interacting partner of Pigeonpea flowering genes

In order to perform its function in vivo, proteins interact with other protein and such interactions play key role in growth and development of the organism. To get an overview of possible interacting partner of flowering genes of Pigeonpea, their orthologs in Arabidopsis were used in string protein database. As shown in Figure 5A, Frigida is found to interact with Early flowering, terminal flowering 1, Frigida essential 1 and leafy proteins. Frigida essential 1 showed interaction with terminal flowering 1, Frigida, TATA binding factor protein, and Frigida like 1 (Figure 5A). Similarly, Frigida like 1 showed interacting partners such as Frigida, Frigida essential 1, Mad box affecting flowering 1, Flowering locus etc. (Figure 5B). Frigida like 2 was found to interact with mad box affecting flowering 1, flowering locus c, and HUA2 (Figure 5B). Important interacting partners of Terminal flowering1 were Apetala1 and leafy, whereas Terminal flowering 2 showed interactions with LIF 2, ICU2, and EMF2 (Figure 5C). Further, using the string database conservation of selected flowering genes and their interacting partners have been studied. Results shows that the genes and their interacting partners are largely conserved among flowering plants (Figure 5C).

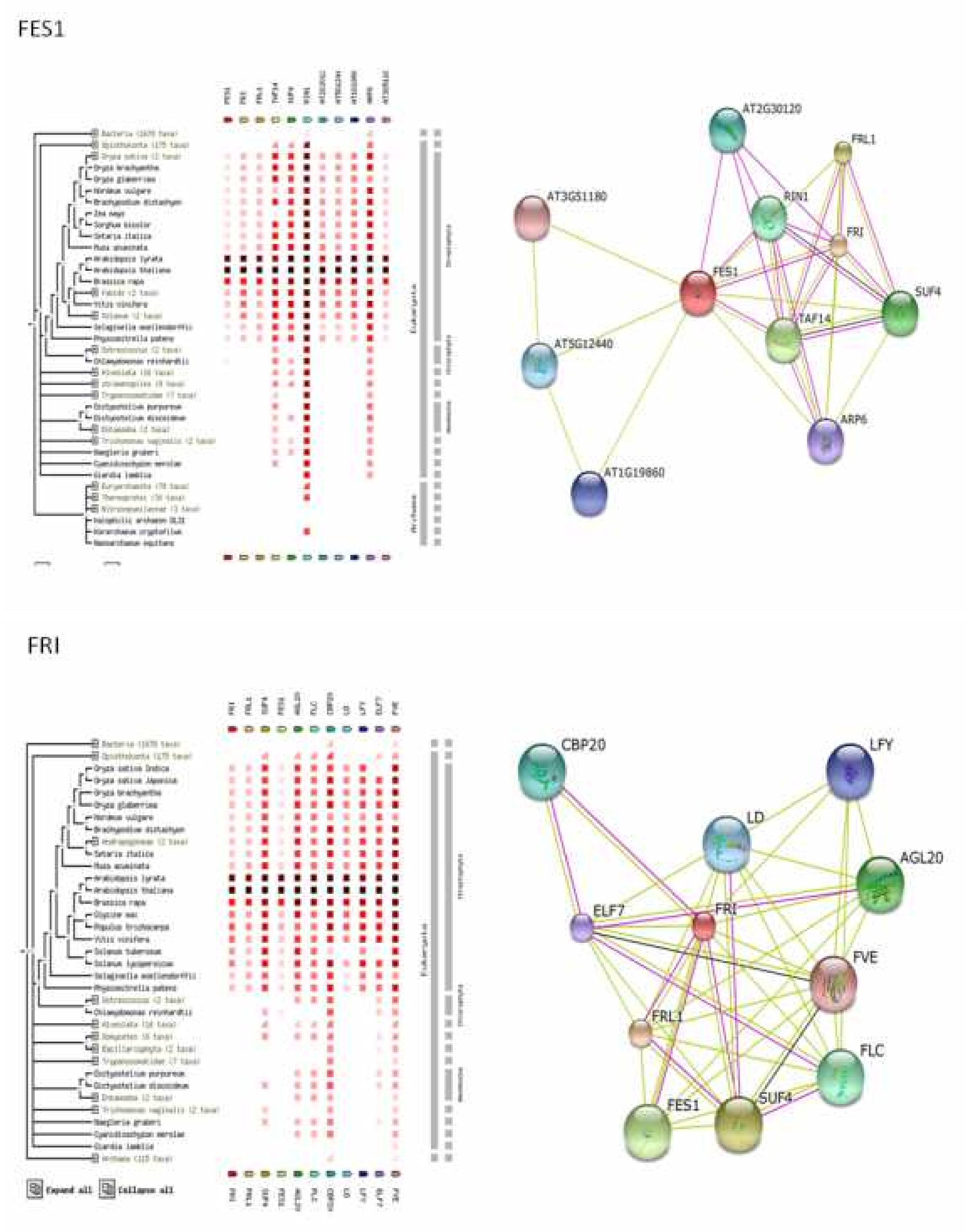


Figure 5A: Conservation of Frigida Essential 1 and Frigida across flowering plants and their functional partner in different organism as predicted by STRING database.

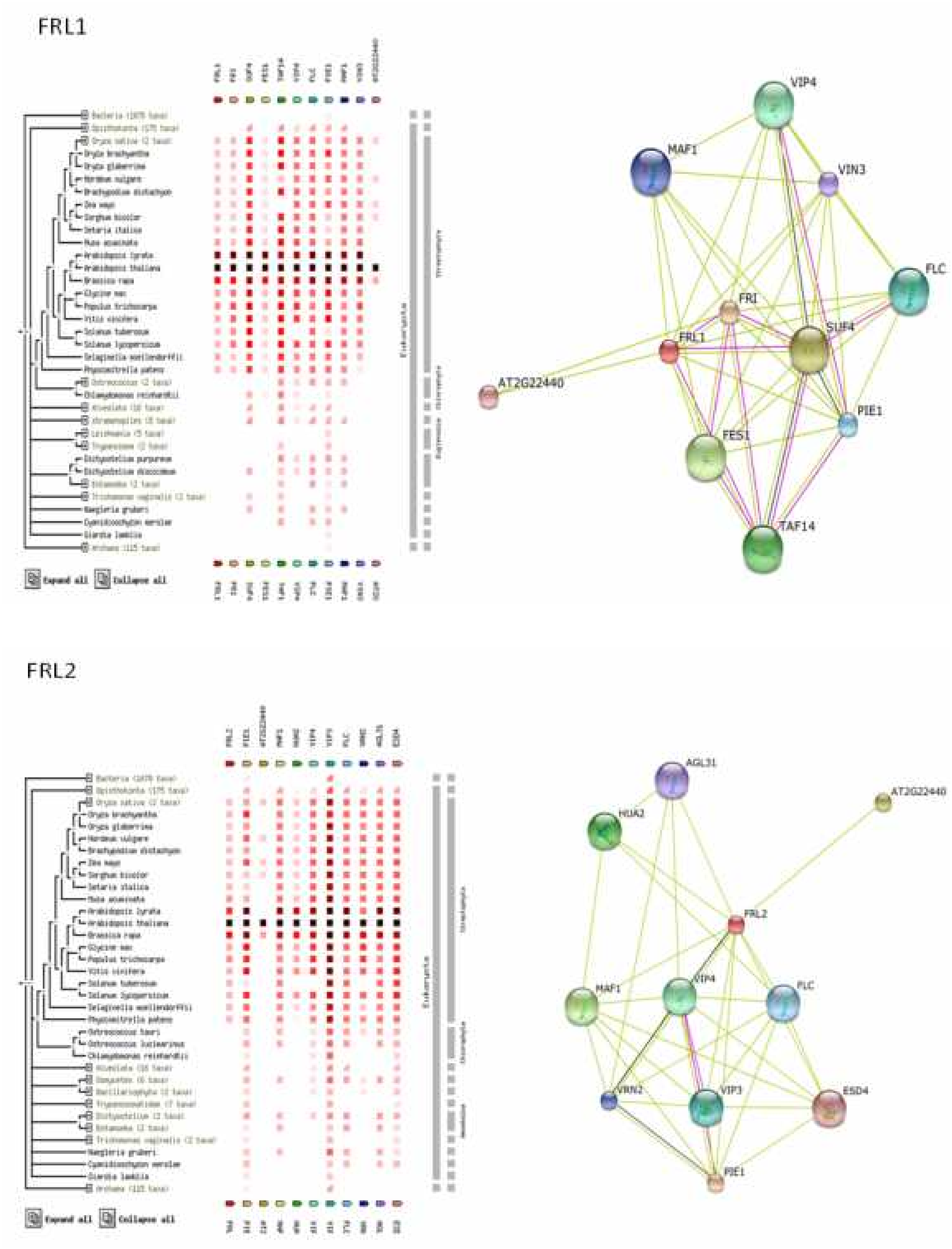
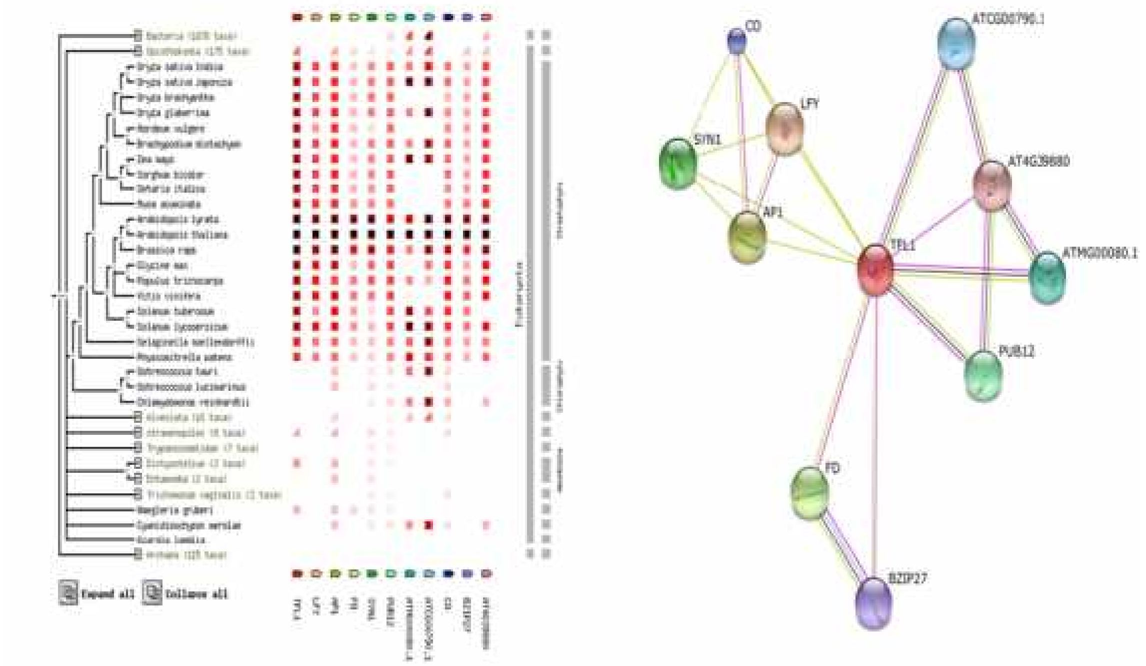


Figure 5B: Conservation of Frigida Like 1 and Frigida Like 2 across flowering plants and their functional partner in different organism as predicted by STRING database.

