A study of C - repeat binding factors (CBF) associated with low temperature tolerance locus in winter wheat.

A Thesis Submitted to the College of

Graduate Studies and Research

In Partial Fulfillment of the Requirements

For the Degree of Doctor of Philosophy

In the Department of Plant Sciences

University of Saskatchewan

Saskatoon, Saskatchewan

By

Parul Jain

PERMISSION TO USE

In presenting this thesis in partial fulfillment of the requirements for a Postgraduate degree from the University of Saskatchewan, I agree that the Libraries of this university may make it freely available for inspection. I further agree that permission for copying of this thesis in any manner, in whole or in part, for scholarly purposes may be granted by professor or professors who supervised my thesis work or, in their absence, by the Head of the Department or the Dean of the College in which my thesis work was done. It is understood that any copying or publication or use of this thesis or parts thereof for financial gain shall not be allowed without my written permission. It is also understood that due recognition shall be given to me and to the University of Saskatchewan in any scholarly use which may be made of any material in my thesis.

Requests for permission to copy or to make other use of material in thesis in whole or part should be addressed to:

Head of the Department of Plant Sciences
51 Campus Drive
University of Saskatchewan
Saskatoon, Saskatchewan, Canada
S7N 5A8

ABSTRACT

Winter wheat has several advantages over spring varieties, higher (25 % more) yield, efficient use of spring moisture, reduction of soil erosion by providing ground cover during the fall and early spring, rapid initial spring growth to out - compete weeds and circumvent the peak of Fusarium head blight infections by flowering early. Winter wheat is planted in early autumn when it germinates and developing seedlings acclimate to cold. The crown survives under snow cover and in spring rapidly grows into a vigorously growing plant for grain to be harvested in summer. However, the harsh Canadian prairie winters require that winter wheat has increased cold hardiness and improved winter survival to reduce losses from sudden cold snaps during winter and spring.

Low temperature (LT) tolerance is one of the major components of cold hardiness**.** Genetic mapping studies have revealed a major quantitative trait locus (*Fr-A2*) at wheat chromosome 5A which can explain at least 50 % of LT tolerance in wheat. Physical mapping of 5A LT QTL in a hardy winter wheat cv Norstar revealed a cluster of at least 23 C - repeat binding factors (CBF) coinciding with peak of *Fr-A2* QTL. The objective of this study is biochemical, and molecular characterization of CBF co - located at *Fr-A2* to identify key CBF participating in conferring LT tolerance in winter wheat.

A comparative analysis of CBF gene cluster at the *Fr-A2* collinear region among Poaceae members showed an expansion in the number of CBF genes with increased LT tolerance. Rice, a cold sensitive member, had only three CBF genes, whereas cold hardy winter wheat cv Norstar has 23 CBF genes. Amino acid sequence - based cluster analysis of complete CBF genes, or their major functional components such as the AP2 - DNA binding domain and C - terminal trans - activation domain, divide Norstar CBF into Pooideae specific clades. However, analyses of Norstar CBF amino acid sequences of different functional groups revealed a shift in clade members. These results suggest divergence of CBF functions which could lead to possible differences / similarity in the regulon activated by a CBF in a specific group.

The 15 CBF genes from winter wheat cv Norstar were expressed in *E. coli* to produce recombinant TrxHisS - CBF fusion proteins in adequate quantities for structural and functional assays. All CBF fusion proteins could be recovered in the *E. coli* soluble phase of cell extract,

except that the CBF17.0 fusion protein could only be recovered with 6 M urea extraction. Eleven of the 15 CBF fusion proteins were very stable in heat (98 ^{o}C), 10 % SDS and 6 M urea treatment. The five other CBF members were very labile under native conditions, but were stable in *E. coli* cell extracts or when extracted under denaturing conditions. Most of the CBF recombinant proteins in denaturing gel electrophoresis migrated slower than expected from their predicted molecular mass, based on amino acid sequence. The slow migration could be associated to their elongated protein structure as determined by dynamic light scattering (DLS). CBF 12.2 and CBF 17.0 were highly resistant to denaturation and retained their secondary structure in these conditions as determined by circular dichroism (CD) spectra. The high stability of these two CBF proteins may be important for cold acclimation or maintenance of cold hardiness in wheat.

CBF proteins are transcription factors that bind to the dehydration-responsive element / C-repeat element (DRE / CRT) motif (CCGAC). Ten of the 15 Norstar recombinant CBFs whether purified under native or denaturing conditions showed *in vitro* binding to the CRT motif. Within hours of cold exposure (4 ^{o}C) the native CBF increased their affinity to CRT interaction which could be due to changes in the CBF secondary structures. Some of the CBF for binding preferred the core GGCCGAC motif while others preferred TGCCGAC. Similarly binding assays with truncated CBF revealed that for some CBF proteins, the second signature motif (DSAWR) and remaining C - terminal were not needed, while for others a considerable portion of the C - terminal region was needed for binding. Norstar CBF 12.1 has a memory of cold experience, and upon exposure to cold, has a high and immediate affinity to CRT elements. A homolog CBF12.2 in less cold - hardy winter wheat cv Cappelle - Desprez had a non - functional protein due to a R $\rightarrow$ Q substitution in a highly conserved residue within the AP2 domain. Several of the cv Norstar CBFs showed increased activity under LT and denaturing conditions, which may be the reason for the greater cold hardiness in Norstar.

In conclusion, detailed and extensive analyses of CBF in this study characterized their structure and function relationships, which are important for understanding and improving LT tolerance in plants. The identification of specific CRT binding motifs and two CBFs which were very stable under adverse conditions may be prime candidates for further study to improve LT tolerance in plants.

ACKNOWLEDGEMENTS

I sincerely thank my supervisor Dr. Ravindra N. Chibbar for his guidance, encouragement and his trust throughout the entire research project. My sincere thanks to Dr. Monica Båga for her guidance and critical inputs during my research. Also, my sincere thanks to my advisory committee members, namely, Dr. Brian Fowler, Dr. Bruce Coulman, Dr. Rosalind Bueckert and Dr. Sean Hemmingsen for their advice and constructive suggestions. I am grateful to Dr. R. Sammynaiken and Mr. Jason Maley, Saskatchewan Structural Sciences Centre, University of Saskatchewan, for their help in protein characterization studies.

Financial support during my studies from Department of Plant Sciences, Canada Research Chairs program, Genome Canada, and Western Grains Research Foundation and NSERC Collaborative Research and Development grant (CRDPJ 406114-10) is gratefully acknowledged. I want to thank all my colleagues / friends in Molecular Crop Quality group for their help during my studies. I would like to thank my family for their support and encouragement.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF ABBREVIATIONS

ABA	:	Abscisic acid
ABRE	:	ABA responsive element
AP2	:	Apetala2
BR	:	Brassinosteroid
CBF	:	C - repeat binding factor
CD	:	Circular dichroism
CK	:	Cytokinin
CAMTA	:	Calmodulin binding transcription activator
CO	:	Constans
COR	:	Cold - regulated
CRT	:	C - repeat
CRY	:	Cryptochrome
DH	:	Doubled - haploid
DHN	:	Dehydrin
DLS	:	Dynamic light scattering
DRE	:	Dehydration responsive
DREB	:	Dehydration responsive binding element
EMSA	:	Electophoretic Mobility Shift Assay
EPS	:	Earliness *per se*
ERF	:	Ethylene - Responsive Element
FLN	:	Final leaf number
Fr	:	Frost resistance
FSI	:	Field survival index
FT	:	Flowering locus T
GA	:	Gibberellic acid
GABA	:	Gamma aminobutyric acid
HOS	:	High expression of osmotically responsive genes
ICE	:	Inducer of CBF expression

IPTG	:	Isopropyl β - D - 1 - thiogalactopyranoside
JA	:	Jasmonic acid
LC-ESI-MS	:	Liquid chromatography electrospray ionisation tandem mass spectrometry
LD	:	Long day
LEA	:	Late Embryogenesis Abundant
LT	:	Low temperature
LT_{50}	:	Median lethal temperature
LTRE	:	LT - responsive element
MAS	:	Marker assisted selection
Mb	:	Mega base
Mmt	:	Millon metric ton
NO	:	Nitrous oxide
PAGE	:	Polyacrylamide gel electrophoresis
PCR	:	Polymerase chain reaction
PHY	:	Phytochrome
PPD	:	Photoperiod
QTL	:	Quantitative trait locus
R / FR	:	Red / Far - red
ROS	:	Reactive oxygen species
SA	:	Salicylic acid
SAM	:	Shoot apical meristem
SB	:	Super broth
SD	:	Short day
SDS	:	Sodium dodecyl sulfate
SL	:	Strigolactone
UV	:	Ultra - violet
VRN	:	Vernalization

LIST OF APPENDICES

CHAPTER 1

INTRODUCTION

1.1 Winter wheat - low temperature (LT) tolerance

Earth can be divided into three main climate zones: tropical, temperate, and polar which can be further separated into smaller zones, each with its own typical climate determining diversity of flora and fauna. Plants adapted to a certain region show specific responses to environmental cues like light, nutrition, temperature and biotic challenges that allow the species to grow and reproduce in that particular ecological niche. Thus, large variations exist among plant species in their morphology, lifespan, reproductive cycles and response to biotic and abiotic stresses. One of the major limitations restricting plant expansion to new areas is their tolerance to extreme temperatures. For plants adapted to arctic regions like the Arctic poppy (*Papaver laestadianum* (Nordh.) and Purple saxifrage (*Saxifraga oppositifolia* L.), growth occurs only during the short polar summer when day - lengths are long. The arctic plants can photosynthesize at extremely low temperatures and have very high frost tolerance (LT_{50} = - 40 to - 90 ^{o}C) allowing survival during long winters. Plants growing in temperate climates complete their life cycle during the short summer season, but are less winter - hardy than arctic species. In contrast to arctic and temperate plants, tropical species grow throughout the year and have no or very little resistance to cold or freezing temperatures. For example, rice can only withstand a 15 ^{o}C temperature before chilling damage is observed (Jiang *et al*., 2011).

To develop temperate crop plants originating from warmer climates, lines with better adaptation to the new environment are developed by plant breeding. For wheat adapted to the Canadian climate zones, traits like winter - hardiness, greater yield potential, grain quality, and resistance to diseases and insects have been improved during the last century (Fowler, 1999). These improvements were possible because the germplasm used in the crosses carried extensive genetic variability for genes affecting the desired traits. With regard to winter - hardiness, the greatest gain in frost - resistance was obtained during the 1980s, when cultivars with 30 % greater winter - hardiness were developed (Fowler, 2002). However, only slight incremental gains in cold hardiness have been made thereafter. Like many crop traits, a genome diversity bottleneck (Fowler *et al*., 1992; Dubcovsky and Dvorak, 2007) was likely reached for cold - hardiness

genes; alternatively, the trait is extremely complex and the likelihood to assemble all favorable alleles into one line may be very low by conventional breeding. Thus, a complete identification and characterization of genetic and / or factors contributing to cold hardiness could open up new ways to improve frost tolerance in winter wheat by marker - assisted selection (MAS).

Temperate region cereals like wheat, barley and rye commonly encounter freezing temperatures when overwintering as well as during their active growth season. Winterkill or frost - damage during formation of reproductive organs or seeds for both spring and winter cultivars can result in total loss of plant productivity. For winter wheat, the cold - hardiness in freshly established plants in the autumn is low (LT_{50} = - 3 °C), but increases upon exposure to low but non-freezing temperatures by cold acclimation (Fowler *et al*., 1996). Many cold - regulated (*cor*) genes are induced during this process leading to physiological, biological and molecular changes in the plants and ultimately protection against freezing damage (Thomashow, 1999). In contrast to low temperature (LT) in the autumn, cold spells in the spring, at anthesis or seed filling do not trigger cold acclimation (Fowler *et al*., 1996; Mahfoozi *et al*., 2001).

Unless covered by a protective snow - cover winter - kill of winter wheat is a common occurrence on the Canadian Prairies due to inadequate cold - hardiness accumulated in the autumn and / or rapid exhaustion of LT tolerance during winter (Fowler *et al*., 1996). Even the most cold - hardy Canadian cultivars like Norstar (LT_{50} = - 22 °C; Fowler *et al*., 1996) often do not survive during severe winters without snow cover. However, Norstar cultured cells have the ability to endure very low temperatures (LT_{50} = - 32.5 °C) upon abscisic acid (ABA) treatment (Chen and Gusta 1983), which suggests improved LT tolerance may be possible in intact plants. Another possibility to enhance winter survival in winter wheat is to utilize cold - hardiness genes expressed in winter rye, which is a close relative of wheat. The hardiest winter rye show LT_{50} values down to - 34 °C (Fowler *et al*., 1996) and survive well on the Prairies during winter. Transfer of the cold - hardiness character from winter rye to winter wheat is not simple. Among all hexaploid and octaploid winter triticale hybrids developed from winter rye x winter wheat crosses, the cold - hardiest progeny showed lower LT tolerance than the wheat parent (Fowler, 2002). Thus, LT tolerance appears to be very complex and requires a detailed genetic analysis to be completely understood.

Both spring and winter wheat will acclimate to the stage of vegetative / reproductive transition. Spring wheat will acclimate to that stage as well as winter wheat, but winter wheat takes longer because of the vernalization requirement. Maintenance of winter - hardiness in wheat is largely associated with growth habit (spring versus winter) where winter lines are much more frost - resistant than spring lines (Fowler, 2008). The winter growth habit requires exposure to low but non - freezing temperature for several weeks in the autumn to transition the shoot apical meristem from vegetative growth to reproductive competence, a process denoted as vernalization (Chourd, 1960; Fowler *et al*., 1996a, b). Vernalization occurs simultaneously with cold acclimation and saturation coincides with timing of maximum LT tolerance (Fowler *et al*., 1996b).

Genetic mapping of winter wheat survival has identified two major quantitative trait loci (QTL) on long arm of chromosome 5A denoted *Frost Resistance-A1* (*Fr-A1,* formerly *Fr-1*) and *Fr-A2* (Galiba *et al*., 1995, 2009; Vágújfalvi *et al*., 2003), of which *Fr-A1* may coincide with the vernalization *vrn-A1* locus (Sutka and Snape, 1989). In Norstar x winter Manitou doubled haploid lines with common *Fr-A1* genotype, more than 50 % of the variation in LT tolerance is explained by *Fr-A2,* (Båga *et al*., 2007). The homologous *Fr-A2* loci in barley (*Fr-H2*), diploid wheat (*Fr-A^{m}2*), rye (*5R*) and *Brachypodium* (Bd4) are all associated with LT tolerance (Vágújfalvi *et al*., 2003, 2005; Miller *et al*., 2006; Li *et al*., 2012). Characterization of *Fr-A2* in diploid wheat led to the identification of C - repeat binding factor (CBF) gene cluster within the locus (Vágújfalvi *et al*., 2003; Miller *et al*., 2006). CBFs are transcription factors belonging to the AP2 / ERF superfamily first identified in *Arabidopsis* as important regulators of COR genes during cold acclimation (Stockinger *et al*., 1997). The number of CBF genes mapped to *Fr-A2* exceeds 23 in hexaploid Norstar (Båga, unpublished), whereas at least 11 and 20 CBF genes are present in diploid wheat and barley, respectively (Skinner *et al*., 2005; Miller *et al*., 2006). The presence of a large CBF cluster at *Fr-A2* locus, suggests that allelic variation for one or several CBF genes underlie the 5A QTL observed for the Norstar x winter Manitou population (Båga *et al*., 2007). Whilst other regulons are likely to have a large role in LT tolerance, data from genetic mapping suggests CBF regulon has a central role in winter wheat. The long term goal, although beyond this study, is to develop a systems biology model to understand how LT tolerance is acquired and maintained in wheat. This will aid development of strategies to produce improved

winter - hardy genotypes suited for the Canadian Prairies.

1.2 Hypothesis

For the study of the Norstar CBFs encoded by *Fr-A2*, the hypothesis of this study is:

The function of one or several CBFs encoded from *Fr-A2* has an effect on LT tolerance in winter wheat.

1.3 Objectives

To test the hypothesis, the objectives of the study were:

1. To perform cluster analysis of CBFs of selected Poaceae family members.
2. To produce recombinant CBFs encoded from *Fr-A2* and analysis of their *in vitro* DNA - binding properties.
3. To study the effect of temperature on CBF function and stability *in vitro*.

CHAPTER 2

LITERATURE REVIEW

2.1 Wheat

Cereals provide 50 percent of the total global calories provided by food products with wheat products contributing to one - fifth in human diet (FAO Statistical Yearbook 2012). Wheat grain also contributes to animal feed and several bio - based industrial products. Bread, cookies, cakes, breakfast cereals and pasta are the main wheat products consumed by nearly 35 % of the world population. Most of the calories in wheat grain are derived from the carbohydrate - rich endosperm, which also contains proteins which confer good dough - making qualities. The presence of valuable minerals, vitamins and fiber add additional value to wheat - based food products. Most of the cultivated wheat consists of allohexaploid bread wheat (*Triticum aestivum* L.) (2n = 6x = 42) composed of three genomes (BBAADD) and tetraploid (2n = 4x = 28) durum wheat (*Triticum turgidum* L.) with two genomes (BBAA). A small quantity of the hulled varieties *e.g.* diploid einkorn (*Triticum monococcum* L.; A^mA^m), tetraploid emmer (*Triticum turgidum* L.; BBAA) and hexaploid spelt (*Trticum aestivum* L.; BBAADD) are produced mainly for the health food markets.

2.1.1 Origin and domestication of wheat

The important food crops wheat, rice, barley and maize belong to the grass family Poaceae. All cultivated hexaploid and tetraploid wheats are derived from wild diploid progenitors originating from the hilly region of the Fertile Crescent. This region is bordered by the Mediterranean coast and plains of the Tigris and Euphrates basins, and stretches throughout the Syrian Desert, extending to central Israel and Jordan (Feldman, 2000). Wheat hybridization was most likely initiated by crosses between *Triticum urartu* Thum. Ex Gandil. (source of A genome) and *Aegilops speltoides* (Tausch) Gren. or a closely related species (source of B genome) (Feldman *et al*., 1995; Talbert *et al*., 1998). The resulting tetraploid wheat, *Triticum turgidum* L. (Thell) (AABB), then hybridized with *Aegilops tauschii* Coss (source of D genome) to produce hexaploid bread wheat with three genomes (BBAADD; Feldman *et al*., 1995). The appearance of hexaploid wheat lies within the corridor stretching from Armenia in Transcaucasia to the southwest coastal area of the Caspian Sea in Iran (Dubcovsky & Dvorak, 2007). Due to the

abundance of genetic variation for traits such as growth habit, day length response, resistance to abiotic and biotic stresses, hexaploid wheat may be polyphylogenetic, and thus arose more than once (Talbert *et al*., 1998; Dubcovsky and Dvorak, 2007).

Through the development of hexaploid wheat, a very large genome was created and present day wheat has a 16,700 Mb / 1C genome (Bennet *et al*., 2000). About 90 % of the wheat nuclear genome consists of repetitive DNA, of which 88 % are retroelements and 12 % are transposon - like sequences (Li *et al*., 2004).

From 3,000 to 4,000 B.C. wheat became part of human civilization and gradually underwent selection for various traits like non - brittle spike, free - threshing grains, erect growth, non - dormant seeds, uniform and rapid germination, increased tillering, reduced plant height, increased yield and adaptation to new environments (Feldman, 2000). The rate of wheat spread from Fertile Crescent to Europe, Africa and Asia was estimated to one km per year, thus it took about 5,000 years to reach China (Feldman, 2000). Hexaploid wheat was introduced to Mexico by Spainards in the 15th century and was taken to Australia at the end of 17th century (Feldman, 2000). Wheat production in Canada started in the early 17th century and the first recorded attempt at growing wheat in Saskatchewan took place between 1753 and 1756 in the Carrot River Valley, located S.E. of Nipawin. Red Fife is Canada's oldest wheat whose commercial production likely started in the 1880s in Peterborough, Ontario. Selkirk settlers from Scotland spread the wheat cultivation westward to reach Manitoba in the early 19th century and subsequently wheat rapidly became the most important crop grown on the Prairies (DePauw, 2011). In 1919, wheat was grown on four million hectares in Saskatchewan.

2.1.2 Winter wheat production in Canada

Winter wheat represents 12 % (3 Mmt) of total wheat produced in Canada (25 Mmt) (Statistics Canada, 2011). Most of the winter wheat production occurs in Ontario (62 %), followed by Manitoba (15 %), Saskatchewan (14 %) and Alberta (9 %) (Statistics Canada, 2011). Only seven percent of total winter wheat production in Canada originates from Saskatchewan (Statistics Canada, 2011).

Winter wheat grown on the Canadian Prairies is generally seeded into standing stubble from a previous crop such as canola, flax, or spring barley. Upon seeding in late August to early September, the seedlings develop a few leaves before ground freezes and overwintering is

initiated. Much of the leaves and roots die during the winter season due to frost damage, but surviving crown tissue located below soil surface will ensure re - growth in the early spring. Due to an early out - growth, winter wheat efficiently utilizes spring moisture, out - competes many weeds like wild oats, and often evades peaks of Fusarium head blight and midge infestations occurring later in the season. Maturity is reached in early August, about two weeks ahead of spring-seeded wheat, and yields are generally higher than in spring wheat. In good years, winter wheat yield reaches 4 metric tons / hectare as compared to 2.4 metric tons / hectare for spring wheat (Statistics Canada, 2011).

Winter wheat has become a desirable component of conservation farming systems as it reduces soil erosion due to no - till seeding and providing ground cover in the late fall and early spring. In addition, the input cost for pesticides and herbicides are lower compared to spring wheat production. From an environmental standpoint, the winter wheat fields present undisturbed nesting grounds for waterfowl in the spring, and therefore, play an important role for Canadian wildlife.

Despite all advantages with winter wheat production, the production level on the Canadian Prairies is low, due to unreliable winter survival. During the 1984 - 85 winter Saskatchewan experienced the worst wheat winterkill in decades, which led to near total crop loss (Fowler, 2002). The plateau reached in LT tolerance improvement could be due to a complex genetic trait and / or limited genetic variability for LT tolerance genes in the available germplasm (Fowler, 2012).

2.2 Assessment of cold tolerance

2.2.1 Winter survival in the field

Based on winter survival data collected from field trials subjected to different levels of LT stress a comparative measure of cultivar winter-hardiness potential in different lines can be calculated. This measure of winter - hardiness potential is known as Field Survival Index (FSI) (Fowler, 2002). Fowler and Gusta (1979) developed FSI based on the relative winter hardiness of winter wheat cultivars tested in more than 60 trials over a five - year period which shows that not all winter wheat cultivars possess same ability to survive winter. For example, FSI for winter wheat cultivar Cappelle - Desprez is 306, Norstar 514, winter barley cultivar Dicktoo is 355 and

winter rye cultivar Puma value is 735. This difference in FSI represents the expected percent difference in field survival; for example Dicktoo has a 49 % (355 - 306 = 49 %) higher winter survival potential than Cappelle - Desprez. Similarly, winter - hardy rye cultivar Puma, which survives winter well on the Prairies, has 429 % and 221 % higher winter survival potential than winter wheat cultivars Cappelle - Desprez and Norstar, respectively.

2.2.2 Freezing tolerance tests

The lethal temperature at which 50 % of the plants die due to freezing (LT_{50}) is often used to measure cold tolerance in wheat, barley and rye lines. These determinations are generally done in place of FSI assessments, which are time-consuming as they generally require multiple - year studies (Fowler *et al*., 1981). One of the methods to assess freezing damage is by estimating electrolyte leakage from cells by conductivity measurements (Sukumaran *et al*., 1972). The data is expressed as LT_{50} values, which represents temperature at which 50 % of the electrolytes have leaked compared to the value for 100 % leakage obtained by freezing leaf sample at – 80 ^{o}C for an hour (Zhang and Willison, 1987; Hawkins *et al*., 2002).

Another method for LT_{50} determinations is based on LT tolerance upon freezing tests and was initially developed by Zech and Pauli (1960) and refined by Fowler *et al*. (1981). Sometimes freezing survival is measured upon removal of the shoots and roots from cold - acclimated plants and the freezing tests are conducted on the remaining crown region placed in a damp sponge (Olien, 1984; Livingston, 1996; Herman *et al*., 2006) or moist sand (Fowler *et al*., 1981) before exposure to different freezing temperatures. LT_{50} values determined on cold-acclimated crowns placed in moist sand show large differences between spring (- 2 to - 6 ^{o}C) and winter lines (- 10 to 23 ^{o}C) (Fowler, 2008). Among the winter cultivars with the lowest LT_{50} values is Norstar (LT_{50} = - 22 ^{o}C); only superseded in cold - hardiness by a few Siberian accessions with LT_{50} values in the - 23 to - 26 ^{o}C range (Fowler, 2002). The hardiest winter barley, cultivar Hohentrum, has a relatively low LT tolerance (LT_{50} = - 17 ^{o}C) as compared to Norstar, whereas winter rye cultivar Puma is extremely hardy (LT_{50} = - 34 ^{o}C). The LT_{50} values determined for 36 cereal lines correlate very well ($r = 0.95$, $p \leq 0.001$) with FSI values determined over multiple field trials (Fowler *et al*., 1981).

2.3 Cold and freezing injury in plants

2.3.1 Causes of freezing damage

When the growth temperature is lowered, the biosynthetic activities of the plant are reduced and normal physiological processes may be inhibited leading to permanent injuries and/or death in chilling - sensitive plants. Chilling - tolerant but freezing - sensitive plants are generally able to survive temperatures slightly below zero but are severely damaged upon ice formation within tissues. Initially freezing occurs extracellularly due to higher freezing point and higher presence of active ice nucleators as compared to cytoplasm. Ice formation in the apoplast reduces its water potential, which leads to withdrawal of water from the cells and subsequently cellular dehydration. Therefore, freezing stress on a cellular level is always accompanied by dehydration stress. The symptoms of freezing injury are a water - soaked appearance of the plant tissue due to leakage of ions and organic solutes, and the inability to regain turgor (Palta, 1977). In cold - acclimated plant cells, freezing causes cells to dehydrate and shrink, while the plasma membrane forms exocytotic extrusions that are reincorporated into the plasma membrane during rehydration and the cells are able to swell to their original size without lysis (Steponkus *et al*., 1988; Uemura *et al*., 1984, 1995).

The plasma membrane is the primary site of freezing injury, and membrane damages are mainly due to dehydration that occurs during freeze - thaw cycles or lacerations caused by ice crystals. Cold exposure also causes changes in membrane lipid composition (Yoshida *et al*., 1984) mainly due to lipid hydrolysis. One form of freezing damage is caused by the formation of lipid hexagonal II phases in regions where the plasma membrane and the chloroplast envelope are closely apposed (Steponkus *et al*. 1988).

Freezing sensitive mutants of *Arabidopsis* named *sfr* (sensitive to freezing) have reduced freezing tolerance after cold acclimation (Warren *et al*., 1996). Function of individual SFR genes in LT tolerance is not known but it has been suggested that they encode components of the cold - induced response or genes that are not transcriptionally regulated by cold but have important role in LT tolerance. *Arabidopsis SFR2* is essential for freezing tolerance and encodes a galactolipid remodeling enzyme of the outer chloroplast membrane (Thorlby *et al*., 2004; Fourrierr *et al*., 2008; Moellering *et al*., 2010), changing the membrane lipid composition thereby stabilizing membranes during LT. *SFR3* mutants carry a missense mutation in acetyl - CoA carboxylase

(*ACC1*) gene, resulting in altered fatty acid composition and freezing sensitivity in *Arabidopsis* plants (Amid *et al*., 2012). The mutations makes leaf cuticles more permeable, resulting in higher water loss as compared to wild type plants (Riederer and Schreiber, 2001; Amid *et al*., 2012). Further, reduction in major constituent of wax including alkanes, long - chain primary alcohols and fatty acid in young leaves of *Arabidopsis* plants grown under LT results in damaged cuticle and thereby LT sensitive plants (Amid *et al*., 2012).

2.3.2 Factors affecting ice nucleation

Pure liquid water when cooled at atmospheric pressure does not freeze immediately at 0 ^{o}C but freezes at about - 40 ^{o}C. The water molecules form clusters with similar molecular arrangement of ice crystals leading to a metastable state referred as supercooled water. Conversion from this metastable state to stable phase (ice) is initiated by the process of nucleation, which is the first occurrence of very small volume of the new stable phase. Homogenous nucleation takes place when only water drops form the freezing embryo while the heterogeneous nucleation takes place when embryos form on the surface of some foreign material with which the metastable / supercooled water is in contact. The presence of foreign material lowers the freezing temperature between - 2 ^{o}C to - 15 ^{o}C. These ice nuclei grow in size leading to ice crystal formation.

There are many reports of proteins that enhance (ice nucleation proteins) and prevent (antifreeze proteins) ice crystal growth. Several bacterial and fungal species such as *Pseudomonas fluorescens*, *P. syringae*, *P. viridiflava*, *Erwinia herbicola*, *E. ananas*, *E. uredovora*, *Xanthomonas campestris* and strains of *Fusarium* have an active role in ice nucleation. Ice formation catalyzed by these microbes occurs at temperatures between - 2 ^{o}C to - 3 ^{o}C, thereby causing frost damage to many crops. The ice nucleating proteins (INP) responsible for the ice nucleation are present on the outer membrane of some of these microbes and constitute 120 - 180 kDa polypeptides with contiguous repeats of a consensus octapeptide (Ala *et al*., 1993).

To prevent ice nucleation or control ice crystal size, cold tolerant plants, insects and fishes secrete antifreeze proteins (AFPs) into the apoplast in plants, and in body fluid in fishes and insects, where AFPs bind irreversibly to the surface of embryonic ice crystals to block ice growth (Griffith *et al*., 1993). INP and AFP work together to control the ice formation in plants where INP seeds the extracellular ice formation and AFP slows down the growth of ice crystal (Griffith

et al., 1993; 2004). AFPs are only induced by cold and their expression in wheat is positively correlated with freezing tolerance (Tremblay *et al*., 2005). Winter rye has also been shown to accumulate AFPs in apoplast of leaves and crown tissue in response to cold (Antikainen *et al*., 1997). Ice - recrystallization inhibition proteins (IRIPs) control the ice formation in apoplastic space and therby minimize the dmage to cell membranes (Li *et al*., 2012). The plants ability to control location of ice formation and its spread within plant tissue may be as important as efficient cold acclimation for freezing-tolerance in plants (Gusta and Wisniewski, 2013).

No consensus structure or sequence has been reported for AFP from different plants. Perennial ryegrass AFP folds into a β - roll with two ice binding domains on opposite sides of the protein (Griffith and Yaish, 2004). AFPs in winter wheat contain two or eight leucine - rich repeats (LRR) (Tremblay *et al*., 2005), whereas the LRR content in carrot AFPs is much higher (about 74 %; Meyer *et al*., 1999). The lipid transfer protein 1 in winter rye has anti - freeze activity (Doxey *et al*., 2006) like most pathogenesis related (PR) proteins that show exclusive accumulation at low temperature (e.g. certain chitinases; Hon *et al*., 1995). Also, accumulation of various low molecular weight antifreeze compounds like glycerol or sorbitol stabilizes the liquid state of the cell. Flavonol glycosides have been reported to have anti - freeze activity in xylem parenchyma cells of katsura tree (*Cercidiphyllum japonicum*) (Kasuga *et al*., 2008). Due to the amphipathic nature, flavanols can partition between the lipid phase of the membrane (Scheidt *et al*., 2004) and thereby protect the membrane from freezing (Hoekstra and Golovina, 2002).

2.4 General strategies to avoid frost damage

Plant tissues encountering LT have two general strategies to survive LT stress; either avoidance or tolerance to freezing (Sakai and Larcher, 1987). Plants can avoid freezing mainly by super - cooling of tissue water in xylem tissues (Ishikawa, 1984). However, this mechanism has limited value since it occurs in special organs such as seeds, overwintering buds or xylem ray parenchyma cells in many hardwood trees and shrubs (Sakai and Larcher, 1987). Other tolerance mechanisms involve control of formation of ice crystals at certain sites within plant tissues (Single, 1964; Sakai and Larcher, 1987). Therefore, tolerance is the only way by which the plant can survive freezing stress and various strategies have been adopted by plants. Evergreen conifers undergo dormancy to survive LT whereas many other plant species have the ability to increase their degree of LT tolerance in response to low, non - freezing temperatures by the phenomenon

known as cold acclimation (Sakai and Larcher, 1987; Thomashow, 1999). Winter cereals such as wheat continue to grow at a lower rate and establish themselves throughout the cold acclimation process (Fowler and Charles, 1979; Pocock *et al*., 2001; Savitch *et al*., 2002). During this process, photosynthetic capacity and efficiency is maintained resulting in accumulation of high amount of photoassimilates before and during the cold acclimation process (Öquist and Huner, 2003). For winter wheat, the actual degree of LT tolerance acquired depends on the species, developmental stage and duration of stress (Limin and Fowler, 2006).

2.5 Perception of low temperature

2.5.1 Cold sensing in non - plant species

All living organisms sense and respond to temperature changes. Even primitive organisms like bacteria show cold acclimation, which allows adaptation to various aqueous cold environments (Médigue *et al*., 2005). Trehalose, glycine betaine and carnitine are a few compatible solutes accumulating in bacteria in response to cold shock (Kandror *et al*., 2002; Ko *et al*., 1994; Bayles and Wilkinson, 2000; Becker *et al*., 2000). Another cold - inducible factor in *E. coli* is the major cold shock protein (CSP), CspA, which is induced immediately following a temperature drop and accumulates up to 10 % of total soluble cellular proteins. Bacterial CSPs destabilize the secondary structures of RNA molecules and function as RNA chaperones regulating transcription and translation. These proteins are considered to be the most ancient form of RNA binding proteins and are represented in eukaryotes by proteins containing an RNA - binding domain called the cold shock domain (CSD). CSD proteins identified in plants are AtGRP2 and AtGRP2b of *Arabidopsis thaliana*, (Karlson and Imai, 2003), NtGRP of *Nicotiana tabacum*, (Kingsley and Palis, 1994), and WCSP1 of *Triticum aestivum*, (Karlson *et al*., 2002). Wheat WCSP1 has a three - domain structure composed of (1) an N - terminal CSD (with two internal conserved consensus RNA binding domains), (2) a glycine - rich region, and (3) three C - terminal zinc fingers (Nakaminami *et al*., 2006). WCSP1 functions in unwinding dsDNA (Nakaminami *et al*., 2006) while AtCSP2 unwinds the RNA duplex (Sasaki *et al*., 2007).

In prokaryotic organisms, the membrane acts as an interface between the external and internal environments and is one of the primary sensors of cold (Rowbury, 2003). A two - component system, commonly known as the phosphor - transfer pathway and consisting of a

membrane - bound sensory kinase and a cytoplasmic response regulator is involved in cold signal transduction in bacteria (Suzuki *et al*., 2000; Aguilar *et al*., 2001). In *E. coli* K - 12, the four classical methyl - accepting chemotaxis proteins (MCPs) Tsr, Tar, Trg and Tap act as thermosensors (Maeda and Imae, 1979; Mizuno and Imae, 1984; Nara *et al*., 1991). The exact mechanism of temperature - dependent alteration of the MCP - signaling state is not fully understood, but changes in fatty acid desaturation, fatty acid isomerization and composition of carotenoids appear to be the common modes to modulate membrane fluidity upon exposure to low temperatures. In addition, changes in DNA topology (Eriksson *et al*., 2002), RNA secondary structures (Lai, 2003; Narberhaus *et al*., 2006), translation and alteration in protein conformation (Gulig *et al*., 1993) could also act as temperature sensors in bacteria.

In higher organisms like humans, cold ($\sim 20\ ^{o}C$) promotes calcium influx, possibly through the direct opening of calcium - permeable ion channels located in cutaneous receptors in skin (Reid and Flonta 2001). Transient receptor potential (TRP) family of ion channels is the primary molecular transducers of thermal stimuli within the mammalian somatosensory system (Clapham, 2003). TRPM8 is located in sensory neurons and activated by cold temperatures (Brauchi *et al*., 2004; Latorre *et al*., 2011). Upon activation, Na^{+} and Ca^{2+} ions enter the cells that lead to depolarization of sensory neurons and the generation of an action potential leading to the feeling of cold.

2.5.2 Cold sensing in plants

One of the more attractive models for cold sensing in plants is based on low temperature recognition by receptors located at the plasma membrane followed by relay of signal to cytosol triggering a cascade of events leading to cold response (Xiong *et al*., 2002). Experiments conducted in *Arabidopsis* suggest that the signal can be initiated not only by direct perception of cold but also through cold induced systemic induction of gene expression (Gorsuch *et al*., 2010c). Thereby, a tissue like the shoot apex can be prepared for cold by sensing temperature in leaves or roots.

In plants, the low temperature leads to many biochemical and physiological changes, any one of which could act as recognition point for temperature sensing. Some of these biochemical changes involve modifications in the membrane composition leading to alteration in membrane fluidity (Orvar *et al*., 2000), increased calcium influx (Chinnusamy *et al*., 2006; Kaplan *et al*.,

2006), changes in the organization of the cytoskeleton (Orvar *et al*., 2000), production of reactive oxygen species (ROS; Ruelland *et al*., 2009), and chromatin remodeling (Stockinger *et al*., 2001).

It is well known that membrane fluidity is affected by temperature change where an increase in temperature renders membranes more fluid (Alonso *et al*., 1997), whereas a temperature decrease rigidifies the membrane. The fluidity of membranes is maintained by modulating desaturation of membrane phospholipids. The fatty acid composition of *Arabidopsis* leaves grown at 15 ^{o}C has higher accumulation of polyunsaturated fatty acids like trienoic acid than leaves grown at 35 ^{o}C (Penfield, 2008). Although the evidence suggests that physical state of membrane may be important in low temperature perception, but the actual molecules that sense temperature changes in plants have not yet been identified (Smallwood and Bowles, 2002). Similar to bacteria, two - component systems in *Arabidopsis*, ATRR1 and ATRR2, have been implicated in cold response (Urao *et al*., 1998). The transcripts encoding the response regulators are significantly accumulated in *Arabidopsis* roots under low - temperature stress and disappear at higher temperatures (Urao *et al*., 1998).

Ca^{2+} influx in cytosol, either due to activation of Ca^{2+}- channels by membrane rigidification or through increased levels of ROS, is an important initial event in temperature perception (Monroy *et al*., 1997). The spatial and temporal patterns of Ca^{2+} influx in response to particular stimuli generate specific Ca^{2+} signatures (DeFalco *et al*., 2010). The information contained in these signatures is interpreted by the Ca^{2+} sensors through Ca^{2+} binding proteins (CBP) (Kaplan *et al*., 2006). The three main classes of sensors known in plants are calmodulin (CaM) / calmodulin like proteins (CMLs), calcium - dependent protein kinases (CDPKs) and calcineurin B - like proteins (CBLs). The CBP proteins participate in calcium directed signal transduction by undergoing a conformational change when Ca^{2+} binds (Clapham, 2007). The Ca^{2+} activated CBPs regulate target proteins, some of which are protein kinases like Ca^{2+}-dependent protein kinase (CDPK) (Cheng *et al*., 2002) which relay the "cold signal" resulting in cold acclimation (DeFalco *et al*., 2010).

The cytoskeleton components such as microfilaments and microtubules are attached to plasma membrane and hold it under tension. Microfilaments are also attached to ion channels. Dynamic inter-conversion of F - actin and G - actin regulate ion channels in plasma membrane which in turn control osmoregulation (Drobak *et al*., 1999). Microtubules can transmit signals

from receptors to nucleus, since they span the distance from nucleus to plasma membrane (Gundersen and Cook, 1999). Studies using the microfilament stabilizer, jasplakinolide, during cold acclimation show a decrease of cold - induced Ca^{2+} influx, accumulation of *cas30* transcript and development of freezing tolerance in *Arabidopsis* (Orvar *et al*., 2000). Conversely, when treated with actin microfilament destabilizer, *cas30* accumulation and Ca^{2+} influx occur in the absence of cold but freezing tolerance is not affected. These studies suggest that low temperature induces cytoskeleton rearrangement and most likely consists of depolymerization followed by repolymerization in an altered pattern (Orvar *et al*., 2000).

Excessive production of ROS during cold stress leads to DNA, protein, and lipid damage. This damage is generally more severe in cold - sensitive species since they often have a weaker antioxidant system than cold - tolerant species (Kocova *et al*., 2009). Another role of ROS generated under cold stress is thought to be in signal transduction (Suzuki *et al*., 2012).