TFL1



TFL2

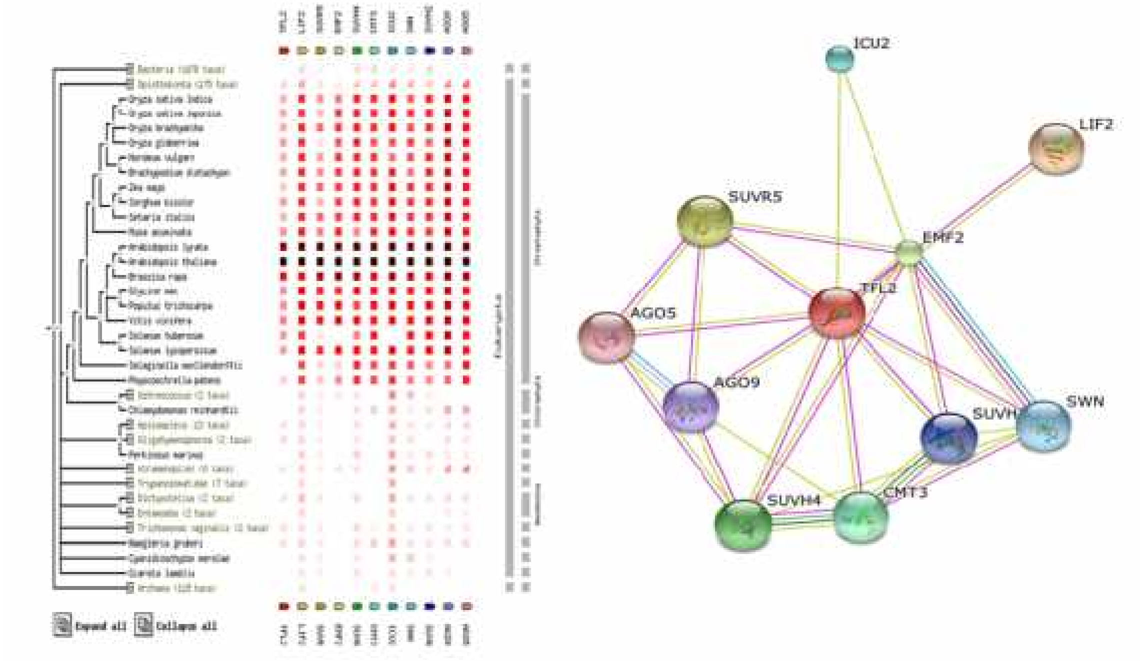


Figure 5C: Conservation of Terminal flowering 1 and Terminal flowering 2 across flowering plants and their functional partner in different organism as predicted by STRING database.

4.2 Cloning and characterization of ICE1 from Pigeonpea

In response to cold stress, plants show expression of cold-regulated (COR) genes which help to cope up cold stress. The COR genes harbor C-repeat at their promoter element and their expression is activated by C-repeat (CRT)-binding transcription factors (CBF). ICE1 (Inducer of CBF Expression 1) a MYC-type basic helix–loop–helix transcription factor is considered to be a central regulator of cold response as it is shown to regulate expression of CBF3 and downstream COR genes (Chinnusamy et al. 2003). In present works bioinformatics approaches have been used to understand ICE1 in temperate, tropical and subtropical plants. Methodology followed in the study has been depicted in Figure 6.

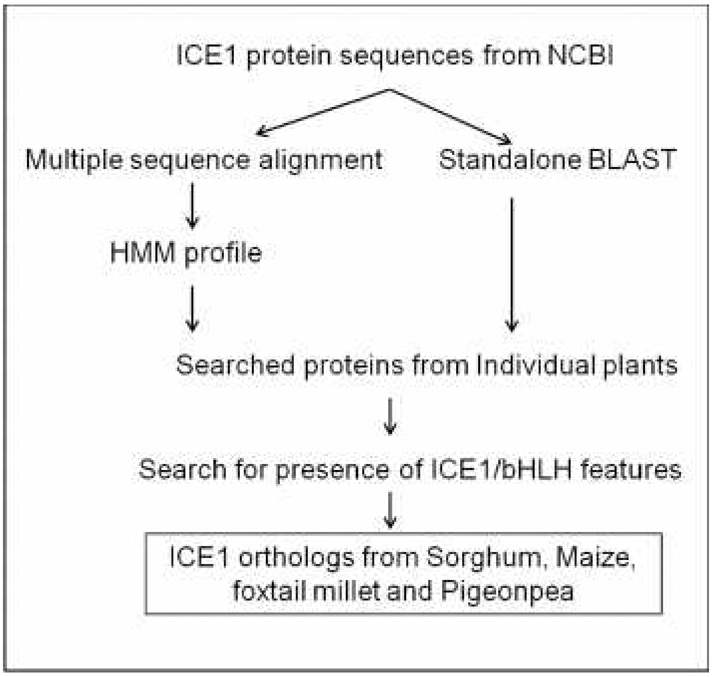


Figure 6: Flow diagram showing steps followed for identification of ICE1 orthologs in Sorghum, Pigeonpea, Foxtail millet and Maize.

ICE1 proteins from temperate (*Brassica rapa*, *Raphanus sativus*, *Arabidopsis thaliana*, *Solanum lycopersicum*, *Solanum tuberosum*, *Malus domestica*, *Triticum aestivum*, *Zea mays*, *Fragaria vesca*), subtropical (*Glycine max*, *Cucumis sativus*, *Cucumis melo)* and tropical crop plants *(Oryza sativa*, *Citrus sinensis*, *Gossypium arboretum*, *Setaria italic*, *Cajanus cajan*, *Sorghum bicolor*, *Ricinus communis*, *Theobroma cacao*) were retrieved from NCBI. HMM profile was build using ICE1 protein sequences from different plants. ICE protein sequences and HMM profiles were independently used as a query to search protein sequences of plants in question. In case of Pigeonpea total contigs, from whole genome sequence project were used for gene prediction program Augustus. Predicted protein sequence were searched employing standalone blast using ICE1 protein as query as well as using HMM profile. Resulting sequences were searched presence of BHLH features and homology to ICE1. Following these approaches putative ICE1 sequences were identified in *C. cajan, Z. mays, S. italica*, and *S. bicolor*. Multiple sequence alignment of ICE1 from 19 crop plants shows conservation in ICE1 in C-terminal regions whereas N-terminal regions show significant variations (Figure 7A). The identified ICE1 protein sequences showed-

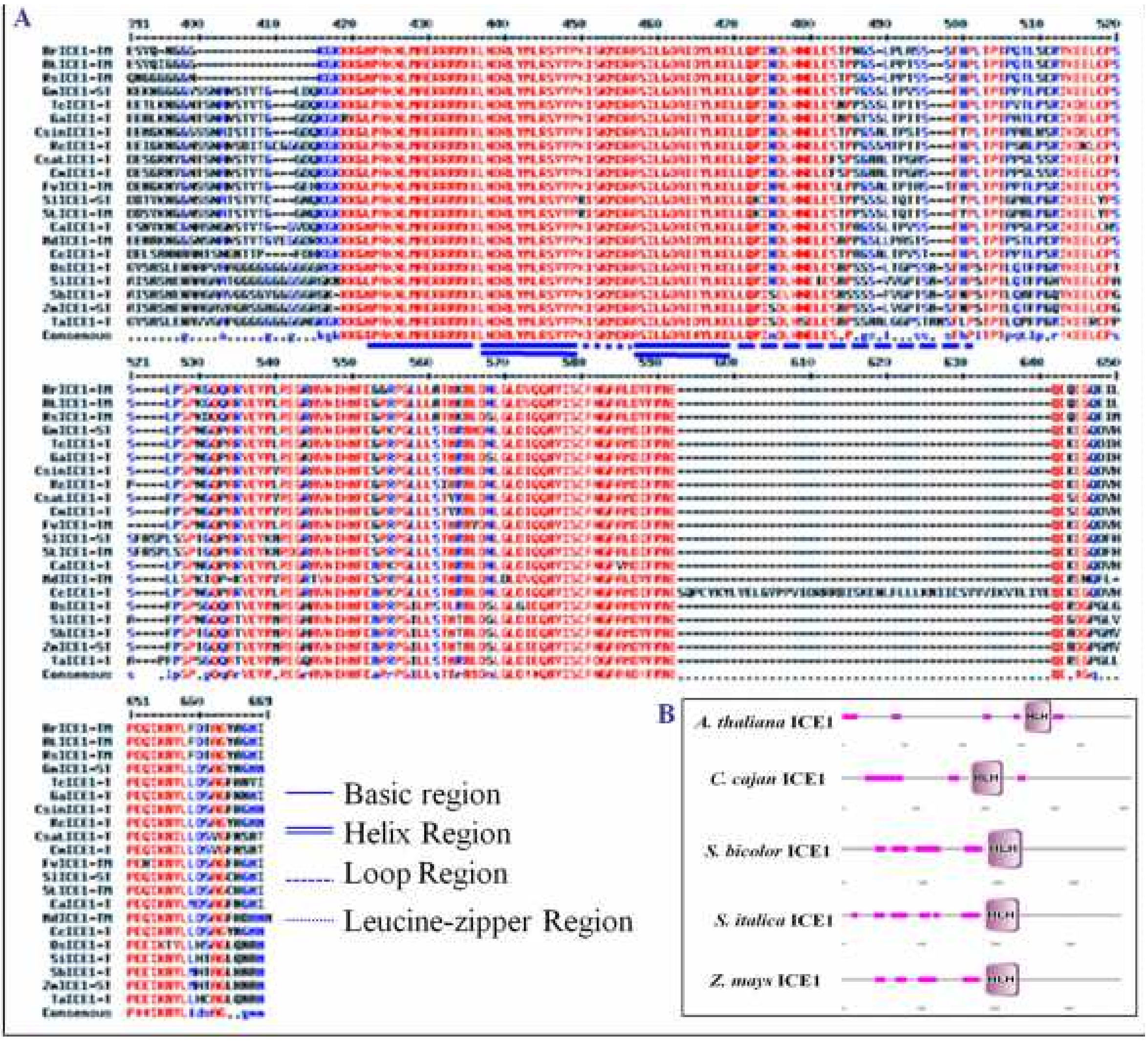


Figure 7. Amino acid sequence alignment of ICE1 from plants from different climatic conditions (A). T, ST, TM after sequence name indicate Tropical, Sub-tropical and Temperate climate, respectively. B. Domains within ICE1 identified from Pigeonpea, Sorghum, Foxtail millet and Maize. Pink bold lines indicate low complexity region, box with HLH indicates Helix-Loop-Helix domain.Signature Helix-Loop-Helix domain (Figure 7B).