DNA - nucleosome composition has been implicated to have a role in cold perception in *Arabidopsis*, where H2A.Z histone protein variant plays a role in temperature perception (Kumar and Wigge 2010). H2A.Z is found in the nucleosome flanking the transcription start site of many temperature sensitive genes and regulates transcription by preventing DNA methylation (Zilberman *et al.*, 2008; Conerly *et al.*, 2010). The protein is present immediately downstream of promoters of temperature responsive genes and at lower temperatures creates a physical barrier between RNA polymerase II and the gene promoter. H2A.Z - containing nucleosomes have more tightly wrapped DNA (Thambirajah *et al.*, 2006) and the degree of unwrapping may be responsive to temperature suggesting a direct mechanism by which temperature may influence gene expression (Kumar and Wigge, 2010) as RNA Pol II waits for local unwrapping of DNA from nucleosomes before extending transcription (Hodges *et al*., 2009).

2.6 Effect of LT stress on photosynthesis

Plants use photosystems to perceive changes in day length, light quality and intensity which are accompanied with changes in the temperatures (Ensminger *et al*., 2006; Kocova *et al*., 2009). A change in temperature leads to imbalance between the energy absorbed by photosystem I and II and energy utilized in various metabolic reactions in the plant, i.e. it creates an imbalance between the energy source and sink. Photo - physical and photo - chemical reactions of trapping

light energy and converting it into redox potential by components of photosystems are not affected by change in the temperature. On the other hand, the biochemical reactions that convert this redox potential into stable reducing energy in form of NADPH and chemical energy into ATP, is very sensitive to temperature. A decrease in temperature slows this biochemical conversion resulting in disconnection of the two photosystems and generation of ROS (Baker 1994).

Plants have developed different strategies to combat disruption in energy balance upon cold exposure. In evergreens, cold acclimation triggers a down - regulation of photosynthesis, induction of protective proteins, adjustment of antenna size and composition and increase in the dissipation of excess energy by non - photochemical quenching (Öquist and Huner, 2003). Modification in the structure and localization of chloroplast (Krol *et al*., 2002; Tanaka, 2007), rearrangement of protein complexes of PSII and PSI as well as thylakoid membrane polypeptide (Ottander *et al*., 1995; Vogg *et al*., 1998; Krol *et al*., 2002; Ivanov *et al*., 2006; Verhoeven *et al*., 2009) are some of the characteristic changes leading to decrease in rate of net photosynthesis during cold acclimation (Öquist and Huner, 2003). Also, effects on excitation energy transfer between protein complexes of PSII and PSI, down - regulation of PSII photochemistry during winter and quick recovery during spring has been known in various pine species for a long time (Ottander *et al*., 1995; Vogg *et al*., 1998; Ivanov *et al*., 2001, 2006; Porcar - Castell *et al*., 2008). The imbalance between energy absorbed and metabolic activity results in excess PSII excitation and photoprotection of the photosynthetic apparatus in overwintering plants is required. As a major photoprotective step, high capacity for non - photochemical quenching of the xanthophyll cycle pigment has been suggested (Adams and Demmig-Adams, 1994; Ottander *et al*., 1995; Savitch *et al*., 2002; Ivanov *et al*., 2006). Upregulation of PsbS protein that binds with chlorophyll and xanthophyll cycle pigments forming light harvesting complexes (Ottander *et al*., 1995; Savitch *et al*., 2002), as well as phosphorylation of PSII reaction center protein D1, is correlated with high energy quenching and dissipating excess light energy absorbed as heat in evergreen plants.

In contrast to evergreens, a downregulation of photosynthesis does not occur in wheat during cold acclimation (Griffith and McIntyre, 1993; Wanner and Junttila, 1999; Savitch *et al*., 2002; Svensson *et al*., 2006; Franklin and Whitelam, 2007; Franklin, 2009) and the amount of

photoassimilate accumulated is not reduced (Öquist and Huner, 2003). However, plants are exposed to photoinhibiton of PSII during cold periods, which leads to formation of ROS (Ruelland *et al*., 2009). Reduction of xanthophyll cycle pigment, de - epoxidation and associated non - photochemical quenching provide some protection against photoinhibiton in *Arabidopsis* (Havaux and Kloppstech, 2001).

For successful cold acclimation the plant must be able to adjust the photosynthetic apparatus to function at a lower temperature (Wanner and Junttila, 1999) and it has been extensively studied in LT tolerant cultivars of wheat, barley and rye (Öquist and Huner, 2003; Ensminger *et al*., 2006). The net carbon uptake is not affected by cold in these plants and some varieties show increased photosynthetic activity at lower temperature resulting in growth and development during the cold acclimation process.

2.7 Cold acclimation

A plant's response to cold stress can be divided into four phases: the early shock phase, acclimation phase, resistance phase and de-acclimation phase (Larcher, 2003). Plants not adapted to withstand the stress of shock phase respond with a sudden spike of stress - related gene expression which occurs within minutes or hours upon perception of "cold signal". The response is generally transient and peaks within one to four hours and often involves genes encoding transcription factors and calcium binding proteins (Mahajan and Tuteja, 2005). During the acclimation phase, plant readjusts its metabolism to function under the stress condition by inducing late response genes (e.g. COR, LEA). These genes are often activated by the early factors. In contrast to early response genes, the late response genes show increased expression after several hours of cold exposure and more sustained expression profiles. The proteins encoded by the late response genes have roles in increasing osmolytes, antioxidants, chaperones and dehydrins, which all function to relieve cold stress symptoms. Finally, maximum stress tolerance in plants is achieved in the resistance phase (Larcher, 2003)

Winter cereals have low LT tolerance in the early fall but it increases during cold acclimation in the late fall to early winter (Sakai and Larcher, 1987; Thomashow, 1999). The process is cumulative and it is initiated when the temperature falls below a certain threshold value (Fowler, 2008). Cereals vary from 5 to 17 ^{o}C in their threshold induction temperature, which is

higher in winter than in spring genotypes (Fowler, 2008). For example, the threshold induction temperature for winter wheat cultivar Norstar is 15°C, whereas spring wheat cultivar Manitou does not start to cold acclimate until temperature has dropped to around 8 °C (Fowler, 2008). Among the winter - hardiest winter cereals, winter rye has the highest threshold induction temperature (e.g. cv Puma, 17 °C), followed by winter wheat (e.g. cv Norstar, 15 °C), and winter barley (e.g. cv Kold, 11 °C) (Fowler, 2008). An inverse relationship exists between threshold temperature and level of LT tolerance attained at end of cold acclimation (Fowler, 2008), because a longer cold - acclimation period contributes towards higher LT - tolerance (Fowler *et al*., 1999).

The maximum LT - tolerance that can be obtained in winter cereals is largely determined by the duration and rate of cold acclimation (Fowler *et al*., 1999). When plants are grown at constant temperatures, the most rapid changes in LT - tolerance occur during the initial stages of acclimation, and acclimation rate is inversely related to acclimation temperature (Fowler *et al*., 2004). Cold acclimation in wheat is relatively slow compared to *Arabidopsis*, for which increased cold tolerance can be observed after 30 minutes of LT exposure (Gilmour *et al*., 1988). Optimal LT - tolerance in Canadian winter wheat cultivars is generally obtained after four to seven weeks of cold acclimation and occurs when minimum morning and maximum afternoon soil temperatures are below 7 and 10 °C, respectively (Fowler, 2002). The LT tolerance obtained through cold acclimation is reversible and lost upon return to warm non - acclimating temperatures. In *Arabidopsis*, deacclimation is slower than acclimation; within one or two days after return to normal growth temperature, *Arabidopsis* plants lose the attained LT - tolerance and may be killed by subsequent freezing at - 7 °C (Wanner and Junttila, 1999). Similarly, winter wheat plants acclimated at 10 to 2 °C for six weeks lose LT - tolerance within two to five days at 20 to 15 °C but have the ability to reacclimate quickly (within 7 to 10 days) (Gusta and Fowler, 1976). Resistance to deacclimation during transient warm spells and the ability to reacclimate when LT re - occurs are important for winter survival, particularly during fluctuating winter conditions (Rapacz, 2002; Kalberer *et al*., 2006).

Studies with *Arabidopsis* mutants with altered cold acclimation response have led to the identification of freezing tolerant mutant *esk1* (*eskimo1*) that can survive at - 8 °C without cold acclimation (Xin and Browse, 1998). The wildtype gene encodes ESK1, which is a negative regulator of the proline biosynthesis gene, pyrroline - 5 - carboxylate synthetase (P5CS). The

esk1 mutation does not affect expression of four major DRE / CRT class of COR genes, i.e. COR6.6, COR15A, COR47 and COR78, which shows that ESK1 is not involved in the CBF - mediated pathway of cold acclimation (Xin and Browse, 2000). Thus, the cold acclimation process is complicated with parallel or branched pathways.

2.7.1 Developmental traits affecting cold acclimation

The developmental stage of the plant influences level of LT - tolerance in winter cereals, where maximum cold acclimation is obtained in vegetative tissues, whereas reproductive tissues develop little or no cold hardiness (Fowler *et al*., 1996b; Mahfoozi *et al*., 2001). For wheat and barley tissues, the crown located below soil surface is less susceptible to frost damage than leaves (Chen *et al*., 1983). Among leaves, younger leaves develop more LT - tolerance than older leaves during cold acclimation. However, most of the leaves will die during winter and the ultimate assurance for winter survival lies with regrowth from crown tissue which has developed under low temperature (Huner *et al*. 1993).

The maximum cold acclimation is achieved when plants have reached vernalization saturation (Fowler *et al*., 1996a, b). The transition from vegetative to reproductive competence in flowering plants is controlled by environmental conditions and internal developmental signals (Colasanti and Sundaresan, 2000; Mouradov *et al*., 2002; Simpson and Dean, 2002; Boss *et al*., 2004). The transition is marked by switch at the shoot apical meristem (SAM), which ceases production of leaf primordia and intiates development of floral primordial. If light conditions are permitting (section 2.3.1.2), reproductive organs start to emerge at the margins of SAM.

Through intensive studies primarily in *Arabidopsis* during the last three decades, five partially independent pathways regulating floral transition have been identified (Fig. 2.1): The light quality (Cerdan and Chory, 2003), vernalization (Henderson and Dean, 2004), photoperiod (Hayama and Coupland, 2003), autonomous (Boss *et al*., 2004) and gibberellin (GA) (Boss *et al*., 2004) pathways, converge on a small set of genes that integrates the signals and initiate development of flower meristems followed by floral organs at the shoot apex. The action of the pathways are highly redundant and no single mutation has been identified that can prevent flowering in all conditions. The vernalization and photoperiod pathways respond to external environmental signals for flowering transition whereas the autonomous and GA pathways act independent of external signals. Homologous genes to flowering pathway genes known in

Arabidopsis are present in wheat and barley, but their role is not fully understood (Trevaskis *et al*., 2007).

2.7.1.1 The role of vernalization

The progenitors of modern wheat in the Fertile Crescent were likely all of winter type (Feldman, 2000), and thus required vernalization to induce flowering (Wang *et al*., 1995). Spring lines subsequently evolved through multiple independent events, where the vernalization requirement for flowering was eliminated (Feldman, 2000). Genetic studies of spring and winter wheat have identified natural variation in their vernalization requirement that is mainly controlled by *Vrn* loci (Takahashi and Yasuda, 1971) In wheat, four *Vrn* loci, *Vrn-1* (Takahashi and Yasuda, 1971), *Vrn-2* (Dubcovsky *et al*., 1998), *VRN-3* (Yan *et al*., 2006) and *Vrn-4* (Yoshida *et al*., 2010) are known. The homoeologous VRN1 genes encode MADS-box transcription factors (Danyluk *et al*., 2003; Trevaskis *et al*., 2003; Yan *et al*., 2004b; Murai *et al*., 2003; Loukoianov *et al*., 2005; Shitsukawa *et al*., 2007) and are located on homoeologous group 5 chromosomes (*Vrn-A1*on 5A, *Vrn-B1* on 5B, *Vrn-D1*on 5D) in hexaploid wheat (Galiba *et al*., 1995; Snape *et al*., 1997; Tóth *et al*., 2003). A dominant allele at any of the three *Vrn-1* loci (*Vrn-A1, Vrn-B1* or *Vrn-D1*) leads to spring habit characterized by flowering irrespective of LT exposure. All winter lines carry recessive *VRN-1* alleles *(vrn-A1, vrn-B1, vrn-D1)* and require weeks of LT to fulfil the vernalization requirement to reach floral transition (Brule-Babel and Fowler, 1988). Four large independent deletions within the first intron and / or a deletion in the promoter of *Vrn-A1* are associated with spring habit phenotypes (Yan *et al*., 2004b, 2006; Fu *et al*., 2005). Spring habit wheats show constitutive expression of *Vrn-A1* and the gene is expressed in winter wheat only upon vernalization saturation (Dubcovsky *et al*., 1998).

In *Triticum monococcum*, the *VRN-A2* locus is located at the distal end of the long arm of chromosome 5A, approximately 60 cM below *VRN-A1* (Dubcovsky *et al*., 1998; Vágújfalvi *et al*., 2003). VRN-2 is a strong repressor of flowering in non - vernalized wheat plants grown under long-day (LD) conditions (Yan *et al*., 2004). *VRN-2* encodes protein with a zinc finger motif and a CCT (CONSTANS, CONSTANS - LIKE, and TIMING OF CAB1 - 1) domain (Yan *et al*., 2004b). Allelic variation in *VRN-2* has been found in diploid wheat and barley only (Distelfeld *et al*., 2009) and the dominant allele of *VRN-2* is associated with winter and recessive allele with spring growth habit (Yan *et al*., 2004a).

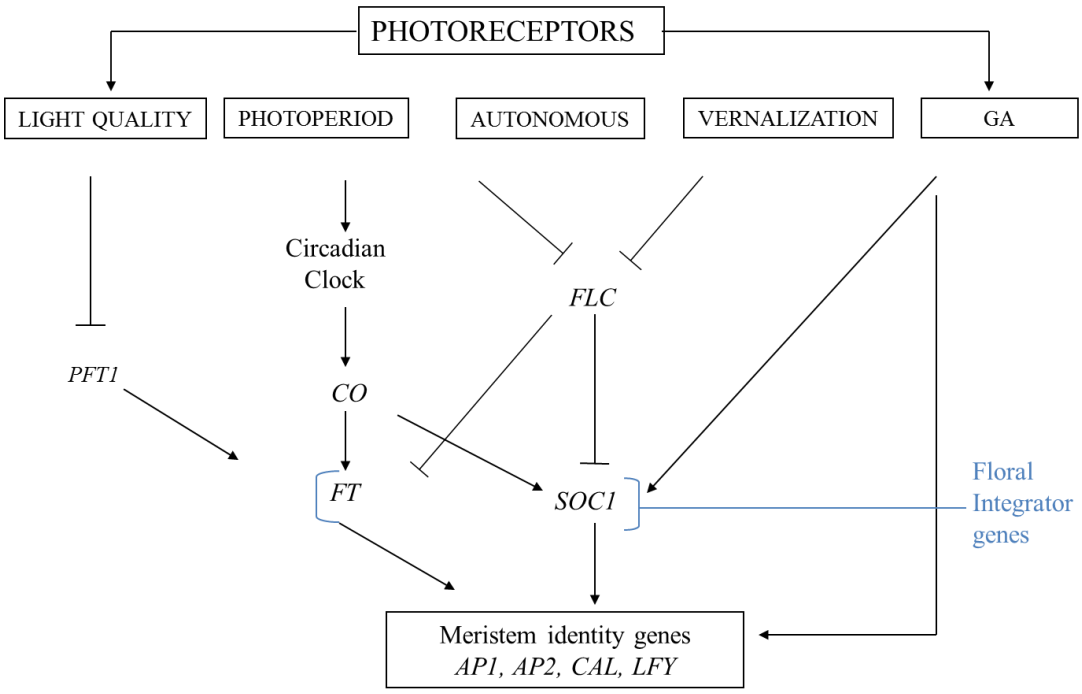


Figure 2.1 Five pathways control flowering time in *Arabidopsis thaliana*.

AP1 = APETALA1; AP2 = APETALA2; CAL = CAULIFLOWER; CO = Constans; FLC = Flowering Locus C; FT = Flowering locus T; GA = Gibberellic acid; LFY = LEAFY; PFT1 = Phytochrome and Flowering Time1; SOC1 = Supressor of overexpression of CO1. (Redrawn from Ausin *et al*., 2005).

VRN-2 is down-regulated whereas *VRN-1* is up - regulated by vernalization (Yan *et al*., 2004a; Galiba *et al*., 2009). *Vrn-3* (formerly called *VRN-B4*) is located on chromosome 7B and encodes a homologue to the flowering integrator *FLOWERING LOCUS T* (*FT*) in *Arabidopsis* (Yan *et al*., 2006). *Vrn-3* is induced in leaves under LD and functions as a flowering promoter when overexpressed in winter wheat (Yan *et al*., 2006). *Vrn-D4* locus (formerly called *VRN-4* or *VRN-D5)* is located to the centromeric region of chromosome 5D, but the underlying gene remains unknown (Yoshida *et al*., 2010). The *Vrn-B1*, *Vrn-D1* and *Vrn-D4* have small residual response to vernalization and their role in vernalization is not fully understood (Yoshida *et al*., 2010; Zhang *et al*., 2012).

The exact mechanism of the epistatic interaction between *VRN1* and *VRN2* in temperate cereals is unclear, but based on findings of different research groups, a model has been proposed (Fig. 2.2). The model suggests that, *VRN1* is positively regulated by *VRN3* which is up - regulated by long days (Fig. 2.2). However, in non - vernalized plants grown under long day, *VRN3* is down - regulated by *VRN2*. Upon vernalization and short days, *VRN2* becomes down-regulated. To understand this regulatory network it is proposed that *VRN1* is initially in a repressed state (incompetent to induce flowering) and vernalization gradually establishes competence to flower. Overexpression of *VRN3* in transgenic winter wheat plants can induce flowering even in the presence of recessive *vrn1* alleles and the absence of vernalization (Yan *et al*., 2006). Once *VRN1* reaches a certain threshold, it represses *VRN2*, thereby releasing *VRN3* and leading to further up - regulation of *VRN1* transcript levels. These interactions result in an irreversible positive feedback loop that leads to flowering. It is not fully understood how *VRN2* can repress *VRN1* independently of *VRN3* (Galiba et al., 2009).

2.7.1.2 The role of light conditions

The role of light during cold acclimation process is mainly in photosynthetic carbon fixation, which is necessary for the accumulation of sucrose and other compatible solutes (Wanner and Junttila, 1999; Savitch *et al*., 2002). Vernalization and photoperiod genes are controlled through separate pathways in winter cereals but eventually converge to influence the expression of LT induced genes and activate genes involved in flower development (Trevaskis *et al*., 2007). The photoperiod sensitivity in winter cereals allows expression of LT induced genes in the fall when days are short (Mahfoozi *et al*., 2000). Upon vernalization saturation plants become

reproductive competent but the short - day (SD) condition prevents the developmental switch from reproductive competent phase to flowering, thereby maintaining a higher level of expression of cold - responsive genes (Fowler *et al*., 1996, 2001; Mahfoozi *et al*., 2000).

The photoperiod response associated changes encountered during floral transitions are also an important adaptation in both winter and spring wheat. In wheat, presence of dominant alleles of *Ppd-1* (*Ppd-A*1, *Ppd-B*1 and *Ppd-D1*) on the short arm of the group 2 chromosomes result in photoperiod insensitivity to long day (LD) (Law *et al*., 1978; Snape *et al*., 2001). Most winter wheat cultivars grown in temperate climates carry one or several recessive *ppd* genes to provide adequate photosensitivity to delay flower development until the risk of freezing temperatures is overcome in the spring.

The diurnal and seasonal variations in ratio of light wavelengths making up the white light also effect floral transition (Ballare, 1999). These signals are transmitted by the red and far - red photoreceptors (phytochrome; PHY) and the blue / UV - A light receptors (cryptochromes; CRY1 and CRY2; Ahmad and Cashmore, 1993). *Aradidopsis* plants grown at 16 ^{o}C under 12 h photoperiod with low R : FR light ratio, show increased transcript levels of COR15a, COR15b, and KIN1, targets genes of CBF than in plants grown at a high R : FR light ratio (Franklin and Whitelam, 2007). The combination of low R : FR ratio with a relatively low temperature mimics the field conditions in autumn. This connection between light quality and cold response suggests that plants can integrate multiple environmental signals to anticipate seasonal changes and turn on adaptive responses.

2.7.1.3 The role of *earliness per se* genes

LT - tolerance in winter wheat is affected by timing to floral transition (Limin and Fowler, 2002), and it is measured by counting the final leaf number (FLN) which indicates timing of floral transition at shoot apical meristem (Robertson *et al*., 1996). In winter wheat, Norstar alleles at regions on chromosomes 1B, 2B, 4A, 5A, 6A and 7A are associated with a higher FLN, and thus contribute to longer cold acclimation and higher cold tolerance (Båga *et al*., 2009). Several of these loci coincide with mapped earliness *per se* (*eps*) loci in other wheat populations (Scarth and Law, 1983; Hoogendoorn, 1985; Miura and Worland, 1994). The *eps* loci are also referred to as earliness *per se*, ear - emergence *per se*, earliness in narrow sense, intrinsic earliness, or basic developmental rate, and affect flowering time independently of vernalization and photoperiod

cues (Hoogendoorn, 1985). However, the genes underlying these loci have not been identified in wheat.

2.7.1.4 The role of phytohormones

Plant growth regulators (phytohormones) are involved in every aspect of plant growth and development and play key roles in stress responses (Fujita *et al*., 2006; Kosová *et al*., 2012). Therefore, it is not surprising that phytohormone levels are affected when plants are exposed to LT (Kosová *et al*., 2012). Cold - induced signaling likely act via secondary messengers like Ca^{2+}, IP_3 or ROS, which subsequently induce changes to the phytohormone levels (Clapham, 2007, DeFalco *et al*., 2010). Through increased or decreased levels, the phytohormones amplify the "cold signal" or lead to induction of a second set of signaling pathways. Based on studies to date, the emerging picture is an extensive cross - talk between phytohormone and cold signaling pathways (Fujita *et al*., 2006).

The phytohormone abscisic acid (ABA) was initially suggested to have a role in cereal cold response by studies done on cultured winter wheat (cv Norstar), winter rye (cv Cougar) and bromegrass cells (Chen and Gusta, 1983). These experiments demonstrated a drastic increase in freezing tolerance by 7.5 $x10^{-5}$ M ABA treatment under normal conditions (20 ^{o}C). For wheat cells, applied ABA can decrease LT_{50} values from - 8.0 to - 32.5 ^{o}C within five days and a similar response is seen in rye cultured cells (- 9.0 to - 30.0 ^{o}C), whereas bromegrass cultured cells show a slightly lower improvement in freezing tolerance (- 8.0 to - 25.0 ^{o}C). ABA influence on frost tolerance was later confirmed by studies of *Arabidopsis* ABA-insensitive mutant *abi-1* and ABA-deficient mutant *aba-1*, which differ in level of impaired cold acclimation (Koornneef *et al*., 1982; Heino *et al*., 1990; Mäntylä *et al*., 1995). Addition of ABA exogenously does not restore cold acclimation in *abi-1* mutant, whereas the treatment restores wildtype phenotype in *aba-1* plants. Interestingly, the *aba-1* mutant can also be rescued by exposure to drought demonstrating overlapping responses between cold and drought - induced pathways (Mäntylä *et al*., 1995). *Arabidopsis* LOS5 / ABA3 mutants (*frs1*) with reduced ABA biosynthesis also show lower cold acclimation ability (LT_{50} = - 5 ^{o}C) as compared to wild type (LT_{50} = - 8 ^{o}C) (Llorente *et al*., 2000). ABA3 encodes a putative molybdenum cofactor (MoCo) sulfurase that functions in the last step of ABA biosynthesis, confirming the observation that many cold responsive genes are also responsive to ABA (Seki *et al*., 2002; Rabbani *et al*., 2003).

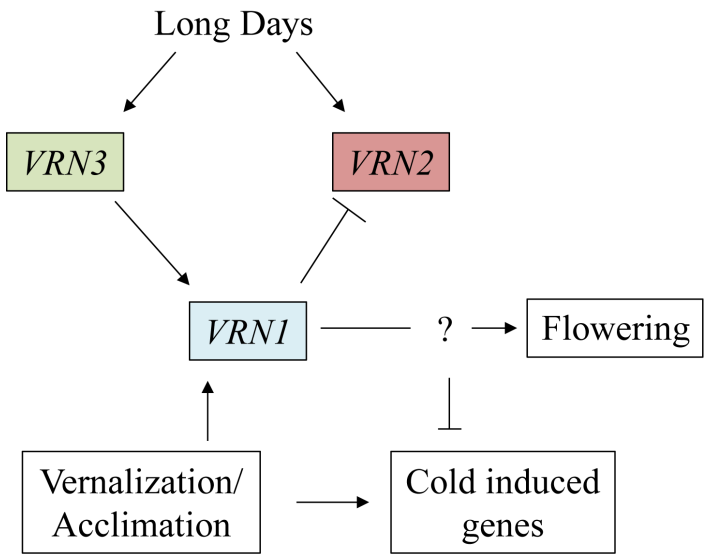


Figure 2.2 Hypothetical model proposed for interaction between vernalization and cold induced genes.
VRN = vernalization. (Redrawn from Galiba *et al*., 2009).

Silencing of EARLY RESPONSIVE TO DEHYDRATION 15 (ERD15), a small acidic protein that functions as a negative regulator of ABA response in *Arabidopsis*, results in ABA hypersensitivity and increased drought and freezing tolerance while overexpression results in decreased ABA response and stress tolerance (Kariola *et al*., 2006). A four - fold increase in ABA concentrations occur in *Arabidopsis* in response to cold (15 h at 4 ^{o}C day / 2 ^{o}C night), whereas drought can induce a 20 - fold increase in ABA concentration (Lang *et al*., 1994). Thus, the role of ABA in *Arabidopsis* cold acclimation is not as important as its role in drought tolerance.

Many COR genes have a *cis* - regulatory ABA - responsive element (ABRE) composed of a conserved G - box motif (CACGTGGC), that is activated by ABA induced proteins resulting in enhanced cold acclimation (Choi *et al*., 2000). Among the ABRE - binding factors are bZIP transcription factors, which are members of the ABF / AREB subfamily. ABF1 - ABF4 have regulatory function in ABA and stress responses (Uno *et al*., 2000) and overexpression enhances ABA sensitivity and tolerance to chilling, freezing, heat and oxidative stress in *Arabidopsis* (Kang *et al*., 2002; Kim *et al*., 2004).

Other phytohormones have not been studied in cereals as extensively as ABA. Effect of LT on auxin shows inhibition of gravity response and root growth in *Arabidopsis* (Fukaki *et al*., 1996; Wyatt *et al*., 2002; Shibasaki *et al*., 2009). Further, using auxin signaling mutants, *axr1* and *tir1*, and microscopy of living cells, it was reported that cold stress inhibits intracellular trafficking of auxin efflux carriers PIN2 and PIN3, and thereby auxin transport (Shibasaki *et al*., 2009). Cold stress in hexaploid wheat cultivars reduces indole - 3 - acetic acid (IAA) concentrations during intial phase of cold stress, followed by an increase in IAA content (Kosova *et al*., 2012).

DELLA proteins are repressors of gibberellic acid (GA) signaling and accumulate in response to LT (Achard *et al*., 2008). Moreover, LT decreases the amount of bioactive GA due to increased *GA 2 - oxidase* expression leading to GA - inactivation, dwarfism and late flowering (Achard *et al*., 2008). Similar reduction in bioactive GA and increase of inactive hydoxylated forms of GA was observed in the leaves of winter wheat cultivar Samanta during the initial phase of cold stress (Kosova *et al*., 2012). However, an increase in active GA in crown and leaves during cold acclimation was noted and correlated with growth of plants to achieve adequate LT

tolerance (Kosova *et al*., 2012).

Studies for the role of cytokinin (CK) in cold tolerance show that overexpression of the CK biosynthetic gene (*ipt*) promotes cold tolerance in *Arabidopsis* (Guo *et al*., 2010). Consistent with this, induction of negative regulators of cytokinin signaling, type - A *Arabidopsis* response regulators (ARRs), by cold treatment results in decreased LT - tolerance (Jeon *et al*., 2010) and can be associated with transient reduction of photosynthesis in *Arabidopsis* (Nagele *et al*., 2011). Similar to GA, initial decrease and subsequent increase in CK content, in both crown and leaves of winter wheat cultivar Samanta was observed and associated with increased photosynthesis contributing to energy required for growth and development during cold acclimation (Kosova *et al*., 2012).

Ethylene is another important phytohormone studied in relation to cold tolerance. An increased ethylene production occurs in rye plants when exposed to cold temperature (Yu *et al*., 2001) and is associated with an accumulation of anti - freeze proteins in apoplasts of leaves. In the absence of cold, ethylene (1 μL L^{-1}) treatment of winter rye plants increases anti - freeze activity in leaves (Yu *et al*., 2001). In contrast, little ethylene is produced in *Arabidopsis* upon exposure to cold, although genes involved in ethylene signaling are rapidly induced (Fowler and Thomashow, 2002). Overexpression of ethylene response factor TERF2 confers cold tolerance by modulating expression of *cor* genes, physiological adjustments, ethylene biosynthesis and the ethylene signaling pathway in rice seedlings, tobacco and tomato under cold stress (Zhang and Huang, 2010; Tian *et al*., 2011). Recently, it was reported that ethylene decreases LT - tolerance in *Arabidopsis* by repressing CBF pathway as well as type - A ARRs, whereas blocking both the biosynthesis as well as signaling of ethylene increases LT - tolerance (Shi *et al*., 2012). In contrast to earlier report where type - A ARRs were suggested to function as a negative regulator of cold stress signaling independent of CBF pathway, mainly through inhibition of ABA response pathway (Jeon *et al*., 2010), type - A ARRs genes were found to be positive regulators of LT tolerance (Shi *et al*., 2012). In *Aradidopsis* crosstalk between the CK and ethylene response pathways during LT exposure seems to occur through type - A ARRs (Shi *et al*., 2012).

Recently, hormonal changes during cold stress response in winter wheat were reported by Kosova *et al*. (2012). During the initial alarm phase of cold acclimation, ABA content increases and bioactive giberellins and cytokinins decreases resulting in growth suppression. Other stress

related phytohormones like jasmonic acid and salicylic acid decrease, whereas ethylene levels increase and can be associated with ROS generation. Following acclimation phase, ABA level is maintained while the growth - promoting hormones auxin, cytokinin and giberellins slightly increase indicating the adjustment of plant growth and development at low temperature. Under prolonged exposure to low temperature, plants enter into a resistance phase when all hormone levels decrease except for maintenance of high ABA concentration and a slight increase in jasmonic acid. At this stage, plant growth and development is suppressed and plants are developing resistance to multiple stresses.

2.8 Transcriptome adjustment during cold acclimation

The key genes involved in regulation of LT - tolerance were first isolated and characterized in *Arabidopsis* (Stockinger *et al*., 1997; Gilmour *et al*., 1998; Liu *et al*., 1998; Haake *et al*., 2002; Chinnusamy *et al*., 2003). These regulatory genes act by ABA - dependent and ABA - independent mechanisms. Two major *cis* - acting promoter elements: the ABA - responsive element (ABRE) function in ABA - dependent and the dehydration responsive element / C - repeat (DRE / CRT) functions in ABA - independent gene expression during LT and osmotic stress responses.

In addition to LT, cold responsive genes often respond to other stresses such as drought, salt and ABA (Thomashow, 1999; Yamaguchi - Shinozaki and Shinozaki, 2006). A microarray assay in *Arabidopsis* identified 299 drought - inducible genes, 54 cold - inducible, 213 high salinity stress - inducible and 245 ABA - inducible genes (Seki *et al*., 2001, 2002a, 2002b, 2003). More than half of drought - inducible genes overlap with high salinity - and ABA - inducible genes, while one - tenth drought - inducible genes overlap with cold - inducible genes, indicating significant cross talk between different stress responses and the complexity of the abiotic stress processes (Kreps *et al*., 2002).

Transcript level of genes involved in photosynthesis, cell wall, lipid and nucleotide metabolism are down regulated while genes associated with carbohydrates, amino acid, secondary metabolism, signal transduction including genes encoding for kinases, phosphatases, calcium trafficking - related proteins, transcription regulation genes including CBF genes, or genes encoding for protein with protective role like chlorophyll a/b binding protein show an up -

regulation with cold acclimation in *Arabidopsis* (Hannah *et al*., 2006). Raffinose synthesis pathway enzymes including galactinol synthase, raffinose synthase and myoinositol phosphate synthase family genes are induced by LT (Usadel *et al*., 2008) resulting in accumulation of monosaccharides and disaccharides like glucose, sucrose, galactinol and raffinose (Cook *et al*., 2004; Hannah *et al*., 2006; Usadel *et al*., 2008). Raffinose seems to have a role in ROS scavenging (Nishizawa *et al*., 2008). In contrast to *Arabidopsis*, where sucrose synthesis genes are up - regulated while invertase family genes are down - regulated by LT, invertase family in wheat is up - regulated and shows a weak association with cold tolerance (Artuso *et al*., 2000; Vargas *et al*., 2007). Genes involved in lipid metabolism are generally down - regulated by LT (Hannah *et al*., 2006). Similarly, most of genes involved in secondary metabolism especially related to flavonoid, anthocyanin, terpenoids and phenylpropanoids, are up - regulated by LT in *Arabidopsis* (Kaplan *et al*., 2007).

During cold acclimation in wheat, the crown and leaf transcriptome responds differentially to LT (Winfield *et al*., 2010; Ganeshan *et al*., 2011). RNA extracted during different stages of cold treatment from two winter and one spring wheat cultivars was used to study global gene expression using Affymetrix Genechip Wheat Arrays (Affymetrix Inc. Santa Clara, CA, USA), revealed 22.4 to 28.4 % transcripts differentially expressed between leaves and crown tissue (Winfield *et al*., 2010). In another approach cDNA - AFLP analysis with RNA isolated from a set of reciprocal near isogenic lines for the *VRN-A1* locus of a spring habit (cv Manitou - *VRN-A1*) and cold hardy winter wheat (cv Norstar - *vrn-A1*) also revealed in crown tissue a higher number (4,074) of transcript derived fragments and a lower number (2,757) in leaf tissues (Ganeshan *et al*., 2011). In overwintering wheat plants it is the crown that survives the winter and in spring gives rise to plant roots and leaves to restablish the plant. Therefore, it is expected that major gene action takes place in the crown tissue. Transcriptome analysis of cold acclimated crown tissue revealed that expression of 423 genes altered significantly, with the ratio of up - regulated to down - regulated genes being 9 : 1 (Skinner, 2009). A number (580) of LT regulated wheat genes, identified by wheat microarray analysis, include 130 genes involved in signalling, or are regulatory genes (transcription factors, protein kinases, ubiquitin ligases and GTP, RNA and calcium binding proteins). These regulatory genes are the same 262 genes from *Arabidopsis* (Monroy *et al*., 2007). Recently, genome - wide gene expression analysis of LT gene

regulation in winter wheat Norstar showed a 25 - fold change in the expression of *VRN-A1*, Rubisco large subunit, and several cold - responsive genes like *cor14a*, *cor14b*, *lea76*, *wcor518* and dehydrins (Laudencia-Chingcuanco *et al*., 2011). Also, in the same study, it was shown for the first time that protamine, hypothesized to play role in chromatin reorganization and histone accessibility, is induced by LT in wheat (Laudencia - Chingcuanco *et al*., 2011). Similarly, in barley seedlings transcriptome readjustment in response to LT show up - regulation of similar signaling candidate genes including calcium binding protein, *Vrn1*, *glucan synthase - like 3 genes* (Greenup *et al*., 2011).

2.9 Metabolome reprogramming during LT stress

Cold acclimation leads to numerous biochemical and physiological changes like accumulation of cryoprotective solutes like sucrose, proline, and raffinose, changes in the membrane lipid composition, rearrangement of cytoskeleton, and accumulation of defensins like endo - chitinase, and β - 1, 3 - glucanase (Thomashow, 1999, Dörffling *et al*., 2009; Gorsuch *et al*., 2010). Cold acclimation results in modification of the thylakoid membrane influencing the electron - transport, increase in abundance of sucrose synthesizing enzymes and Calvin cycle enzymes like Rubisco (Savitch *et al*. 2002; Goulas *et al*., 2006; Laudencia - Chingcuanco *et al*., 2011).

In leaves developed under LT, the carbon metabolism is reprogrammed and routed to sugar phosphate pool and free sugars (Kaplan *et al*., 2004). Carbohydrate metabolism was found to be a prominent component in metabolome readjustment during cold stress, as carbohydrates are not only the energy source but also function as carbon precursors, storage, transport and signaling molecules (Rolland *et al*., 2006; Wormit *et al*., 2006). Sucrose and trehalose are considered as components of disaccharide sensing system in plants. To function as signal molecule, sucrose is transported to the cell and cleaved either by invertase or sucrose synthase (Rolland *et al*., 2006; Iordachescu and Imai, 2008). Trehalose, a non - reducing disaccharide consisting of two glucose units (α – D - glucopyranosyl - 1, 1 - α - D - glucopyranoside), is present in wide range of organisms ranging from bacteria, fungi, plants, as well as insects and other invertebrates (Elbein *et al*., 2003) and can function as an energy source, osmoprotectant or membrane protectant (Crowe *et al*., 1984; Müller *et al*., 1995; Elbein *et al*., 2003). Trehalose - 6 - phosphate, a precursor of trehalose has a regulatory function in sugar metabolism, growth and development in

plants. Plants generally express very low levels of trehalose but trehalose may increase in response to abiotic stresses and confer increased drought, salt and cold tolerance (Cook *et al*., 2004). Fructans help to stabilize the membrane by forming complexes with phosphate and choline groups present in lipids thereby reducing the water loss and also stimulate production of cryoprotectants (Valluru and Van den Ende, 2008; Valluru *et al*., 2008). The size of hexose phosphate pool rapidly increases following the shift to LT while di - hexose pool increased steadily with time (Kaplan *et al*., 2004).

Metabolism of certain amino acids (*e.g.* proline, cysteine) and polyamine synthesis increase under LT (Davey *et al*., 2009). GABA (γ - aminobutyric acid) is synthesized in the cytosol by decarboxylation of glutamine and is considered an important cryoprotective solute in wheat and barley (Mazzucotelli *et al*., 2006). Studies in *Arabidopsis* have shown that both availability of glutamate as well as presence of glutamate decarboxylase enzyme are associated with cold tolerance and GABA biosynthesis (Guy *et al.*, 2008). In addition, secondary metabolites like flavonoids, anthocyanins, terpenoids etc., positively correlate with cold tolerance in *Arabidopsis* (Kaplan *et al*., 2007).

Metabolite - profiling using GC - MS metabolite profiling approach on *Arabidopsis* Col - 0 plants after transfer to 4 ^{o}C for 4 days or 40 ^{o}C for 4 h, (Kaplan *et al*., 2004) reported marked alteration in metabolite profile, in agreement with Cook *et al*., (2004) who concluded that the metabolome of *Arabidopsis* is extensively readjusted in response to LT and that the CBF cold response pathway plays a prominent role in this reconfiguration. Kaplan *et al*., (2004) reported LT influence metabolism extremely as 63 % (311 of total 497) metabolites studied show altered profile under LT, 60 % (186 of 311) being LT specific response and 40 % (125 of 311) metabolite levels were affect by both low and high temperatures, indicating significant cross talk between different stress responses.

2.10 COR genes in winter wheat

In various plants expression of a number of COR genes regulated by CBFs, have been characterized (Yamaguchi - Shinozaki and Shinozaki, 2006). Many of these *COR* belong to LEA (Late Embryogenesis Abundant) family and are referred to as COR / LEA gene family according to Thomashow (1999). *Lea* encode proteins that were first characterized in cotton as hydrophilic

proteins synthesized in vegetative tissues in response to dehydration (Bray, 1993). Since then, homologs of this protein have been identified in many plants. All LEA proteins have high glycine content, are highly hydrophilic and show a low secondary structure (Garay - Arroyo *et al*., 2000). LEA proteins are divided into three major groups based on the sequence similarity and structure (Bray, 1993; Wise, 2003; Battaglia *et al*., 2008). Group I LEA proteins are found only in plants and are unstructured in solution (Battaglia *et al*., 2008). Group III proteins have been discovered in organisms other than plants including prokaryotes and nematodes (Solomon *et al*., 2000; Dure, 2001; Browne *et al*., 2002). Group II LEA proteins, called dehydrins (DHNs) are found in wide range of organisms including plants, algae, yeast and cyanobacteria. In plants DHNs accumulate in dehydrating tissues like seeds during maturation or in vegetative tissues under abiotic stress like LT or drought (Close, 1996, 1997; Allagulova *et al*., 2003).