Phylogenetic analysis of ICE1 from different plants shows grouping largely as per its lineage and climatic conditions *per se* have not shown any significant impact in groupings. However, apple (*M. domestica*), a temperate fruit tree has been clustered together with other temperate crops such as Brassica, Arabidopsis and Sugar beet

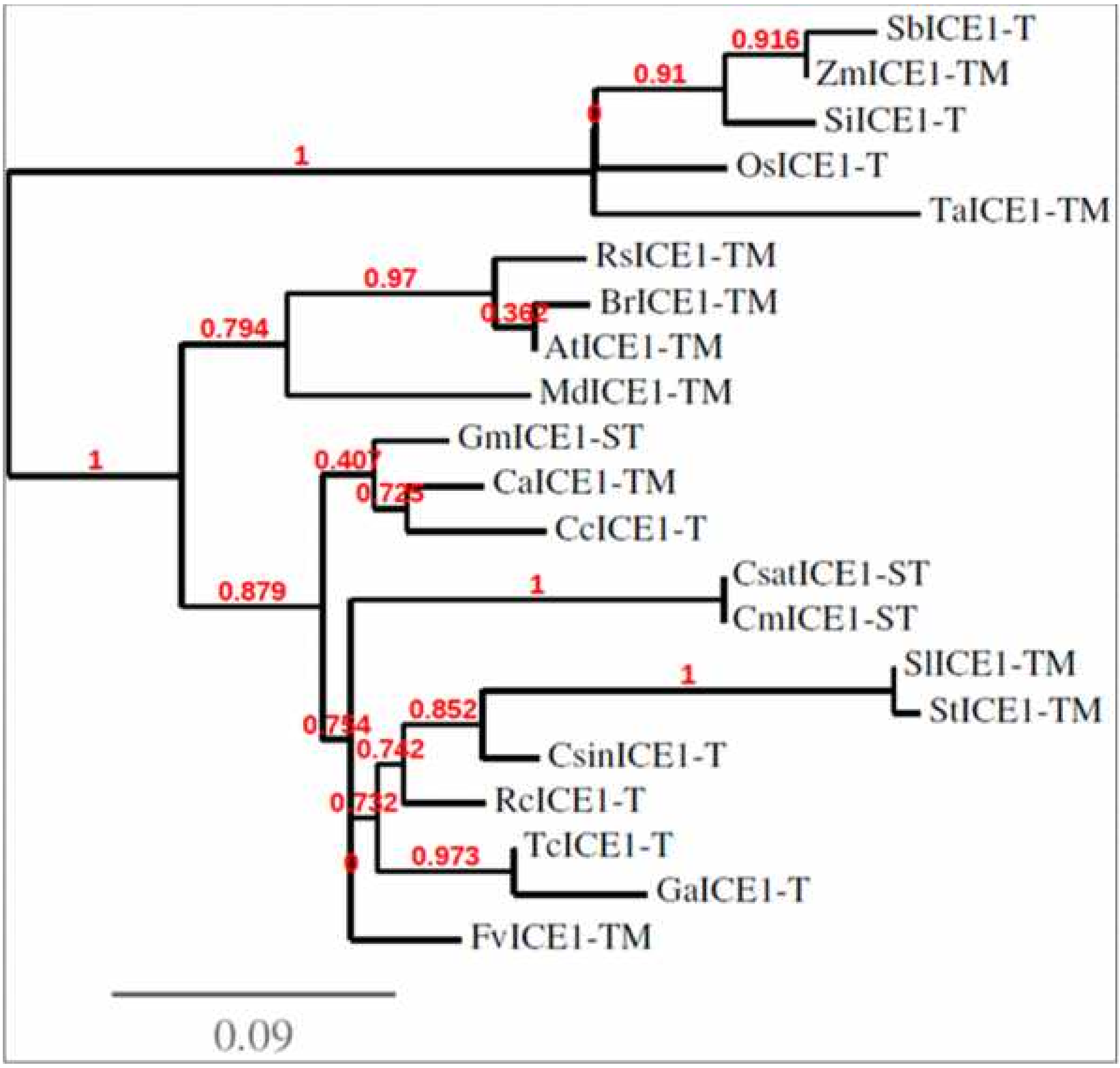
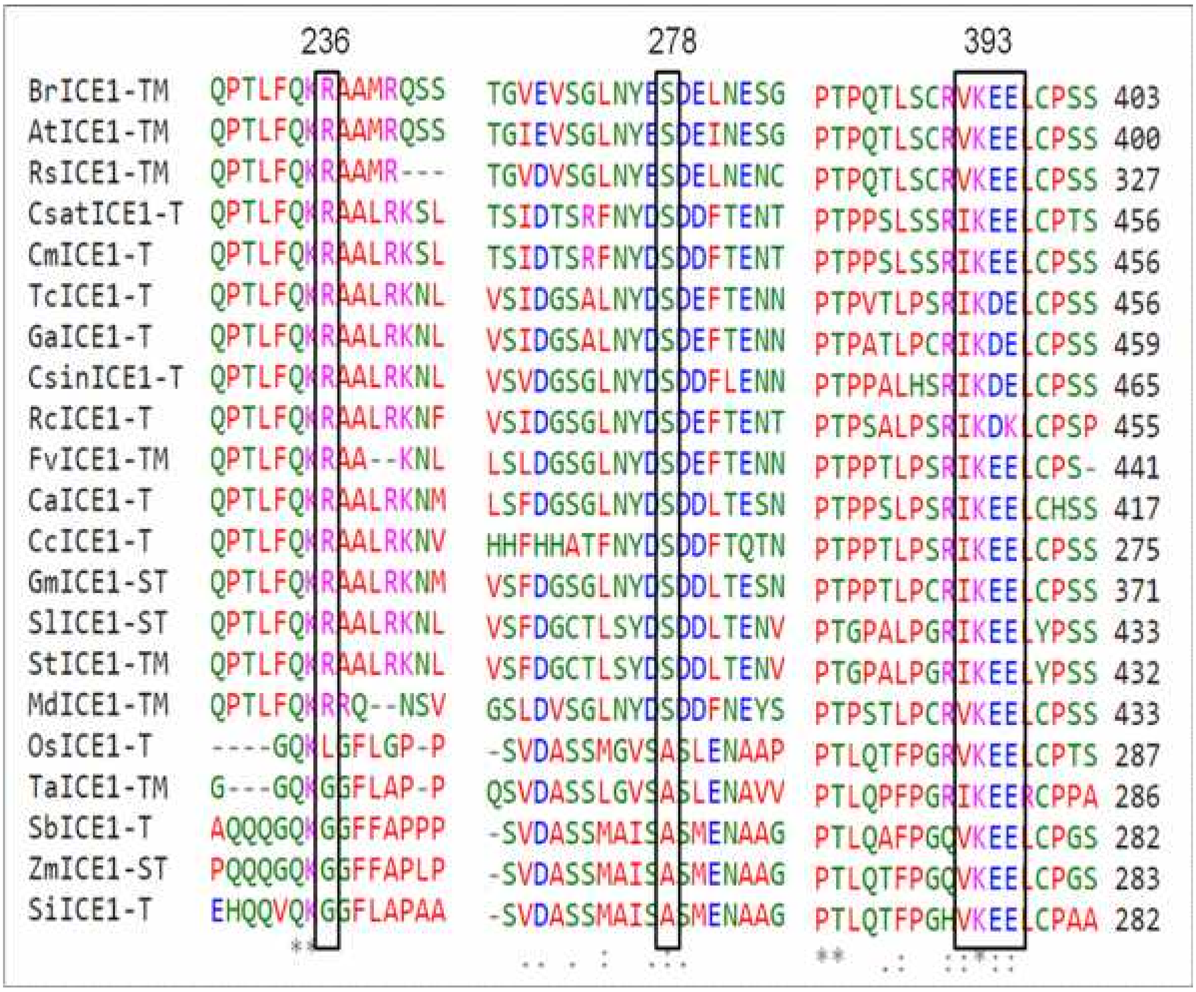


Figure 8. Phylogenetic relationship between ICE1 from different crop plants. Phylogenetic tree was built following maximum likelihood method using PhyML. Values at branch points indicate confidence index of approximate likelihood-ratio test (aLRT)

The SUMOylation site which is known in Arabidopsis is found conserved in all the studied plants, however, additional sites have also been identified in several plants. AtICE1 is known to be phosphorylated at Ser 278 by OST1 (Ding et al., 2015) and this site is conserved across temperate, tropical and subtropical plants, however, is not present in cereals including *O. sativa*, *T. aestivum*, *Z. mays*, *S. bicolor* and *S. italica* (Figure 9).



```
                         236                  278                      393
BrICE1-TM    QPTLFQKRAAMRQSS  TGVEVSGLNYESDELNESG  PTPQTLSCRVKEELCPSS 403
AtICE1-TM    QPTLFQKRAAMRQSS  TGIEVSGLNYESDEINESG  PTPQTLSCRVKEELCPSS 400
RsICE1-TM    QPTLFQKRAAMR---  TGVDVSGLNYESDELNENC  PTPQTLSCRVKEELCPSS 327
CsatICE1-T   QPTLFQKRAALRKSL  TSIDTSRFNYDSDDFTENT  PTPPSLSSRIKEELCPTS 456
CmICE1-T     QPTLFQKRAALRKSL  TSIDTSRFNYDSDDFTENT  PTPPSLSSRIKEELCPSS 456
TcICE1-T     QPTLFQKRAALRKNL  VSIDGSALNYDSDEFTENN  PTPVTLPSRIKDELCPSS 456
GaICE1-T     QPTLFQKRAALRKNL  VSIDGSALNYDSDEFTENN  PTPATLPCRIKDELCPSS 459
CsinICE1-T   QPTLFQKRAALRKNL  VSVDGSGLNYDSDDFLENN  PTPPALHSRIKDELCPSS 465
RcICE1-T     QPTLFQKRAALRKNF  VSIDGSGLNYDSDEFTENT  PTPSALPSRIKDKLCPSP 455
FvICE1-TM    QPTLFQKRAA--KNL  LSLDGSGLNYDSDEFTENN  PTPPTLPSRIKEELCPS- 441
CaICE1-T     QPTLFQKRAALRKNM  LSFDGSGLNYDSDDLTESN  PTPPSLPSRIKEELCHSS 417
CcICE1-T     QPTLFQKRAALRKNV  HHFHHATFNYDSDDFTQTN  PTPPTLPSRIKEELCPSS 275
GmICE1-ST    QPTLFQKRAALRKNM  VSFDGSGLNYDSDDLTESN  PTPPTLPCRIKEELCPSS 371
SlICE1-ST    QPTLFQKRAALRKNL  VSFDGCTLSYDSDDLTENV  PTGPALPGRIKEELYPSS 433
StICE1-TM    QPTLFQKRAALRKNL  VSFDGCTLSYDSDDLTENV  PTGPALPGRIKEELYPSS 432
MdICE1-TM    QPTLFQKRRQ--NSV  GSLDVSGLNYDSDDFNEYS  PTPSTLPCRVKEELCPSS 433
OsICE1-T     ----GQKLGFLGP-P  -SVDASSMGVSASLENAAP  PTLQTFPGRVKEELCPTS 287
TaICE1-TM    G---GQKGGFLAP-P  QSVDASSLGVSASLENAVV  PTLQPFPGRIKEERCPPA 286
SbICE1-T     AQQQGQKGGFFAPPP  -SVDASSMAISASMENAAG  PTLQAFPGQVKEELCPGS 282
ZmICE1-ST    PQQQGQKGGFFAPLP  -SVDASSMAISASMENAAG  PTLQTFPGQVKEELCPGS 283
SiICE1-T     EHQQVQKGGFLAPAA  -SVDASSMAISASMENAAG  PTLQTFPGHVKEELCPAA 282
```

Figure 9: Sequence conservation/variation at important motifs in ICE1 protein from different plants. Numbers mentioned above sequence alignment indicate important amino acid positions with reference to Arabidopsis ICE1. Amino acid, Arginine (R) on 236th position is crucial for ICE1 activity in cold, Serine (S) on 278th position is known phosphorylation site, Lysine (K) on 393rd position is known SUMOylation site.

Prediction of sub-cellular localization of ICE1 from most of the studied plant species showed localization in nucleus, however, ICE1 from *Triticum aestivum, Zea mays, Setaria italica,* and *Sorghum bicolor* showed significant chloroplast as well as nuclear localisation signal suggesting novel additional function in these plant species (Table 4.5).

PLANTS	**AMINO ACID**	**PI**	**Size (Dalton)**	**LOCALIZATION**
Brassica rapa	497	5.11	53049.47	Nuclear-3.866
Raphanus sativus	421	5.41	45795.48	Nuclear-3.474
Arabidopsis thaliana	494	5.31	53538.93	Nuclear-4.004
Solanum lycopersicum	531	5.38	58337.14	Nuclear-4.232
Solanum tuberosum	530	5.31	57936.63	Nuclear-3.82
Malus domestica	526	5.09	57345.52	Nuclear-4.419
Triticum aestivum	381	5.04	39435.30	Nuclear-1.903 Chloroplast-2.331
Zea may	377	4.77	39444.35	Nuclear-1.823 Chloroplast-1.889
Fragaria vesca	535	5.81	58210.02	Nuclear-3.839
Glycine max	465	6.08	51333.91	Nuclear-4.373
Cucumis sativus	550	5.60	60123.20	Nuclear-3.989
Cucumis melo	550	5.60	60174.25	Nuclear-3.92
Oryza sativa	381	5.76	39776.83	Nuclear-3.06
Citrus sinensis	559	5.27	61213.05	Nuclear-3.879
Gossypium arboreum	553	6.04	60834.77	Nuclear-4.273
Setaria italica	376	5.12	39175.05	Nuclear-1.687 Chloroplast-1.976
Cajanus cajan	418	9.24	46309.93	Nuclear-3.864
Sorghum bicolor	376	4.73	39435.30	Nuclear-2.256 Chloroplast-1.722
Ricinus communis	549	5.69	60114.38	Nuclear-4.068
Theobroma cacao	550	5.73	60736.76	Nuclear-4.186

Table 4.6: Important features of ICE1 from plants of different climatic regions

Further, interactions of Arabidopsis ICE1 with other proteins and their conservation/divergence across plant kingdom have been studies using string database. As shown in Figure 10, ICE1 and its key interacting partners have been observed across plant kingdom.

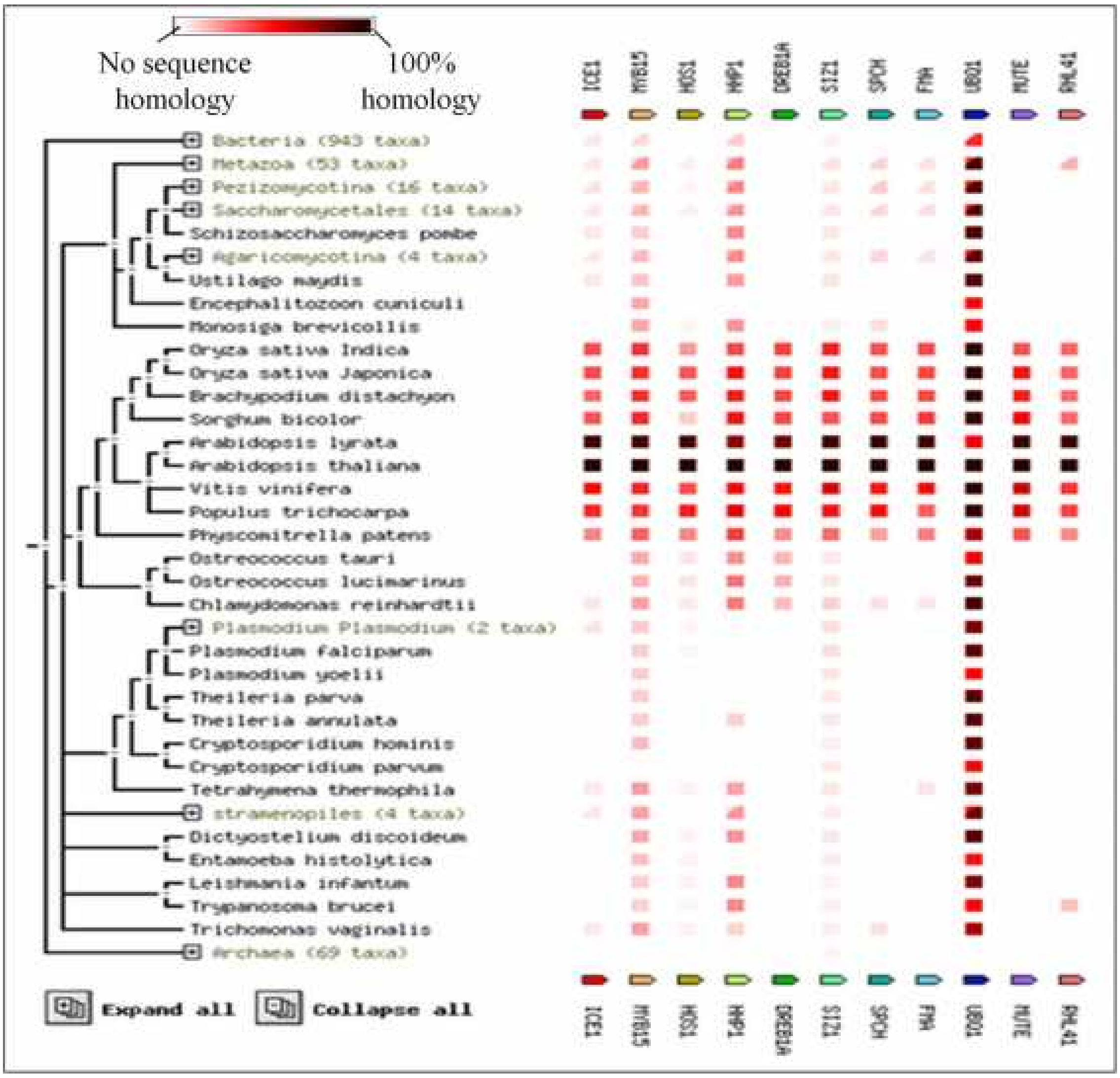


Figure 10: Conservation of ICE1 and its functional partner in different organism as predicted based on STRING database

Expression analysis of flowering genes in Pigeonpea

Expression of flowering genes were studied using pigeonpea SRA (Sequence Read Archives) data available at SRA archives at NCBI. RNAseq data was available for two different cultivars, Asha and Upas120 from pooled RNA of leaf, stem and root tissues. Expression analysis was performed using CLC genomics workbench. As shown in figure 11, all the genes showed expression of all but one gene, CcTFL1.

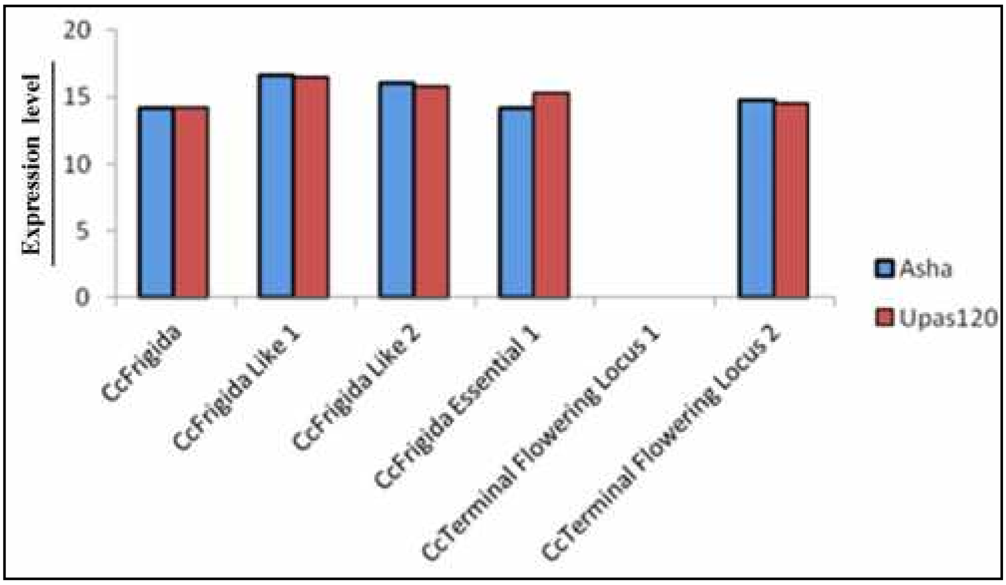


Figure 11: Expression analysis of flowering related genes in two cultivars of Pigeonpea. Expression represents log2 of RPKM value for SRA data.