All DHNs are characterized by a conserved, lysine - rich 15 - amino acid domain, EKKGIMDKIKEKLPG, named the K - segment, present near the C - terminus (Close 1996, 1997). Other conserved sequences in DHNs are: a stretch of four to ten serine residues (the S - segment); one to three copies of the T/VDEYGNP consensus motif (the Y - segment) near the N - terminus; and less conserved regions, usually rich in polar amino acids (the Φ - segments). The number and order of the Y -, S - and K - segments are used to classify DHN sub - classes : Kn, SKn, YnSKn, YnKn, and KnS. Plant genomes usually have more than one copy of dehydrin genes belonging to different sub - classes. DHNs are heat stable and localized in cytosol, nucleus, mitochondria, vacuole, and the plasma membrane (Close 1996, 1997). 'Moonlighting' properties of DHN in plants include being a chaperone, an ion - binder, an antifreeze, cryoprotectant and radical - scavenger (Hanin *et al*., 2011). Cold - inducible dehydrin promoters contain more than one CRT / DRE element (Thomashow, 1999; Ruelland *et al*., 2009), ABRE elements and MYB and MYC elements, thereby indicating that dehydrins are involved in both ABA - independent as well as ABA - dependent pathways.

Out of 13 dehydrin genes identified in barley, only *Dhn5*, *Dhn8* and *Dhn13* are cold-inducible (Choi *et al*., 1999; Tommasini *et al*., 2008). *Dhn5* located on 6HL encodes a slightly acidic K_9 dehydrin and has several CRT / DRE, MYC and two ABRE elements in its promoter. Thus, it can be induced not only by cold but by ABA as well (Choi *et al*., 1999). Orthologues of barley *Dhn5* gene are present on group 6L chromosome arms in wheat and belong to *wcs120*

gene family (Houde *et al*., 1992). The five members, *wcs40*, *wcs99*, *wcs120*, *wcs180* and *wcs200*, and encode K_n type proteins. *Wdhn13* is another member of *wcs120* gene family and located on homeologous group 7 chromosomes in wheat. The gene is induced predominantly by LT and encodes K_3 dehydrin (Ohno *et al*., 2003), with three K segments.

LT regulates the expression of wheat *wcs120* which encodes the 50 kDa WCS120 protein (Houde *et al*., 1995), which is a member of the LEAII group of polypeptides (Houde *et al*., 1992). WCS120 is highly hydrophilic, remains soluble upon boiling and is composed of a K - segment that is repeated six times. The WCS120 protein is expressed more abundantly in the crown tissue and leaves compared to roots (Houde *et al*., 1992). Differential freezing tolerance in crown > leaf > root tissues (Chen *et al*., 1983) could be due to difference in abundance of WCS120 (Houde *et al*., 1992). In winter wheat, the accumulation of both *wcs120* mRNA and protein are closely correlated with the ability to develop freezing tolerance (Limin *et al*., 1995). Quantitative real time PCR analyses has confirmed that very early in the cold acclimation process cold hardy Norstar crowns accumulated higher amount of *wcs120* transcripts compared to its near isogenic lines derived from reciprocal crosses with spring wheat cultivar Manitou (Ganeshan *et al*., 2008). The higher amount of transcripts also resulted in higher WCS 120 polypeptide in crowns of cold hardy Norstar. The *wcs120* transcripts and WCS 120 polypeptide were also higher in Norstar leaves but the increase compared to its near isogenic lines was less compared to crown tissues (Ganeshan *et al*., 2008).

Other LT regulated genes in wheat include *wcs200*, *wcs66*, *wcs*19, *wcor14*, *wcor15* and *wcor410* (Hughes and Dunn, 1996). The WCS19 gene is regulated by LT but requires light for maximum induction (Chauvin *et al*., 1993). Over - expression experiments have shown that wheat *wcs*19 increases the LT tolerance in *Arabidopsis* (NDong *et al*., 2002). Wheat and related wild genomes possess multiple copies of *wcor15* homologues (Takumi *et al*., 2003). The *wcor15* gene encodes a chloroplast - targeted protein and has a conserved CRT / DRE - like sequence motif in their promoter (Takumi *et al*., 2003). In LT - tolerant winter cultivars, the accumulation of *wcor15* transcripts is greater than in LT - sensitive spring cultivars (Takumi *et al*. 2003; Kobayashi *et al*. 2004). The *wcor410* gene encodes an acidic dehydrin protein which accumulates in the vicinity of the plasma membrane of cells where freeze - induced dehydration occurs. The dehydrin is thought to prevent destabilization of the plasma membrane that occurs during 'freeze-

induced dehydrative conditions (Danyluk *et al*., 1998). Like *wcor15*, the expression of WCOR410 is closely correlated with the capacity to develop freezing tolerance (Danyluk *et al*., 1998). In hexaploid wheat, *wcor410* belongs to a family of three homologous members (*wcor410a*, *wcor410b* and *wcor410c)* which have been mapped to the homologous group 6 chromosomes (Danyluk *et al*., 1998). The WLT10 protein belongs to a cereal - specific COR protein family (Ohno *et al*., 2001). LT tolerant winter wheat cultivars accumulate *wlt10* mRNA more rapidly than the spring cultivar (Ohno *et al*., 2001). The expression of *wlt10* mRNA correlates with the maximum level of freezing tolerance attained (Ohno *et al*., 2001). The *wpi6* gene encodes a plasma membrane protein, WPI6, which belongs to the BLT101 protein family (Imai *et al*., 2005). Some more *cor* genes identified in wheat include *wrab17*, *wrab19*, *wcor80*, *wcor615*, *wcor726*, and *wcor 825* (Houde *et al*., 1992, Danyluk *et al*., 1994, Sarhan *et al*., 1997, Tsuda *et al*., 2000, Tsvetanov *et al*., 2000, Ohno *et al*., 2003). Increased accumulation of HSP with a chaperone function was observed in winter wheat during cold acclimation (Vitamvas *et al*., 2012).

2.11 Genetic mapping of cold - tolerance traits in wheat

The ability to cold - acclimate is a quantitative trait involving expression of a large number of genes in plants (Thomashow 1999). Initial studies of cold acclimation in wheat using monosomic and chromosome substitution lines associated at least 10 out of the 21 chromosome pairs with cold - tolerance (Sutka, 1994). The combined effect of the cold - induced gene products influences the level of LT tolerance (Fig. 2.3). At the genetic level, LT tolerance is affected by two major QTLs on the long arm of chromosome 5A: *Frost Resistance-A1* (formerly *Fr-1*) (Sutka and Snape, 1989; Snape *et al*., 1997; Tóth *et al*., 2003) and *Fr-A2* (Galiba *et al*., 1995, 2009; Vágújfalvi *et al*., 2003) located 30 cM apart. *Fr-A1* locus co - segregates with the *VRN-1* locus and may be the same locus. A dominant allele at both the loci confers the spring habit, whereas dominant *Fr-A1* with another recessive *vrn1* allele confers the winter habit (Sukta and Snape, 1989). Using diploid wheat mutants carrying deletions in *VRN1*, it has been reported that *Fr-1* and *VRN-1* are the same; it is the pleotropic effect of *VRN1* rather than *FR-1* being a separate gene (Dhillon *et al*., 2010). The *Fr-A2* QTL is associated with a cluster of CBF genes (Vágújfalvi *et al*., 2003, 2005; Francia *et al*., 2004; Miller *et al*., 2006; Båga *et al*., 2007; Campoli *et al*., 2009; Knox *et al*., 2011).

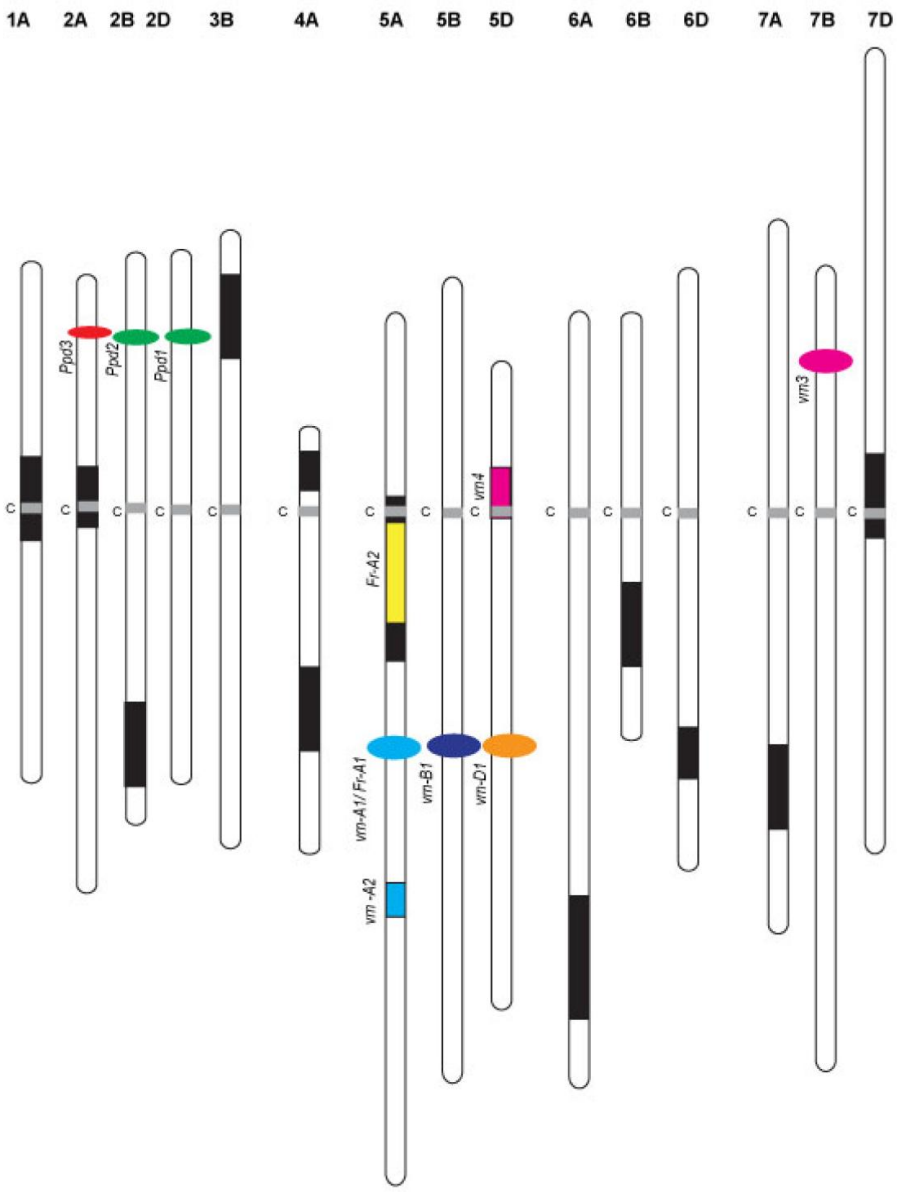


Figure 2.3 Summary of genomic regions associated with winter survival.

Fr = Frost resistance locus; *Ppd* = Photoperiod response gene; *vrn* = vernalization locus; ▌- QTLs associated with floral transition, length of vegetative growth, and flowering time (Båga *et al*., 2009); - *Ppd-A1* (Law *et al*., 1978); - *Ppd-B1 Ppd-D1* (Welsh *et al*., 1973; Snape *et al*., 2001); - *vrn-A1/ Fr-A1*(Galiba *et al*., 1995); - *vrn-B1* (Tóth *et al*., 2003); - *vrn-D1* (Snape *et al*., 1997); - *vrn-A2* (Dubcovsky *et al*., 1998); - *vrn3* (Yan *et al*., 2006); - *vrn4* (Yoshida *et al*. 2010); - *Fr-A2* (Vágújfalvi *et al*. 2003; Båga *et al*., 2007); C centromere

CBFs were first identified in long - day dicot *Arabidopsis* plants (Stockinger *et al*., 1997), for their role as transcriptional activator proteins that bind to C - repeat / dehydration responsive (CRT / DRE) element (Baker *et al*., 1994) present in multiple copies in different LT regulated or cold regulated (*cor*) genes (Baker *et al*., 1994; Yamaguchi - Shinozaki and Shinozaki, 1994; Wang *et al*., 1995). Induction of *COR* by CBFs is one of the major changes during cold acclimation (Thomashow, 1999). *Fr-A2* is 30 cM upstream of *Fr-A1* in diploid wheat (*T. monococcum*) (Vágújfalvi *et al*., 2003). Hexaploid wheat (*T. aestivum*) has two loci located on chromosome 5A, one in the same chromosome region as the *T. monococcum Fr-A2* locus and the other one closely linked to *Fr-A1*. Up to 40 % variance for frost tolerance can be explained due to in difference at *Fr-A2* QTL in barley (*Fr-H2*), diploid wheat (*Fr-A^{m}2*) and hexaploid wheat (*Fr-A2*) mapping populations (Francia *et al*., 2004; Vágújfalvi *et al*., 2003; Båga *et al*., 2007). Similarly, frost - resistance (*Fr*) loci on chromosomes 5B and 5D are closely associated with the *Vrn-B1* (Tóth *et al*., 2003) and *Vrn-D1* (Snape *et al*., 1997), respectively.

The involvement of CBF genes in natural variation of LT tolerance has been demonstrated in *Arabidopsis* recombinant inbred lines derived from a cross between LT sensitive Cape Verde Islands (Cvi) and LT tolerant Landsberg *erecta* (L*er*) accessions (Alonso - Blanco *et al*., 2005). In this population, a cluster of CBF genes on chromosome 4 co - segregates with frost tolerance QTL. A deletion of 1.6 kb in the promoter region of Cvi *AtCBF2* was implicated to be the reason for the difference in the LT tolerance between the two parental accessions (Alonso - Blanco *et al*., 2005). Extensive studies in *Arabidopsis* to understand the role of each CBF gene and their downstream gene network has indicated that *AtCBF1* and *AtCBF3* have additive effect and redundant activities (Novillo *et al*., 2007). In diploid wheat, *T. monococcum*, allelic variation in *TmCBF12* was associated with difference in LT tolerance between frost tolerant accession G3116 and frost sensitive DV92 (Knox *et al*., 2008). However, in hexaploid wheat it is still not known if the number of CBF is important or whether certain CBF genes have more important role in LT tolerance than others.

2.12 The AP2 / ERF superfamily

The CBF / DREBs belong to the APETALA2 / Ethylene - Responsive Element - Binding Factor (AP2 / ERF) superfamily, which is one of the largest groups of transcription factors in plants (Sakuma *et al*., 2002). One or two AP2 domains are carried by members of the

superfamily, which can be divided into four subfamilies: (DREB, ERF, AP2, and RAV) (Sakuma *et al*., 2002; Dietz *et al*., 2010; Figs. 2.4, 2.5). Each subfamily is distinguished by their number of AP2 domains, other DNA - binding domains and presence / absence of a WLG sequence (Magnani *et al*., 2004). Sequence variation of AP2 domain divides the ERF family into CBF / DREB and ERF - like subfamilies. Based on complete protein sequences the CBF / DREB subfamily is further separated into four groups (Fig. 2.5) and, in ERF - like subfamily nine groups can be distinguished (Nakano *et al*., 2006).

Within the AP2 / ERF superfamily, the different sub families recognize different DNA binding motifs. The ERF family members bind to GCC box (Ohme - Takagi and Shinshi, 1995). Second AP2 domain of the AP2 family members binds to TTTGTT or AACAAA (Dinh *et al*., 2012) and RAV family consensus for DNA binding sites is CAACA for AP2 domain and CACCTG for B3 domain (Kagaya *et al*., 1999). The AP2 domain of the CBF / DREB subfamily members binds C - repeat (CRT) motif CCGAC (Thomashow *et al*., 1998) present in one or several copies in the promoters of many COR genes such as *cor15a* (Baker *et a.,l* 1994) and *lti78* (Nordin *et al*., 1993). The element is also found in *B. napus bn115* (Jiang *et al.*, 1996) and wheat *wcs120* (Ouellet *et al*., 1998), where it is known as Low Temperature Responsive Element (LTRE). Yamaguchi - Shinozaki and Shinozaki (1994) identified a 9 - bp DNA element, TACCGACAT, in promoter region of *Arabidopsis rd29A* gene and named it Dehydration Responsive Element (DRE) and showed that it was able to induce gene expression in response to dehydration and low temperature. Liu *et al*., (1998) identified that DREB1A, DREB2A and DREB1B/CBF1 interact with DRE in response to cold and dehydration stress.

AP2 transcriptions factors are present in humans, fish, insects, and plants, and are also found in lower organisms like bacteria, viruses and bacteriophages (Hilger *et al*., 2000; Zhao *et al*., 2001; Feng and Williams, 2003; Magnani *et al*., 2004; Wang *et al*., 2004). The viral and prokaryotic AP2 proteins have an HNH domain positioned adjacent to the AP2 domain and proteins function as homing endonucleases (Magnani *et al*., 2004). AP2 proteins of higher organisms do not carry the HNH domain and the AP2 domain is followed by a central bsic region and HLH motif at the C - terminus. Plant AP2 domain genes are believed to have been obtained during endosymbiosis with bacteria or virus resulting in lateral gene transfer of HNH

endonuclease gene into plant cells (Magnani *et al*., 2004). AP2 / ERF domain is believed to be plant specific (Riechmann and Meyerowitz, 1998; Krizek, 2003).

CBFs are characterised by basic residues rich N - terminal nuclear localization sequence (Medina, 1999), a highly conserved central AP2 / ERF domain (Stockinger *et al*., 1997) and an acidic C - terminal domain (Fig. 2.6) that may act as transcriptional activation region (Wang *et al*., 2005). The AP2 DNA - binding domain consists of 60 - 70 amino acids flanked by signature motifs CMIII - 3 (PKK / RPAGRxKFxETRHP) and CMIII - 1 (DSAWR; Jaglo *et al*., 2001). The CBF promoters share less similarity than the coding regions, indicating that the CBF genes are differentially regulated or have been shaped by different evolutionary forces (Medina *et al*., 1999, 2011), which may include binding sites for different regulatory factors (Chinnusamy *et al*., 2003).

The 3D structure of AP2 - DNA binding domain of CBF1 protein from *Arabidopsis* consists of three - stranded anti - parallel β - sheet and an α - helix packed approximately parallel to β - sheets (Allen *et al*., 1998). Arginine and tryptophan residues in the β - sheet contact eight consecutive base pairs in the major groove of DNA containing the CRT motif. The target DNA bends slightly at the central CG step, thereby allowing the DNA to follow the curvature of the β - sheet (Allen *et al*., 1998).

The CBF genes in *Arabidopsis* are intronless and encode highly similar proteins with approximately 88 % amino acid sequence identities and 91 % similarities between them (Thomashow *et al*., 1998). CBF1 / DREB1B, CBF3 / DREB1A, and CBF2 / DREB1C are located in a tandem array within a 8.7 kb region on chromosome 4 in *Arabidopsis* (Thomashow *et al*., 1998) and in all likelihood arose from gene duplication events followed by selection (Medina *et al*., 2011). The *AtCBF1 - 3* is expressed by cold, drought and salt stress (Thomashow *et al*, 1998; Yamaguchi - Shinozaki *et al*., 1998; Xue, 2003; Skinner *et al*., 2005). *AtCBF1 - 2* is induced within 15 min of cold exposure (Gilmour *et al*., 1998), but expression is transient. After reaching maximum level at 3 h cold exposure, *CBF1 - 2* transcript levels decline and become undetectable after 24 h (Zarka *et al*., 2003). The half - life of the *CBF* transcripts is estimated to be 7.5 min on return to warm temperature (Zarka *et al*., 2003). Overexpression of *AtCBF1 - 3*

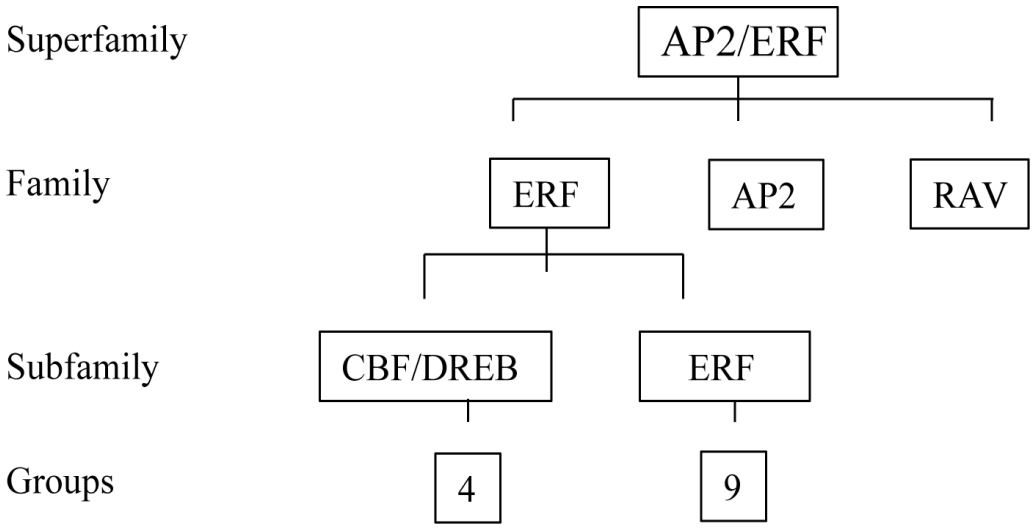


Figure 2.4 Schematic representation of AP2/ERF superfamily.

AP2 = APETALA2; CBF / DREB = C - Repeat binding Factor / Dehydration Responsive Element Binding Factor; ERF = Ethylene Responsive Factor; RAV = Related - to - ABI3 / VP1; Phylogenetic groups = 4 for CBF / DREB and 9 for ERF sub - family (Nakano *et al*., 2006).

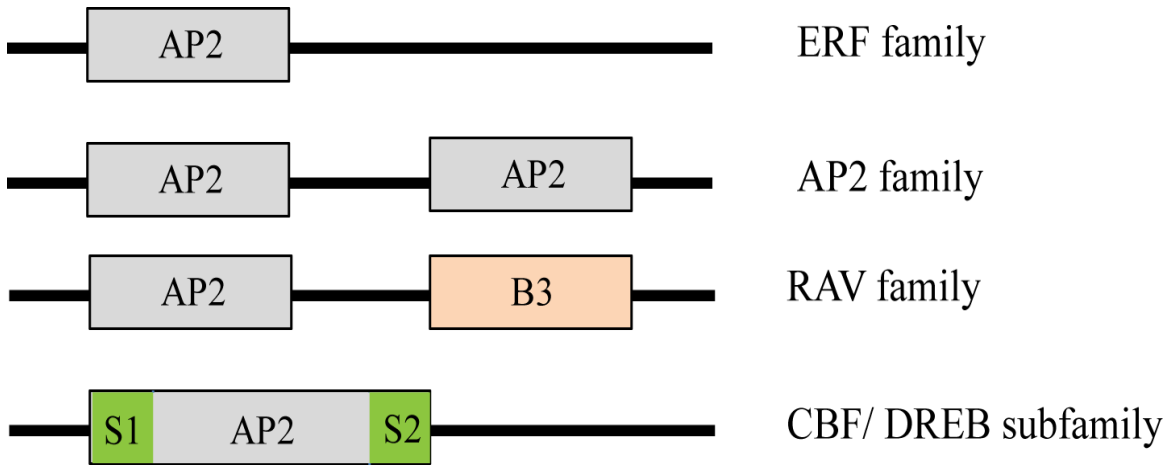


Figure 2.5 DNA-binding domains of AP2, RAV and ERF families and CBF / DREB subfamily. AP2 = APETALA2; CBF / DREB = C - Repeat binding Factor / Dehydration Responsive Element Binding Factor. B3 = Basic region 3 specific for plant RAV family (Kagaya *et al*., 1999); ERF = Ethylene Responsive Factor; RAV = Related - to - ABI3 / VP1; S1 and S2 represent signature sequences specific for plant CBF / DREB subfamily (Jaglo *et al*. 2001). (Redrawn from Medina *et al*., 2011).

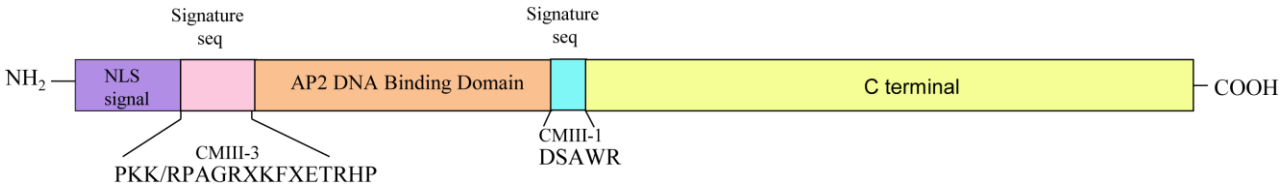


Figure 2.6 General structure of CBF.

N-terminal nuclear localization sequence (NLS; Medina *et al*., 1999), central AP2-DNA binding domain (Stockinger *et al*., 1997) flanked with signature sequences (CMIII - 3 and CMIII - 1; Jaglo *et al*., 2001), and an acidic C - terminal domain that may act as transcriptional activation region (Wang *et al*., 2005).

leads to constitutive expression of COR genes, elevated levels of proline and sugars and improved freezing tolerance in non - acclimated plants (Jaglo - Ottosen *et al*., 1998, Haake *et al*., 2002; Gilmour *et al*., 2000). *Arabidopsis* CBF1 protein interacts with *AtGCN5* and *AtADA2* which are homologues of yeast histone acyltransferase GCN5 and transcriptional adapters ADA2 and ADA3, respectively (Stockinger *et al*., 2001). Overexpression of CBF1 results in H3 acetylation and unwinding of COR gene at promoter region. Overexpression of truncated CBF2 resulted in increased H3 acetylation but no unwinding of DNA, indicating different donaims of CBF protein have a different role. Also, it was observed that GCN5 or ADA2b are not required for H3 acetylation but are necessary for DNA unwinding at COR gene promoter (Pavangadkar *et al*., 2010). Thus, the function of certain CBFs may involve protein complexes and modulation of chromatin structure.

A cluster of CBF genes are found in wheat and barley *Fr-A2* associated with frost - resistance (Vágújfalvi *et al*. 2003, 2005; Francia *et al*. 2004). At least 11 CBF genes are present at 0.8 cM *Fr-A^{m}2* region on chromosome 5A^m in diploid wheat (Miller *et al*., 2006). Likewise, a group of 20 CBF genes are located at *Fr-H2* in barley (Francia *et al*., 2004; Skinner *et al*., 2005), and > 23 CBF genes at *Fr-A2* in hexaploid wheat (Vágújfalvi *et al*., 2003, 2005; Badawi *et al*., 2007; Båga *et al*., unpublished). Overexpression of CBF genes from wheat, barley, and rice in *Arabidopsis* enhanced LT tolerance under non - stress conditions (Dubouzet *et al*., 2003; Skinner *et al*., 2005), thus confirming the cereal CBF gene family has an important role in LT tolerance.

2.12.1 Transcriptional regulation of CBF expression

The *Arabidopsis* CBF genes are under the control of several transcription factors (Fig. 2.7), some of which are induced during the early phase of cold acclimation. Inducer of CBF expression 1 (*ICE1*) is a MYC - like bHLH transcriptional activator that binds specifically to the MYC recognition sequence (CANNTG) in the *Arabidopsis CBF3* promoter (Chinnusamy *et al*., 2003). Two potential binding sites for ICE1, *ICEr1* and *ICEr2* (inducer of CBF expression region 1 and 2), are also present in the *Arabidopsis CBF2* promoter (Zarka *et al*., 2003). Several MYC binding sites including the ICE1 - like inducers (*TaICE41 and TaICE87*) are also present in the *TaCBFIVd-B9* promoter of hexaploid wheat (Badawi *et al*., 2008). *ICE1* is constitutively expressed, but is only able to induce *CBF3* expression upon exposure to cold (Chinnusamy *et al*., 2003). Overexpression of wheat *TaICE41* and *TaICE87* in *Arabidopsis* increases CBF and COR

gene expression leading to enhanced LT - tolerance but only after exposure to LT (Badawi *et al*., 2008). This suggests that similar to *Arabidopsis* ICE1, wheat ICE1 needs to be activated to be functional.

MYB15 is a R2R3 - MYB transcription factor that acts as negative regulator of CBF by binding to the *MYB* recognition sequence (GGTAGGT or TTGGTG) (Romero *et al*., 1998) in CBF promoters to down - regulate *CBF* expression (Agarwal *et al*., 2006). The MYB15 gene is induced by cold and physically interacts with ICE1. Transgenic *Arabidopsis* overexpressing *MYB15* show less induction of CBFs under LT and thereby less LT - tolerance (Agarwal *et al*., 2006).

Further effort to understand the *cis* - acting elements and *trans* - acting factors involved in expression led to identification of seven conserved DNA motifs (CM1 to CM7) in the promoter of *Arabidopsis CBF2*. Not all CMs have a role in CBF expression regulation. CM4 and CM6 act as negative regulators, whereas CM2 has both negative and positive regulatory effects (Doherty *et al*., 2009). Members of the calmodulin binding transcription activator (CAMTA) family of transcription factors bind to the CM2 motif (CCGCGT). The CAMTA family has six members in *Arabidopsis* (*AtCAMTA1* - *AtCAMTA6*) (Bouché *et al*., 2002). AtCAMTA3 is a positive regulator of *CBF2* expression in response to cold and *camta1camta3* double mutant show decreased freezing tolerance (Doherty *et al*., 2009). *Arabidopsis* PIF7 (Phytocrome Interacting Factor 7) is a HLH transcription factor that binds to the G - box motif (CACGTG) and acts as transcriptional repressor of CBF1 / DREB1B and CBF2 / DREB1C under circadian control to prevent CBF accumulation causing growth retardation under unstressed conditions (Kidokoro *et al*., 2009; Fowler *et al*., 2005). Similarly, a fluctuation in expression of hexaploid wheat CBF genes occurs in response to diurnal cycle, with slower induction during mornings and rapid induction during evenings (Badawi *et al*., 2007).

2.12.2 Post - transcriptional regulation

CBF expression is also regulated by post ranscriptional process such as mRNA export to cytoplasm from nucleus**.** A DEAD box RNA helicase encoding gene, *LOS4* (low expression of osmotically responsive genes 4), identified in *Arabidopsis* is localized in the nuclear rim and is required for RNA export. The *LOS4* mutant plants are chilling sensitive with reduced induction of CBFs and their target genes, providing evidence that RNA export is involved in regulation of

cbf expression (Gong *et al*., 2002). Additional evidence supporting the post transcriptional regulationof *CBF* expression is provided by *Arabidopsis AtNUP160* which encodes a putative homolog of the human nucleoporin Nup160 protein. *Arabidopsis* nucleoporin AtNUP160 / SAR1 is also localized at nuclear rim and controls RNA export and influences seedling growth, flowering time, and cold stress tolerance. The plants carrying mutant *atnup160* show reduced induction of *CBF3* under cold treatment, impaired mRNA export from the nucleus and are highly sensitive to cold stress, indicating that mRNA export might have a significant role in regulating *CBF* expression (Dong *et al*., 2006).

2.12.3 Translational and post-translational regulation

Protein synthesis is yet another major step in the regulation of CBF expression and function. *LOS1* gene encodes translation elongation factor2 - like protein and *LOS1 - 1* mutants show an increased expression of *CBF* genes in response to cold but reduced expression of down-stream COR genes *RD29A*, *COR47*, *COR15A* and *KIN1*. Cold induced protein synthesis is blocked, suggesting that *CBF* expression is controlled by feedback inhibition by its protein or by the downstream gene products (Guo *et al*., 2002).

HOS1 (high expression of osmotically responsive genes) encodes a protein with modified RING finger ubiquitin E3 ligase (Dong *et al*., 2006). HOS1 is known to regulate both cold acclimation and vernalization. It is localized in the cytosol at normal temperature, but accumulates in the nucleus in response to cold (Lee *et al*., 2001), where it negatively regulates expression of both CBF and COR genes. A recessive *hos1* mutant shows increased induction of CBF and their down - stream COR genes. HOS1 facilitates the ubiquitination of ICE1, resulting in decrease of *CBF* and COR genes induction in response to cold (Dong *et al*., 2006). SIZ - 1 mediated sumoylation of ICE1 results in decreased expression of MYB15 and promoting the CBF3 / DREB1A expression in *Arabidopsis* (Miura *et al.*, 2007) indicating that sumoylation might be stabilizing the ICE1 protein and / or increasing its activity required of activating expression of *CBF* and repression of *MYB15*. A potential sumolyation site has been reported in *TaICE41 and TaICE87* (Minty *et al*., 2001) which is similar to *AtICE1*, suggesting regulation of CBF expression in wheat might be similar to *Arabidopsis.*

2.13 The CBF regulon

CBF genes activate a cascade of genes involved in cold acclimation process. The CBFs and down - stream activated genes are referred to as the CBF regulon (Gilmour *et al*., 1998). Comparisons of the transcriptomes of cold-acclimated and non - acclimated plants indicated that 12 % of COR genes in *Arabidopsis* are members of the CBF regulon (Fowler and Thomashow, 2002). The CBF LT pathway is conserved and components of this pathway are present in dicots and monocots. Both LT tolerant cereals such as wheat, barley and rye, as well as in LT sensitive plants such as tomato, rice, and maize carry various components of the CBF regulon (Jaglo *et al*., 2001; Dubouzet *et al*., 2003; Takumi *et al*., 2003; Kume *et al*., 2005; Oh *et al*., 2007). Studies have indicated that the CBF regulon of cold - sensitive tomato is less diverse in function as compared to that of *Arabidopsis* (Zhang *et al*., 2004*).*

In *Arabidopsis*, the CBF regulon is controlled by light quality (Franklin *et al*., 2007), where a low red to far - red (R : FR) ratio upregulates the CBF genes and their downstream COR genes at 16 ^{o}C but not 22 ^{o}C. Remarkably, this low R : FR - induced increase in *CBF* expression is dependent on the circadian rhythm and confers freezing tolerance at 16 ^{o}C. The combination of low R : FR ratio with a relatively low temperature mimics the field conditions in autumn. This connection between light quality and cold response suggests that plants can integrate multiple environmental signals to anticipate seasonal changes and turn on adaptive responses. Post - acclimation transcriptome analyses of winter wheat have revealed a total of 423 genes that were altered. 68 genes were upregulated fivefold and 17 of these 68 (25 %) genes encode CBF, WRKY or other Zn - finger proteins, i.e. proteins that are likely to be involved in transcription regulation (Skinner, 2009). These results suggest that LT tolerance is a complex process with transcription factors playing a key role. Interestingly, some of the CBF target genes encode STZ/ZAT10, RAP2.1, RAP2.6 and various other transcription factors indicating that the CBF regulon is composed of several sub - regulons (Fowler *et al*., 2002).

In *Arabidopsis*, a non - CBF transcription factor, ZAT12, has been identified which functions independently from the CBF pathway (Vogel *et al*., 2005). Constitutive expression of the ZAT12 regulon causes a small increase in LT - tolerance while down - regulating the expression of CBF genes, and thus, ZAT12 may be a negative regulator of the CBF pathway (Vogel *et al*., 2005).

In summary, the literature reviewed above and a decade of work from our group has shown that improving low - temperature tolerance is an essential component to develop cold hardy winter cereals. Low temperature tolerance has been located to wheat chromosome 5A (Båga *et al*., 2007) where at least 23 *CBF* are located. A study of the structure, function relationship of these *CBF* in conferring low temperature tolerance will help in improving our understanding cold tolerance in winter cereals.

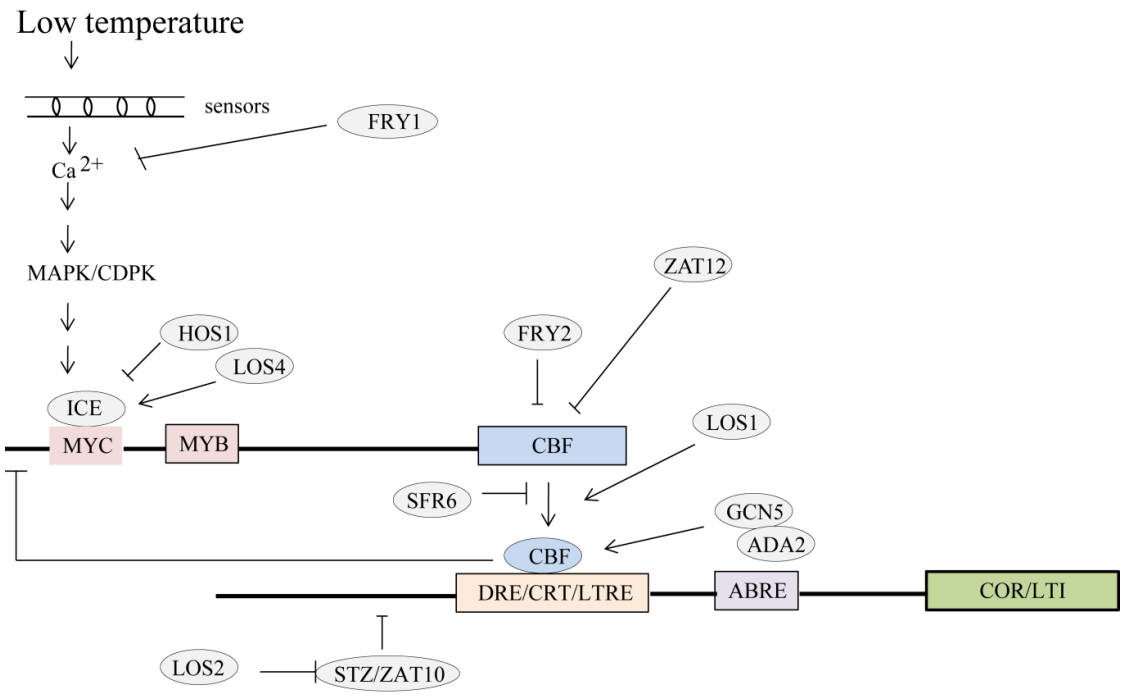


Figure 2.7. A summary of regulation of *CBF* expression in *Arabidopsis.*

ABRE = ABA responsive element; ADA = Transcriptional adapter; CDPK = Calcium dependent protein kinase; COR = Cold responsive; CRT = C-repeat element; DRE = Derhydration responsive element; FRY = FIERY; GCN = General control non repressible protein; HOS = High expression of osmotically responsive gene; ICE = Incducer of CBF; LOS = Low expression of osmotically responsive gene; LTI = Low temperature induced; LTRE = Low temperature responsive element; MAPK = Mitogen activated protein kinase; MYB = Myeloblastosis; MYC = Cellular homolog of MYB; SFR = Sensitive to freezing; STZ = Salt tolerance zinc finger; ZAT = Zinc finger protein.

CHAPTER 3

CLUSTER ANALYSIS OF C - REPEAT BINDING FACTORS (CBFs) OF SELECTED POACEAE FAMILY MEMBERS

3.1 Abstract

To understand the evolution and function of CBF, nucleotide sequences of 130 CBF genes identified in monocots have been classified into 10 groups (Badawi *et al*., 2007). Six of these groups (IIIc, IIId, IVa, IVb, IVc and IVd) were Pooideae-specific. However, to gain a better understanding of the functional relatedness of CBFs from *Fr-A2* locus of *T. aestivum* cv Norstar, amino acid sequence - based cluster analysis was done. All the fifteen CBFs identified in this study were members of Pooideae - specific groups (IIId, IVa, IVb, IVc and IVd). Expression studies identified that these five Pooideae - specific groups displayed higher constitutive and low temperature induced expression in the winter cultivar (Badawi *et al*., 2007). There is a shift of group members based on functional domain amino acid sequences of the CBFs. Broadly, they remain members of Pooideae - specific groups, but individual members change. This shifting in groups indicated that there is a divergence of CBF functions although the classification is based on the sequence similarity and co - localization on the same loci in the genome. The CBF protein function could be expected to differ or overlap leading to possible differences / similarity in COR genes activated by the CBFs of a specific group.

3.2 Introduction

Most genes from the three domains of life (Bacteria, Archaea and Eukarya) belong to gene families where members share common motifs or domains. Many large families or super - families are grouped based on similar motifs. Several databases such as, Pfam database (http://pfam.sanger.ac.uk/), UniProt (http://www.uniprot.org/); SWISS - MODEL Repository (http://swissmodel.expasy.org/repository/) and many more, are based on amino acid sequence similarity to deduce potential functional domain in a protein and propose it to a family based on its functional characteristic. Cluster analysis is usually performed by comparing multiple sequences of either DNA or proteins. The members of a group or clade are monophyletic, meaning that they can be simply cut away from a large tree with a single cut and are more related to each other than to members of other clades. There are a number of criteria proposed that may

be used to construct a cluster tree, but four of the most commonly used are maximum parsimony, maximum likelihood, minimum evolution and algorithmic methods i.e. UPGMA and neighbor - joining (reviewed by Holder and Lewis, 2003). Unrooted trees illustrate the relatedness of the nodes without making assumptions about ancestry. The neighbor join method proposed by Saitou and Nei (1987) finds pairs of units (neighbors) that minimize the total branch length at each stage of clustering, starting with a star - like tree. This method does not assume that all lineages evolve at the same rate.

Cluster analyses have identified some gene families which are in steady state as exemplified by rRNA and histone families (Ohta, 1990, 2000). Other families are constantly evolving as exemplified by protein kinases, MYB family and APETALA2 (AP2) / ethylene responsive factor (ERF) superfamily (Sappl *et al*., 2004; Harbak *et al*., 2003; Nakano *et al*., 2006). *Arabidopsis* MYB protein family consists of more than 130 members that have sequence similarity within the DNA binding domain but show several sequence variations in the flanking regions (Sappl *et al*., 2004). Other than functioning as transcription regulators, MYB proteins in also regulate secondary metabolism, control cellular morphogenesis and disease resistance and several other plant processes (Jin and Martin, 1999; Kranz et al., 1998). Based on the number of adjacent repeats in the MYB domain (one, two or three), MYB proteins are classified into three subfamilies (Rosinski and Atchley, 1998), which are further divided into 22 sub - groups (Kranz et al., 1998). Functionally conserved genes cluster together in the dendrogram for example, AtMYB91/AS1 and AmMYBPHAN both negatively regulate KNOX (KNOTTED) expression in organ primordia (Byrne *et al*., 2000) and they cluster together on the same clade in a dendrogram (Stracke *et al*., 2001). Similarly, protein kinase superfamily from *Arabidopsis* also consists of more than 1000 members (Harbak *et al*., 2003). Cluster analysis revealed several distinct clades in a dendrogram (Wang *et al*., 2003). Further characterization of sub - group members reveal that functionally related genes grouped together on same sub - group for example all calcium dependent protein kinases form one sub-group with several clades and were diverse from sub-group comprised of phosphoenolpyruvate carboxylase kinase members (Hrabak *et al*., 2003). Similarly, in *Barchypodium distachyon* cluster analysis of ice - recrystallization inhibiton proteins (IRIPs) and fructosyltransferase (FSTs), two proteins with different function in low temperature stress tolerance, split into two separate clusters (Li *et al*., 2012).

The CBF genes contributing to cold or frost-resistance in plants are members of the CBF / DREB subfamily within the AP2 / ERF subfamily (Sakuma *et al*., 2002; see Fig 2.4). The *Arabidopsis* CBF/DREB subfamily is likely derived from one ancestral gene that was initially involved in a segmental genome duplication followed by two segmental and multiple tandem duplications to generate six CBFs (Novillo *et al*., 2007; Mao and Chen, 2012). The tandem duplicated genes *CBF1* / *DREB1B*, *CBF3* / *DREB1A*, and *CBF2* / *DREB1* are positioned within a 8.7 kb region on chromosome 4 (Thomashow *et al*., 1998), CBF5 on chromosome 5 is part of a segmental duplication of the CBF1 - 3 region, and *DDF1* and *DDF2* arose from the initial segmental duplication followed by a tandem duplications and are located on chromosome 1 (Mao and Chen, 2012). CBFs 1 - 3 are all produced upon cold stress (Thomashow *et al*., 1998; Xue, 2003; Skinner *et al*., 2005), CBF4 is drought - inducible (Haake *et al*., 2002), and the dwarf and delayed flowering (*DDF*) 1 and 2 are were initially shown to be salt - inducible and have a role in gibberellin homeostasis (Magome *et al*., 2004, 2008). In later studies DDF1 was shown to be activated by multiple abiotic stresses including heat, cold, drought and salinity (Kang *et al*., 2011) and DDF1 increase cold tolerance when overexpressed in *Arabidopsis* (Jaglo *et al*., 1998, 2001; Liu *et al*., 1998; Gilmour *et al*., 2000; Kang *et al*., 2011). Thus, the *Arabidopsis* CBF / DREB subfamily is an example where some of the family members have expanded their repertoire of functions or obtained new functions over time due to adaptation and selection.

More than 20 CBF genes are located at a major cold - hardiness locus *Fr-2* on group 5 chromosomes in hexaploid wheat and diploid barley (Francia *et al*., 2004; Skinner *et al*., 2005; Båga *et al*. unpublished). Although, only 12 CBF genes have been proposed for rye *Fr-2* so far (Campoli *et al*., 2009), more genes are likely to be present. The *Fr-2* corresponding region in *Brachypodium distachyon* carries a tandem array of nine CBF genes, which all but one (*Bradi4g35650*) is cold - inducible (Li *et al*., 2012). In the relatively cold - sensitive rice, the *Fr-2* syntenous region is located on chromosome 9 and carries only three CBF genes in tandem (Fig.3.1; Dubouzet *et al*. 2003; Skinner *et al*. 2005). Two of these CBF genes are cold - inducible, whereas the third gene appears to be a pseudogene (Mao and Chen, 2012). Thus, through evolution, some genes lose function (pseudogenes), maintain a complementary function (sub - functionalization) or obtain a new function (neo - functionalization).

An initial cluster analysis of the Poaceae CBF polypeptides divided them into three monophyletic clades, HvCBF1, HvCBF3 and HvCBF4 (Skinner *et al*., 2005). All of the HvCBF3 and HvCBF4 group genes co - localized to cold - hardiness locus *Fr-H2* (Skinner *et al*., 2007). A later cluster analysis of Poaceae CBF genes divided the CBF genes into ten monophyletic groups, I - V, with groups III and IV being divided further into four sub - groups each (Badawi *et al*., 2007). Many of the wheat CBF genes from the IIId sub - group are induced by low temperature (Vágújfalvi *et al*., 2005; Danyluk *et al*., 2007). Similarly several group IV barley members have a role in cold acclimation (Skinner *et al*., 2006). However, four barley CBF genes (*HvCBF3*, *HvCBF10A*, *HvCBF10B* and *HvCBF13*) mapped to *Fr-H2* are non - responsive to cold (Skinner *et al*., 2006) and orthologues genes in wheat are also not induced by cold (Badawi *et al*., 2007).

To determine the relationship between CBFs encoded from *Fr-A2* locus in cold - hardy hexaploid wheat cultivar Norstar, 15 CBF protein sequences were selected for cluster analysis in this study. This analysis focused on the complete CBF protein sequence and two functional domains of the protein separately to evaluate any pattern between function and clustering. The selected CBF genes were analysed for the cladistic relation to other know CBFs from monocots.

3.3 Material and Methods

3.3.1 CBF sequences

A total of 130 protein sequences for CBFs from the Poaceae family were analyzed in this study. Ten sequences were from the Ehrhartoideae subfamily represented by *Oryza sativa*, nine sequences from Panicoideae subfamily represented by three from *Panicum virgatum*, three sequences from *Sorghum bicolor*, and three sequences from *Zea mays* and 112 sequences from Pooideae subfamily represented by *Avena sativa*, *Brachypodium distachyon*, *Festuca arundindcea*, *Lolium perenne*, *Triticum aestivum*, *Triticum monococcum*, *Hordeum vulgare*, and *Secale cereale* of Triticeae tribe. Accession numbers for the genes and proteins and source of sequences are listed in Table 3.1.

3.3.2 Cluster analysis of CBF sequences

The CBF protein sequences were analyzed using full - length sequences, the amino-terminal region including first signature sequence alone, the AP2 domain alone and the C - terminal sequence including the second signature sequence alone. All multiple sequence

alignments were generated using ClustalW2 at EBI ClustalW server, (http://www.ebi.ac.uk/Tools/msa/clustalw2/) with default parameters (Thompson *et al*., 1994) and T-Coffee and Expresso (Tree-based Consistency Objective Function for alignment Evaluation; Notredame *et al*., 2000) using web based server at (www.tcoffee.org). Cluster analyses were done using the Molecular Evolutionary Genetics Analysis (MEGA) version 4.0 software (Tamura *et al*., 2007; (http://www.megasoftware.net). The unrooted tree construction neighbor join method of Saitou and Nei (1987) and minimum - evolution (Rzhetsky and Nei, 1993) method were used to generate the dendrogram. The unrooted tree was verified by a bootstrap test with 500 replications and 79,859 seed size.

3.3.3 Hydrophobic cluster analysis (HCA)

A web - based interface (http://bioserv.impmc.jussieu.fr/hca-form.html) was used for HCA based on the principle that hydrophobic amino acids are not randomly distributed but have a tendency to form clusters. To display this clustering in a recognizable pattern, HCA wraps the linear amino acid sequence as an alpha - helix around a cylinder (this does not necessarily imply the native protein adopts an alpha - helical configuration). The cylinder is then cut parallel to the axis, unrolled, and displayed on a 2D surface. Hydrophobic residues (V, I, L, M, Y, F, W) separated by four or more non - hydrophobic residues, or a Proline are placed into distinct clusters. Sequence stretches of P, G, D, N, S between clusters mainly correspond to loops.

3.4 Results

3.4.1 Comparative analysis of co - linear regions at CBF - locus in Poaceae

A comparison of CBF clusters on group five chromosomes of Triticeae tribe species *Triticum aestivum*, *Triticum monococcum*, *Hordeum vulgare* and *Secale cereale* to colinear regions of Japonica rice cv. Nipponbare (Chr. 9) and *Brachypodium distachyon* diploid accession 'Bd21' (Chr. 4) reveals a large differences in the number of CBF genes (Fig. 3.1). Rice, a representative of subfamily Ehrhartoideae, has within the 10 kb syntenous region only three CBF genes (Goff *et al*., 2002), one of which may be a pseudogene (Mao and Chen, 2012). *Brachypodium distachyon*, a representative of the Poaceae subfamily Pooideae, carries nine CBF genes within a 40 kb region. However, the largest number of CBF genes at '*Fr-2*' has been identified for the Triticeae species wheat, barley or rye despite lack of a complete DNA sequence

Table 3.1 List of selected monocotyledon CBFs used in cluster analysis.

CBF Gene [a]	Gene accession number	Protein accession number	Stress tolerance	References[b]
		Ehrhartoideae		
Oryza sativa				
OsCBFI-1F	AY785897	AAX23723	Cold, Drought, Salt	**17**
OsCBFIa-1G	AK060550	BAG87488	Cold	**7**
OsCBFIa-1E	AY785896	AAX23722	Cold	**17**
OsCBFII-1C	AY327040	AAP92125	Cold, Drought	**16**
OsCBFIII-1D	AY785895	AAX23721	-	**17**
OsCBFIII-1I	NM_001068950.1	NP_001062415.1	Not Cold	**16**
OsCBFIII-1J	NM_001068949.2	NP_001062414.2	Not Cold	**16**
OsCBFIIIa-1A	AF300970	AAN02486	Cold	**5**
OsCBFIIIb-1H	AP008215	BAF25625	pseudo gene	**14**
OsCBFIV-1B	AY785894	AAX28958	Cold	**17**
		Pooideae		
Avena sativa				
AsCBFIIId-12	AM071409	CAJ21278	-	**3**
AsCBFIIId-16A	AM071406	CAJ21276	-	**3**
AsCBFIIId-16B	AM071407	CAJ21277	-	**3**
Brachypodium distachyon				
Bradi1g77120	XM_003562122	XP_003562170	Cold	**9**
Bradi1g57970	XM_003561396	XP_003561444	-	**9**
Bradi2g60331	XM_003567408	XP_003567456	Cold	**9**
Bradi2g60340	XM_003567409	XP_003567457	Cold	**9**
Bradi3g57360	XM_003578416	XP_003578464	-	**9**

Table 3.1 cont.

Bradi4g35570	XM_003576699	XP_003576747	Cold	**9**
Bradi4g35580	XM_003576700	XP_003576748	Cold	**9**
Bradi4g35590	XM_003576701	XP_003576749	Cold	**9**
Bradi4g35600	XM_003578416	XP_003578464	Cold	**9**
Bradi4g35610	XM_003578417	XP_003578465	Cold	**9**
Bradi4g35620	XM_003578418	XP_003578466	Cold	**9**
Bradi4g35630	XM_003578419	XP_003578467	Cold	**9**
Bradi4g35640	XM_003576700	XP_003576748	Cold	**9**
Bradi4g35650	XM_003578420	XP_003578468		**9**
Festuca arundinacea				
FaCBFIIIa-6	AJ717399	CAG30550	-	**25**
FaCBFIVa-2	AY423713	AAQ98965	-	**12**
Hordeum brevisubulatum				
HbCBFIVa-2	DQ250027	ABB54457		**22**
Hordeum vulgare				
HvCBF2B	AF442489	AAM13419	Cold	**24**
HvCBF11	AY785890	AAX23718		
HvCBF15A	EU593541	ACC63531		**8**
HvCBF15B	EU593539	ACC63529		**8**
HvCBFIa-1	AY785836	AAX23683	Drought	**17**
HvCBFII-5	AY785855	AAX23698	-	**17**
HvCBFIIIa-6	AY785860	AAX23701	Cold	**17**
HvCBFIIIc-3	AY785845	AAX23692	-	**17**
HvCBFIIIc-13	DQ095158	ABA01492	-	**17**
HvCBFIIIc-10A	AY785882	AAX23711	-	**17**
HvCBFIIId-12	DQ095157	ABA01491	Cold	**17**
HvCBFIVa-2A	AY785841	AAX23688	Cold	**17**

Table 3.1 cont.

HvCBFIVc-14	DQ095159	ABA01493	Cold	**17**
HvCBFIVd-9	AY785878	AAX23707	Cold	**17**
HvCBFIVd-4A	AY785849	AAX28949	Cold	**17**
Loilim perenne				
LpCBF2	AB258393	BAF36838		**19**
LpCBFIIIa-6	AY960831	AAX57275	-	**23**
Secale cereale				
ScCBFIa-11	EU194240	ABY59777	Cold	**4**
ScCBFII-5	EU194241	ABY59778	-	**4**
ScCBFIIIa-6	EU194242	ABY59779	Not cold	**4**
ScCBFIIIc-10	EU194243	ABY59780	Not cold	**4**
ScCBFIIIc-3A	EU194244	ABY59781	Not cold	**4**
ScCBFIIIc-3B	EU194245	ABY59782	-	**4**
ScCBFIIId-12	EU194246	ABY59783	Cold	**4**
ScCBFIIId-15	EU194247	ABY59784	Cold	**4**
ScCBFIIId-19	EU194248	ABY59785	Cold	**4**
ScCBFIVa-2A	EU194249	ABY59786	Cold	**4**
ScCBFIVa-2B	EU194250	ABY59787	-	**4**
ScCBFIVb-20	AF370728	AAL35759	Cold	**6**
ScCBFIVd-9A	AF370729	AAL35760	Cold	**6**
ScCBFIVd-9B	AF370730	AAL35761	-	**6**
Triticum aestivum				
TaCBFIa-A11	EF028751	ABK55354	-	**1**
TaCBFII-5	EF028752	ABK55355		**1**
TaCBFII-5.2	EF028753	ABK55356	Not cold	**1**
TaCBFII-5.3	EF028754	ABK55357	-	**1**
TaCBFIIIa-6	EF028755	ABK55358	-	**1**
TaCBFIIIa-6.2	EF028756	ABK55359	Cold	**1**

Table 3.1 cont.

TaCBFIIIa-D6	EF028757	ABK55360	-	**1**
TaCBFIIIc-3	EF028758	ABK55361	-	**1**
TaCBFIIIc-3.2	EF028759	ABK55362	-	**1**
TaCBFIIIc-D3	EF028760	ABK55363	Cold, Freezing	**1, 17**
TaCBFIIIc-B10	EF028761	ABK55364		**1**
TaCBFIIId-B12	EF028763	ABK55366	Cold	**1**
TaCBFIIId-A15	EF028764	ABK55367	-	**1**
TaCBFIIId-15.2	EF028765	ABK55368	Cold	**1**
TaCBFIIId-A19	EF028766	ABK55369	Cold	**1**
TaCBFIIId-B19	EF028767	ABK55370	-	**1**
TaCBFIIId-D19	EF028768	ABK55371	-	**1**
TaCBFIVa-2.2	EF028770	ABK55373	-	**1**
TaCBFIVa-2.3	EF028771	ABK55374	-	**1**
TaCBFIVb-21	EF028775	ABK55378	-	**1**
TaCBFIVb-A20	EF028772	ABK55375	-	**1**
TaCBFIVb-B20	EF028773	ABK55376		**1**
TaCBFIVb-D20	EF028774	ABK55377	Not cold	**1**
TaCBFIVb-D21	EF028776	ABK55379	Cold	**1**
TaCBFIVc-B14	EF028778	ABK55381	Cold	**1**
TaCBFIVd-4	EF028780	ABK55383	-	**1**
TaCBFIVd-9	EF028782	ABK55385	-	**1**
TaCBFIVd-B4	EF028781	ABK55384	-	**1**
TaCBFIVd-B9	EF028783	ABK55386	-	**1**
TaCBFIVd-D9	EF028784	ABK55387	Cold	**1**
TaCBFIVd-B22	EF028786	ABK55389	Cold	**1**
TaCBFIVd-D22	EF028787	ABK55390	-	**1**
TaCBFIVa-A2 (*TaCBF2.1*)	EF028769	ABK55372	Cold	**1**

Table 3.1 cont.

TaCBF2.2	JF758493	AEE00130	-	**2**
TaCBFIVd-4.1 (*TaCBF4.0*)	EF028780	ABK55383	-	**1**
TaCBF9.0	JF758499	-	-	**2**
TaCBF12.1	JF758493	AEE00129	-	**2**
TaCBFIIId-12.1 (*TaCBF12.2*)	EF028762	ABK55365	Cold, Freezing	**1, 17**
TaCBFIVc-14.1 (*TaCBF14.1*)	EF028777	ABK55380	Freezing	**1,17**
TaCBFIVc-14.3 (*TaCBF14.2*)	EF028779	ABK55382	-	**1**
TaCBFIIId-15.2 (*TaCBF15.0*)	EF028765	ABK55368	Freezing	**1,17**
TaCBF17.0	JF758499	-	-	**2**
TaCBF19.1	JF758498	-	-	**2**
TaCBF19.2	JF758493	-	-	**2**
TaCBF20.0	JF758492	AEE00126	-	**2**
TaCBF21.0	JF758492	AEE00127	-	**2**
TaCBFIVd-A22 (*TaCBF22.0*)	EF028785	ABK55388	-	**1**
Triticum monococcum				
TmCBF3	AY951949	AAY32553	-	**15**
TmCBF9	AY951945	AAY32563	-	**15**
TmCBF12	AY951944	AAY32557	Cold	**15**
TmCBF14	AY951948	AAY32552	Cold	**15**
TmCBF15	AY951944	AAY32556	Cold	**15**
TmCBFII-5	AY951947	AAY32551	Not cold	**15**
TmCBFIIIb-18	AY951946	AAY32550	-	**15**

Table 3.1 cont.

TmCBFIIIc-10	AY951950	AAY32554	-	**15**
TmCBFIIIc-13	AY951951	AAY32555	-	**15**
TmCBFIIId-16	AY951944	AAY32558	-	**15**
TmCBFIIId-17	AY951945	AAY32564	-	**15**
TmCBFIVa-2	AY951945	AAY32560	Not Cold	**15**
TmCBFIVd-4	AY951945	AAY32562	Not Cold	**15**
		Panicoideae		
Panicum virgatum				
PvCBFIa-11	DN144490	-	-	**20**
PvCBFII-5	DN145297	-	-	**20**
PvCBFIIIa-6	DN143145	-	-	**20**
Sorghum bicolor				
SbCBFIa	JN853584	AFP33239	-	**16**
SbCBFII-5	AY785898	AAX28959	-	**17**
SbCBFIIIa-6	AY785899	AAX28960	-	**17**
Zea mays				
ZmCBF1	NM_001146976	NP_001140448	-	**10**
ZmCBF4	NM_001177010	NP_001170481	Cold	**10**
ZmCBFIIIb-1A	AF450481	AAN76804	Cold	**20**

[a] Gene names used in this study are in red and previously used names are in black font.

[b] **1** Badawi *et al*., 2007; **2** Båga *et al*.; **3** Brautigam *et al*. 2005; **4** Campoli *et al*., 2009; **5** Dubouzet *et al*., 2003; **6** Haake *et al*., 2001; **7** Kikuchi *et al*., 2001; **8** Knox *et al*., 2010, **9** Li *et al*., 2012; **10** Liu *et al*., 2004; **11** Liu *et al*., 2006; **12** Lv *et al*., 2003; **13** Mao and Chen, 2012; **14** Matsumoto *et al*., 2009; **15** Miller *et al*., 2006; **16** Quanhong *et al*., 2003; **17** Skinner *et al*., 2005, 2009; **18** Soderlund *et al*., 2009; **19** Tamura *et al*., 2007; **20** Tobias, 2005; **21** Vágújfalvi 2005; **22** Wang *et al*., 2009; **23** Xiong *et al*. 2006; **24** Xu, 2003; **25** Yang *et al*., 2005.

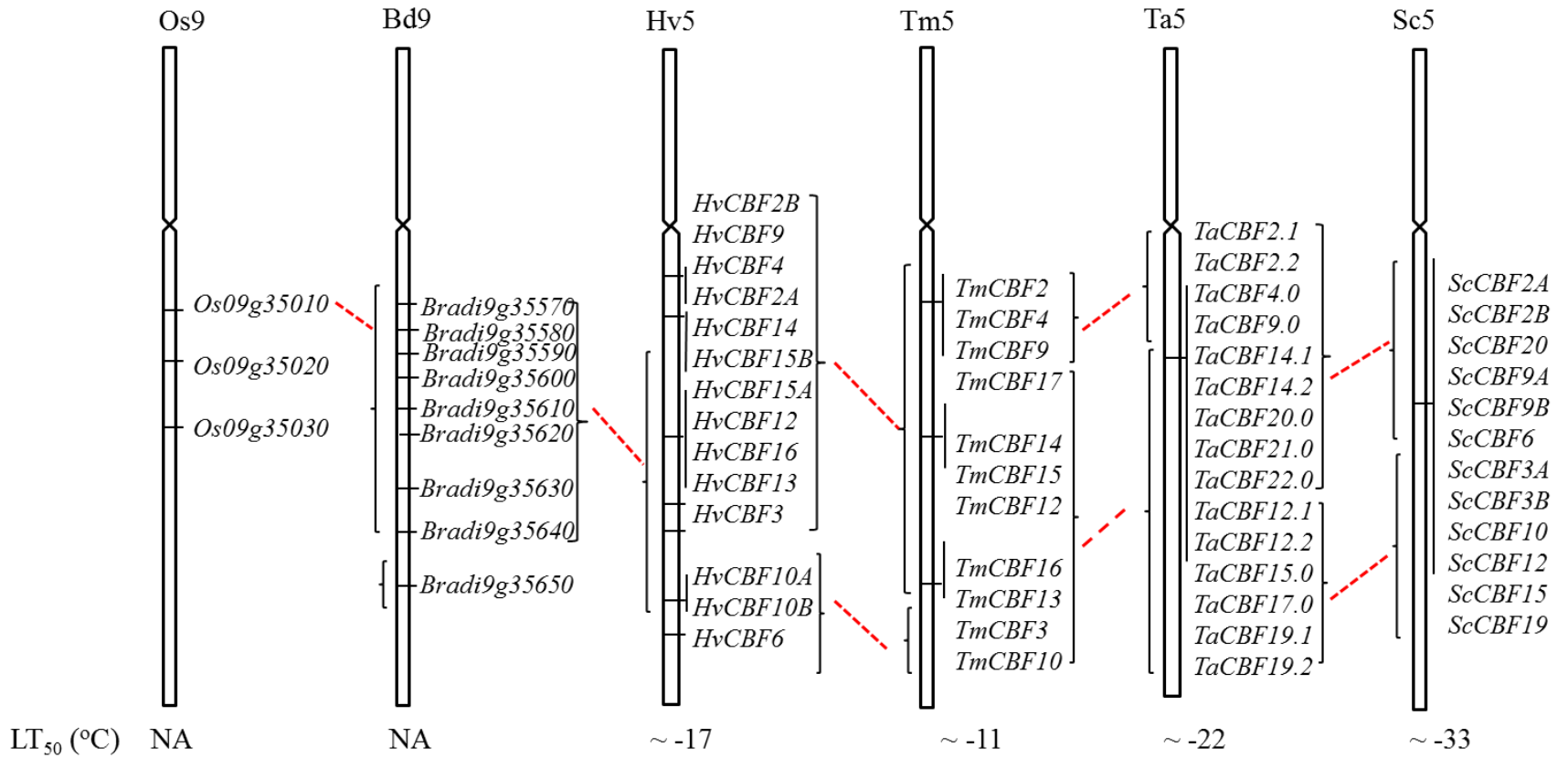


Figure 3.1 CBF genes among the *Fr-A2* colinear region from close relatives of Poaceae family members (rice, *Brachypodium*, barley, rye and wheat).
Gene numbering for Brachypodium and rice is derived from Phytozome (http://www.phytozome.net). Barley gene order was derived from Francia *et al*., 2007 and Knox *et al*., 2010. Distance among genes are not to scale. Rye gene numbering is derived from Campoli *et al*., 2007 and wheat numbering is from Miller *et al*., 2006, Båga *et al*. (unpublished, gene order for Ta5 has not been determined).

of the region. Nevertheless, at least 20 CBF genes exist at *Fr-H2* locus of *Hordeum vulgare* cv Dicktoo (Choi *et al*., 2002, Xue, 2002; Francia *et al*., 2004; Skinner *et al*., 2006), ≥ 11 CBF genes at *Fr-A^{m}2* of *T. monococcum* (Miller *et al*., 2006) and at least 23 at *Fr-A2* of *T. aestivum* cv Norstar (Båga *et al*., unpublished). For rye cv Puma, 12 CBF genes based on EST sequences were mapped to chromosome arm 5RL (Campoli *et al*., 2009; Jaglo *et al*., 2001) and presumed to represent *Fr-2* CBF genes. Since only one report of rye CBF genes is available, the number of CBF genes at rye *Fr-2* may be much larger than 12.

3.4.2 Amino acid - level cluster analysis of CBFs

To compare protein sequences of CBFs produced by *Triticum aestivum* cv Norstar to other monocot CBFs, a comprehensive search of the literature and GenBank databases for CBF sequences from the Poaceae family was conducted. From this search, a total of 130 sequences with representatives within the Ehrhartoideae, Panicoideae, and Pooideae subfamilies were selected for analysis (Table 3.1). The sequences were compiled into three datasets composed of full - length CBF sequences, amino - terminal sequence including first signature sequence, AP2 - DNA binding domain and C - terminal *trans* - activation domain including second signature sequence, respectively. For each analysis, only full - length sequences were included in the analyses; thus, protein sequences derived from 5' and 3' truncated DNA sequences of *Secale cereale* and *Panicum virgatum* (see Table 3.1) were included only for AP2 - DNA binding domain analyses. Each data set was aligned with ClustalW, T - Coffee and Expresso multiple sequence alignment server and used to generate cluster tree. No difference was seen between the clusters obtained using different alignment and / or by diferent clustering algorithm (data not shown).

Cluster analysis based on amino acid sequences of 130 monocotyledon CBFs show that the CBFs cluster into four distinct monophyletic groups (A - D) which could be further divided into four subgroups (I to IV) each with two to five distinct clades (a - e) (Fig. 3.2). The group A shows distinct clade (Fig. 3.2) containing CBFs from Poales subfamilies Erhartoideae and Pooideae including two rice (*OsCBF1a-1E* and *OsCBFIa-1G*) and two barley (*HvCBF1-1* and *HvCBF11*) and single wheat (*TaCBFIa-11*) CBF. Group A CBF proteins have been reported to show maximum homology with ancestral type dicot CBFs suggesting their closer relationship

(Skinner *et al*., 2005). The group B consists of proteins from all three subfamilies of Poales, but has only one rice protein (*OsCBFII-1C*). These were initially classified as group A (Skinner *et al*., 2005) but were later divided by Badawi *et al*. (2007) due to structural differences.

Group C clustered into several distinct subgroups, named as IIIa, IIIb, IIIc and IIId. Subgroups IIIa and IIIb contained all three Poaceae family members (Oryazceae, Pooideae and Panicoideae) with a single Brachypodium member each (*Bradi4g35650* and *Bradi1g77120* respectively) while IIIc and IIId contained members of Pooideae subfamily only. These differences suggest that IIIa and IIIb members were present before divergence of the subfamilies, as previously observed (Badawi *et al*., 2007).

Interestingly, subgroup IIIc contains CBF from wheat and barley and a single *Brachypodium* member (*Bradi4g35630*). Barley members of this and none of the Norstar CBFs, identified in this study, are members of this subgroup. Subgroup IIId can be further divided into five distinct clades. Clade IIId - 1 contains only *Brachypodium* members. Clade IIId - 2 consists of two oat (*AsCBFIIId-12* and *AsCBFIIId-16B*), one diplod wheat (*TmCBFIIId-16*) and one perennial ryegrass (*LpCBF2*) members. CBF15 homologs from diploid and hexaploid wheat as well as barley clustered on clade IIId - 3 and CBF12 and 19 homologs clustered on clade IIId - 5. CBF17 from diploid and hexaploid wheat form a separate clade, IIId - 4. Six *Triticum aestivum* cv Norstar CBFs from present study were present within subgroup.

The analysis of group D reveals that it contains mainly members of wheat and barley with four CBFs (*FaCBFIVa-2*, *ScCBFIVd-9A*, *ScCBFIVd-9B*, and *ScCBFIVb-20*) from *Festuca arundiacea* and *Secale cereale*. The group has been further divided into four subgroups IVa, IIId only. IVb, IVc and IVd. Nine *Triticum aestivum* cv Norstar CBFs from present study were present in group IV, with subgroups IVa, IVb, IVc having two CBFs each and IVd having three CBFs. All the subgroups show quite compact clustering with the exception of few outliers like *OsCBFIII-1D*, *OsCBFIV-1B*, *OsCBFI-1F* and *FaCBFIVa-2* could not be grouped in any of the subgroups. Also, *TmCBFIVd-4* which was earlier grouped with IVd (Badawi *et al*., 2007) could be placed with IVc members in this study.

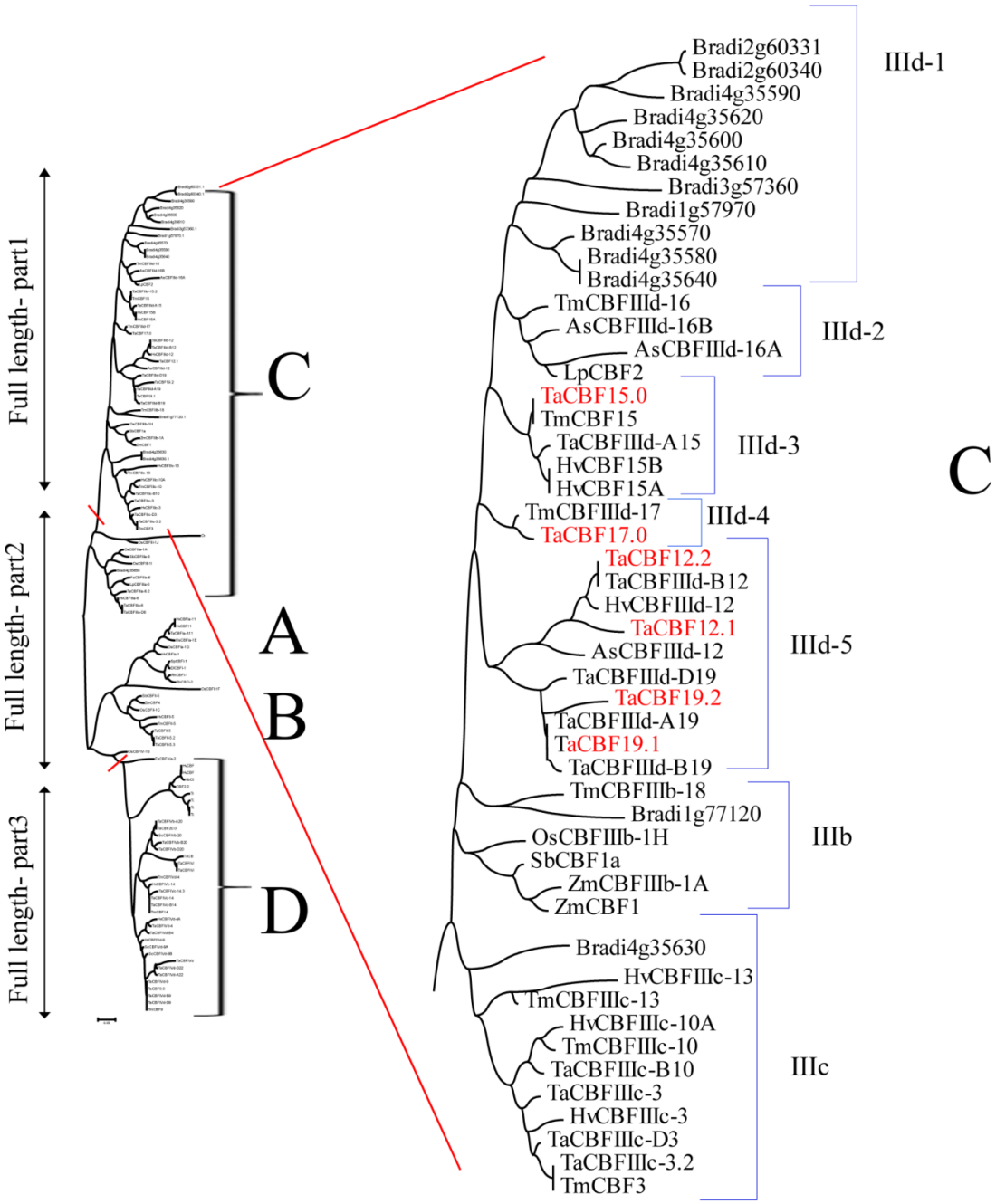


Figure 3.2 Cluster analysis based on full length CBF amino acid sequences (Part1 enlarged)

Norstar CBFs are in red font and other species CBFs are in black font. Data set is listed in Table 3.1. Sequences were aligned using ClustalW2 (Appendix Table 3.1). A neighbor join tree was derived from the alignment using the Poisson Correction parameter. CBFs belonging to specific monoplyletic groups are marked by open bracket.

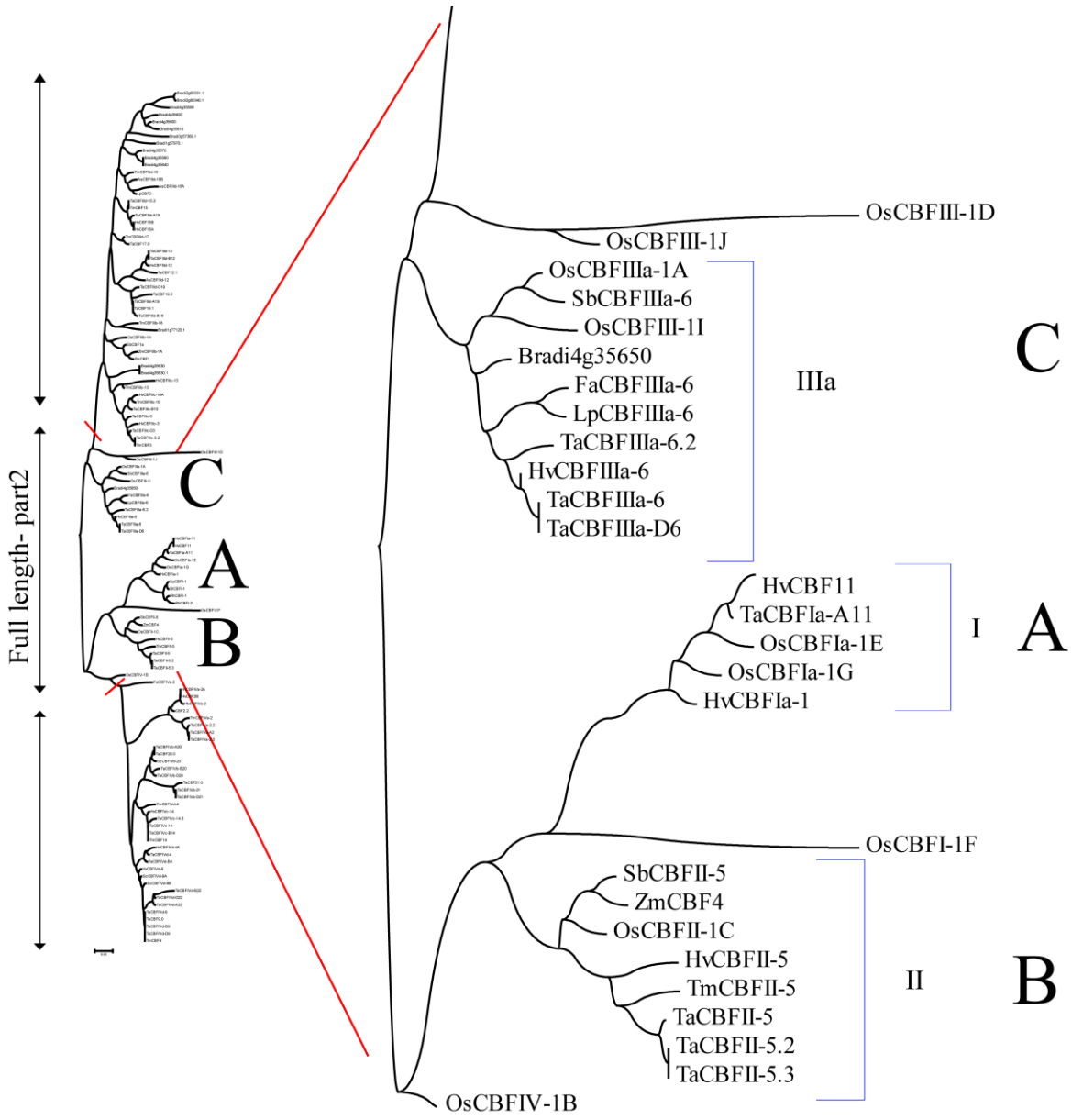


Figure 3.2 Cluster analysis based on full length CBF amino acid sequences (Part2 enlarged).