DNA Isolation from different pigeonpea verities

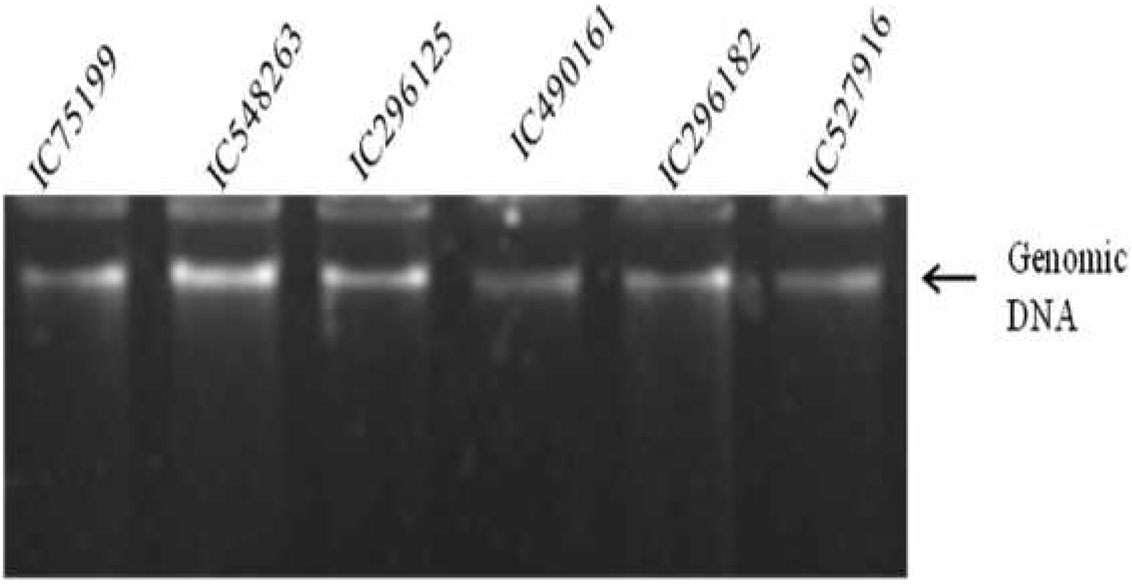


Figure 12: DNA was isolated from Pigeonpea varieties and checked 1.5% agarose gel (Figure No 12). This will be used for PCR amplification of identified flowering genes.

CHAPTER V

DISCUSSION

DISCUSSION

In silico characterization of flowering genes from Pigeonpea

Flowering is one of the most important traits of plant growth and has major impact on crop yield in agriculture. This feature has achieved further importance in changing climatic conditions. In present work attempts have been made to identify flowering related genes from Pigeonpea. The identified Pigeonpea (*Cajanus cajan*) genes were *CcFrigida, CcFrigida Like1, CcFrigida Like2, CcFrigida Essential1, CcTerminal Flowering1,* and *CcTerminal Flowering2*. The selection of candidate genes during in the present study was based on the previous information (Roux et al. 2006). These genes have been shown to play significant role in flowering trait regulation in several plants. All the identified genes showed presence of important features which have been known for the selected genes in questions. Further, there exist variations in Pigeonpea flowering genes to that of model plant Arabidopsis showing explaining phylogenetic divergence and distinct nature of flowering. Expression analysis using SRA data clearly suggest that the identified genes are genuine and transcriptionally active. Low/no expression of a few genes could be due to their specific temporal as well as spatial expression and the tissue used for transcriptome sequencing might be distinct to what these showed expression.

Importance of the considered genes in flowering trait has already been established. There are several loss-of-function mutation in FRIGIDA confer early flowering. Genetic approaches have found that FRIGIDA (FRI) gene as the main promoter of the early flowering (Werner et al. 2005). A locus, FRIGIDA-ESSENTIAL 1 (FES1), which like FRI is specifically required for the up regulation of FLC expression. Terminal flowering locus has been shown to control flowering time in Arabidopsis as well as several other plants (Liu et al. 2010; Tian et al. 2010). Other than model plants, Arabidodbs different TFL1 homolog’s control two distinct aspects of plant development in pea, whereas a single gene, TFL1, performs both functions in Arabidopsis (Fabrice Foucher, 2003).

Flowering and determinacy have also been studied in Pigeonpea. CcTFL1 was shown to be an important candidate gene for determinacy in Pigeonpea through candidate gene sequencing analysis, mapping, QTL analysis together with comparative genomics and expression profiling. (Reyazul Rouf Mir, 2014). In pigeonpea, the occurrence of sequence variability in terms of number of SNPs, SNP frequency, nucleotide diversity and number of haplotypes among seven candidate genes strongly indicate the occurrence of different evolutionary constraints. With the availability of genome sequence of several crop plant, and advances in bioinformatics analysis opens new avenues to decipher the genetic elements involved in the control of complex traits such as flowering.

ICE1 and regulation of cold stress response

Cold temperatures lead to transcription of the CBF genes transcription factors, which in turn activate the transcription of genes containing the DRE/CRT promoter element (Chinnusamy et al. 2003). In present, ICE1 have been identified in five crop plants (Pigeonpea, maize, sorghum and foxtail millets) for the first time. The identified genes showed conservation of important feature as well as variations in few plants especially in cereals. Although there was no specific features restricted to plants from climatic zone (tropical, sub-tropical and temperate), variations such as presence/absence of phosphorylation sites does exist in cereals and dicots. Overall, present work shows that cold stress regulation by ICE1 is likely to be conserved across temperate, tropical as well as sub-tropical crop plants as there are no climate specific features observed. However, variations in ICE1 protein sequences in cereals at important motifs/sites suggest possible involvement of distinct cold signalling pathways in cereals towards regulation of ICE1. Further, these distinct features also hint at novel functions of ICE1 in cereals, which await to be unravelled.

CHAPTER VI

CONCLUSION AND SUMMARY

CHAPTER VI

CONCLUSION AND SUMMARY

Conclusion: In this study, six flowering related genes have been identified from Pigeonpea for the first time and studied for their important properties and expression pattern. This work would pave way for cloning of these genes from Pigeonpea and detailed characterization to identify their specific role in flowering regulation in this important pulse crop. In-silico identification of ICE1 and Pigeonpea, sorghum, foxtail millet and maize will also help molecular biology to clone this gene from respective crops and elaborate. This will further, help understanding on role of ICE1 and cold stress regulations in different crop plants.

Summary

In the present work, six flowering related genes from Pigeonpea have been identified using bioinformatics approaches. Identified genes were putative FRIGIDA, FRIGIDA LIKE 1, FRIGIDA LIKE 2, FRIGIDA ESSENTIAL 1, TERMINAL FLOWERING 1, and TERMINAL FLOWERING 2 which have been shown to play important role in flowering regulation in Arabidopsis. In-silico characterizations of identified genes were also performed along with phylogenetic analysis. Further, most of the identified genes show active transcription as studied from the RNAseq expression analysis in tow varieties of Pigeonpea. Additionally, ICE1 (Inducer of CBF Expression 1), an important regulator of cold stress have also been identified from Pigeonpea, sorghum, foxtail millet and maize. ICE1 from 18 plants were studied for climate specific variations. Although no climate specific variation observed, the present work showed important variation between cereals and other crop plants.

References

Ahsan R, Islam M. *In vitro* antibacterial screening and toxicological study of some useful plants (*Cajanus Cajan*) Euro J Sci Res. 2009; 41:227–32.

Amaya, I., Ratcliffe, O.J., and Bradley, D.J. (1999). Expression of *CENTRORADIALIS* (*CEN*) and *CEN*-like genes in tobacco reveals a conserved mechanism controlling phase change in diverse species. Plant Cell 11, 1405–1418.

Ambasta SP. 4th ed. New Delhi: National Institute of Science Communication; 2004. The useful plants of India; pp. 94–5.

Ariyanayagam, R.P., A.N. Rao and P.P. Zaveri (1995) Cytoplasmic-genic male-sterility in interspecific mating of *Cajanus*. *Crop Science* 35: 981-985.

Banfield, M.J., and Brady, R.L. (2000). The structure of *Antirrhinum* CENTRORADIALIS protein (CEN) suggests a role as a kinase regulator. J. Mol. Biol. 297, 1159–1170.

Benedict, C. et al. (2006) Consensus by democracy. Using meta-analyses of microarray and genomic data to model the cold acclimation signalling pathway in Arabidopsis. Plant Physiol. 141, 1219–1232.

Bradley, D., Carpenter, R., Copsey, L., Vincent, C., Rothstein, S., and Coen, E. (1996). Control of inflorescence architecture in *Antirrhinum*. Nature 379, 791–797.

Bradley, D., Ratcliffe, O., Vincent, C., Carpenter, R., and Coen, E. (1997). Inflorescence commitment and architecture in Arabidopsis. Science 275, 80–83.

Bradley, D., Ratcliffe, O., Vincent, C., Carpenter, R., and Coen, E. (1997). Inflorescence commitment and architecture in Arabidopsis. Science 275, 80–83.

Chinnusamy, V. et al. (2006) Gene regulation during cold acclimation in plants. Physiol. Plant. 126, 52–61.

Cronk, Q.C.B. (2001). Plant evolution and development in a post-genomic context. Nat. Rev. Genet. 2**,** 607–619.

Dong, C.H. et al. (2006) A putative Arabidopsis nucleoporin AtNUP160 is critical for RNA export and required for plant tolerance to cold stress. Mol. Cell. Biol. 26, 9533–9543.

Fabrice Foucher, Julie Morin, Juliette Courtiade, Sandrine Cadioux, Noel Ellis, Mark J. Banfield, and Catherine Rameau (2003). *DETERMINATE* and *LATE FLOWERING*

Force, A., Lynch, M., Pickett, F.B., Amores, A., Yan, Y.-l., and Postlethwait, J. (1999). Preservation of duplicate genes by complementary, degenerative mutations. Genetics 151, 1531–1545.

Foucher, F. et al. (2003) DETERMINATE and LATE FLOWERING are two TERMINAL FLOWER1/CENTRORADIALIS homologs that control two distinct phases of flowering initiation and development in pea. Plant Cell 15, 2742–2754.

Gene FRIGIDA.