Norstar CBFs are in red font and other species CBFs are in black font. Data set is listed in Table 3.1. Sequences were aligned using ClustalW2 (Appendix Table 3.1). A neighbor join tree was derived from the alignment using the Poisson Correction parameter. CBFs belonging to specific monoplyletic groups are marked by open bracket

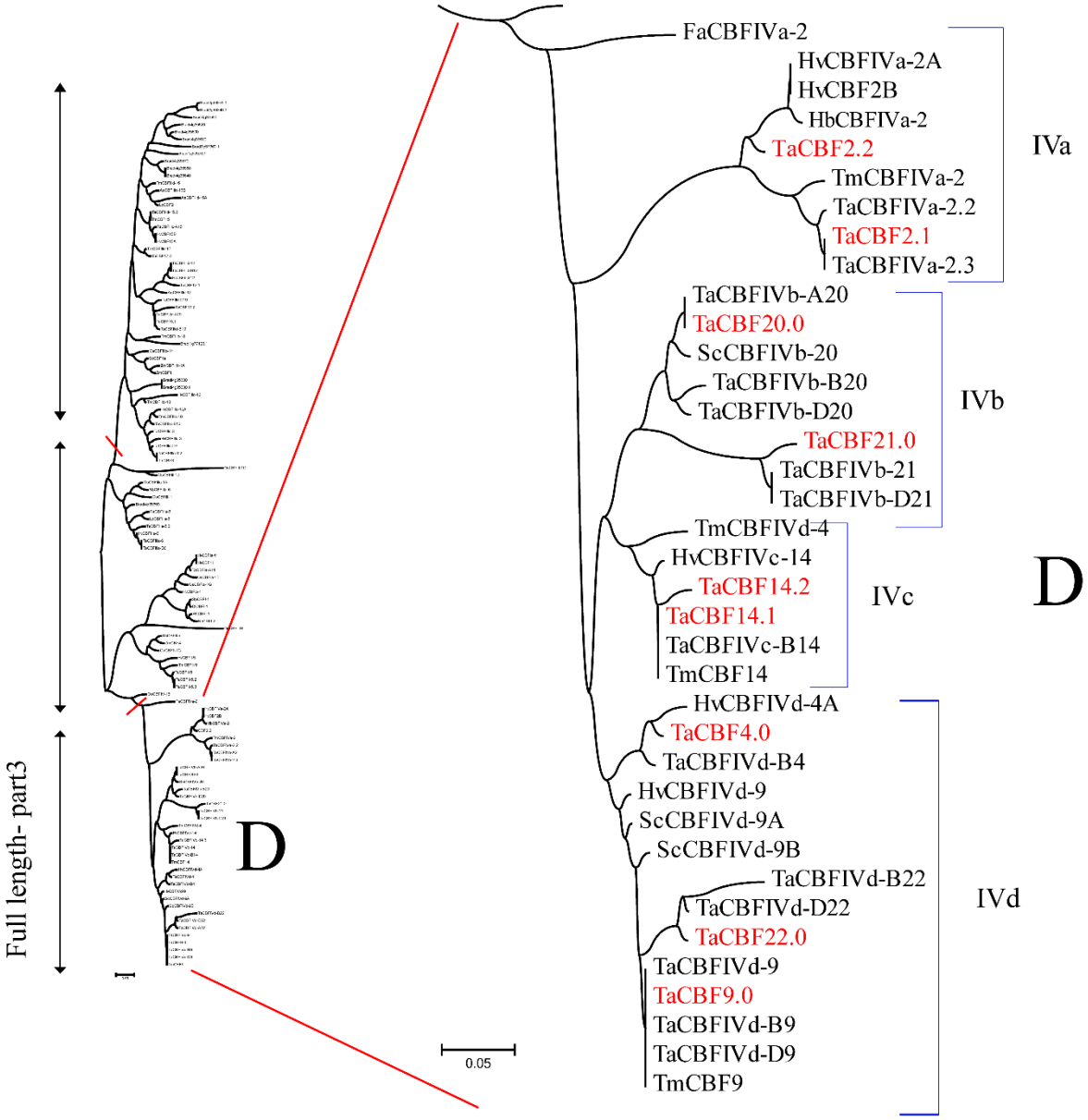


Figure 3.2 Cluster analysis based on full length CBF amino acid sequences (Part3 enlarged).

Norstar CBFs are in red font and other species CBFs are in black font. Data set is listed in Table 3.1. Sequences were aligned using ClustalW2 (Appendix Table 3.1). A neighbor join tree was derived from the alignment using the Poisson Correction parameter. CBFs belonging to specific monoplyletic groups are marked by open bracket

3.4.3 Cluster analyses of amino terminal region

Alignment of the amino terminal peptide, 35 - 61 amino acid long, preceding the AP2-DNA binding domain, shows presence of putative nuclear localization signal within the first signature motif (CMIII - 3) as a cluster of basic amino acids (KRPxxRxKxKxxR) as suggested Medina *et al*. (1999). Also, cluster of serine residues preceding the CMIII - 3 motif is seen in all CBFs, ranging from 10 - 35 % of the peptide. Group I and III members show higher frequency of serine residues than group II and IV members (Appendix Table 3.2).

3.4.4 Cluster analysis on the basis of AP2 - DNA binding domain sequence

The complete CBF protein can be divided into three domains, N - terminal nuclear localization domain, a central AP2 - DNA binding domain and a C - terminal trans - activation domain. Compared to complete protein sequence, cluster analysis based on AP2 - DNA binding domain only (Fig. 3.3), showed some variations, which suggest important consequences in the overall bioactivity of the protein. Group A consisted of CBF from Poales. Group B consists of the same members as seen on basis of full length of CBF protein cluster analysis. In the group C, sub - group IIIa splits into two clades and OsCBFIII-1 D which was an outlier based on complete sequence, fits into clade IIIa - 2 with CBF6 homologs from *Festuca arundinacea*, and *Lolium perenne*. *PvCBFIIIa-6* clustered on this clade as well. Rest of the group members remain the same. Group IIIb also splits into two clades, with clade IIIb - 1 containing *TmCBFIIIb-18* and *Bradi1g77120*, and IIIb - 2 having *OsCBFIIIb-1H*, *ZmCBF1*, *ZmCBFIIIb-1A*, and *Bradi4g35630*. Sub - group IIIc remains undisturbed with same clade members with addition of *ScCBFIIIc-3A* and *ScCBFIIIc-3B* (these genes had partial sequence available and were included in AP2 domain analysis only). Interestingly, the group IIId CBF members, separate into four different clades when analyzed on the basis of AP2-DNA binding domain sequence only. Clade IIId - 1 continued to contain only *Brachypodium* members and clade IIId - 2 included CBF 15 and CBF16 homologs from Triticeae. CBF17 homolgs, previously member of clade IIId - 4 on basis of full length sequence, forms a clade IIId - 3, with CBF19 homologs when analyzed on the basis of AP2 DNA binding domain only. All Triticeae CBF12 homologs form clade IIId - 4. In group D, the wheat and barley members remain the same with addition of rye CBFs (*ScCBFIVd-9A*, *ScCBFIVd-9B*, *ScCBF2* and *ScCBF3*).

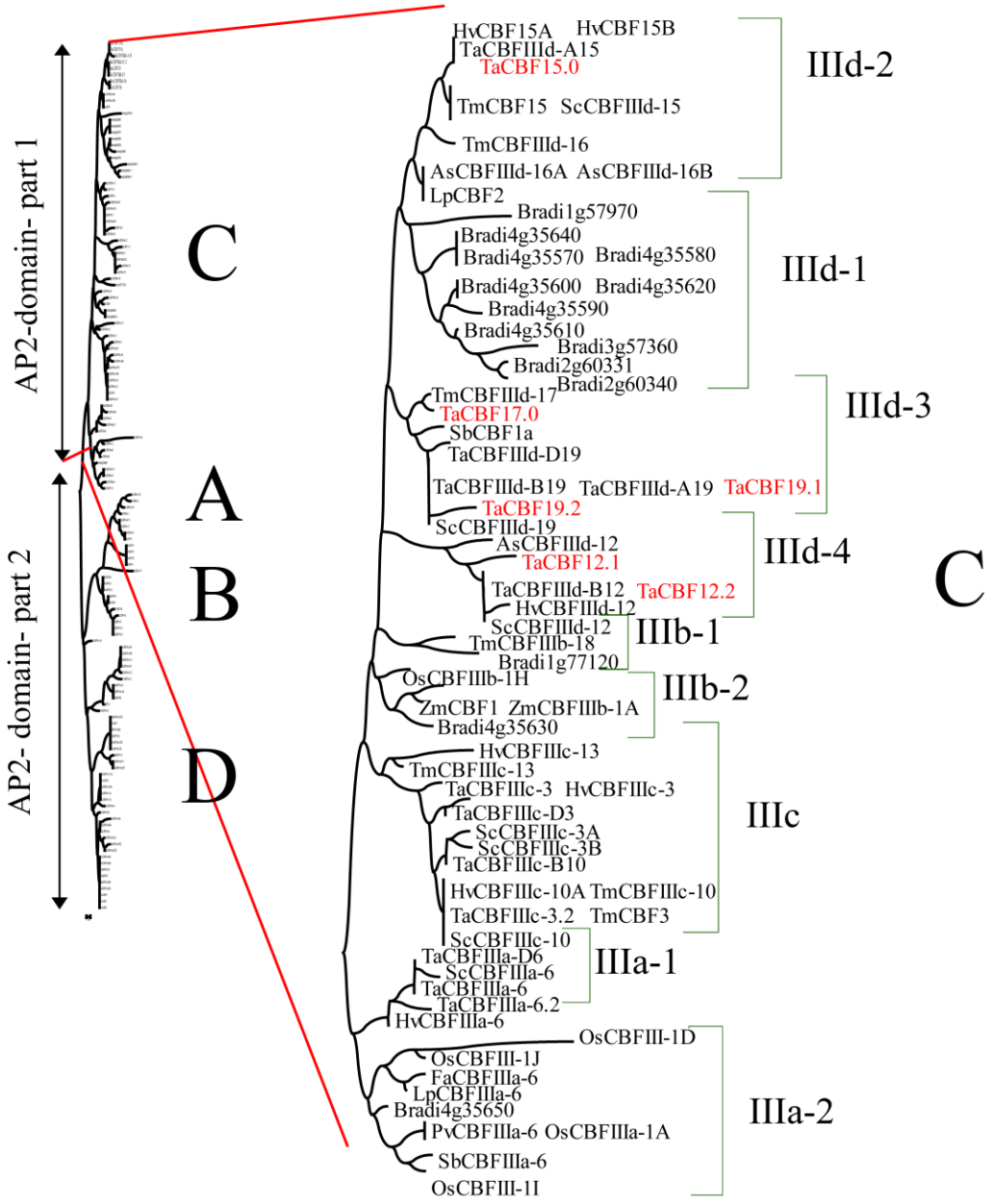


Figure 3.3 AP2-DNA binding domain sequence based cluster analysis. (Part1 enlarged)

Norstar CBFs are in red font and other species CBFs are in black font. Data set is listed in Table 3.1. Sequences were aligned using ClustalW2 (Appendix Table 3.3). A neighbor join tree was derived from the alignment using the Poisson Correction parameter. CBFs belonging to specific monoplyletic groups are marked by open bracket.

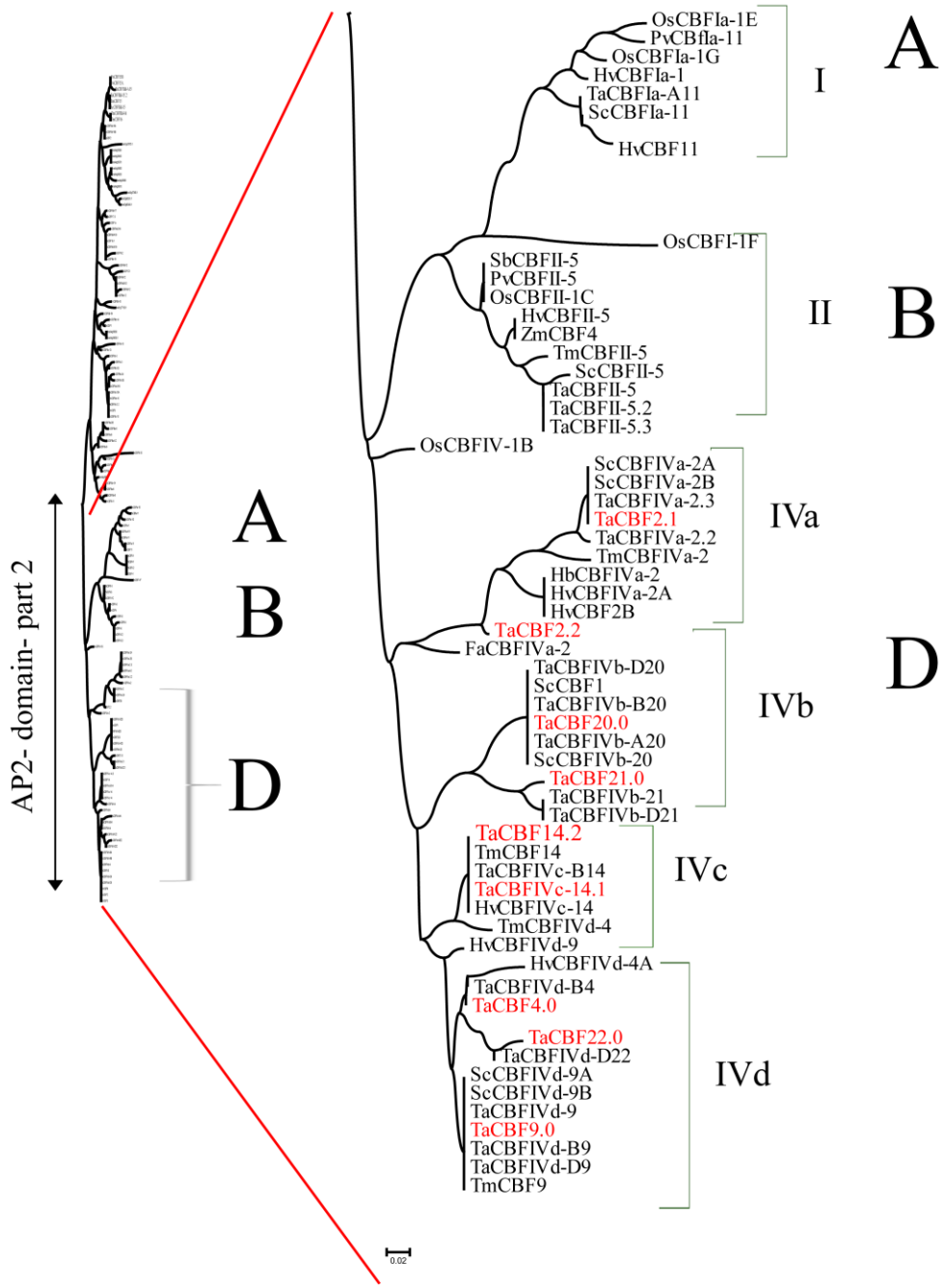


Figure 3.3 AP2-DNA binding domain sequence based cluster analysis. (Part1 enlarged)

Norstar CBFs are in red font and other species CBFs are in black font. Data set is listed in Table 3.1. Sequences were aligned using ClustalW2 (Appendix Table 3.3). A neighbor join tree was derived from the alignment using the Poisson Correction parameter. CBFs belonging to specific monoplyletic groups are marked by open bracket.

3.4.5 Cluster analysis on the basis of C - terminal trans - activation domain sequence

The C - terminal sequence is much more variable between and within the groups. Analysis based on the C - terminal sequence only, showed shift in distribution of the group members (Fig. 3.4). Comparison of groups based on complete sequence with the C - terminal based groups show that the groups that contained members from all three Poales subfamilies Oryzaceae, Panicoideae and Poodideae, namely group I and II, form a clear monophyletic group. Sub-group IIIa also remained undisturbed with same members. Sub - group IIIb members did not form close clades as was observed on the basis of AP2 - DNA binding domain and full length amino acid sequences. *TmCBFIIIb-18* and *Bradi1g77120* remain as separate clade, and *OsCBFIIIb-1H* grouped with clade IIId - 4 members. Members of clade IIIb - 2 remain the same. Sub - group IIIc remains same in all the three sequence analysis except *Bradi4g35630* moves to clade IIIc from IIIb - 2 from analysis on the basis of AP2 - domain sequence. Members of clade IIId could be further divided into four separate clades.

Clade IIId - 1 had fewer *Brachypodium* members as compared to the clade based on full length and AP2 domain sequences. Clade IIId - 2 had only CBF15 homologs from diploid and hexaploid wheat and barley. Clade IIId - 3 has CBF17 homologs from diploid and hexaploid wheat along with *Brachypodium* and oat CBFs. Clade IIId - 4 is similar to clade IIId - 5 from full length sequence analysis, having CBF12 and CBF19 homologs. Variation seen in groups IVa is same as that seen on basis of AP2 domain only, i.e. wheat and barley members remain the same while *FaCBFIVa-2* forms a separate clade. *TaCBF21.0* of group IVb, on the basis of full length and AP2 domain sequence, moves to separate clade in sub group IVc, when analyzed on the basis of C - terminal sequence and other members of the sub - group remain the same. Members of group IVd include only CBF9 and CBF22 variants from wheat, rye and barley. CBF20 variants from hexaploid wheat form two separate clades, IVb - 1 having only *TaCBFIVb-B20* and *TaCBFIVb-D20* and other clade IVb - 2 having *TaCBFIVb-A20* and *TaCBF20.0* as its members. *ScCBFIVb-20* could not be placed in any clade.

3.4.6 Bioinformatic analysis of wheat CBFs

Norstar *TaCBF14.1* AP2 - DNA binding domain and the flanking signature sequences of group C and *TaCBF12.2* of group D were considered as representative sequences for respective group members.

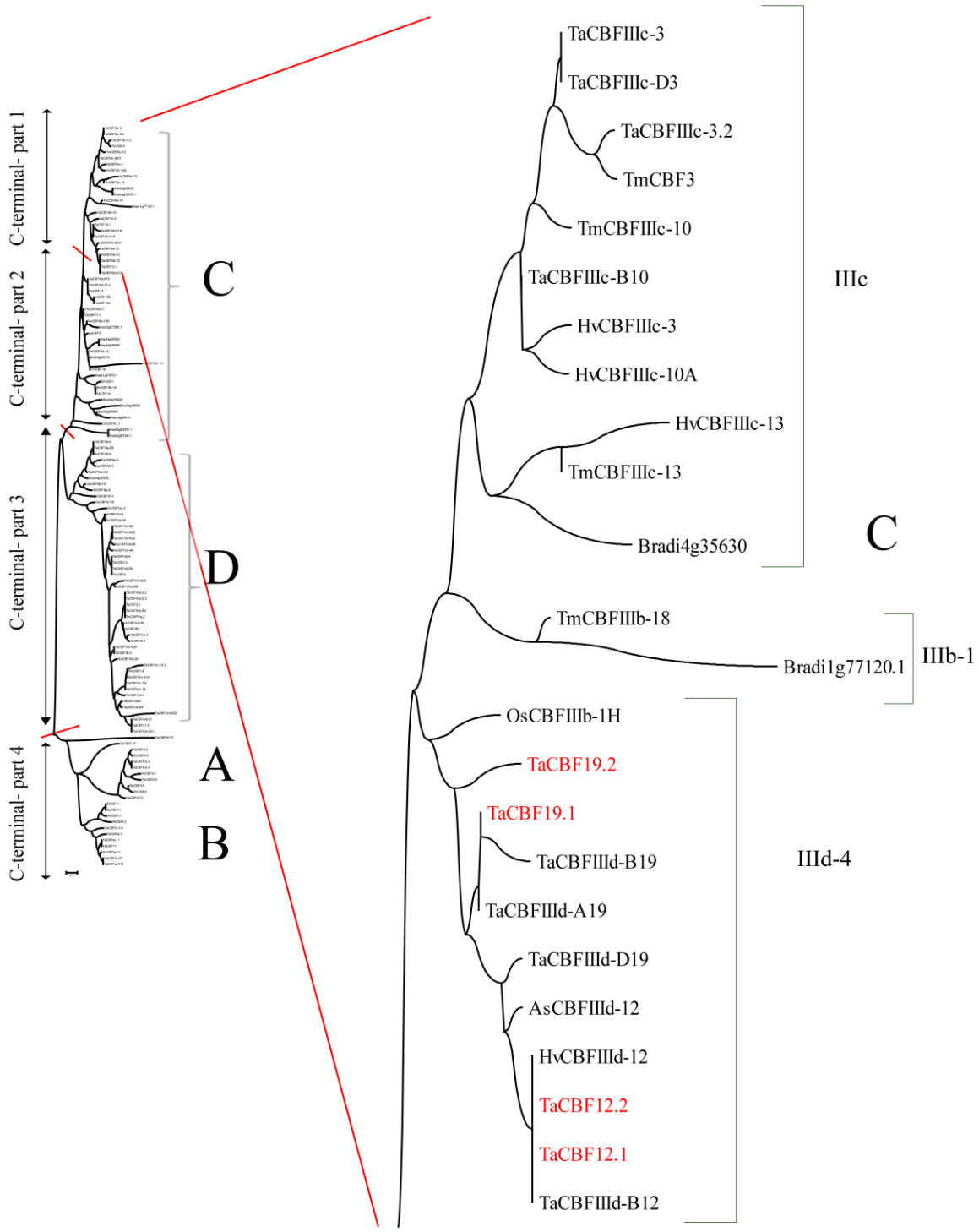


Figure 3.4 Cluster analysis based on C - terminal sequences. (Part 1 enlarged).

Norstar CBFs are in red font and other species CBFs are in black font. Data set is listed in Table 3.1. Sequences were aligned using ClustalW2 (Appendix Table 3.4). A neighbor join tree was derived from the alignment using the Poisson Correction parameter. CBFs belonging to specific monoplyletic groups are marked by open bracket

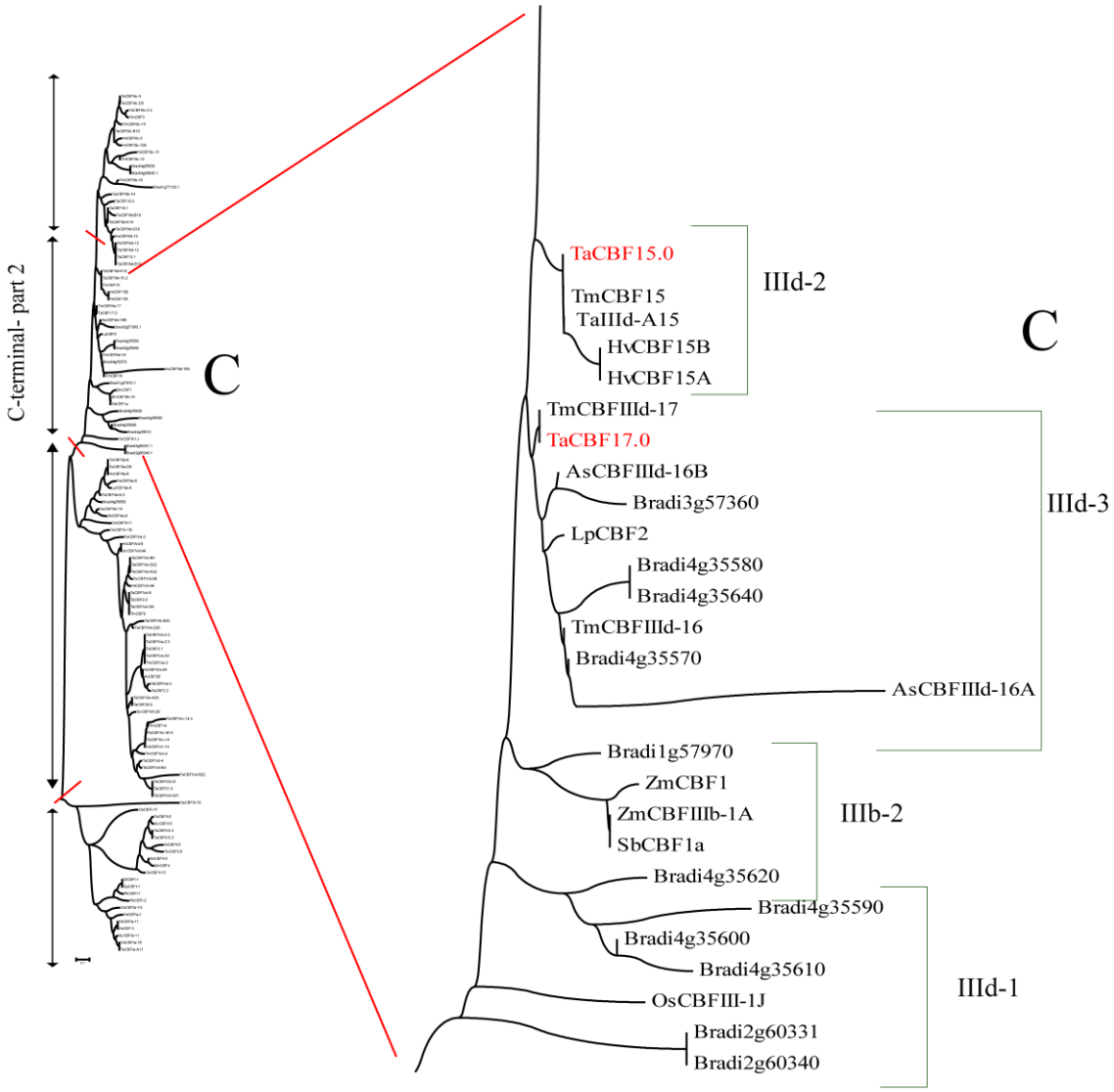


Figure 3.4 Cluster analysis based on C - terminal sequences. (Part 2 enlarged).

Norstar CBFs are in red font and other species CBFs are in black font. Data set is listed in Table 3.1. Sequences were aligned using ClustalW2 (Appendix Table 3.4). A neighbor join tree was derived from the alignment using the Poisson Correction parameter. CBFs belonging to specific monoplyletic groups are marked by open bracket

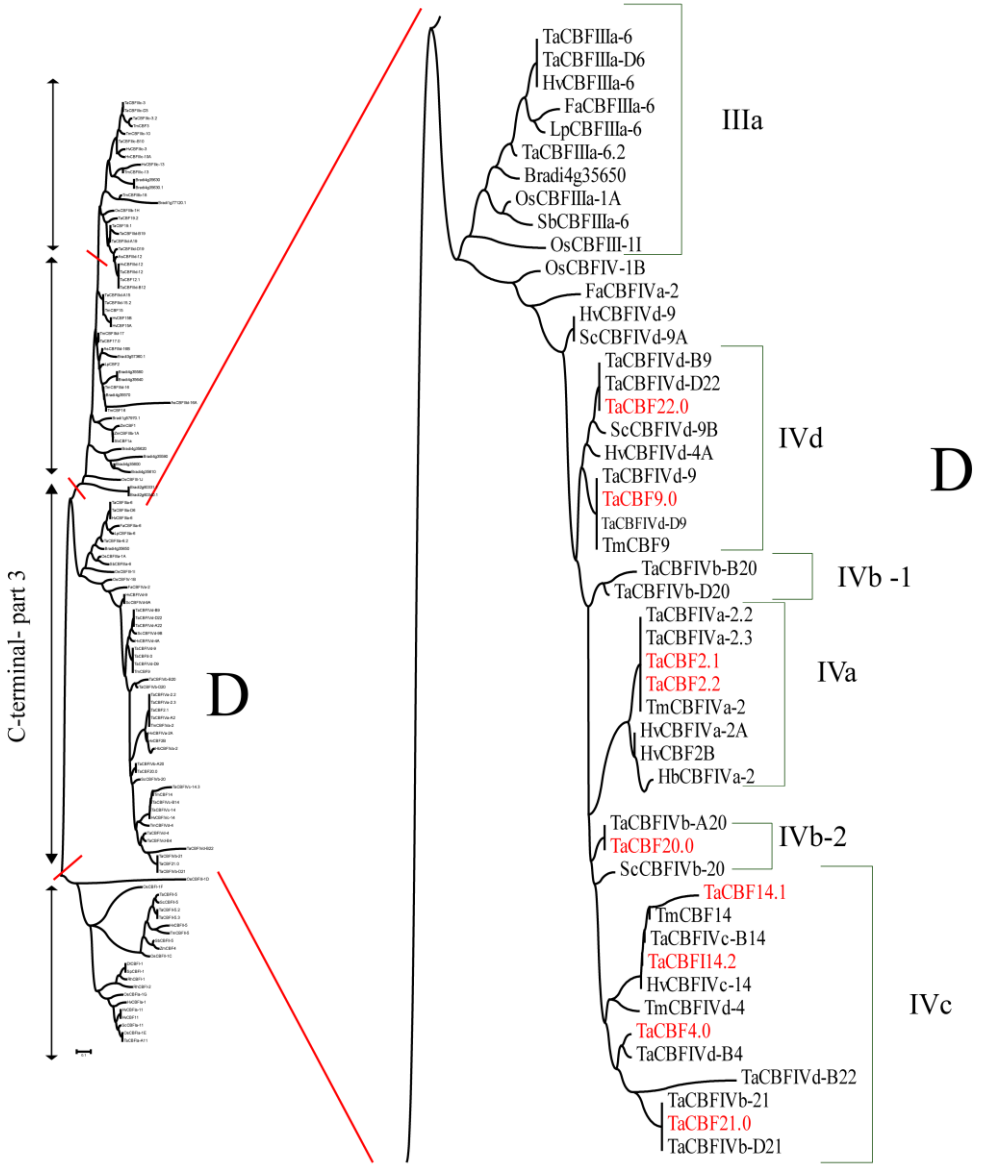


Figure 3.4 Cluster analysis based on C - terminal sequences. (Part 3 enlarged).

Norstar CBFs are in red font and other species CBFs are in black font. Data set is listed in Table 3.1. Sequences were aligned using ClustalW2 (Appendix Table 3.4). A neighbor join tree was derived from the alignment using the Poisson Correction parameter. CBFs belonging to specific monoplyletic groups are marked by open bracket

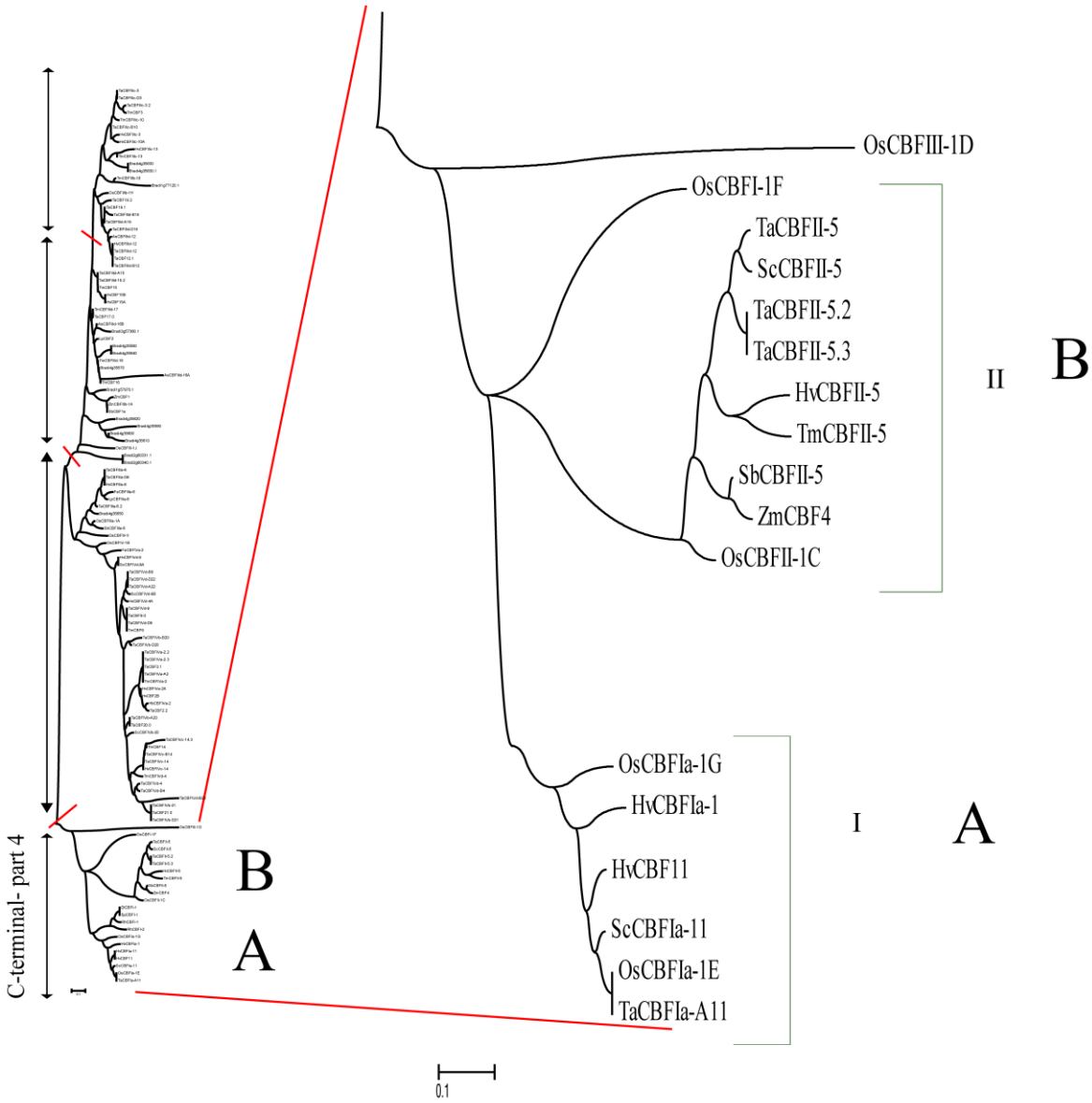


Figure 3.4 Cluster analysis based on C - terminal sequences. (Part 4 enlarged).

Norstar CBFs are in red font and other species CBFs are in black font. Data set is listed in Table 3.1. Sequences were aligned using ClustalW2 (Appendix Table 3.4). A neighbor join tree was derived from the alignment using the Poisson Correction parameter. CBFs belonging to specific monoplyletic groups are marked by open bracket

Alignment of rye and Norstar CBFs reveal six rye CBFs are in group C and five in group D (Fig. 3.5). Group D sequence analysis shows that there some difference between the group members. Norstar *TaCBF15.0* is a perfect match to *ScCBF15* and *ScCBF19.1* is exact match to TaCBF19. DNA contacting residues (Allen *et al*., 1998) are conserved among all the group members. Group C wheat and rye CBF members are expressed for longer time at 6 ^{o}C (Campoli *et al*., 2009).

Group D members of rye and wheat are similar. For every rye CBF there is an exact match present in the Norstar CBFs. *TaCBF 2.1* from Norstar has exactly same sequences as *ScCBF2A* and *ScCBF2B,* Norstar *TaCBF9.0* is same as *ScCBF9A* and *ScCBF9B* and *TaCBF20.0* is same as *ScCBF20*. However, rye has two copies of *ScCBF2* and *ScCBF9*. This subtle difference among the sequences and presence of more copies CBFs might result in overall functional differences and might be the underlying reason for the difference in the low temperature tolerance between rye and wheat.

3.4.7 Bioinformatic analysis of signature sequences

The dicot and monocot CBFs share similar primary domain structure, mainly conserved AP2 - domain and flanking CBF signature motifs (Jaglo *et al*., 2001; Skinner *et al*., 2005). These signature motifs were used to identify monocot CBFs from other AP2 - domain family members. Sequence alignment of CBF from *T. aestivum* cv Norstar shows some variation (Table 3.2) from the reported signature motifs from barley CBF family which is considered a representative of cereals (Skinnner *et al*., 2005). The CMIII - 3 motifs from wheat can be defined for group C as xKRPAGRTKF/LKE/DTRHP and CMIII - 1 as DSAW/CL. Similarly, for group D the motifs can be defined as PKR/WP/RAGRxKxxETRHP and DSAWR. The function of these signature sequences is not knownat this time, these differences might be of significance in influencing the overall functional properties of the protein.

We also compared the AP2 domain sequence from *Triticum aestivum* cv Norstar amino acid sequence from the two groups, C and D and found it to be highly conserved with few differences (Fig. 3.5). *TaCBF 12.2* from group C shows an insertion of two amino acids, whereas *TaCBF19.1* and *TaCBF19.2* show an insertion of three amino acids after α - helix sequence. In addition, group D member *TaCBF2.1* show an insertion of single amino acid between β - sheet two and three. These changes might affect the stability / structure of the protein thereby affecting their functional properties.

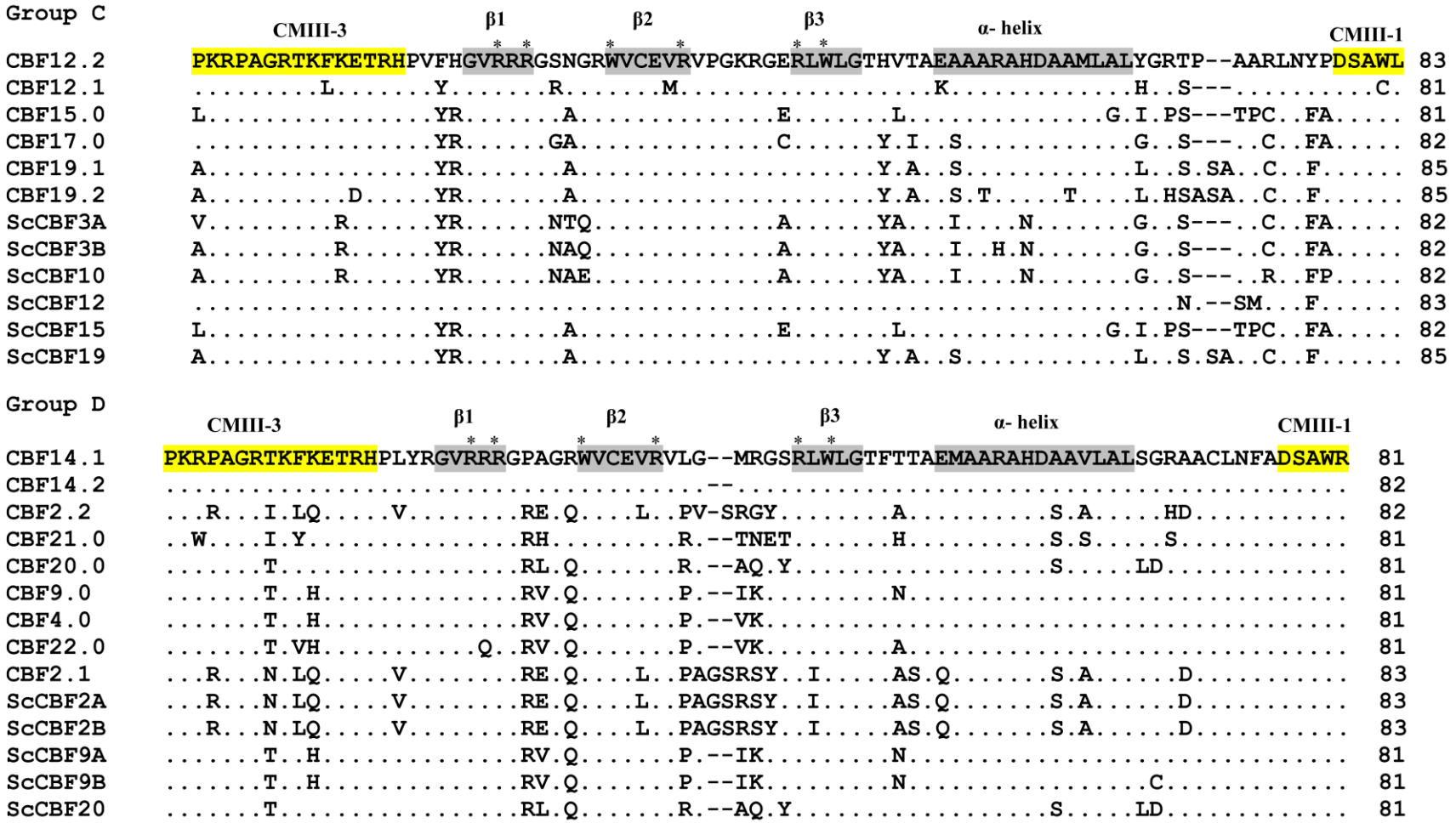


Figure 3.5 Amino acid sequence alignment of AP2-DNA bind domain from wheat and rye CBFs.

AP2 - DNA binding domain and flanking signature sequence (CMIII - 3 and CMIII - 1) from *Triticum aestivum* cv. Norstar and *Secale cereale* cv. Puma were aligned using ClustalW2. α - helix and β - sheet regions making residues are highlighted in grey and conserved signature sequences (CMIII - 3 and CMIII - 1) are highlighted in yellow. Asterisks represent DNA contacting residues within the AP2 domain as identified by Allen *et al*., (1998).

Table 3.2. Consensus and variant amino acids in signature sequences (CMIII - 3 and CMIII - 1) flanking the AP2 - DNA binding domain of CBFs among different species (wheat, rye, barley, rice and *Brachypodium*), with most variation occurring in amino terminal motif (CMIII - 3) of Norstar CBFs. Barley CBF signature sequences are taken as representative for each group.

Group C	**CMIII-3**	**CMIII-1**
HvCBF	$^{P}/_{A}$ K R P A G R T K F $^{K}/_{R}$ E T R H P	D S A $^{W}/_{E}$ L
TaCBF	$^{P}/_{A}/_{L}$ . . . . . K $^{E}/_{D}$	. . . $^{W}/_{C}$ L
ScCBF	$^{P}/_{V}/_{A}/_{L}$	 L
Group D		
HvCBF	P K R $^{P}/_{R}$ A G R $^{T}/_{I}$ K $^{F}/_{L}$ $^{K}/_{H}/_{Q}$ E T R H P	D S A W R
TaCBF	. . $^{R}/_{W}$ $^{P}/_{R}$ . . . $^{T}/_{I}/_{N}$. $^{F}/_{Y}/_{V}/_{L}$	. . . . .
ScCBF	. . . . . $^{N}/_{T}$	. . . . .
Others		
OsCBF	$^{P}/_{K}$ K R P A G R T F F R E T R H P	D S A W $^{R}/_{L}$
BdCBF	P K R P A G R T F F K E T R H P	D S A W L

Hv *Hordeum vulgare*; Ta *Triticum aestivum*; Sc *Secale cereale*; Bd *Brachypodium distachyon*; Os *Oryza sativa*.
(.) same residue; (/) presence of either / or residue.

3.4.8 Hydrophobic Cluster Analysis (HCA)

Hydrophobic cluster analysis reveals differences on the internal surface of secondary structure of proteins and was done previously on CBF protein AP2 - DNA binding domain and the flanking signature sequences (Badawi *et al*., 2007, Mohseni *et al*., 2012). In this study additional CBFs have been identified (Table 3.1) that did not form part of the above mentioned studies. HCA on the newly reported Norstar CBF AP2 - DNA binding domain without the signature sequences alone, revealed five HC (HC1 - 5) clusters that are quite conserved (Fig. 3.6). HC1 and HC5 are conserved through all the clades containing Norstar CBFs, identified in this study. HC2 is elongated in group D, clade IVd, making region between beat sheet one and two more hydrophobic than other clade members. Further, there is difference within the individual members of the clade, for example TaCBF2.1 has an additional residue (V) at the C - terminal of HC2 interrupted by a proline and CBF20.0 has residue (L) at the N - terminal of the HC2. HC3 is extended in group C, it has seven residues, making region between β - sheet three and α - helix more hydrophobic. Further TaCBF17.0 HC3 has yet another extra residue (I) making it an eight residue long cluster. TaCBF12.2 and TaCBF15.0 have additional residue at the C - terminal of HC4 (Y / I). Similarly, TaCBF19.1 and TaCBF19.2 show presence of an extra L in HC4.

Cluster analysis on the basis of C - terminal show a change in grouping of individual members of the clades, indicating that there are differences in the C - terminal region which might result in difference in the functional properties of the CBFs. HCA of the C - terminal region show different length and position of HC and proline (Fig.3.7). In group C HC1 is long with interrupting single or double proline residues compared to group D, where HC1 is short. Similarly HC2 is longer in group C than in group D members. HC3 is longer in group D, particularly TaCBF9.0 and TaCBF22.0 show an extended repeat of VA residues making it more hydrophobic than other CBFs in this study. Several such differences are seen in the entire C - terminal region of all the CBFs (Fig. 3.7). Clusters of four and five proline residues are present after HC9 in group C. HC10 is flanked by proline residues in TaCBF20.0. Another notable difference is presence of more HC (HC11, HC12 and HC13) in clade IIId - 1, IIId - 3 and IVd members. HC11 is long for clade IVd members and short in clade IIId - 1 and IIId - 3 memebrs. It is completely absent in clade IIId - 4 members. TaCBF17.0 has much longer HC12 than clade IIId - 3 and IVd members. HC13 is present in clade IIId - 3 and IVd only.

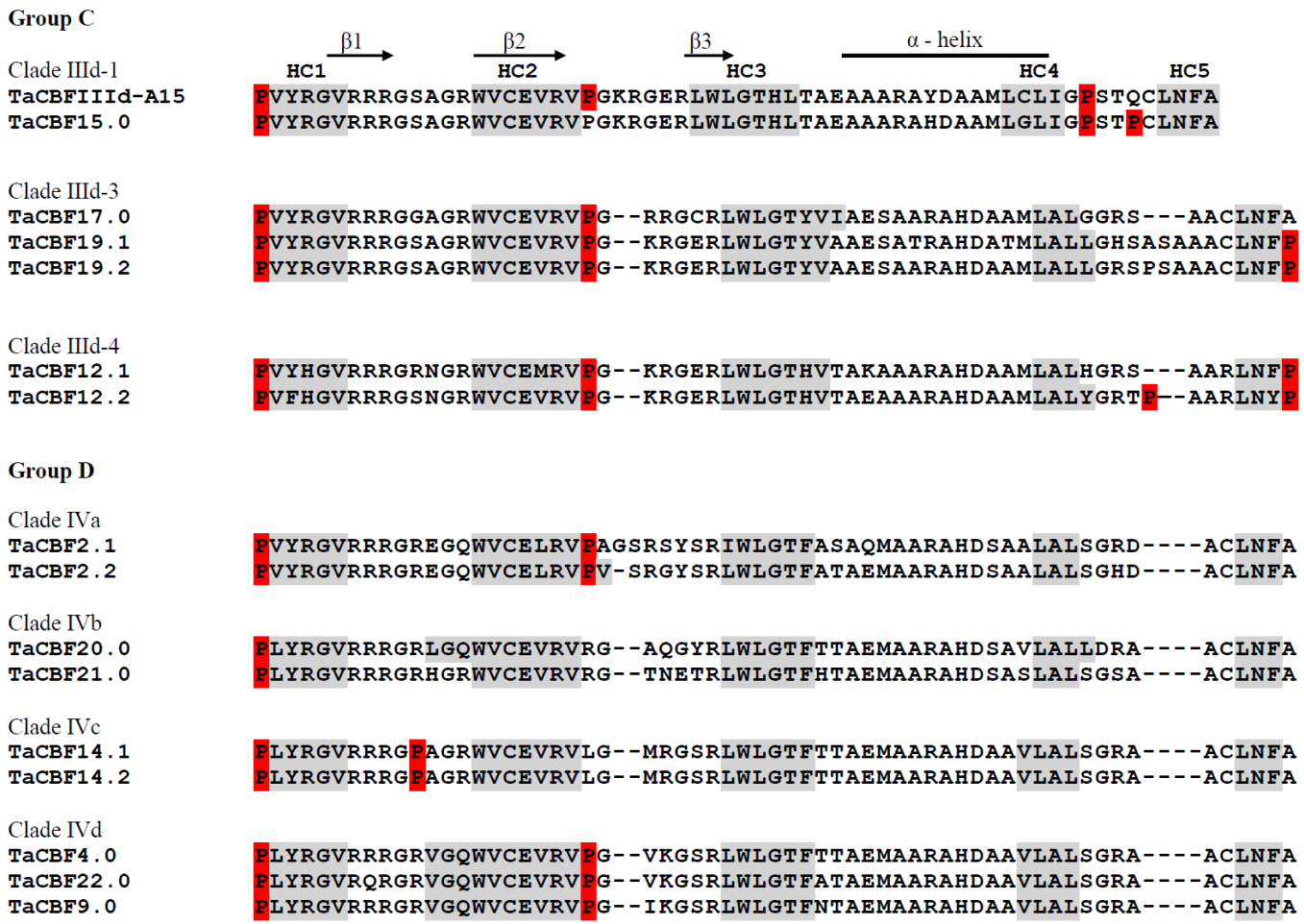


Figure 3.6 Hydrophobic cluster analysis of AP2 - DNA binding domain of CBFs from Norstar.

The amino acid sequences were aligned using ClustalW2 and hydrophobic clusters (HC1 - HC5) are highlighted in grey and proline in red. TaCBFIIId-15 was included in the alignment of clade IIId - 1 as it had a single member TaCBF15.0 from present study. Arrows and line above the alignment indicate the residues making the α - helix and β - sheets within the AP2 - domain.

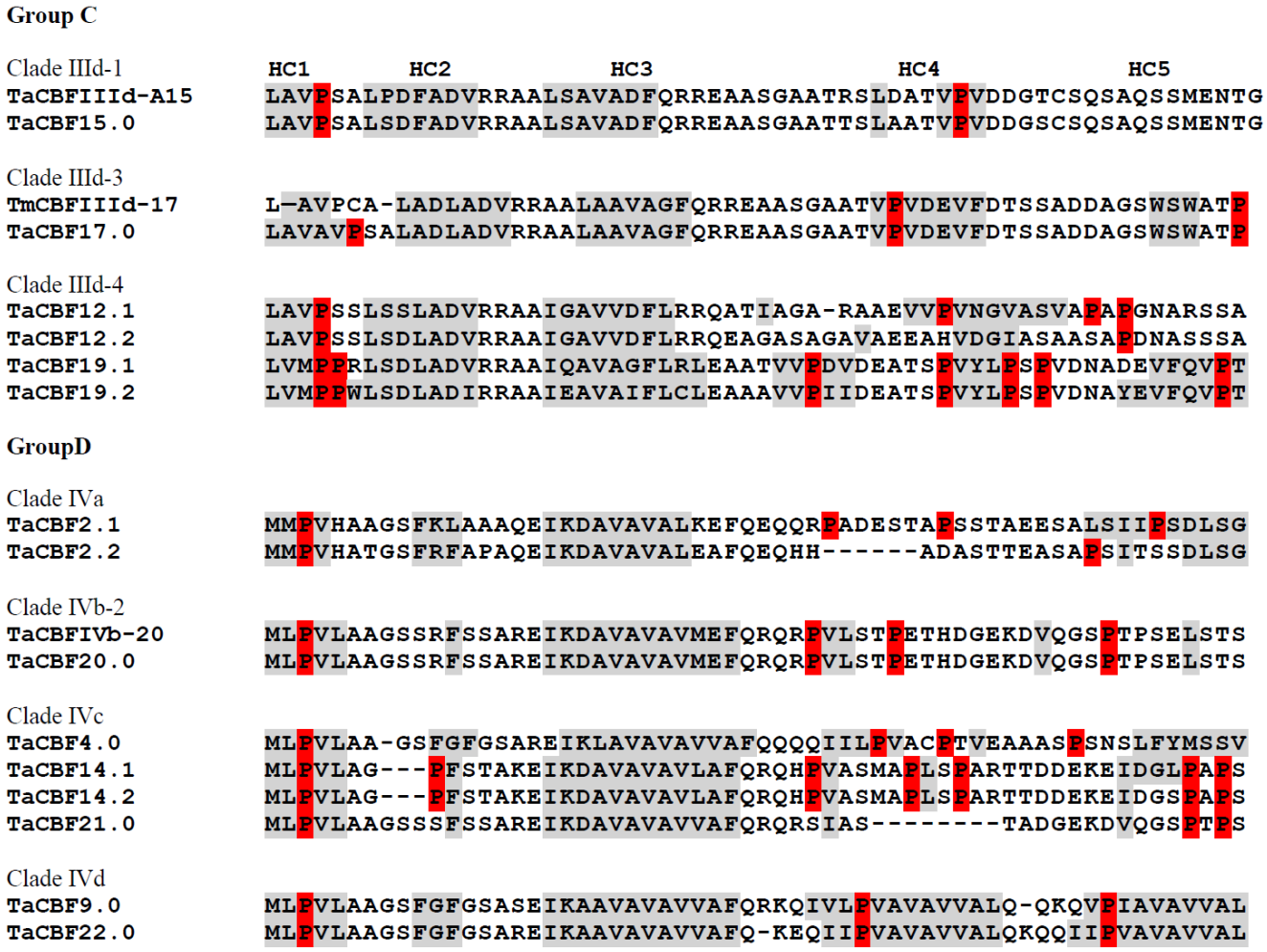


Figure 3.7 Hydrophobic cluster analysis of C - terminal region of CBFs from Norstar.

The amino acid sequences were aligned using ClustalW2 and hydrophobic clusters (HC1 - HC5) are highlighted in grey and proline in red. TaCBFIIId-15 was included in the alignment of clade IIId - 1 and TmCBFIIId-17 was included in clade IIId - 3 as it had a single member TaCBF15.0 and TaCBF17.0, respectively, from present study.

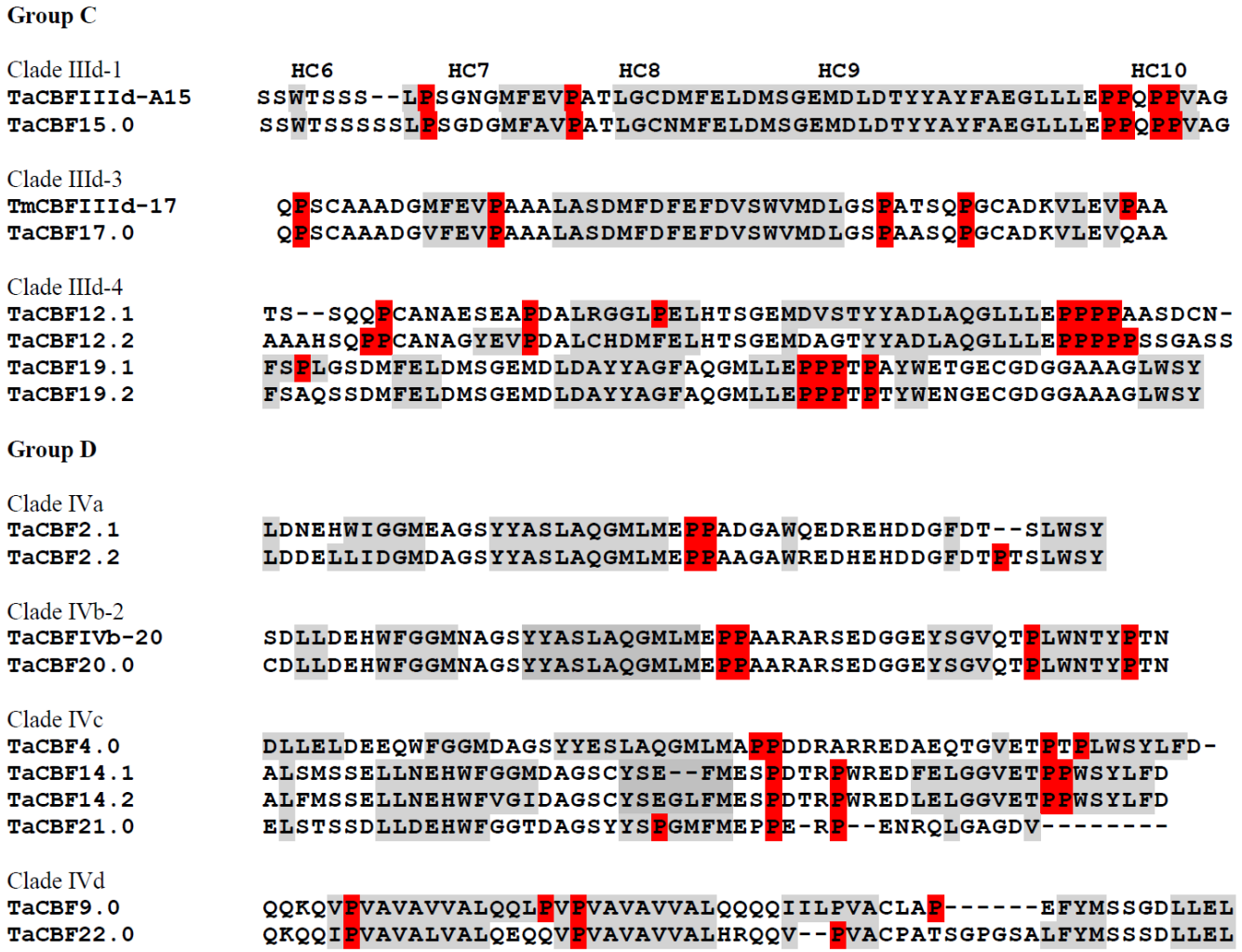


Figure 3.7 Hydrophobic cluster analysis of C - terminal region of CBFs from Norstar.

The amino acid sequences were aligned using ClustalW2 and hydrophobic clusters (HC6 - HC10) are highlighted in grey and proline in red. TaCBFIIId-15 was included in the alignment of clade IIId - 1 and TmCBFIIId-17 was included in clade IIId - 3 as it had a single member TaCBF15.0 and TaCBF17.0, respectively, from present study.

Figure 3.7 Hydrophobic cluster analysis of C - terminal region of CBFs from Norstar.

The amino acid sequences were aligned using ClustalW2 and hydrophobic clusters (HC11 - HC13) are highlighted in grey and proline in red. TaCBFIIId-15 was included in the alignment of clade IIId - 1 and TmCBFIIId-17 was included in clade IIId - 3 as it had a single member TaCBF15.0 and TaCBF17.0, respectively, from present study.

These differences in the hydrophobicity may lead to difference in the folding / stability of protein thereby affecting its interaction with target DNA.

3.5 Discussion

Genome duplications and polyploidization is considered to facilitate survival of species and better adaptation to changing climates (Blanc and Wolfe, 2004; Fawcett *et al*., 2009). During the Eocene - Oligocene cooling period (~ 33.5 – 26.5 million years ago) a rapid expansion happened in the CBF gene family along with Fructosyl transferase (FT) and Ice - recrystallisation inhibition (IRI) protein families (Sandve *et al*., 2008; Sandve and Fjellhelm, 2010). Since then the CBF gene family has undergone an expansion and complexity with evolution and the presence of large number of CBF gene in a single genotype are indicative of the importance of CBFs in cold tolerance (Fig. 3.1). Within the Triticeae tribe, the highest cold - hardiness is noted for rye cv. Puma (LT_{50} = - 33 ^{o}C), followed by hexaploid wheat cv. Norstar (LT_{50} = - 22 ^{o}C; Fowler *et al*., 1996) and winter barley cv Hohentrum (LT_{50} = - 17 ^{o}C). The cold - tolerance among the Pooideae subspecies follows the order of CBF gene number in the representative species *Triticum aestivum*, *Brachypodium distachyon*, and *Oryza sativa*. The number of CBF genes, > 23, in *T. aestivum* cv Norstar at *Fr-A2* locus on chromosome 5A (Båga *et al*., unpublished) is comparable to the number of CBF genes in *H. vulgare* subsp. vulgare genotype Dicktoo (Skinner *et al*., 2005). In contrast only nine and three CBF genes in cold sensitive *Brachypodium distachyon* diploid accession Bd21 and *Japonica* rice cv Nipponbare plants respectively are present at the *Fr-2* collinear region.

However, it is still not known whether low temperature tolerance is influenced by presence of a specific CBF in higher quantities or is it the combined effect of presence of several CBFs. Therefore it is difficult to assess the level of LT tolerance by presence of number of CBF genes alone. An important objective of this study was to characterize CBFs on structural basis and their relatedness to previously characterized CBFs from cold hardy and cold sensitive genotypes. Towards this goal, cluster analysis was conducted for CBF protein in the monocotyledons and analyzed the amino acid changes that could affect the secondary structure of the protein, its function in conferring cold tolerance.

CBF is a transcription factor that has well defined domains with specific functions. A nuclear localization sequence is positioned at the amino terminal end of CBFs (Medina *et al*., 1999) followed by a highly conserved central AP2 / ERF domain flanked by two relatively conserved signature motifs, CMIII - 3 (PKK/RPAGRxKFxETRHP) and CMIII - 1 (DSAWR) (Jaglo *et al*., 2001). The acidic C - terminal region may act as transcriptional activation domain (Wang *et al*., 2005) and often carries a LWSY motif at the end (Skinner *et al*., 2005).The 60 - 70 amino acid long AP2 domain (Fig. 2.6), binds the C - repeat (CRT) regulatory element present in the promoter region of many cold - induced genes (Jaglo *et al*., 2001).

Cluster analysis of complete CBF proteins revealed four distinct groups (A - D), which are further distributed in to 14 subgroups (Fig. 3.2), however, no distinct pattern in relation to cold tolerance was observed. *Triticum aestivum* cv Norstar CBFs at *Fr-A2* locus were present in two major groups C and D and four related clades in group C (IIIa, d - 3, d - 4 and d - 5) and three in D (IVb, c and d). Unrooted dendogram show the relatedness of the leaf nodes, therefore it could be assumed that the orthologs of these groups could have common functions. *Bradi4g35570* - *Bradi4g35640* are induced only by cold stress and not by ABA, drought or salinity (Li *et al*., 2012) and they cluster close to wheat CBFs in subgroup CIIId - 1 (Fig. 3.2), indicating that wheat CBFs from this clade might have retained similar function and likely be induced in cold stress. Similarly, transcript levels of several wheat and barley members of IIId group have been shown to be induced by low temperature, like *TmCBF12*, *TmCBF15*, *HvCBF15*, T*mCBF16* and *HvCBF16* (Knox *et al*., 2008; Stockinger *et al*., 2007; Vágújfalvi *et al*., 2005). Rye group C CBFs are induced at higher level and for longer time at low temperature in cold tolerant cultivars (Campoli *et al*., 2009). Barley as well as rye members of group D have been shown to be up regulated by cold (Stockinger *et al*., 2007; Campoli *et al*., 2009). These five CBF groups have been associated with higher LT inducible capacity of winter wheat cv. Norstar (Badawi *et al*., 2007). Similarly, it has been observed that barley HvCBF1 and HvCBF3 clade members bind to CRT in response to both cold and warm temperature while HvCBF4 - group members function under low temperature only (Skinner *et al*., 2005). These differences are indicative of divergence of CBF gene functions although the classification is based on the sequence similarity and co - localization on the same loci in the genome. Therefore, Norstar CBF

proteins within the group / sub - groups can be expected to have similar function and expected to behave differently among the groups.

CBFs have been shown to have differential affinity to CRT / DRE motif (Xue, 2003; Skinner *et al*., 2005) as well as differences in the C - terminal trans - activation domain between the ten groups (Badawi *et al*., 2007) have been observed. The specific functional relatedness of the CBFs can be further correlated to other known CBFs by analyzing the independent functional domains of the protein. Interestingly, a shift was observed in CBF grouping on the basis of AP2 - DNA binding domain amino acid sequence. Sub - group IIId split into four separate clades. This might suggest that members of one clade i.e. IIId - 4 (TaCBF 12.1 and TaCBF12.2) have preference to particular CRT / DRE motif than the other clades i.e. IIId - 1 (TaCBF15.0), IIId - 3 (TaCBF19.0).

Similarly, a shift was observed between the groups rather than within the group with C - terminal domain cluster analysis. Splitting of group IIId into four separate clades indicates that there might be substantial difference in trans - activation potential of the clade members. Little is known about the activity of C - terminal trans-activation domain, the differences observed in this study suggest that different CBFs interact with different target genes. Taken together, these results indicate divergence of CBF functions although the classification is based on the sequence similarity and co - localization on the same loci in the genome. The CBF protein function could be expected to differ or overlap leading to possible differences / similarity in the regulon activated by the CBFs of particular group.

Cluster analysis on the basis of C - terminal show a change in grouping of individual members of the clades, indicating that there are differences in the C - terminal region that might result in difference in the functional properties of the CBFs. Overall these differences in the length and position of HC, proline residues pattern (considered as cluster breakers), and WY residues (considered to mediate intermolecular interactions) (Callebaut *et al*., 1997), indicate differences in the functional properties of CBFs due to interaction with different proteins. This might also result in difference in protein stability. It also indicates that different selection pressure pattern shaped the fate of individual clade members.

In summary, it is estimated that *Brachypodium* and wheat diverged about 15 Mya while wheat and rice diverged about 50 Mya (Bossolini *et al*., 2007) suggesting that *Brachypodium* is more closely related with wheat (Bortiri *et al*., 2008) than rice. Divergence of wheat from barley is estimated to be 11 Mya and between wheat and rye it is only 5 Mya. Therefore, the orthologous genes are more likely to play the same role in these closely related species. This expansion of CBF genes in cold-hardy Pooideae species occurred during the Eocene - Oligocene cooling period about 33.5 - 26 Mya, when individuals with redundancies in genes involved in cold tolerance were selected (Sandve and Fjellheim, 2010). In the subsequent two chapters, the aim is to identify key structural features of CBF participating in cold tolerance, by conducting a structural and functional analysis of CBF in the cold hardy winter wheat cultivar Norstar.

CHAPTER 4

PROPERTIES OF RECOMBINANT CBF PROTEINS PRODUCED IN *ESCHERICHIA COLI*

4.1 Abstract

A set of 15 CBF genes derived from cold - hardy winter wheat cultivar Norstar were expressed in *Escherichia coli* to produce TrxHisS - CBF fusion proteins to study their properties. Various steps of the prokaryotic protein expression system such as growth conditions, host strain, isopropyl β - D - 1 - thiogalactopyranoside (IPTG) - induction conditions, cell lysis and purification of His-tagged protein by Ni^{2+} affinity chromatography were optimized to maximize yield of native protein. All fusion proteins produced, except TrxHisS - CBF17.0, could be recovered in soluble *E. coli* extracts, whereas TrxHisS - CBF17.0 could only be recovered in the presence of 6 M urea. Ten of the purified TrxHisS - CBF fusions were found to be very stable to heat (98 °C), 10 % SDS, and 6 M urea treatment. The remaining five TrxHisS - CBF fusion proteins became labile upon native purification, but could be maintained relatively stable in *E. coli* extracts or when extracted under denaturing conditions. The TrxHisS - CBF fusions showed variable degrees of slow migration on SDS - PAGE gels, which could be related to elongated protein structures as determined by dynamic light scattering (DLS). Circular dichroism (CD) spectra of TrxHisS - CBF12.2 purified under native conditions revealed only a partial loss of secondary structures when protein was heated from 5 to 95 ^{o}C. Both TrxHisS - CBF12.2 and TrxHisS - CBF17.0 were shown to be highly resistant to denaturation and retained most of their secondary structures in the presence of 6 M urea. The high protein stability observed for CBF12.2 and CBF17.0 may be important properties for cold acclimation or maintenance of frost tolerance in winter wheat.

4.2 Introduction

The CBF genes encoded from Norstar *Fr-A2* are closely related (Chapter 3) and many of them show very similar expression profiles during cold stress (Badawi *et al.*, 2007). Allelic variation in HvCBF14 expression was associated with difference in cold tolerance of European germplasm collection of *H. vulagre* (Fricano *et al.*, 2009). Similarly, between cold resistant and

cold susceptible lines of winter wheat 'Winoka', differential expression of cluster of CBF gene was identified as the underlying cause of difference in cold tolerance (Sutton *et al.*, 2009). As transcriptome data gives some information about temporal and spatial mRNA accumulation, it does not always agree well with *in vivo* protein concentrations (Mooney *et al.*, 2006). Thus, the transcriptome data generated for *Fr-A2* genes does not allow a good prediction of which alleles provide higher freezing tolerance than others. In addition, very small changes in protein primary sequences may have drastic effects on its long - term stability or functionality and the correlation between protein concentration and functional significance may be weak. For example, in *T. monococcum* compared to a frost tolerant accessions G3116, a frost sensitive accession DV92, has a five amino acid deletion in the AP2 domain of the TmCBF12 that results in loss of CRT / DRE binding and enhanced frost - sensitivity (Knox *et al.*, 2008). To - date it is not known if similar differences in functionality exist between closely related Norstar CBF isoforms and those of less hardy winter genotypes. Rather, only a few limited studies of CBF properties have been reported for winter cereals (Skinner *et al.*, 2005; Knox *et al.*, 2008). More detailed information is needed to improve our understanding of the large CBF regulon in wheat.

About 50 - 60 CBF genes are estimated to exist within the hexaploid wheat genome (Mohseni *et al.*, 2012), and many of these are expressed during cold acclimation (Badawi *et al.*, 2007). Due to the abundance of CBF genes and low concentration of transcription factors in general (Harrison *et al.*, 1991), it would be near impossible to purify CBFs individually from plant tissues. Thus, for functional studies, CBFs must be produced as recombinant proteins using a protein expression system.

Recombinant protein expression *in vitro* can be done either in a eukaryotic system or a prokaryotic system. Eukaryotic systems such as the baculovirus - based expression system and yeast cell based systems can carry out complex post - translational modifications. The baculovirus systems are expensive and time consuming as it takes about 2 - 5 weeks before the protein can be purified from the cell culture. Another disadvantage of baculovirus-based systems is the genomic alterations that occur upon serial passage resulting in reduced expression of recombinant protein (reviewed by Krell, 1996). However, a eukaryotic expression systems need to be considered in cases where the expressed protein is toxic in prokaryotic cells and / or requires non-bacterial post - translational modification for activity. In contrast to eukaryotic

systems, prokaryotic expression systems show rapid cell growth, short induction time and generally gives high yield of recombinant protein at low cost. Due to these advantages, several expression systems based on the gram - negative bacterium *E. coli* K12 and B strains have been developed and are commercially available.

Some of the factors affecting yield of foreign proteins in *E. coli* is codon usage, presence of rare codon cluster in introduced gene (Goldman *et al.*, 1994; Kane, 1995; Kurland and Gallant, 1996) and stability of 5' mRNA structures (Kudla *et al.*, 2009). A high frequency of rare codons slows down bacterial growth resulting in reduced protein yield (Klumpp *et al.*, 2012). In some cases it has been possible to improve yield in heterologous systems by optimizing codon usage by genetic engineering (Fuhrman *et al*., 1999; Gustafsson *et al*., 2004).

It is highly desirable for functional studies that the recombinant protein is recovered from *E. coli* soluble cell extract in a native form. However, it is not always a simple task as deduced from many reports of expressed proteins being degraded by *E. coli* proteases (Swamy and Goldberg, 1981) or forming inclusion bodies (reviewed by Marston, 1986). Choice of *E. coli* strain for protein expression and use of efficient protease inhibitor cocktails during protein extraction overcomes much of the problems associated with proteolytic activities (Prouty and Goldberg, 1972). The risk for protein aggregation cannot be predicted from amino acid sequence alone (Thomas and Baynex, 1996), although it is generally considered that a high content of cysteine residues increases the risk for incorrect inter - and intra - disulfide bonds resulting in formation of inclusion bodies (reviewed by Marston, 1986). The recovery of proteins from inclusion bodies is usually achieved by dissolving the aggregates in a chaotropic agent such as urea or thiourea; a treatment that generally results in complete protein denaturation and inactivity. Removal of the denaturing agent by dialysis using an optimized refolding buffer is required to recover active protein from the extract (Thomas and Baynex, 1996). The refolding can be inefficient as many proteins forming inclusion bodies become irreversibly mis - folded.

A few studies have reported successful production of barley, wheat and *Arabidopsis* CBFs using *E. coli* based expression systems (Sakuma *et al.*, 2002; Skinner *et al.*, 2005, Knox *et al.*, 2008). For these CBFs, no particular post - translational modification such as glycosylation or sumoylation seems to be needed for binding to DNA target. As for previous reports, an *E. coli* expression was used in this study to produce of Norstar CBFs for functional studies. The main

objective of this study was to optimize methods for *in vitro* CBF gene expression and recombinant protein purification and dertermine properties of purified proteins.

4.3 Material and Methods

4.3.1 *E. coli* strains

All *E. coli* strains used in study were commercially obtained (EMD Millipore, San Diego, CA, USA) and are listed in Table 4.1. The K12 strain NovaBlue was employed as host during construction of pTrxHisS, strains BL21(DE3)pLysS, BLR(DE3)pLysS, Origami B(DE3)pLysS and Tuner(DE3)plysS were used to produce of TrxHisS - CBF. All four host strains are λDE3 lysogenic and carry a pLysS plasmid. The prophage provides a chromosomal copy of T7 RNA polymerase gene controlled by the lacUV5 promoter, whereas pLysS carries T7 lysozyme gene encoding a natural inhibitor of T7 RNA polymerase. BLR, Origami B and Tuner hosts are all derivatives of BL21 [F– ompT hsdSB(rB– mB–) gal dcm]. BLR is a recA- derivative, whereas Origami B and Tuner strains are thioredoxin reductase (trxB) and glutathione reductase (gor) mutants allowing enhanced disulfide bond formation in the cytoplasm. Tuner has also a lacZY deletion, which eliminates Lac permease activity and allows for uniform IPTG uptake into all cells. The protocol of Hanahan *et al.* (1983) was used for production of chemically competent cells used for transformation of plasmids.

4.3.2 Primer design and PCR amplification of CBF coding sequence

The CBF genes analyzed in this project were derived from BAC clones of *Triticum aestivum* cv Norstar (Ratnayaka *et al.*, 2005; Appendix Table 4.1). Based on the CBF sequences, PCR primers were designed using Primer3 software (Rozen and Skaletsky, 2000) and oligonucleotides (Appendix Table 4.1) were synthesized by Sigma Genosys (Sigma - Aldrich, Oakville, ON, Canada). The primers contained 15 - 12 nucleotide overhangs to allow Ligation Independent Cloning (LIC) of PCR amplified products into expression vector pET-32 Ek/LIC (EMD Millipore, San Diego, CA, USA). The PCR amplifications were conducted in 25 µL reaction volumes containing 1 ng BAC DNA (or 0.1 µg genomic DNA), 1x *Pfu* buffer (20 mM Tris - HCl, pH 8.8, 10 mM $(NH_4)SO_4$, 10 mM KCl, 0.1 % Triton X - 100, 0.1 mg/mL BSA), 2.0 - 3.0 mM $MgSO_4$, 0.2 mM of each dNTP, 1x CES (54 mM betaine, 1.34 mM DTT, 1.34 % DMSO, 11 µg / mL BSA), 1 unit *Pfu* Taq DNA polymerase (ThermoFisher Scientific, Waltham, MA, USA) and 5 µM of forward and reverse primer. The thermocycling program was carried out

in Eppendorf EpgradientS Mastercycler (Eppendorf, Hamburg, Germany) using an initial denaturation at 95 °C for 4.15 min, followed by 32 cycles of 45 sec at 95 °C, 20 sec at annealing temperature (optimized for each primer pair), 2 min at 72 °C and a final cycle of 10 min at 72 ° C. The PCR products were loaded onto 1 % (w / v) agarose gels containing 0.5 µg / mL ethidium bromide and separated by gel electrophoresis. 1x TAE buffer (40.0 mM Tris-acetate, 1.0 mM EDTA, pH 8.3) was used as electrophoresis running buffer and MassRuler DNA ladder Mix (10 - 0.08 kb; ThermoFisher Scientific, Waltham, MA, USA) as molecular weight marker. After electrophoresis, the migration of DNA fragments was visualized using a BioRad Gel Documentation System (Bio - Rad Laboratories, Hercules, CA, USA). Purification of PCR amplified CBF gene fragments from agarose gels was done using the QIAquick Gel Extraction Kit (Qiagen Inc., Hilden, Germany), following the manufacturer's instructions. The concentration of eluted DNA was determined by UV spectroscopy using the DU800 spectrophotometer (Beckman Coulter, Fullerton, CA, USA).

4.3.3 Assembly of CBF expression vector

Assembly of TrxHisS - CBF expression cassette is outlined in Fig. 4.1. To generate vector compatible ends, the purified CBF gene fragment (0.2 pmol) was treated with 1.0 U T4 DNA Polymerase (EMD Millipore, San Diego, CA, USA) at 22 ^{o}C for 30 min in total volume of 20 µL containing 1x T4 DNA Polymerase Buffer (330 mM Tris - acetate pH 7.8, 660 mM potassium acetate, 100 mM magnesium acetate), 2.5 mM dATP, 5 mM DTT. The enzyme was inactivated at 72 ^{o}C for 20 min and annealed with the LIC vector pET-32 Ek/LIC (EMD Millipore, San Diego, CA, USA). The annealing reaction was carried out at 22 ^{o}C for 5 min in 4 µL volume containing 100 ng linearized vector and 0.02 pmol T4 DNA Polymerase - treated PCR fragment. EDTA was added to 6.25 mM final concentration and the reaction mixture was again incubated at 22 ^{o}C for 5 min. Annealed vectors were transformed into *E. coli* NovaBlue host cells following manufacturer's instructions (EMD Millipore, San Diego, CA, USA).

4.3.4 Plasmid purification

E. coli NovaBlue cells harboring pTrxHisS - CBF expression vector were grown in Luria Broth media (10 g / L tryptone, 5 g / L yeast extract, 5 g / L NaCl, pH 7.0 - 7.2) supplemented with 40 µg / mL ampicillin and 12.5 µg / mL tetracycline. The cell culture was grown overnight with agitation at 37 ^{o}C.

Table 4.1. *Escherichia coli* strains used in study.

Strain	Type	Genotype	Resistance	Feature
NovaBlue	K12	*endA1 hsdR17*(r_{K12}^{-} m_{K12}^{+}) *supE44 thi-1 recA1 gyrA96 relA1 lac* F'[*proA*$^{+}$*B*$^{+}$ *lacI qZ*Δ*M15*::Tn*10*]	Tetracycline	Lacks endonuclease acitvity; low risk of homologus recombination; high yield of plasmid DNA; good for initial cloning
BL21(DE3) pLysS	B	F^{-} *ompT hsdS*B(r_B^{-} m_B^{-}) *gal dcm* (DE3) pLysS	Chloramphenicol and Tetracycline	Low protease activity, good for protein expression
BLR(DE3)pLysS	B	F– *ompT hsdS*B(r_B^{-} m_B^{-}) *gal dcm* (DE3) Δ(*srl-recA*)*306*::Tn*10* pLysS	Chloramphenicol and Tetracycline	Low risk of homologus recombination; low protease activity; good for protein expression
Origami B(DE3)pLysS	B	F^{-} *ompT hsdS*B(r_B^{-} m_B^{-}) *gal dcm lacY1 aphC* (DE3) *gor522*::Tn*10 trxB* pLysS	Kanamycin and Tetracycline	Enhances disulphide bond formation, fine-tuning of protein expression by IPTG induction
Tuner(DE3)pLysS	B	F^{-} *ompT hsdS*B(r_B^{-} m_B^{-}) *gal dcm lacY1* (DE3) pLysS	Chloramphenicol	Uniform entry of IPTG into all cells, fine-tuning of protein expression.

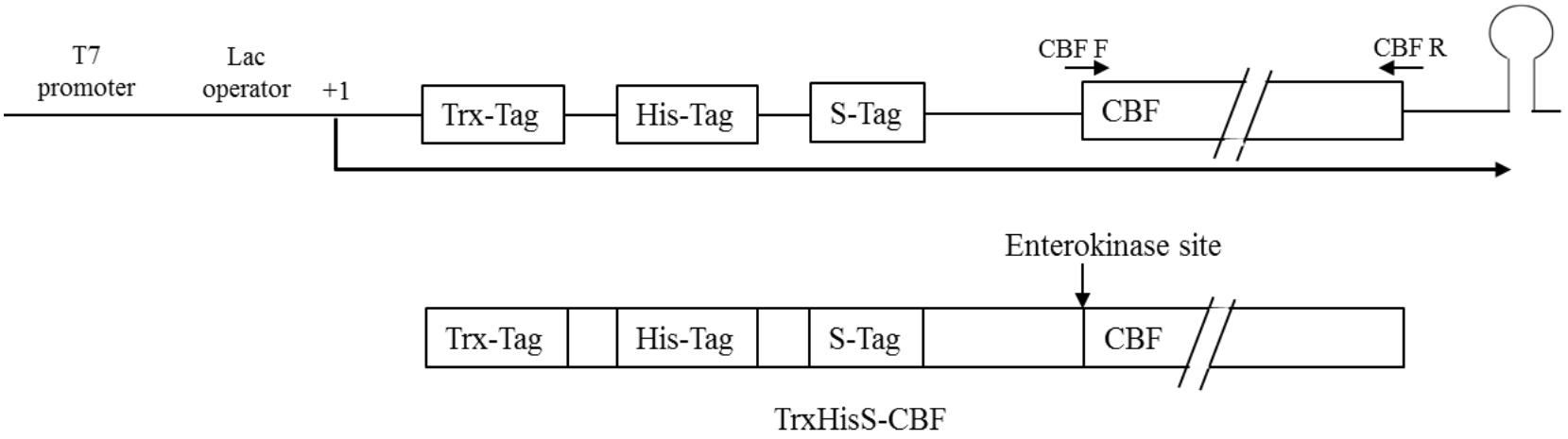


Figure 4.1 Schematic representation of TrxHisS - CBF expression cassette.

TrxHisS - CBF operon under control of T7 promoter and Lac operator expressing TrxHisS - CBF fusion protein is illustrated. CBF = C - repeat binding factor; His = Histidine; Lac = Lactose; S = Ribonuclease S; Trx = Thioredoxin.

Plasmids were extracted from 4 mL overnight cultures using the GeneJET^{TM} Plasmid Miniprep Kit (ThermoFisher Scientific, Waltham, MA, USA) following manufacturer's instructions. The concentration of eluted DNA was determined by UV spectroscopy or estimated after 1 % agarose gel electrophoresis and comparing it with DNA fragment intensity to MassRuler DNA ladder Mix (ThermoFisher Scientific, Waltham, MA, USA).