Hannah, M.A. et al. (2005) A global survey of gene regulation during cold acclimation in Arabidopsis thaliana. PLoS Genet 1, e26.

He, Y. et al. (2004) PAF1 complex-mediated histone methylation of FLOWERING LOCUS C chromatin is required for the vernalizationresponsive, winter-annual habit in Arabidopsis. Genes Dev. 18, 2774–2784.

Hecht, V. et al. (2005) Conservation of Arabidopsis flowering genes in model legumes. Plant Physiol. 137, 1420–1434.

Hofer, J., and Ellis, N. (2002). Conservation and diversification of gene function in plant development. Curr. Opin. Plant Biol. 5, 56–61.

Jensen, C.S., Salchert, K., and Nielsen, K.K. (2001). A *TERMINAL FLOWER1*-like gene from perennial ryegrass involved in floral transition and axillary meristem identity. Plant Physiol. 125, 1517–1528.

Johanson, U., West, J., Lister, C., Michaels, S., Amasino, R. & Dean, C. (2000) *Science* 290, 344–347.

John R. Stinchcombe, Cynthia Weinig, Mark Ungerer, Kenneth M. Olsen, Charlotte Mays.

Kardailsky, I., Shukla, V.K., Ahn, J.H., Dagenais, N., Christensen, S.K., Nguyen, J.T., Chory, J., Harrison, M.J., and Weigel, D. (1999). Activation tagging of the floral inducer *FT*. Science 286, 1962–1965.

Kobayashi, Y., Kaya, H., Goto, K., Iwabuchi, M., and Araki, T. (1999). A pair of related genes with antagonistic roles in mediating flowering signals. Science 286, 1960–1962.

Komeda, Y. (2004) Genetic regulation of time to flower in Arabidopsis thaliana. Annu. Rev. Plant Biol. 55, 521–535.

Kwak M, Velasco DM, Gepts P (2008) Mapping homologous sequences for determinacy and photoperiod sensitivity in common bean (*Phaseolus vulgaris*). J Hered 99:283–291.

Le Corre, V. (2005) Variation at two flowering time genes within and among populations of Arabidopsis thaliana: comparison with markers and traits. Mol. Ecol. 14, 4181–4192.

Le Corre, V., Roux, F. & Reboud, X. (2002) *Mol. Biol. Evol.* 19**,** 1261–1271. How to be early flowering: an evolutionary perspective.

Lee, H. et al. (2001) The Arabidopsis HOS1 gene negatively regulates cold signal transduction and encodes a RING finger protein that displays cold-regulated nucleo-cytoplasmic partitioning. Genes Dev. 15, 912–924.

Lee, J.Y. and Lee, D.H. (2003) Use of serial analysis of gene expression technology to reveal changes in gene expression in Arabidopsis pollen undergoing cold stress. Plant Physiol. 132, 517–529.

Liu B, Watanabe S, Uchiyama T, Kong F, Kanazawa A, Xia Z, Nagamatsu A, Arai M, Yamada T, Kitamura K, Masuta C, Harada K, Abe J (2010) The soybean stem growth habit gene *Dt1* is an ortholog of *Arabidopsis* terminal flower. Plant Physiol 153:198–210.

Liu B, Watanabe S, Uchiyama T, Kong F, Kanazawa A, Xia Z, Nagamatsu A, Arai M, Yamada T, Kitamura K, Masuta C, Harada K, Abe J (2010) The soybean stem growth habit gene *Dt1* is an ortholog of *Arabidopsis* terminal flower. Plant Physiol153:198–210.

Ma, H., and De Pamphilis, C. (2000). The ABCs of floral evolution. Cell 101**,** 5–8.

Maruyama, K. et al. (2004) Identification of cold-inducible downstream genes of the Arabidopsis DREB1A/CBF3 transcriptional factor using two microarray systems. Plant J. 38, 982–993.

Michaels, S.D. et al. (2004) FRIGIDA-related genes are required for the winter-annual habit in Arabidopsis. Proc. Natl. Acad. Sci. U. S. A. 101, 3281–3285.

Mimida, N., Goto, K., Kobayashi, Y., Araki, T., Ahn, J.H., Weigel, D., Murata, M., Motoyoshi, F., and Sakamoto, W. (2001). Functional divergence of the *TFL1*-like gene family in Arabidopsis revealed by characterization of a novel homologue. Genes Cells 6.

Mir RR, Kumar N, Jaiswal V, Girdharwal N, Prasad M, Balyan HS, Gupta PK (2012b) Genetic dissection of grain weight (GW) in bread wheat through QTL interval and association mapping. Mol Breed 29:963–972.

Mir RR, Saxena RK, Saxena KB, Upadhyaya HD, Kilian A, Cook DR, Varshney RK (2012a) Whole-genome scanning for mapping determinacy in pigeonpea (*Cajanus* spp.). Plant Breed 132:472–478.

Mir RR, Saxena RK, Saxena KB, Upadhyaya HD, Kilian A, Cook DR, Varshney RK (2012a) Whole-genome scanning for mapping determinacy in pigeonpea (*Cajanus* spp.). Plant Breed 132:472–478.

Miura, K. et al. (2007) SIZ1-mediated sumoylation of ICE1 controls CBF3/DREB1A expression and freezing tolerance in Arabidopsis. Plant Cell 19, 1403–1414.

Morgante, M. and Salamini, F. (2003) from plant genomics to breeding practice. Curr. Opin. Biotechnol. 14, 214–219.

Mula, M. G. and K.B. Saxena (2010) *Lifting the Level of Awareness on Pigeonpea – a Global Perspective.* ICRISAT, Patancheru, Andhra Pradesh, India.

Noh, Y.S. and Amasino, R.M. (2003) PIE1, an ISWI family gene, is required for FLC activation and floral repression in Arabidopsis. Plant Cell 15, 1671–1682

Orvar, B.L. et al. (2000) early steps in cold sensing by plant cells: the role of actin cytoskeleton and membrane fluidity. Plant J. 23, 785–794.

Pal D, Mishra P, Sachan N, Ghosh AK. J Adv Pharm Technol Res. 2011 Oct; 2(4):207-14.

Palusa, S.G. et al. (2007) Alternative splicing of pre-mRNAs of Arabidopsis serine/arginine-rich proteins: regulation by hormones and stresses. Plant J. 49, 1091–1107.

Pnueli, L., Carmel-Goren, L., Hareven, D., Gutfinger, T., Alvarez, J., Ganal, M., Zamir, D., and Lifschitz, E. (1998). The *SELF-PRUNING* gene of tomato regulates vegetative to reproductive switching of sympodial meristems and is the ortholog of *CEN* and *TFL1*. Development 125, 1979–1989.

Pnueli, L., Gutfinger, T., Hareven, D., Ben-Naim, O., Ron, N., Adir, N., and Lifschitz, E. (2001). Tomato SP-interacting proteins define a conserved signalling system that regulates shoot architecture and flowering. Plant Cell 13, 2687–2702.

Ratcliffe, O.J., Amaya, I., Vincent, C.A., Rothstein, S., Carpenter, R., Coen, E.S., and Bradley, D.J. (1998). A common mechanism controls the life cycle and architecture of plants. Development 125, 1609–1615.

Repinski SL, Kwak M, Gepts P (2012) the common bean growth habit gene *PvTFL1y* is a functional homolog of *Arabidopsis TFL1*. Theor Appl Genet 124:1539–1547.

Reyazul Rouf Mir · Himabindu Kudapa · Sandhya Srikanth · Rachit K. Saxena · Ashutosh Sharma · Sarwar Azam · Kulbhushan Saxena · R. Varma Penmetsa · Rajeev K. Varshney (2014). Candidate gene analysis for determinacy in pigeonpea (*Cajanus* spp.)

Simpson, G.G. and Dean, C. (2002) Arabidopsis, the Rosetta stone of flowering time? Science 296, 285–289

Singh, A.K., N. Singh, S.P. Singh, N.B. Singh and J. Smartt (2006) Pigeonpea In: *Plant Genetic Resources: Food Crops* (eds.) B.S. Dhillon, S. Saxena, A. Agrawal and R.K. Tyagi, Narosa Publishing House, NBPGR, New Delhi, India, pp 223-239.

Solveig S. Halldorsdottir, Michael D. Purugganan, and Johanna Schmitt (2014). A latitudinal cline in flowering time in Arabidopsis thaliana modulated by the flowering time

Thomashow, M.F. (1999) Plant cold acclimation: freezing tolerance genes and regulatory mechanisms. Annu. Rev. Plant Physiol. Plant Mol. Biol. 50, 571–599.

Tian Z, Wang X, Lee R, Li Y, Specht JE, Nelson RL, McClean PE, Qiu L, Ma J (2010) Artificial selection for determinate growth habit in soybean. Proc Natl Acad Sci USA 19:8563–8568.

Two *TERMINAL FLOWER1/CENTRORADIALIS* Homologs That Control Two Distinct Phases of Flowering Initiation and Development in Pea.

Ulrich, H.D. (2005) Mutual interactions between the SUMO and ubiquitin systems: A plea of no contest. Trends Cell Biol. 15, 525–532.

Viswanathan Chinnusamy, Jianhua Zhu and Jian-Kang Zhu (2003).Cold stress regulation of gene expression in plants.

Weinig, C. and Schmitt, J. (2004) Environmental effects on the expression of quantitative trait loci and implications for phenotypic evolution. BioScience 54, 627–635.

Werner, J.D. et al. (2005) FRIGIDA-independent variation in flowering time of natural A. thaliana accessions. Genetics 170, 1197–1207.

Xiong, L. et al. (2001) The Arabidopsis LOS5/ABA3 locus encodes a molybdenum cofactor sulfurase and modulates cold stress- and osmotic stress–responsive gene expression. Plant Cell 13, 2063–2083.

Yamaguchi-Shinozaki, K. and Shinozaki, K. (2006) Transcriptional regulatory networks in cellular responses and tolerance to dehydration and cold stresses. Annu. Rev. Plant Biol. 57, 781–803.

Yan, L. et al. (2003) Positional cloning of the wheat vernalization gene VRN1. Proc. Natl. Acad. Sci. U. S. A. 100, 6263–6268.

Zhang, H. and van Nocker, S. (2002) The VERNALIZATION INDEPENDENCE 4 gene encodes a novel regulator of FLOWERING LOCUS C. Plant J. 31, 663–673.

Zhang, X. et al. (2004) Freezing-sensitive tomato has a functional CBF cold response pathway, but a CBF regulon that differs from that of freezing-tolerant Arabidopsis. Plant J. 39, 905–919.