4.3.5 DNA sequencing and analysis

DNA sequence analysis of plasmid inserts was done by the by the Big Dye Terminator technology at the Robarts Research Institute (The University of Western Ontario, Ontario, Canada). The vector specific Trx - Tag (5'-GTGCACTGTCTAAAGGTCAG-3') and the T7 terminator (5'-GCTAGTTATTGCTCAGCGG-3') primers were used in the sequencing reactions (Fig. 4.1). DNA sequences were analyzed using DNASTAR Lasergene version 7.1.0 software (DNASTAR Inc., Madison, WI, USA).

4.3.6 Analysis of codon usage

To check the codon usage for CBF gene to be expressed in *E. coli*, a web based program "Graphic codon usage analyser 2.0" (Fuhrmann *et al*., 2004) was used with codon table for *E. coli* B - type strain selected. The codon adaptibility of CBF genes from Norstar when expressed in *E. coli* B - type cell was expressed in relative adaptiveness index, which takes into account the number of codons that code for respective amino acid (Sharp *et al*., 1987). For each amino acid, the codon with the highest frequency value is set to 100 % relative adaptiveness and other codons for the same amino acid are scaled accordingly. The index also gives an approximate indication of the likely success of heterologous gene expression. Threshold value of < 10 % indicates very few used codons whereas < 20 % indicates few used codons.

4.3.7 Expression of fusion protein

The *E. coli* strains BLR(DE3)pLysS or BL21(DE3)pLysS harboring expression vector pTrxHisS - CBF were grown at 37 °C or 28 ^{o}C in super broth (SB) medium (25 g / L tryptone, 15 g / L yeast extract, 5 g / L, NaCl, 1 % glucose, pH 7.1-7.2) supplemented with 40 µg / mL ampicillin, 34 µg / mL chloramphenicol and 12.5 µg / mL tetracycline for BLR(DE3)pLysS cells and 40 µg / mL ampicillin, 34 µg / mL and chloramphenicol for BL21(DE3)pLysS cells. After 16 h growth, cells were diluted 1 : 50 into fresh medium with antibiotics and grown until A_{600nm} = 0.5 was reached. The inducer isopropyl β - D - 1 - thiogalactopyranoside (IPTG) was added to a

final concentration of 0.1 - 2.0 mM and the culture were grown for an additional 1 - 4 h to express the TrxHisS - CBF fusion protein. Cell samples collected at different stages of growth and induction were centrifuged at 10,000 x g for 10 min and resulting pellets were stored at - 20 ^{o}C until analysis.

4.3.8 Preparation of *E. coli* cell extracts

Preparation of cell extracts from frozen *E. coli* cell pellets under native conditions was done using an enzymatic or by lysozyme / sonication method. In the enzymatic method, a cell pellet from 100 mL induced culture (1 g) was resuspended in 5 mL BugBuster Master Mix solution (EMD Millipore, San Diego, CA, USA) and incubated on a shaking platform (150 rpm) at 23 ^{o}C for 20 min. Insoluble cell debris was removed by centrifugation at 13,000 x g for 40 min at 4 ^{o}C. The sonication method was performed according to QIAexpressionist protocol (Qiagen Inc., Hilden, Germany). Briefly, a frozen cell pellet of 100 mL cell culture was thawed on ice and resuspended on ice in 5 mL lysis buffer (50 mM NaH_2PO_4, 300 mM NaCl and 10 mM imidazole, pH 8.0) containing 1 mg / mL lysozyme. Cells were sonicated five times with 10 sec bursts at 200 - 300 W, 10 sec cooling between each burst using a sonicator (Branson SONIFIER® 450) equipped with microtip. The native soluble protein extracts prepared were stored on ice for short term or supplemented with glycerol 20 % (v / v) for long term storage at - 20 ^{o}C. To prepare urea-soluble extracts the frozen cell pellet of 50 mL cell culture was suspended in 5 mL urea buffer (6 M urea, 300 mM NaCl, 50 mM sodium phosphate, 10 mM imidazole, pH 8.0). The suspension was centrifuged at 10,000 x g for 30 min to remove insoluble cell material. The soluble cell extract extract was stored on ice for short term and at - 20 ^{o}C for long term storage.

4.3.9 Purification of TrxHisS - CBF

CBF fusion protein was purified using 1 mL Ni-MAC cartridges according to manufacturer's protocol (EMD Millipore, San Diego, CA, USA). Briefly, a 5 mL syringe was filled with deionized water and attached to Ni - MAC cartridge in "drop - to - drop" manner to avoid introducing air bubbles. Cartridge was washed with 5 mL water and equilibrated with 5 mL 1X MAC binding buffer (300 mM NaCl, 50 mM sodium phosphate, 10 mM imidazole, pH 8.0). The cartridge was then loaded with 1 mL soluble *E. coli* extract at a flow rate of 30 drops per min followed by washes with 10 mL 1x MAC bind buffer and 6 mL 1x MAC wash buffer (300 mM NaCl, 50 mM sodium phosphate, 20 mM imidazole, pH 8.0). Proteins bound to column

were eluted with 2 mL 1x MAC elution buffer (300 mM NaCl, 50 mM sodium phosphate, 250 mM imidazole, pH 8.0) and stored on ice for short term and at - 20 ^{o}C for long term.

Purification of His - tagged protein under denaturing conditions was done by loading urea - soluble protein extract (1 mL) onto washed and equilibrated Ni-MAC cartridge followed by washes with buffer B (6 M urea, 300 mM NaCl, 50 mM sodium phosphate, 10 mM imidazole, pH 8.0) and buffer C (6 M urea, 300 mM NaCl, 50 mM sodium phosphate, 20 mM imidazole, pH 8.0). Proteins bound to the cartridge were eluted with 2 mL buffer D (6 M urea, 300 mM NaCl, 50 mM sodium phosphate, 250 mM imidazole, pH 8.0) and stored on ice for short term and at - 20 ^{o}C, with 10 % glycerol, for long term. 50 µL aliquots were analyzed by 12 % SDS - PAGE. Purified samples were dialyzed against 20mM sodium phosphate buffer for 4h at 4 ^{o}C with constant stirring at low speed and stored at 4 ^{o}C for use in the short term.

4.3.10 Recharging of Ni - MAC cartridges

Recharging of cartridges was done following manufacturers (EMD Millipore, San Diego, CA, USA) instructions. Briefly, the cartridge was first washed in six steps: 3 mL 100 mM EDTA, 20 mM sodium phosphate buffer pH 7.5; 1 mL 0.5 M HCl at slow rate; 2 mL 300 mM NaCl; 10 mL 0.5 M NaOH; 2 mL 300 mM NaCl and finally 5 mL deionized water. Recharging of the column was done in five steps: 3 mL 300 mM NaCl, 20 mM sodium phosphate pH 7.5; 5 mL 300 mM NaCl; 5 mL 100 mM NiSO4; 10 mL 300 mM NaCl and finally 3 mL 1x MAC bind buffer (300 mM NaCl, 50 mM sodium phosphate, 10 mM imidazole, pH 8.0). Colums were stored at 4 ^{o}C in 150 mM NaCl, 20 % ethanol.

4.3.11 SDS - PAGE

SDS - PAGE was performed on a V16 - 2 vertical electrophoresis system (Life Technologies Inc, Carlsbad, CA, USA) in accordance with the discontinuous buffer system of Laemmli (1970). The gel had 10 - 14 wells and was 15 x 17 cm^2 large and 1.5 mm thick. The stacking gel contained 5 % acrylamide : bisacrylamide (30 : 0.8) in 0.1 % SDS, 0.5 M Tris - HCl, pH 6.8 and separating gel contained 12 - 15 % acrylamide : bisacrylamide (30 : 0.8) in 0.1 % SDS, 1.5 M Tris - HCl, pH 8.8. Samples were mixed with equal volume of 2x sample dye (0.125 M Tris - HCl, pH 6.9, 10 % SDS, 20 % glycerol, 10 % β - mercaptoethanol, 0.1 % bromophenol blue) and denatured at 100 °C for 5 min before loading. The electrophoresis buffer contained 0.1 % SDS, 0.25 M glycine, 0.025 M Tris - HCl, pH 8.3 and electrophoresis was conducted at 8 mA

constant current for 16 h. The SDS - PAGE gel was stained with Coomassie Brilliant Blue R - 250 (Sigma-Aldrich, St. Louis, MO, USA) according to standard procedure (Sambrook and Russell, 2001) or silver stained. Apparent molecular weight was determined by comparing the migration to SDS - PAGE molecular weight standards obtained from Bio - Rad Laboratories (Bio - Rad Laboratories, Hercules, CA, USA).

4.3.12 Analysis of protein glycosylation

Proteins in cell extracts (160 - 200 µg) containing TrxHisS - CBF12.1, -CBF12.2 and - CBF17.0 fusions were separated on 12 % SDS - PAGE gel. Analysis of protein glycosylation was done using Schiff's base. The gel was fixed in (50 % methanol and 10 % acetic acid) for one hour, followed by protein oxidation by periodic acid (7 g / L in 5 % acetic acid) for 10 min. After rinsing with distilled water, the gel was stained with Schiff's reagent (Sigma - Aldrich, St. Louis, MO, USA) until polypeptides become visible (~ 15 min). Excess stain was removed by 2 - 3 washes with sodium metabisulfite (50 mL / L in acetic acid). After destaining in (5 % methanol + 7.5 % acetic acid) solution for 20 min, the gel was finally washed with distilled water for 5 min and inspected for purple bands indicating protein glycosylation.

4.3.13 Analysis of protein phosphorylation

Samples of purified TrxHisS - CBF proteins (~ 3 µg) in 20 µL total volume were treated with 3 units of calf intestinal phosphatase (CIP; New England Biolabs, Ipswich, MA, USA) for one hour at 37 °C. The reaction was terminated by adding 50 mM EDTA followed by analyzing CIP - treated and untreated samples by 12 % SDS - PAGE. Protein migration was compared after silver staining of gel.

4.3.14 Urea treatment of proteins

The effect of urea on protein migration by SDS - PAGE was analyzed following protocols reported by Mukherjee *et al*., (2005). One set of samples of *E. coli* cell extract containing TrxHisS - CBF12.2 fusion protein was heated at 98 ^{o}C in the presence of 50 mM urea for increasing time (0, 10, 20, 30 and 60 min). A second set of samples were incubated at 98 ^{o}C for 30 min with increasing urea concentration (500 mM, 1 M, 2 M and 4 M). Treated and untreated control samples were resolved by 12 % SDS - PAGE and stained with Coomassie Brilliant Blue R250 (Sigma - Aldrich, St. Louis, MO, USA) according to standard procedure (Sambrook and

Russell, 2001). Untreated control samples were boiled with 1x SDS sample buffer prior to loading on the gel.

4.3.15 Two - dimensional (2D) - gel electrophoresis

A 50 µg sample of TrxHisS - CBF12.2 fusion polypeptide was analysed by 2D - gel electrophoresis using the BioRad Ready StripTM IPG strip system (pH 3 - 10, 7cm) and PROTEAN® IEF cell (Bio - Rad Laboratories, Hercules, CA, USA) following the manufacturer's instructions. Rehydration of IPG strip with the sample was done under active conditions (i.e. with application of 50V during rehydration to facilitate sample uptake by IPG strip) for 12 h followed by focusing for about 9 h using a preset program (linear voltage ramping slope, active rehydration at 20 ^{o}C for 12 h, focusing for 9 h at 20 ^{o}C) . The second dimension electrophoresis was done by 12 % SDS - PAGE, run at constant voltage (200 V) for 55 min. Polypeptides were visualized by silver stained gels.

4.3.16 In - gel digestion

Spots dissected from SDS - PAGE gels were digested using the MassPrep II Proteomics Workstation (Micromass, Manchester, UK) following the procedure described by Sheoran *et al.* (2005). Briefly, excised protein gel bands are cut into ~ 1 mm^3 pieces and placed into 96 - well plates and destained two times for 10 min with 100 µL of 1 : 1 ammonium bicarbonate : acetonitrile (v / v). Proteins were reduced by addition of 50 µL 10 mM DTT (Dithiothreitol), 0.1 M ammonium bicarbonate solution followed by 30 min incubation at 37 ºC. Protein alkylation was achieved by adding 50 µL 55 mM iodoacetamide prepared in 0.1 M ammonium bicarbonate. The alkylation reaction was carried out for 20 min at 37 ºC. Gels are washed with 100 mM ammonium bicarbonate and dehydrated with acetonitrile before being saturated with 25 µL of 6 ng / µL trypsin prepared in 50 mM ammonium bicarbonate; digestion was carried out at 37 ºC for 5 h. Peptides were extracted with 30 µL 0.1 % trifluoroacetic acid, 3 % acetonitrile solution for 30 min. This step was followed by two more extractions with 24 µL 0.1 % trifluoroacetic acid, 50 % acetonitrile for 30 min. The combined extracts were dried in a Labconco speedvac analysis (ThermoFisher Scientific, Waltham, MA, USA). Samples were reconstituted in 40 µL 0.2 % formic acid, 3 % acetonitrile solution and subjected to liquid chromatography electrospray ionisation tandem mass spectrometry (LC - ESI MS) analysis.

4.3.17 LC - ESI - MS spectra collection and data analysis

For LC - ESI - MS analysis, a Quadrupole Time - Of - Flight (Q - TOF) Global Ultima mass spectrometer (Micromass, Manchester, UK) was used. The instrument is equipped with a nano-electrospray (ESI) source and is interfaced with a nanoACQUITY UPLC solvent delivery system (Waters, Milford, MA, USA). The mobile phase was composed from a binary solvent system composed of A (0.2 % aqueous formic acid, 3 % acetonitrile) and B (0.2 % aqueous formic acid, 95 % acetonitrile) solutions. Peptides were desalted with an in - line solid - phase trap column (180 µm × 20 mm) packed with 5 µm Symmetry C18 resin (Waters, Milford, MA, USA) and separated on a 100 µm x 100 mm capillary column (Waters, Milford, MA, USA) packed with 1.7 µm BEH130 C18 resin (Waters, Milford, MA, USA) at a column temperature of 35 ºC. An injection volume of 2 to 5 µL was loaded on the trapping column for desalting at a flow rate of 15 µL / min for 3 min at initial conditions, using A : B 99 : 1 and diverting the flow to waste. After desalting, flow is diverted through the trap column to the analytical column with a linear gradient of 1 - 10 % solvent B at 400 nL / min for 16 min, followed by a linear gradient of 10 - 45 % solvent B delivered with a flow rate of 400 nL / min over 30 min. A fast gradient of 45 - 80 % solvent B for 6 min with flow rate of 800 nL / min was used to clean out the column for subsequent injections followed by equilibrating to initial conditions for 7 min. Typical Q - TOF parameter settings consist of capillary voltage of 3,850 V, cone voltage of 120 V, and source temperature of 80 ºC.

Samples were analyzed using Data Dependant Acquisition (DDA), which consists of the detection of multiple charged positive ions (z = 2, 3, and 4) from an MS survey scan. The MS survey scan range was set from m / z 400 to 1,900 with a scan time of 1 second. Up to three MS / MS scans were triggered from each MS scan event with the signal intensity threshold 16 counts / sec. In MS / MS experiments, data was acquired in continuum mode with a scan time of 1.9 sec and dynamic exclusion of previously detected precursors was set at 2 min; peptides from trypsin and keratin were also excluded from MS / MS data collection.

To obtain high mass accuracy, a compound of known mass is continuously introduced to a second ESI source. This “lock mass” compound is used as a reference during data processing to correct for small adjustments to the mass calibration thereby improving mass accuracy. A baffle is used to select whether the ions from the sample or the reference are allowed to pass into

the mass spectrometer. For these experiments, Leucine Enkephalin (Environmental Resource Associates (ERA), Arvada, CO) with m / z 556.2771 was prepared at a concentration of 80 fmol / µL in 1 : 1 acetonitrile:water with 0.1 % formic acid. Ions from this source were sampled for 1 sec every 20 sec with a flow rate of 1.0 µL / min.

Data was processed with ProteinLynx Global Server 2.4 (PLGS 2.4; Waters, Milford, MA, USA) using .RAW files from LC - ESI - MS and LC - ESI - MS / MS. PKL files were generated using ProteinLynx Global Server 2.4 (PLGS 2.4; Waters, Milford, MA, USA), and subsequently submitted to Mascot (Matrix Science Ltd., London, UK) for peptide search against the NCBI or database hosted by National Research Council of Canada, Ottawa, Canada. In the database search parameters, a maximum of 1 miscleavage was allowed for tryptic digestion. The tolerance for precursor peptide ions was ± 50 ppm and for fragment ions it was ± 0.2 Da. Carbamidomethylation of cysteine was selected as a fixed modification and oxidation of methionine was used as a variable modification. A minimum of two unique peptides were allowed for confident protein identification.

4.3 18 Dynamic light scattering (DLS)

DLS experiments were conducted on a DynaPro MS800 instrument (Protein Solutions Inc., Charlottesville, VA, USA) equipped with 830 nm laser diode operating at 55.5 mW and a scattered light collection angle of 90^{o}. Purified TrxHisS - CBF samples (0.3 mg / mL in 20 mM sodium phosphate buffer, pH 8.0) were centrifuged at 13,000 x g for 20 min, and approximately 12 µL of supernatant was placed in a 20 µL QS1.06 cuvette (b = 1.5 mm; Hellma GmbH & Co. KG, Müllheim, Germany) for analysis. Sample acquisition time was set to 45 sec and at least 25 measurements were collected per analysis. The regularization algorithm, which uses Laplace inversion and non - negative least squares to fit the average of autocorrelation coefficients to find the size distributions that best fits the data without restrictions on the modality of the distribution, found in the Dynamics V5.26.60 (Protein Solutions Inc., Charlottesville, VA, USA) was used to extract translational diffusion coefficients (D_T) from the measurements. Assuming Brownian motion, the equivalent hydrodynamic radius (RH) of a hard sphere model was calculated using the Stokes Einstein equation : $R_H = kT / 6\pi\eta D_T$;

Where k is the Boltzmann constant, T is the temperature, and η is the solvent viscosity.

The following criteria for homogeneity were used :

12 - 18 % polydispersity : homogenous, majority of monomers

20 - 30 % polydispersity : increasing amounts of dimers

> 30 % polydispersity : various non - specific aggregates

Calculations for shape predictions were done using following formula (Erickson, 2009):

S = M / (4205R) and Smax = 0.00361(M 2 / 3), where

S = sedimentation coefficient

R = radius in nanometer

M = molecular mass in Dalton

Smax / S :

For approximately globular protein = 1.2 - 1.3

Moderately elongated protein = 1.5 to 1.9

Highly elongated protein = 2.0 to 3.0

4.3.19 CD spectroscopy and secondary structure analysis

Far - UV CD spectra (200 - 260 nm) of protein sample (0.3 mg / mL, 20 mM sodium phosphate buffer, pH 8.0) were recorded at 22 ^{o}C on PiStar - 180 spectrometer (Applied Photophysics, Surrey, UK) using a 1.0 mm path length quartz cuvette. The instrument was calibrated at 290.5 nm using (1S) - (+) - 10 - camphorsulfonic acid. Spectra were acquired at 0.5 nm steps using a scan rate of 10 nm / min, and entrance / exit slit set at 6 nm. Far - UV CD spectra (200 - 230 nm) for CBF samples (0.3 mg / mL) were acquired at variable temperatures in the range of 5 to 95 ^{o}C. Spectra were collected using a using 1.0 mm quartz cuvette and a 5 ^{o}C temperature ramp, where each temperature was equilibrated for 180 sec and had a tolerance of ± 0.5 ^{o}C. Spectra were acquired at 0.5 nm steps using a scan rate of 10 nm / min, and the entrance / exit slits set at 6 nm. The mean residual ellipticity [Θ], expressed in units of deg.cm^{2}.dmol^{-1}, was calculated by using an average amino acid molecular weight of 110. The secondary structure of intact and truncated TrxHisS - CBF12.2 as well as native and denaturing condition purified TrxHisS - CBF12.2 and CBF17.0 proteins was predicted using CDNN deconvolution software (V2.1) (Böhm *et al.*, 1992). Thermal stability of TrxHisS - CBF12.1 and CBF12.2 were monitored by analyzing molar ellipticity at 222 nm from the recorded temperature scan.

4.4 Results

4.4.1 Analysis of *CBF* codon usage

A total of 18 CBF genes (Appendix Table 4.1) from winter wheat cv Norstar (15 genes), cv Cappelle Desprez (two genes) and genotype Winter Manitou (one gene), were selected for protein expression in *E. coli*. To determine if any of the selected CBF genes had a bias in their codon usage that might affect protein expression in *E. coli*, the CBF coding sequences were analyzed using Graphic codon usage analyzer 2.0 software. The relative codon adaptability analysis was based on codon usage in *E. coli* B type strain and when calculated for TrxHisS tag and the individual CBF genes showed an overall low frequency of rare codons and lack of rare codon clusters (see typical profile represented by CBF12.2 in Fig. 4.2; Table 4.2). Therefore, modification of the *CBF* codons by genetic engineering to better fit tRNA population in *E. coli* B strains seemed unnecessary. Also, the Rosetta2(DE3)/pRARE2 *E. coli* strain, which over expresses seven rare tRNA molecules was excluded from tests of suitable host strains. The codon analysis further showed a low frequency of cysteine codons (Table 4.2) and low conservation of Cys positions along the CBF gene sequences (see Appendix Table 4.1) suggesting cysteine residues may have a minor role in CBF folding.

4.4.2 Assembly of *CBF* expression vectors

The construction of the TrxHisS - CBF expression vectors is outlined in Fig. 4.1 and was initiated by PCR amplification of *CBF* coding regions. As CBF genes are intron less, BAC clones or genomic DNA were used as template in PCR amplifications. To generate CBF gene fragments, 15 - 18 bp allele - specific primers (Appendix Table 4.1) matching 15 - 18 bp of the start and end of *CBF* coding sequence were synthesized with 12 bp (forward primer) or 14 bp overhangs (reverse primer). The overhangs were designed to facilitate production of vector complementary single-stranded ends for directional cloning into pET32-Ek/LIC expression vector using the ligation independent cloning (LIC) technique. The constructed expression vectors were denoted pTrxHisS - CBF and encoded proteins TrxHisS - CBF. The Trx - tag encodes thioredoxin that increases protein solubility and the His- and S-tags can be used for protein purification. A T7 promoter controlled by lac operator / repressor regulates expression of genes inserted into the expression vector; therefore TrxHisS - CBF expression can be induced by addition of IPTG. Another advantage of this expression cassette is that it encodes an enterokinase (Ek) site

positioned immediately before the amino - terminal Met residue of CBF. Cleavage at this site releases the complete CBF from the vector - encoded TrxHisS tags. In addition to full - length TrxHisS - CBF expression cassettes, C - terminal truncated version of TrxHisS - CBF12.1, CBF12.2, CBF15.0 and CBF17.0 were assembled. A control expression vector pTrxHisS was also constructed.

The constructed expression vectors were initially transformed into the K12 strain NovaBlue (Table 4.1). NovaBlue is well suited for plasmid propagation as it carries the endA- mutation eliminating EndA endonuclease that might cause degradation of DNA during or after plasmid preparation. In addition, NovaBlue cells are *recA*$^{-}$ and, therefore, homologous recombination is almost negligible ensuring stable maintenance of vectors. For each CBF expression construct transformed into NovaBlue, three colonies were selected for plasmid propagation followed by DNA sequencing of insert to validate the constructs.

4.4.3 Selection of *E. coli* host cells

E. coli B - type strains (Table 4.1) were selected for this study as they lack Lon and OmpT protease activities and therefore cause less degradation of overexpressed proteins than K12 strains. To identify a suitable B type host strain for TrxHisS - CBF expression, construct pTrxHisS - CBF12.2 was transformed into four different hosts: BLR(DE3)pLysS, BL21(DE3)pLysS, Origami B(DE3)pLysS and Tuner(DE3)pLysS. All hosts carry a compatible plasmid pLysS, which encodes T7 lysozyme, a natural inhibitor of T7 RNA polymerase ensuring low expression from T7 lac promoter under non - induced conditions. The *recA*$^{-}$ BLR(DE3)pLysS strain has the advantage of maintaining more stable expression constructs compared to the other three strains, which are all *rec*$^{+}$. The advantage of Origami B(DE3)pLysS and Tuner(DE3)pLysS strains is lack of permease activity caused by a lacZY deletion allowing greater fine tuning of protein expression by IPTG concentration. In addition, the Origami B(DE3)pLysS strain carries mutated glutathione reductase and thioredoxin reductase genes, which leads to enhanced disulphide bond formation in *E. coli*.

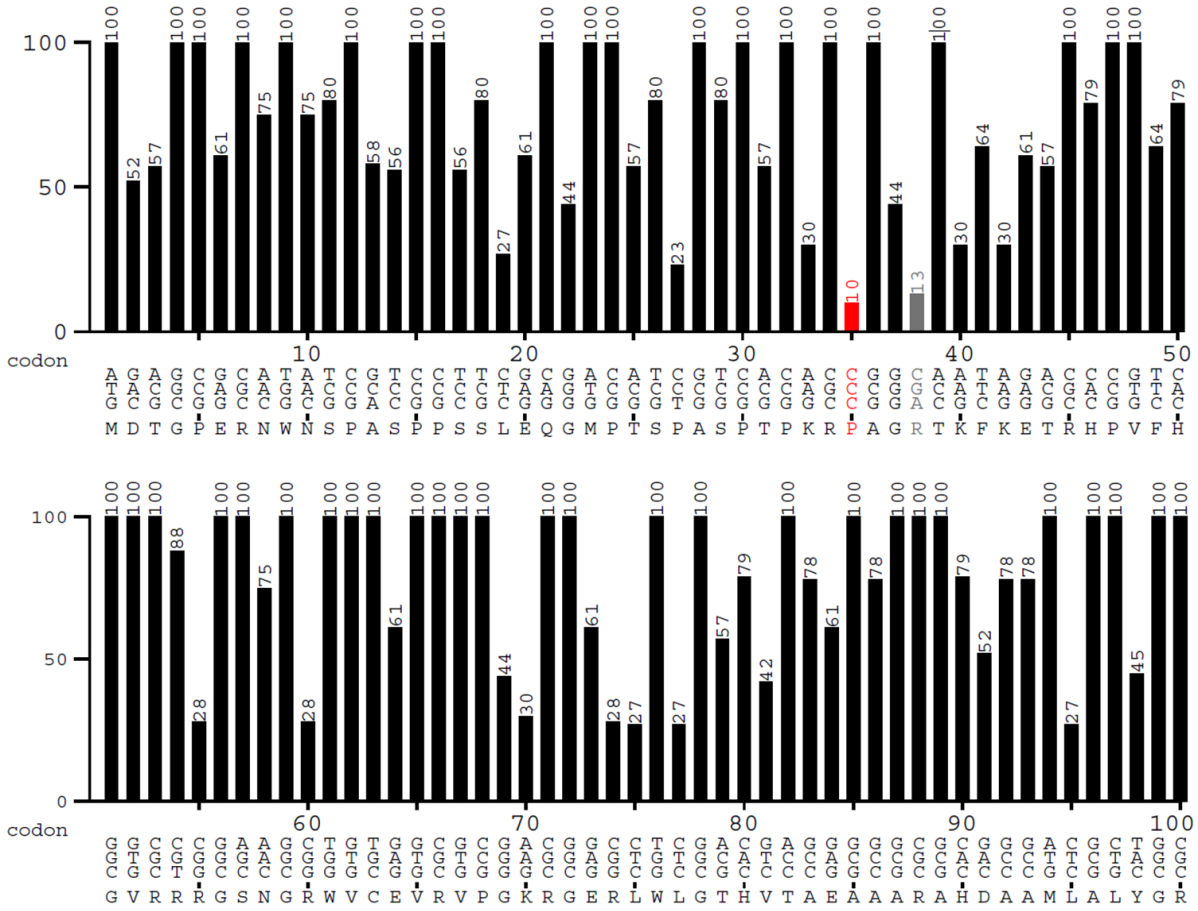


Figure 4.2 Relative codon adaptiveness for *Triticum aestivum* cv Norstar CBF12.2 in *E. coli* type B strain.

Section of graphic output of codon relative adaptiveness along protein sequence. For each amino acid the codon with the highest frequency value is set to 100 % relative adaptiveness. All other codons for the same amino acid are scaled accordingly. Threshold value for very few and few used codons are set at < 10 % and < 20 %, respectively.

Table 4.2 Codon usage relative adaptiveness in *E. coli* B - type cells for wheat CBF gene.

	Codons used < 10 %	Codons used < 20 %	Number of Cystine codons
TrxHisS-tag	0	0	2
CBF2 .1	8	1	1
CBF2 .2	4	1	2
CBF4 .0	5	0	2
CBF9 .0	7	1	4
CBF12 .1	11	0	4
CBF12 .2	5	0	3
CBF14.1	6	1	3
CBF14 .2	8	0	3
CBF15 .0	4	0	5
CBF17.0	7	1	6
CBF19.1	7	1	3
CBF19.2	8	0	3
CBF20.0	7	1	4
CBF21.0	4	1	2
CBF22 .0	7	1	3
CBF12.2a**	5	0	3
CBF12.2b**	5	1	2
CBF12.2***	5	0	3

** Cappelle - Desprez; *** winter Manitou

4.4.4 Test of induction conditions

A rich SB medium containing 1 % glucose and a relatively low incubation temperature (28 °C) were chosen for growth of the four *E. coli* B strains harboring pTrxHisS - CBF12.2. The added glucose increases repression of TrxHisS - CBF expression under non - induced conditions and low temperature reduces formation of inclusion bodies (Kopetzki *et al.*, 1989). The growth rate for all four strains carrying the pTrxHisS - CBF12.2 construct were near identical and logarithmic from A_{600nm} ~ 0.2 until ~ 1.0. To direct TrxHisS - CBF production during the later phase of logarithmic growth, cells were grown to A_{600nm} ~ 0.5 before expression of TrxHisS - CBF12.2 was induced by IPTG addition (Fig. 4.3). The induction efficiency in BLR(DE3)pLys strain carrying pTrxHisS - CBF12.2 was initially tested using IPTG concentrations of 0.1, 0.5., 1.0 and 2.0 mM and varying induction time from 1, 2, and 4 h (Fig. 4.4). Maximal accumulation of fusion protein in total cell pellet was obtained after 4 h of induction with 2.0 mM IPTG; however, an IPTG concentration of 1.0 mM and 2 h gave almost equally high yield. Only a very low level of TrxHisS - CBF12.2 was produced in cells induced for 1 h with 0.1 mM IPTG.

To evaluate all four host strains for efficiency of TrxHisS - CBF12.2 production, protein accumulation upon induction with varying IPTG concentration (0.1 and 0.5 mM) and time (1, 2 and 3 h) were analyzed. TrxHisS - CBF12.2 accumulation in BL21(DE3)plysS was at least two-fold lower for all IPTG concentrations and induction times compared to the other hosts (Fig. 4.4 and 4.5). The strains designed to enhance disulphide bond formation, [Origami B(DE3)pLysS], or fine - tune IPTG - induction [Tuner(DE3)plysS] showed no advantage in TrxHisS - CBF12.2 production over *recA*$^-$ strain BLR(DE3)pLysS, as protein profiles for all IPTG levels and induction times were very similar for these three strains. Thus, for subsequent studies, the BLR(DE3)plysS strain was selected as host for production of all TrxHisS - CBF fusions.

4.4.5 Optimization of soluble protein production

To optimize extraction of soluble protein from IPTG - induced BLR(DE3)plysS cells carrying TrxHisS - CBF, two cell disruption methods were tested. The first method utilized sonication treatment combined lysozyme digestion of cell wall. In the second method a commercially available solution containing a mixture of detergents and nucleases, BugBuster Master Mix, was used to lyze the cells. Analysis of the generated soluble fractions by SDS - PAGE showed that the BugBuster treatment resulted in higher extraction of TrxHisSCBF12.2

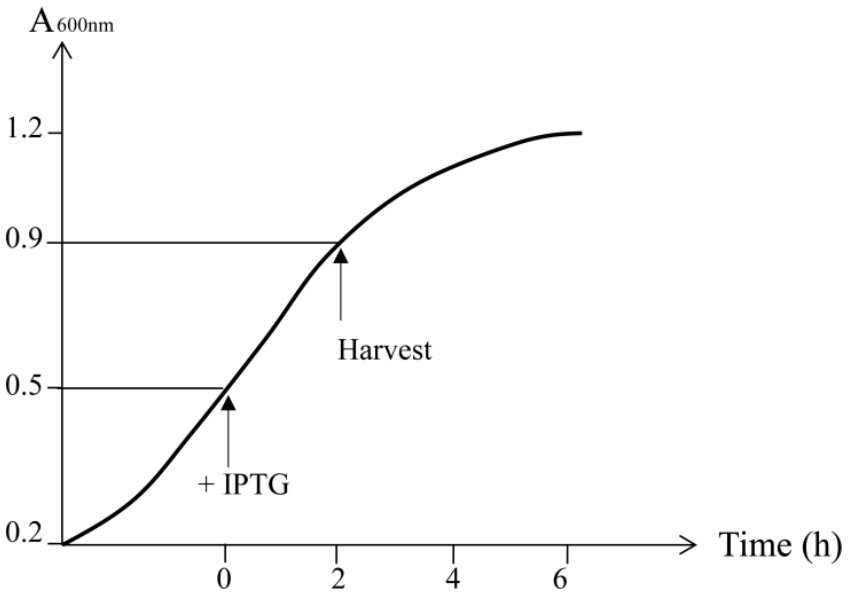


Figure 4.3 Growth curve for BLR(DE3)pLysS cells expressing TrxHisS - CBF12.2 protein. Time points for TrxHisS - CBF12.2 induction and cell harvest are indicated.

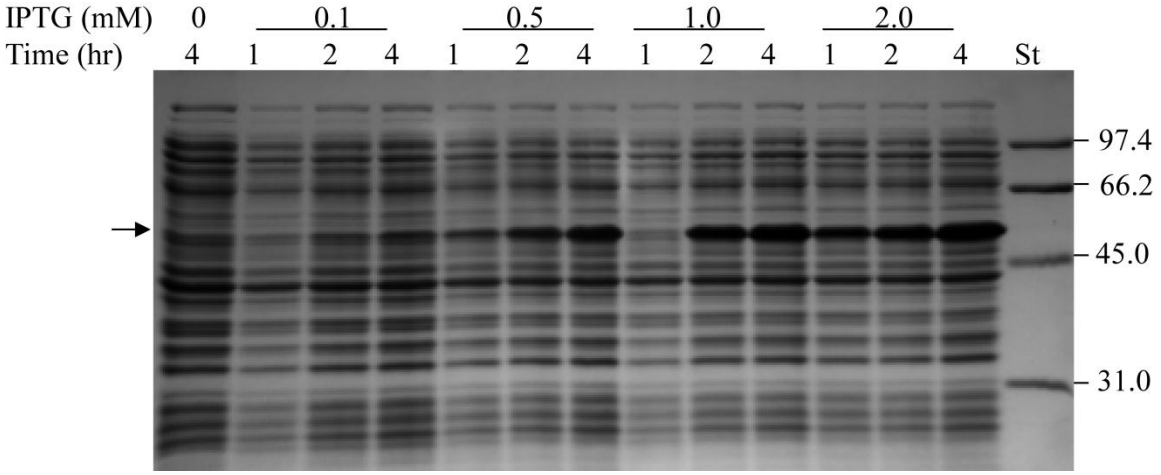


Figure 4.4 Analysis of proteins produced in BLR(DE3)pLys cells harboring pTrxHisS - CBF12.2.
SDS - PAGE of total *E. coli* proteins produced in non-induced and IPTG-induced cultures grown with varying IPTG concentration and time as indicated. Arrow indicates migration of TrxHisS - CBF12.2. Molecular weight (kDa) of standard proteins is indicated to the right.

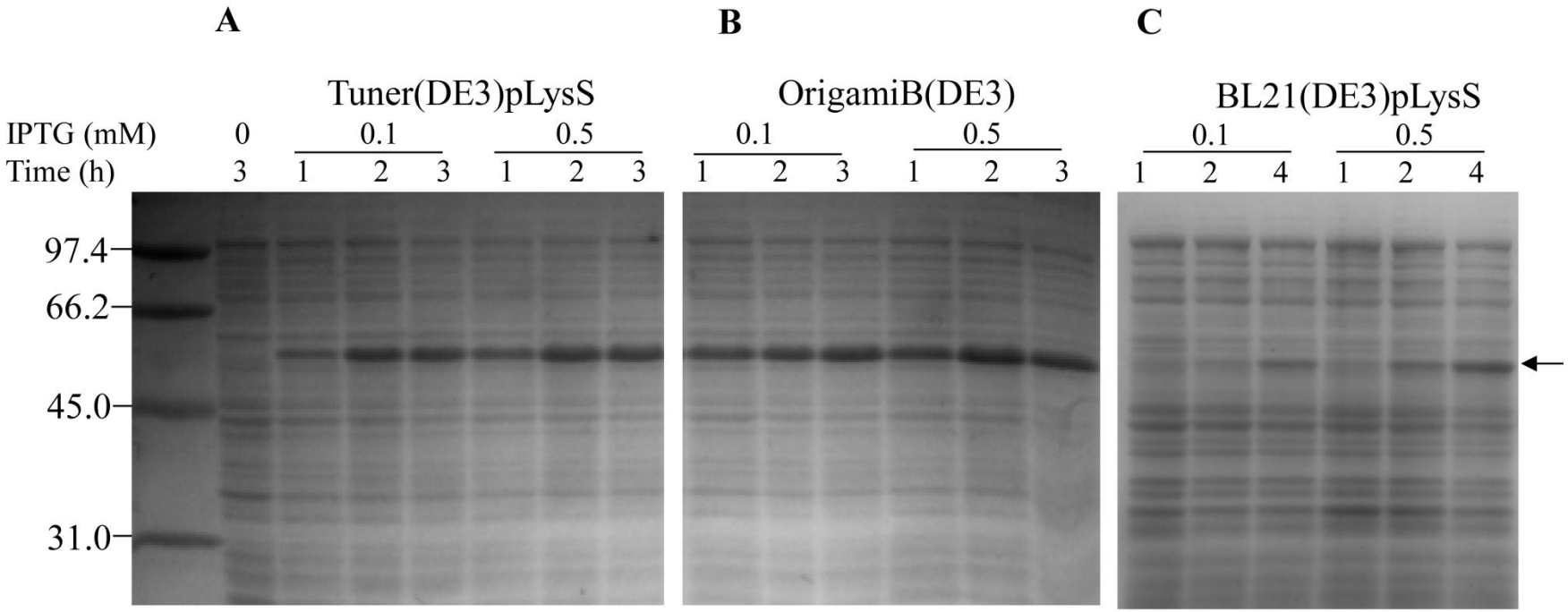


Figure 4.5 Effect of IPTG concentration and induction time on TrxHisS - CBF12.2 production in *E. coli*.

SDS - PAGE analysis of total cell extracts of induced *E. coli* strains Tuner(DE3)pLysS, Origami B(DE3)pLysS, and BL21(DE3)pLysS carrying pTrxHisS - CBF12.2. Samples (~ 40 µL) analyzed were prepared from cultures induced by varying IPTG concentrations and induction times as indicated. Arrow indicates migration of TrxHisS - CBF12.2. Molecular weight (kDa) of standard proteins are indicated to the left.

protein compared to the sonication / lysozyme method (Fig. 4.6A). Therefore, cells were solubilized in solution containing BugBuster for preparation of soluble extracts in the following experiments.

To test conditions for maximal production of soluble recombinant protein, *E. coli* soluble extracts obtained from cultures induced with different IPTG concentration and induction time were compared. Although four hours of induction and 2.0 mM IPTG produced the highest amount of TrxHisS - CBF12 in cell pellet (Fig. 4.4), these conditions did not produce maximal yield of soluble TrxHisS - CBF12. Rather, a shorter induction time of two hours (Fig. 4.6B) and lower IPTG concentration (0.5 mM) maximized recovery of native TrxHisS - CBF12.2 in the soluble fraction (Fig. 4.7). The influence of temperature on production of soluble TrxHisS - CBF12.2 was tested by growing the cells at 37 ^{o}C and 28 ^{o}C and inducing them with 0.5 mM IPTG. Cells grown at 37 ^{o}C reached A_{600nm} ~ 0.9 within one hour of IPTG induction, while the cells grown at 28 ^{o}C needed two hours to reach the same stage. The slower growing culture was found to produce more soluble protein than cultures grown at 37 ^{o}C (Fig. 4.8), which agreed with reports showing higher formation of inclusion bodies at higher culture temperatures.