Figure 13A: **Start codon and stop codon of the Frigida and Frigida Essential Gene predicted by ORF Finder.**

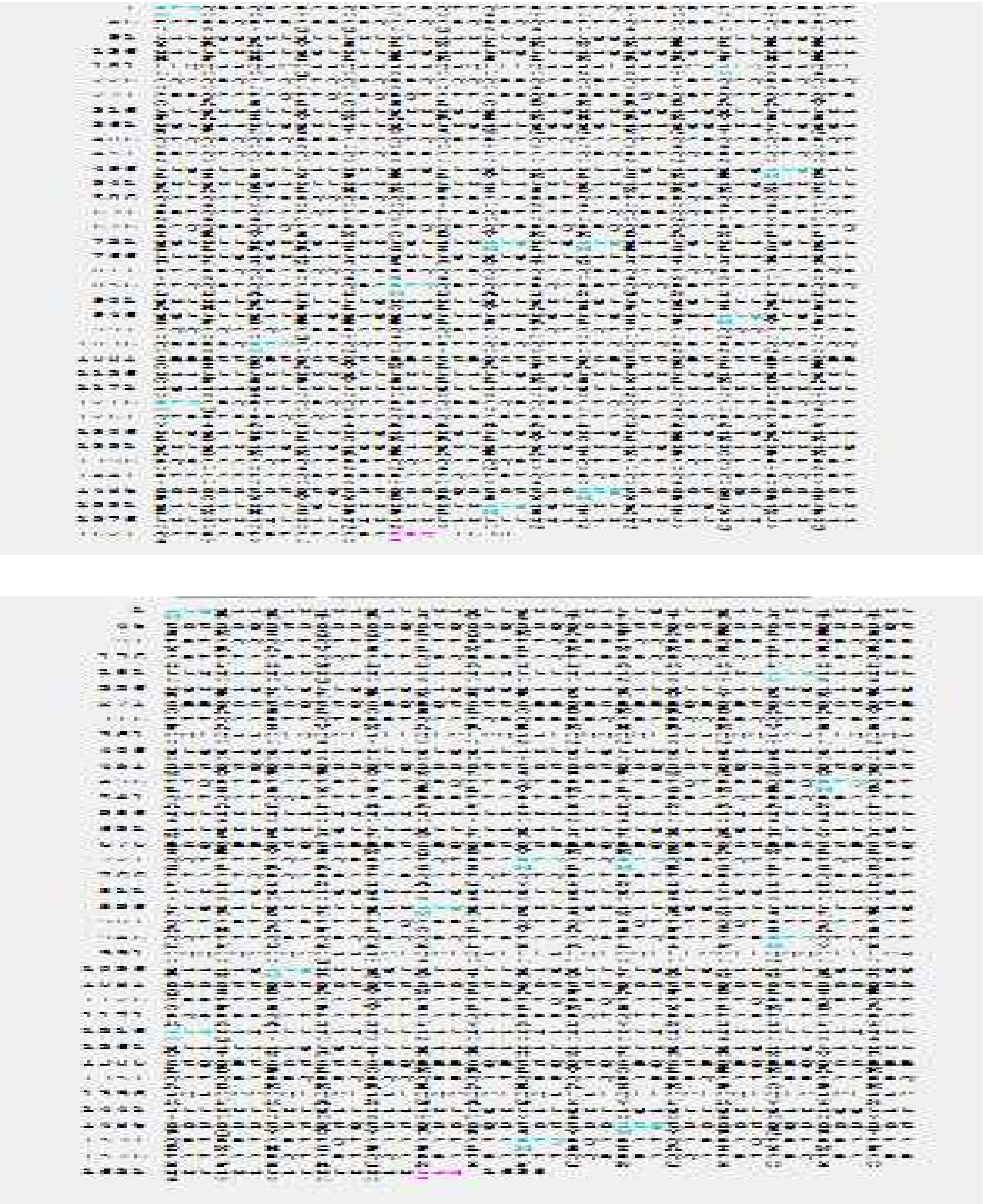


Figure 13B: **Start codon and stop codon of the Terminal Flowering 1 and Terminal Flowering 2 Gene predicted by ORF Finder.**

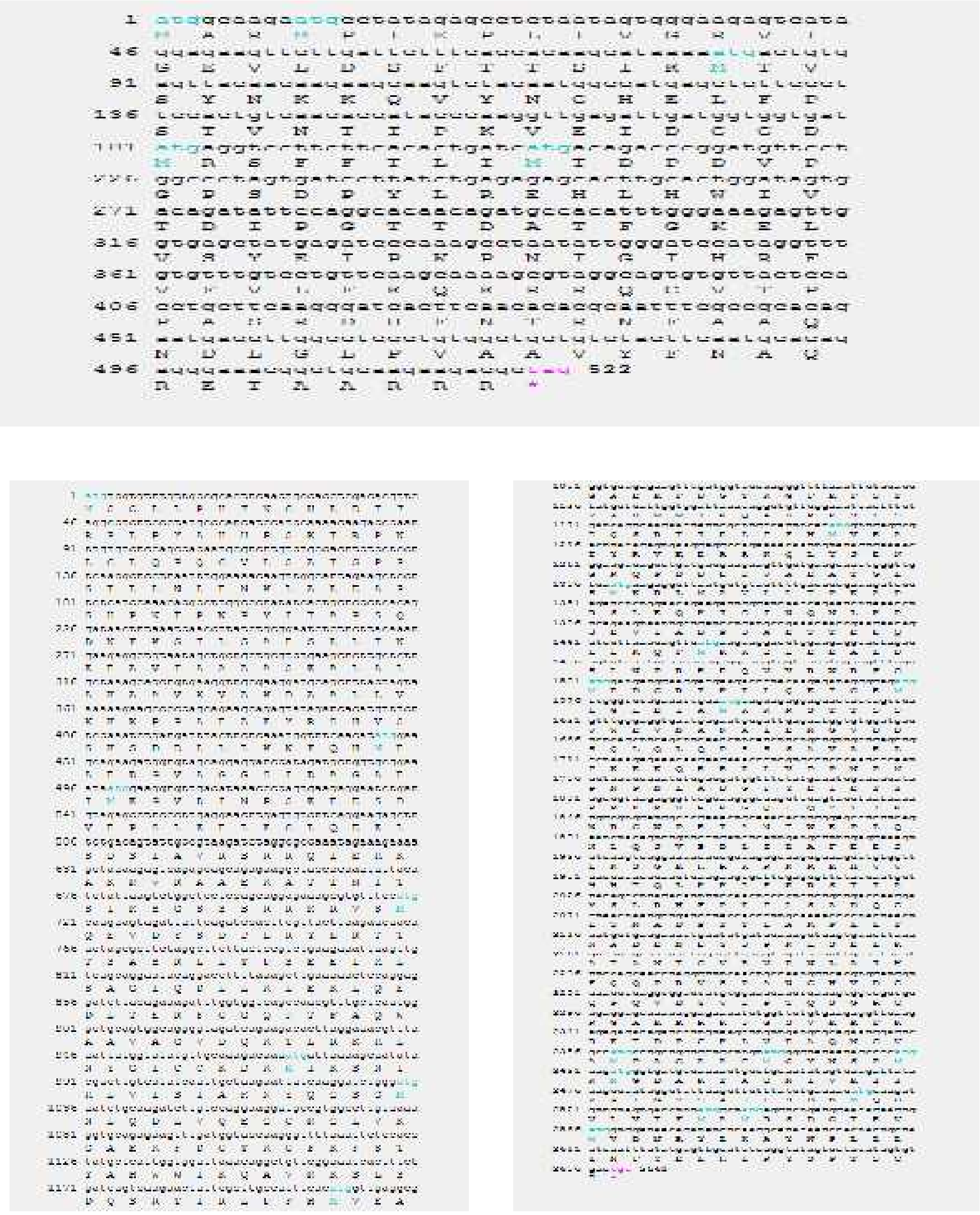


Figure 13C: **Start codon and stop codon of the Terminal Flowering 1 and Terminal Flowering 2 Gene predicted by ORF Finder.**

Brief Insight into ICE1 (Inducer of CBF Expression 1) from Temperate, Subtropical and Tropical Crop Plant to Understand Cold Tolerance

D.P. WANKHEDE*, ANIMESH PATTNAIK[1], J. ARAVIND, SATYA P. MISHRA, NIHAR B. DAS[1], SOMA S. MARLA, K.V. BHAT, SUNDEEP KUMAR, AMIT K. SINGH, VIKENDER KAUR, S. RAJKUMAR AND RAJESH KUMAR
ICAR-National Bureau of Plant Genetic Resources, New Delhi-110112 India
[1]Orissa University of Agriculture and Technology, Bhubaneswar, India
Email:dpwankhede@nbpgr.ernet.in

Cold stress is one of the most crucial stress conditions which have deleterious impact on crop growth and development and thus resulting in significant economic loss due to reduced crop productivity. Plants growing in different climatic conditions show varying degree of cold tolerance owing to differential adaptations to the prevalent climatic conditions. In response to cold stress, plants show expression of cold-regulated (COR) genes which help to cope up cold stress. The *COR* genes harbour C-repeat at their promoter element and their expression is activated by C-repeat (CRT)-binding transcription factors (CBF). In Arabidopsis ICE1 (Inducer of CBF Expression 1) a MYC-type basic helix–loop–helix transcription factor is considered to be a central regulator of cold response as it is shown to regulate expression of CBF3 and downstream *COR* genes. In present works bioinformatics approaches have been used to understand ICE1 in temperate, tropical and subtropical crop plants. ICE1 protein sequences were retrieved from gene bank from different plants through BLAST search using Arabidopsis ICE1. The available ICE1 sequences from diverse plants were aligned using multiple sequence alignment and used to construct HMM profile. The local protein database was searched on the basis of the HMM profile. HMMER and BLAST hits were compared and parsed by manual editing. Thus for the present study, ICE1 sequences were obtained for 18 crop plants including temperate crops, tropical and subtropical plants (*Brassica rapa, Raphanus sativus, Arabidopsis thaliana, Oryza sativa, Glycine max, Solanum lycopersicum, Malus domestica, Solanum tuberosum, Citrus sinensis, Cicer arietinum, Cucumis sativus, Cucumis melo, Cajanus Cajan, Setaria italic, Sorghum bicolor, Triticum aestivum, Zea mays* and *Gossypium arboretum*). Multiple sequence alignments of ICE1 from these diverse plant species gives insights in sequence conservation and divergence at important sites. Further, ICE1 were studied for their gene structure and motif conservation, putative phosphorylation and sumoylation sites. The present bioinformatics work will give a brief insight into ICE1from crop plants across different climatic conditions and its possible role in differential cold stress responses.

Abstract Book- **International Conference on "Low Temperature Science and Biotechnological Advances"**, ICAR-National Bureau of Plant Genetic Resources, New Delhi, India, April 27-30, 2015.

Brief Insight into ICE1 (Inducer of CBF Expression 1) from Temperate, Subtropical and Tropical Crop Plants to Understand Cold Tolerance

D.P. WANKHEDE*, **ANIMESH PATTNAIK**[1] , **SATYA P. MISHRA, J. ARAVIND, NIHAR B. DAS**[1], **SOMA S. MARLA, K.V. BHAT, SUNDEEP KUMAR, VIKENDER KAUR, AMIT K. SINGH, S. RAJKUMAR, AND RAJESH KUMAR**

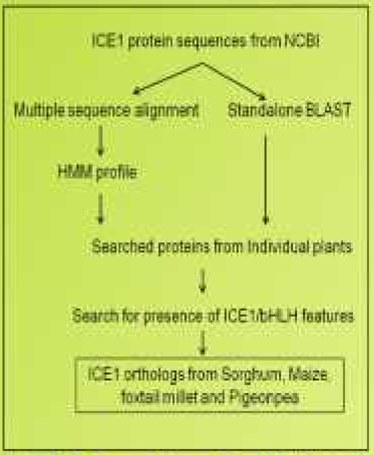
ICAR-National Bureau of Plant Genetic resources, New Delhi 110012, India; 1-Orissa University of Agriculture and Technology, Bhubaneswar, India

*Email: dpwankhede@nbpgr.ernet.in

Introduction: Cold stress is one of the most crucial stress conditions which have deleterious impact on crop growth and development and thus resulting in significant economic loss due to reduced crop productivity. In response to cold stress, plants show expression of cold-regulated (COR) genes which help to cope up cold stess. The COR genes harbour C-repeat at their promoter element and their expression is activated by C-repeat (CRT)-binding transcription factors (CBF). In Arabidopsis ICE1 (Inducer of CBF Expression 1) a MYC-type basic helix–loop–helix transcription factor is considered to be a central regulator of cold response as it is shown to regulate expression of CBF3 and downstream COR genes (Chinnusamy et al. 2003). Does ICE1 from temperate, tropical and subtropical crop plants have same features to regulate cold induced expression of downstream genes? If yes then what are the unique features? Here attempts were made to find an answer to this questions

Methods: ICE1 proteins from temperate (*Brassica rapa, Raphanus sativus, Arabidopsis thaliana, Solanum lycopersicum, Solanum tuberosum, Malus domestica, Triticum aestivum, Zea mays, Fragaria vesca*), subtropical (*Glycine max, Cucumis sativus, Cucumis melo*) and tropical crop plants (*Oryza sativa, Citrus sinensis, Gossypium arboretum, Setaria italic, Cajanus cajan, Sorghum bicolor, Ricinus communis, Theobroma cacao*) were retrieved from NCBI. HMM profile was build using ICE1 protein sequences from different plants. ICE protein sequences and HMM profile were used to search protein sequences of plants in question. In case of Pigeonpea total contigs, from whole genome sequence project were used to for gene prediction program Augustus. Predicted protein sequence were searched employing standalone blast using ICE1 protein as query as well as using HMM profile. Resulting sequences were searched presence of BHLH features and homology to ICE1.

Results: Putative ICE1 sequences were identified in *C. cajan, Z. mays, S. italica*, and *S. bicolor*. Multiple sequence alignment of ICE1 from 19 crop plants shows conservation in ICE1 in C-terminal regions whereas N-terminal regions show significant variations. Phylogenetic tree of ICE1 shows grouping largely as per its lineage and climatic conditions *per se* have not shown any significant impact in groupings. However, apple (*M. domestica*), a temperate fruit tree has been clustered together with other temperate crops such as Brassica, Arabidopsis and Sugar beet. The known SUMOylation site in Arabidopsis is found conserved in all the studied plants, however, additional sites have also been identified in several plants. AtICE1 is known to be phosphorylated at Ser 278 by OST1 (Ding et al., 2015) and this site is conserved across temperate, tropical and subtropical plants, however, is not present in cereals including *O. sativa, T. aestivum, Z. mays, S. bicolor* and *S. italica*. In silico prediction of phosphorylation sites show presence of Casine Kinase II target sites. Prediction of sub-cellular localization mostly shows localization in nucleus, however, ICE1 from *Triticum aestivum, Zea mays, Setaria italica,* and *Sorghum bicolor* showed significant chloroplast as well as nuclear localisation signal suggesting novel additional function in these plant species.


ICE1 protein sequences from NCBI

Multiple sequence alignment

Standalone BLAST

HMM profile

Searched proteins from individual plants

Search for presence of ICE1/bHLH features

ICE1 orthologs from Sorghum, Maize foxtail millet and Pigeonpea

Figure 1. Flow diagram showing steps followed for identification of ICE1 orthologs in Sorghum, Pigeonpea, Foxtail millet and Maize.

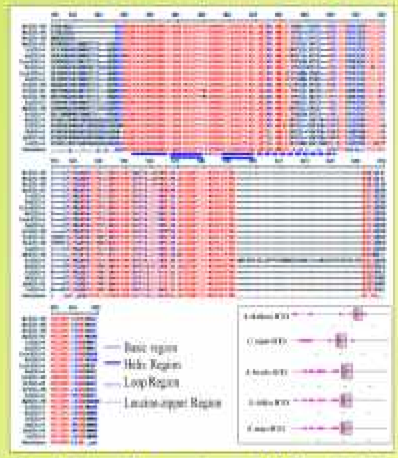


Figure 2A. Amino acid sequence alignment of ICE1 from plants from different climatic conditions. T, ST, TM after sequence name indicate Tropical, Sub-tropical and Temperate climate, respectively. B. Domains within ICE1 identified from Pigeonpea, Sorghum, Foxtail millet and Maize. Pink bold lines indicate low complexity region, box with HLH indicates Helix Loop Helix domain.

Plants	Amino acid	PI	Mol.Wt	Sub-cellular Localization
Brassica rapa	497	5.11	53049.47	Nuclear-3.800
Raphanus sativus	421	5.41	45795.40	Nuclear-3.474
Arabidopsis thaliana	494	5.31	53558.95	Nuclear-4.004
Solanum lycopersicum	551	5.38	58377.14	Nuclear-4.222
Solanum tuberosum	538	5.21	57936.61	Nuclear-3.82
Malus domestica	524	5.09	57743.52	Nuclear-4.419
Triticum aestivum	383	5.06	39433.36	Nuclear-1.983 Chloroplast-2.231
Zea mays	377	4.77	39444.35	Nuclear-1.823 Chloroplast-1.169
Fragaria vesca	533	5.81	58230.02	Nuclear-3.439
Glycine max	485	6.08	51333.31	Nuclear-4.273
Cucumis sativus	598	5.66	66023.29	Nuclear-3.089
Cucumis melo	596	5.60	66074.25	Nuclear-3.92
Oryza sativa	383	5.55	39776.83	Nuclear-3.46
Citrus sinensis	559	5.27	61213.07	Nuclear-3.679
Gossypium arboreum	555	6.06	60031.77	Nuclear-4.273
Setaria italica	376	7.12	39673.83	Nuclear-1.487 Chloroplast-1.578
Cajanus cajan	418	5.21	44209.93	Nuclear-3.864
Sorghum bicolor	378	4.71	39432.36	Nuclear-2.236 Chloroplast-1.722
Ricinus communis	548	5.09	60524.58	Nuclear-4.068
Theobroma cacao	558	5.73	60736.76	Nuclear-4.186

Table 1. Important features of ICE1 from plants of different climatic regions.

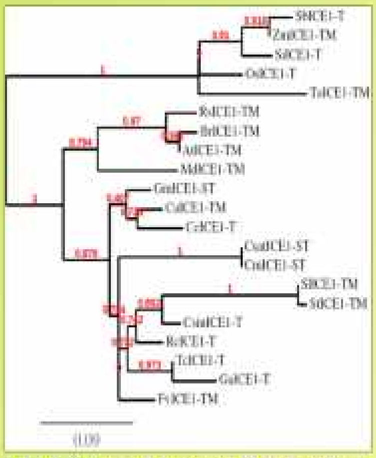


Figure 3. Phylogenetic relationship between ICE1 from different crop plants. Phylogenetic tree was built following maximum likelihood method using PhyML. Values at branch points indicate confidence index of approximate likelihood-ratio test (aLRT)

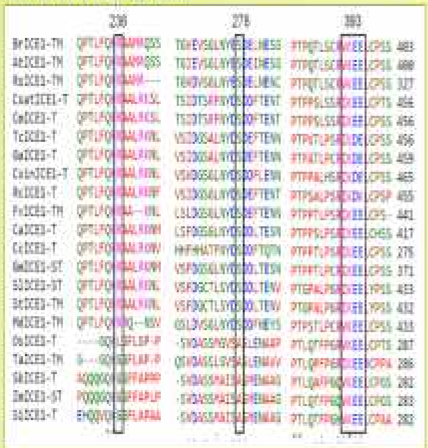


Figure 4. Sequence conservation/variation at important motifs in ICE1 protein from different plants. Numbers mentioned above sequence alignment indicate important amino acid positions with reference to Arabidopsis ICE1. Amino acid, Arginine (R) on 236th position is crucial for ICE1 activity in cold, Serine (S) on 278th position is known phosphorylation site, Lysine (K) on 393rd position is known SUMOylation site.

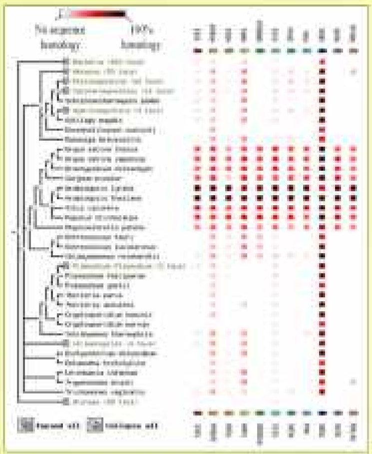


Figure 5. Conservation of ICE1 and its functional partner in different organism as predicted by STRING is a database.

Conclusion: Present work shows that cold stress regulation by ICE1 is likely to be conserved across temperate, tropical as well as sub-tropical crop plants as there are no climate specific features observed. However, variations in ICE1 protein sequences in cereals at important motifs/sites suggest possible involvement of distinct cold signalling pathways in cereals towards regulation of ICE1. Further, these distinct features also hints at novel functions of ICE1 in cereals, which awaits to be unravelled.

References

Chinnusamy et al. Genes & Development, 17:1043–1054 (2003)
Ding et al. Developmental Cell, 32, 278–289 (2015)