The initial optimization experiments showed maximal production of native TrxHisS - CBF12.2 in BRL(DE3)pLysS cells was obtained with 28 ^{o}C culture temperature, induction with 0.5 mM IPTG for two hours and use of Bugbuster to make the soluble extract. These conditions resulted in good accumulation of fusion protein in *E. coli* cell pellets (Fig. 4.9). However, the amounts of soluble protein varied between the different recombinant proteins with TrxHisS - CBF12.2 showing the highest yield. In stark contrast, TrxHisS - CBF17.0 showed good accumulation in cell pellet, but could not be recovered in soluble phase (Fig. 4.10). In an attempt to produce TrxHisS - CBF17 in soluble phase, a very short induction time (one hour) and very low IPTG concentration (0.1 mM) was tested. However, this treatment did not reduce sequestration of TrxHisS - CBF17 in inclusion bodies (Fig. 4.11). A truncated version of TrxHisS - CBF17 with 177 amino acids removed from C - terminal end also became trapped in inclusion bodies when expressed in *E. coli*.

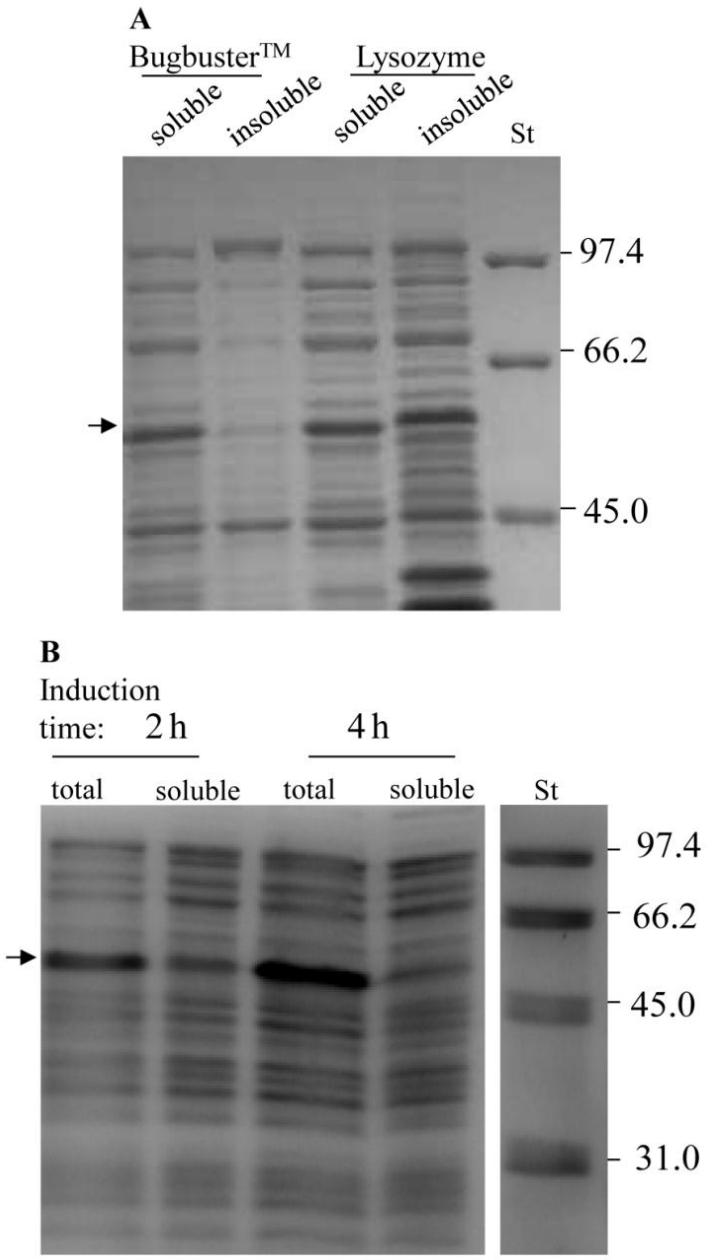


Figure 4.6 Effect of extraction method on TrxHisS - CBF12.2 yield in soluble extract.

(A) SDS - PAGE analysis of soluble and insoluble (pellet) proteins in 40 µL aliquots upon E. coli extraction using enzymatic (BugBuster) or lysozyme + sonication (Lysozyme) method. Migration of TrxHisS - CBF12.2 is shown by arrow.Molecular weight of standard polypeptides is indicated to the right. **(B)** Soluble fraction with different induction time for expression of CBF12.2, soluble cell extract made by the BugBuster® method. Arrow indicate migration of TrxHisS - CBF12.2. Molecular weight (kDa) of standard proteins is indicated on right side of last lane.

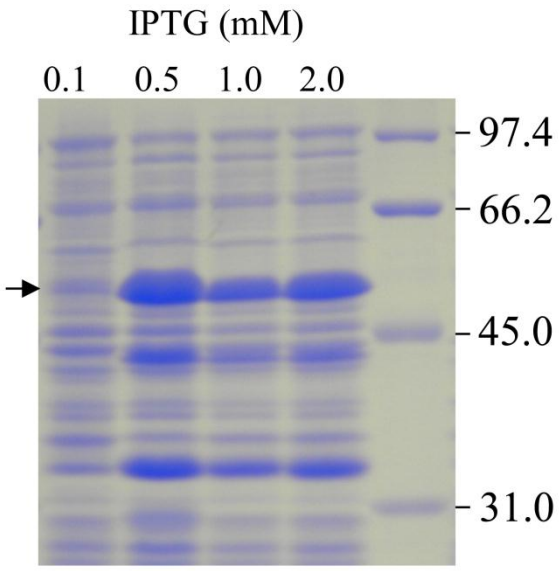


Figure 4.7 Effect of IPTG concentrations on TrxHisS - CBF12.2 solubility.

SDS - PAGE of *E. coli* BRL(DE3)pLysS/pTrxHisS - CBF12.2 soluble extracts (500 µg) prepared from cells induced for 2 h with different IPTG concentrations. Arrow indicates migration of TrxHisS - CBF12.2. Molecular weight (kDa) of standard proteins is shown to the right.

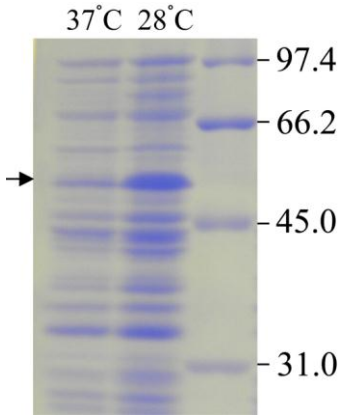


Figure 4.8 Effect of growth temperature on TrxHisS - CBF12.2 solubility.

SDS - PAGE analysis of BRL(DE3)pLysS/pTrxHisS - CBF12.2 soluble extracts (40 µL) soluble extracts prepared from cultures induced with 0.5 mM IPTG for 2 h at 37 °C or 28 ° C. Arrow indicates migration of TrxHisS - CBF12.2. Molecular weight (kDa) of standard proteins is shown to the right.

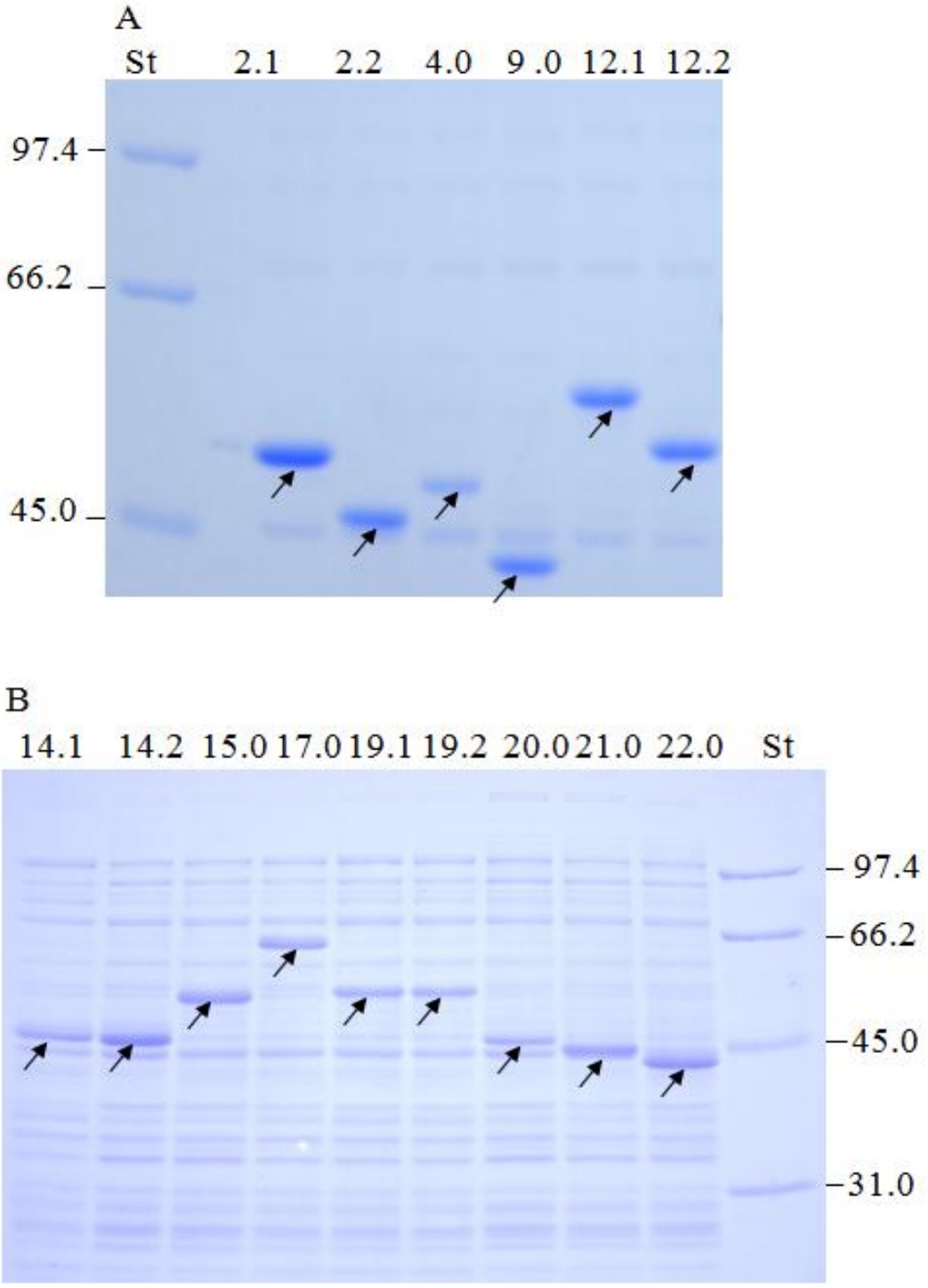


Figure. 4.9 SDS - PAGE analysis of IPTG - induced BLR(DE3)pLysS/pTrxHisS - CBF cultures.

Samples of cell pellets of 0.5 mL 2 h, 0.5 mM IPTG - induced cells grown at 28 °C were analyzed by 12 % SDS - PAGE and stained with Coomassie Brilliant Blue R250. Number above lanes indicate TrxHisS - CBF variant expressed. Arrows indicated migration of TrxHisS - CBFs. Molecular weight (kDa) of standard proteins are shown to the left and right.

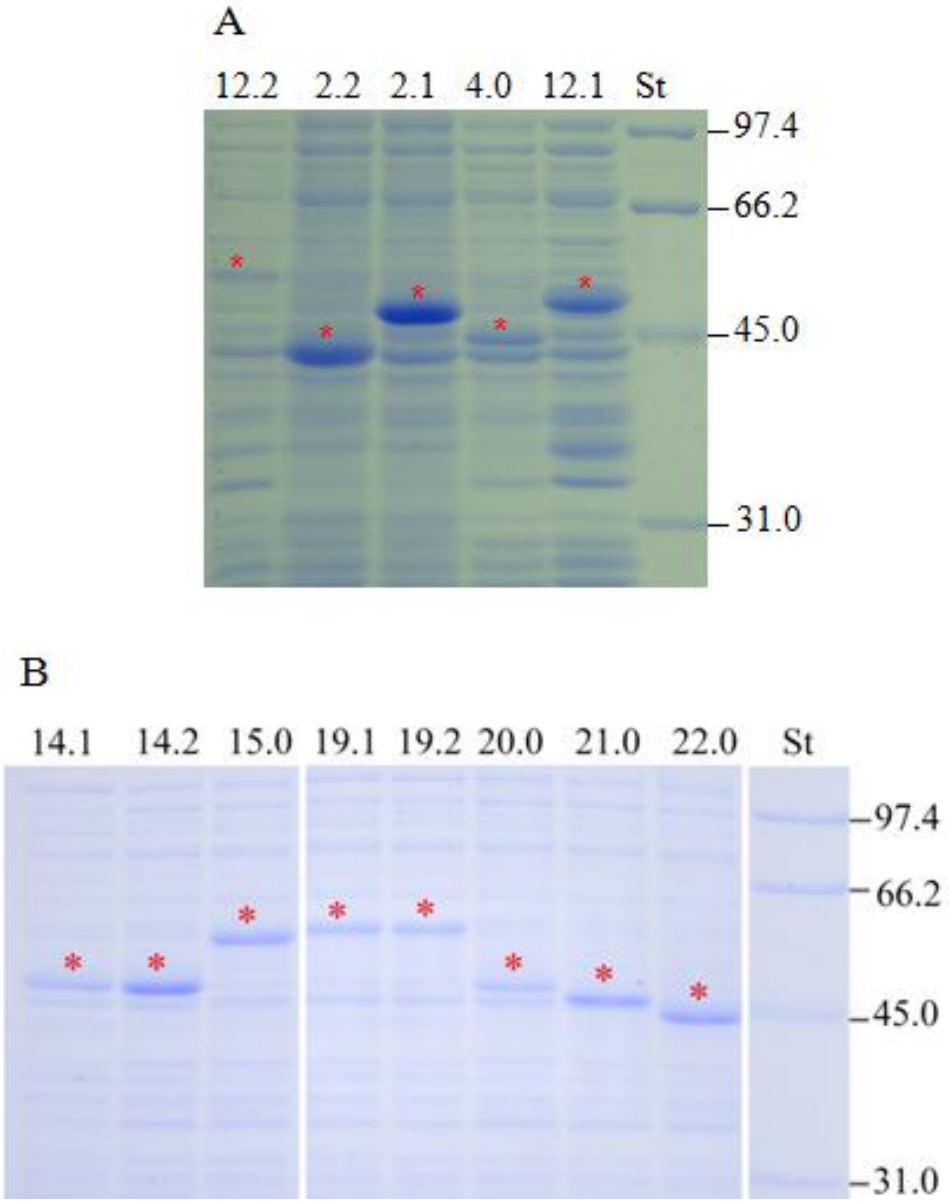


Figure 4.10 SDS - PAGE analysis of BRL(DE3)pLysS/TrxHisS - CBF soluble extracts.

Samples of soluble fractions obtained by BugBuster method from cultures grown at 28 °C and induced with 0.5 mM IPTG for 2 h. Numbers above lanes indicate CBF isoform expressed and analyzed. Asterix indicates position of the expressed protein. Molecular weight (kDa) of standard proteins are shown to the right.

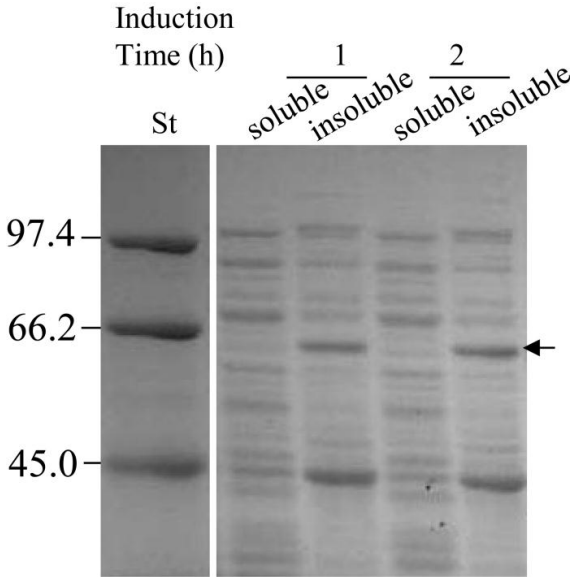


Figure 4.11 Analysis of TrxHisS - CBF17.0 solubility in IPTG - induced *E. coli* cells.

SDS - PAGE analysis of BugBuster soluble and insoluble fraction of BRL(DE3)pLysS/pTrxHisS - CBF17.0 cells inducted with 0.1 mM IPTG for one or two hours. Arrow indicates migration of TrxHisS - CBF17.0. Molecular weight (kDa) of standard proteins is shown to the left.

4.4.6 Purification of TrxHisS - CBF

To study the properties of expressed recombinant proteins in *E. coli*, it is desirable to purify it from host cell soluble proteins. This can be achieved by metal affinity chromatography utilizing the exposed His - tag, which binds to resin charged with divalent nickel ions. Addition of a soluble competing chelator such as imidazole or EDTA is then used to release bound His - tagged proteins from the resin. Ni - MAC cartridges (EMD Millipore, San Diego, CA, USA) were chosen for purification of TrxHisS - CBF12.2 fusion under native and denaturing conditions according to manufacturer's instructions. The Ni - MAC cartridges are packed with fractogel matrix which is a methacrylate - based polymeric resin with a pore size of 40 - 90 µm and an inert hydrophilic surface. These properties result in low non - specific binding and high protein binding capacities. Purification of TrxHisS - CBF12.2 under non-denaturing conditions at 4 °C and at room temperature from 100 mL induced cell culture gave similar yields of ~ 1.6 mg protein from 40 mg total soluble protein (about 4 %). When the same recombinant polypeptide was prepared under denaturing conditions, yield was two times higher; 4.4 mg from 64 mg cell extract (7 %). The increase in yield under denaturing conditions was likely due to efficient solubilization of TrxHisS - CBF12.2 aggregates by urea. Analysis of purified TrxHisS - CBF12.2 native protein by SDS - PAGE revealed a major polypeptide with an apparent molecular mass of 53 kDa (Fig. 4.12), which corresponded closest to expected mass of expressed protein (43 kDa). In addition, a few smaller (30 - 35 kDa) polypeptides were observed and these products were more abundant when purification was done at room temperature as compared to 4 °C (Fig. 4.12). In contrast, purification done under denaturing conditions resulted only in a single major product of 53 kDa (Fig. 4.13). A single major polypeptide was also obtained when the 17 kDa TrxHisS tag was purified from BLR(DE3)pLysS cells harboring pTrxHisS (Fig. 4.17A). Thus, the 30 - 35 kDa polypeptides obtained during native purification of TrxHisS - CBF12.2 were unlikely to represent co-purified host proteins, but were rather derived from TrxHisS - CBF12.2.

4.4.7 Analysis of purified TrxHisS - CBF

To verify the identity of the 53 and 35 kDa polypeptides obtained upon native purification of TrxHisS - CBF12.2 (Fig. 4.12), the bands were excised from SDS - PAGE gel and subjected to LC - ESI - MS analysis. Six of the peptides generated from the larger 53 kDa band showed highly significant ($p < 0.05$) matches to thioredoxin tag, AP2 and C - terminal domains of

TrxHisS - CBF12.2, respectively. (Fig. 4.14; Table 4.3). These results confirmed that the 53 kDa polypeptide was a full-length TrxHisS - CBF12.2. Peptides derived from the smaller 35 kDa polypeptide band showed strong matches to the thioredoxin tag and AP2 domain of TrxHisS - CBF12.2 (Fig. 4.14). The lack of peptides matching the C - terminal end suggested that the 35 kDa protein band was a TrxHisS - CBF12.2 product truncated at the C - terminal end. Since the 35 kDa polypeptides were reduced when purification was done at low (4 °C) temperature and absent when done under denaturing conditions (Fig. 4.13), it was concluded that premature termination of translation did not cause accumulation of truncated TrxHisS - CBF12.2. As protease activities generally decline with temperature and are inactivated by 6 M urea, incomplete silencing of proteolytic activities during extraction and/or Ni - MAC affinity chromatography was more likely to have caused the reduced recovery of full - length TrxHisS - CBF12.2.

To reduce the protease activity during purification of TrxHisS - CBF12.2, protease inhibitor concentration during preparation of cell extract was increased ten - fold. A much reduced proteolytic activity was noted upon purification of TrxHisS - CBF12 (Fig. 4.15); unfortunately, several attempts to repeat the experiment failed and the amount of inhibitors appeared to be of less significance.

In another purification experiments, Ni - MAC columns recharged with freshly prepared $NiSO_4$ were used. Interestingly, use of the recharged columns resulted in a very clean TrxHisS - CBF12.2 preparation where one major 53 kDa protein band was observed and truncated variants of TrxHisS - CBF12.2 were near absent (Fig. 4.16). Repeated purifications using recharged columns consistently showed very low amount TrxHisS - CBF12.2 truncated products and suggested that the Ni - MAC columns as the source of the apparent protease activity in previous experiments. Thus, for the purification of all TrxHisS - CBFs under native conditions, recharged columns were used (Fig. 4.17).

4.4.8 TrxHisS - CBFs migration in SDS - PAGE

It was observed from the SDS - PAGE analysis of total cell lysate that some TrxHisS - CBF polypeptides migrate slower than predicted from their amino acid sequences (Fig. 4.10; Table 4.4). The deviations were highest for TrxHisS - CBF12.2, 14.2, 17.0 and 19.2, for which apparent molecular masses were 8 - 15 kDa higher than their predicted values.

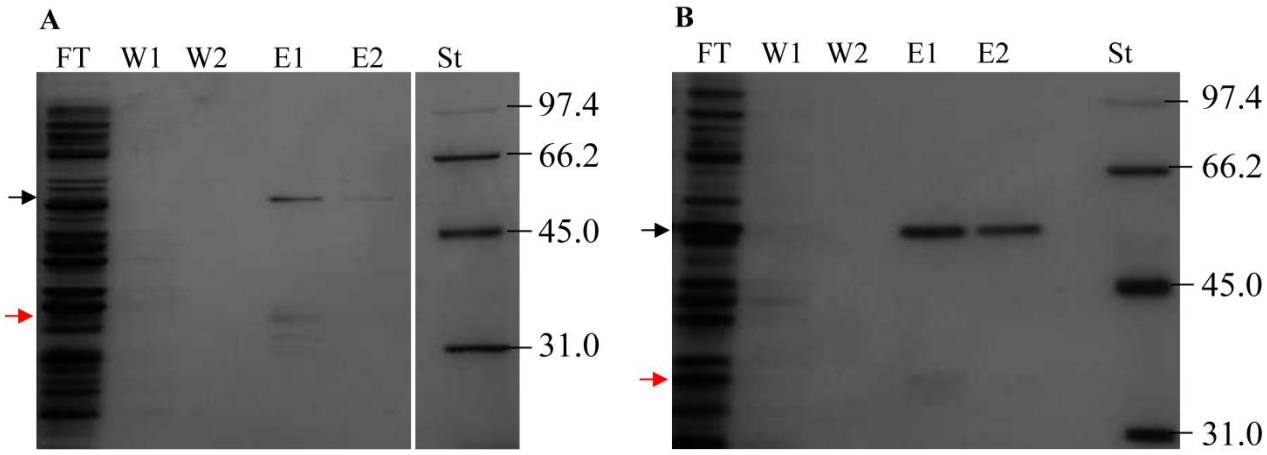


Figure 4.12 Purification of TrxHisS - CBF12.2 under native conditions.

SDS - PAGE of fractions obtained during Ni - MAC chromatography *E. coli* cells soluble extract prepared from BRL(DE3)pLysS/pTrxHisS - CBF12.2. Fractions purified at (A) room temperature and (B) 4 °C are shown. The samples analysed were 40 µl - aliquots of flow through (FT), wash fraction 1 (WB1), wash fraction 2 (WB2), eluted fraction 1 (E1) and eluted fraction 2 (E2). Migration of 53 kDa full - length TrxHisS - CBF12.2 (black arrow) and ~ 31 - 35 kDa co - purified products (red arrow) are indicated. Molecular weight (kDa) of standard proteins are shown to the right.

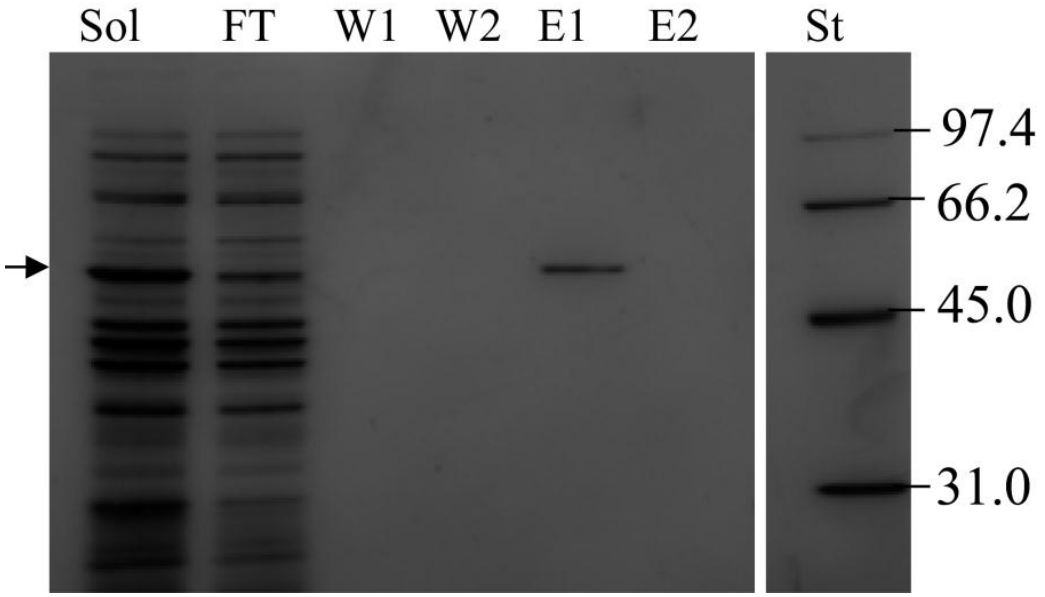


Figure 4.13 Purification of TrxHisS - CBF12.2 under denaturing conditions.

SDS - PAGE analysis of 10 µL total soluble extract of BRL(DE3)pLysS/pTrxHisS - CBF12.2 cells (Sol) and 10 µL flow through (FT), 40 µL of wash fractions 1 (WB1), 40 µL of wash fractions 2 (WB2), 40 µL of eluted fraction 1 (E1) and eluted fraction 2 (E2) obtained from Ni - MAC chromatography of soluble extract. Arrow indicates migration of 53 kDa full - length TrxHisS - CBF12.2. Molecular weight (kDa) of standard proteins is shown to the right.

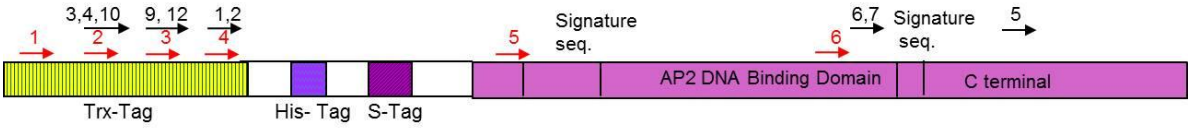


Figure 4.14 Schematic illustration peptide matches to TrxHisS - CBF12.2.

Arrows indicated peptides generated upon trypsin digest of 53 kDa TrxHisS - CBF12.2 and 35 kDa truncated TrxHisS - CBF12.2 (Table 4.5; Fig. 4.12) and identified by LC - ESI - MS and Mascot searches. Black arrows indicate peptides obtained from 53 kDa polypeptide and red arrow indicate peptides obtained from 35 kDa polypeptide.

Table 4.3 Details of Mascot search results for complete and small peptides.

Small molecular mass peptide results

	Accession No.	Peptide name	Sequence
1	gi|230777	Chain A, Thioredoxin, *E. coli*	SDKIIHLTDDSFDTDVLK
2	gi|149241884	Chain A, Thioredoxin	MIAPILDEIAEEYQGK
3	gi|84393642	Thioredoxin, *Vibrio. splendidus*	LNIDQNPGTAPK
4	gi|157833855	Thioredoxin - S2, *E. coli*	GIPTLLLFK
5	gi|63098613	CRT / DRE binding factor 12, *T. monococcum*	NWNSPASPPSSLEQGMPTSPASPTPK
6	gi|63098613	CRT / DRE binding factor 12, *T. monococcum*	LWLGTHVTAEAAAR

Significance threshold $p < 0.05$

Complete TrxHisS - CBF12.2 protein results

1	gi|24987897	Chain A, Thioredoxin	GIPTLLLFK
2	gi|157833855	Chain A, Thioredoxin, *E. coli*	GIPTLLLFK
3	gi|59710664	Thioredoxin, *Vibrio fischeri*	MIAPILDEIADEYEGK
4	gi|84393642	Thioredoxin - S2, *E. coli*	MIAPILDEIANEYEGK
5	gi|63098613	CRT / DRE binding factor 12, *T. monococcum*	AAIGAVVDFLR
6	gi|63098613	CRT / DRE binding factor 12, *T. monococcum*	AHDAAMLALYGR
7	gi|75706706	CBF12, *H. vulgare*	AHDAAMLALYGR
8	gi|136429	Trypsin, Precursor	LSSPATLNSR
9	gi|238755776	Thioredoxin 1, *Yersinia ruckeri*	LNIDENPGTAPK
10	gi|271502411	Thioredoxin, *Dickeya dadantii*	MIAPILDEIADEYEGK
11	gi|28317	unnamed protein product, *Homo sapiens*	ALEESNYELEGK
12	gi|86148270	Thioredoxin, *Vibrio* sp.	LNIDENPGTAPK

Significance threshold $p<0.05$

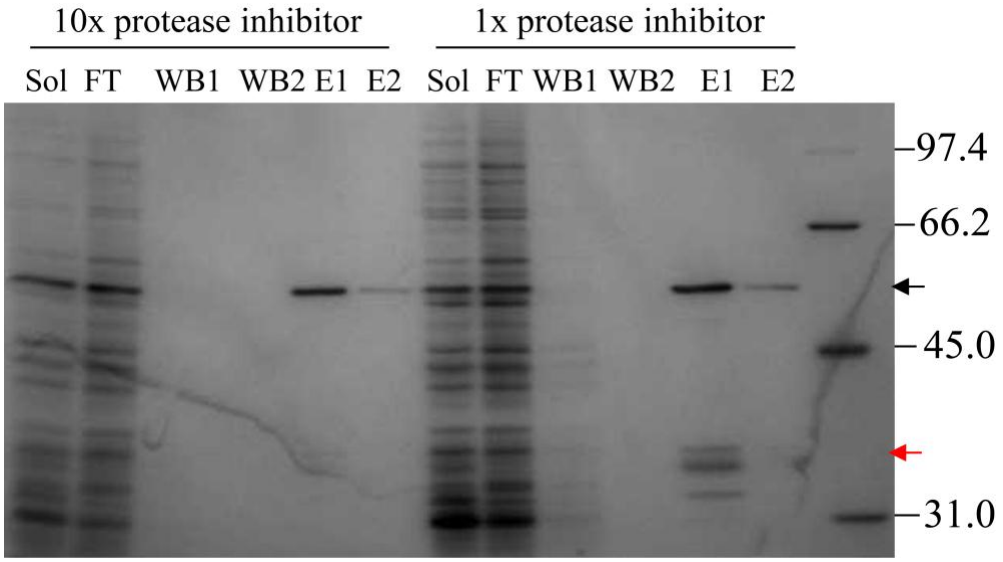


Figure 4.15 Effect of protease inhibitor concentration during purification of TrxHisS - CBF12.2.

SDS - PAGE analysis of TrxHisS - CBF12.2 fractions obtained from native purification using 10x and 1x Sigma protease inhibitor cocktail. Samples analyzed were soluble extract (Sol), flow through (FT), wash buffer 1 (WB1), wash buffer 2 (WB2), eluted fraction 1 (E1) and eluted fraction 2 (E2). Migration of 53 kDa full - length and 31 - 35 kDa truncated co - purified polypeptides are indicated by black and red arrows respectively. Molecular weight (kDa) of standard proteins is shown to the right.

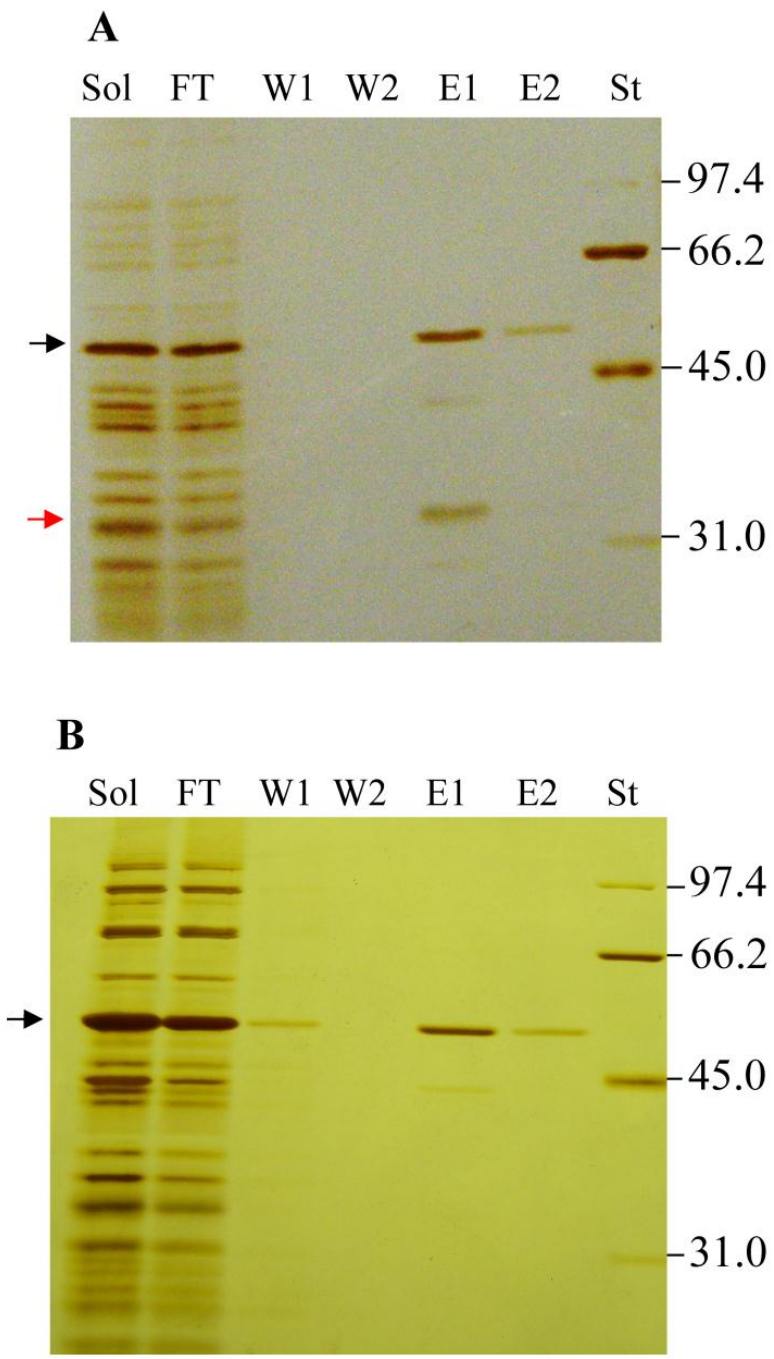


Figure 4.16 Purification of TrxHisS - CBF12.2 under native conditions.

(A) Original Ni - MAC column and (B) Recharged Ni - MAC column. Migration of full - length (black arrow) and truncated TrxHisCBF12.2 (red arrow) are indicted. Molecular weight (kDa) of standard proteins is shown to the right.

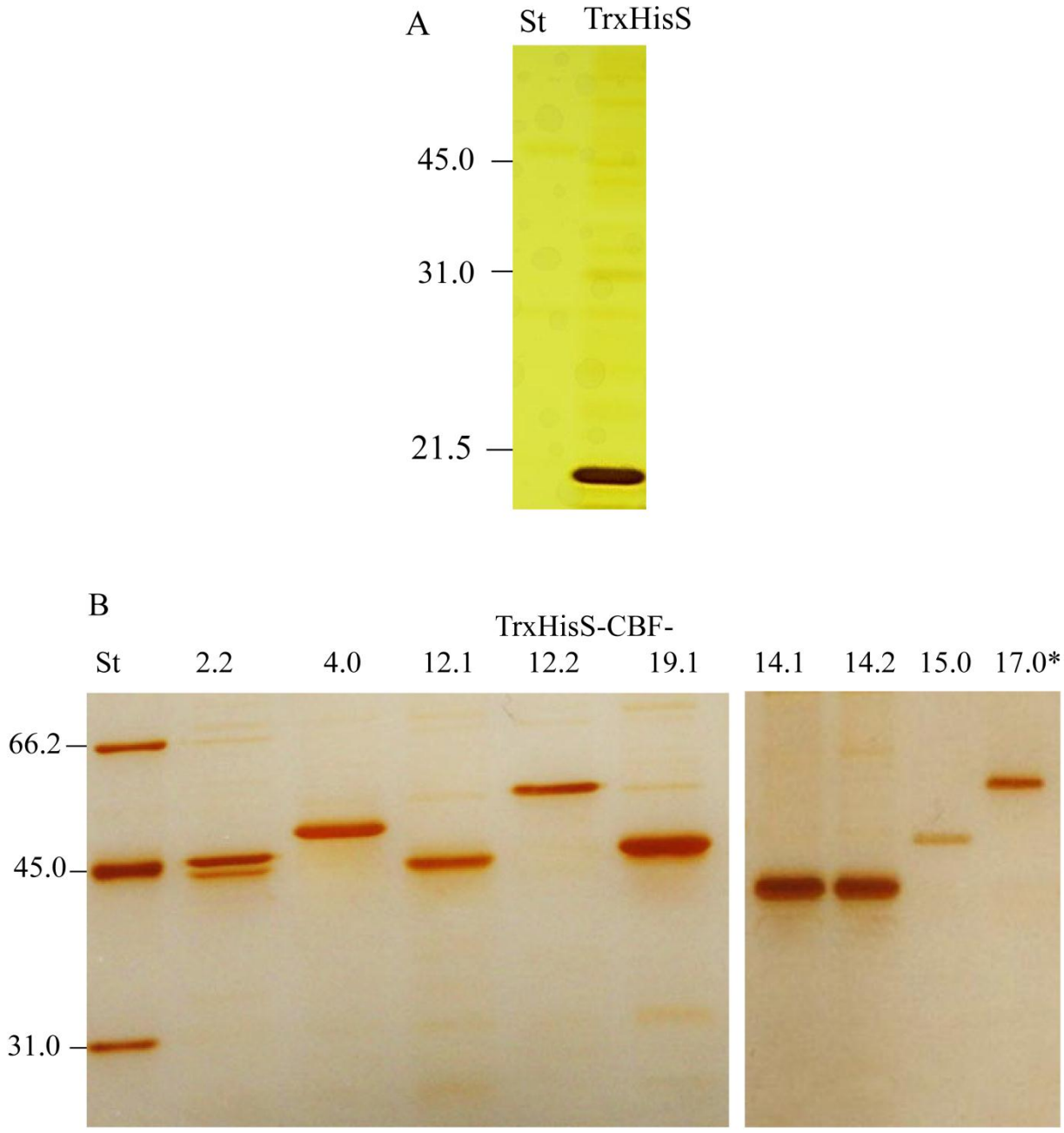


Figure 4.17 Purified (A) TrxHisS and (B) TrxHisS - CBF proteins under native conditions.

Numbers above each lane indicate the CBF variant. TrxHisS - CBF17.0 was purified under denaturing condition. Molecular weight (kDa) of standard proteins are shown on the left.

For TrxHisS - CBF2.2, CBF4.0 and the TrxHisS tag alone, the migration seemed normal as their estimated masses were 1 - 4 kDa different, which is within the error of determination (Figs. 4.9, 4.17; Table 4.4). Re - sequencing of vector isolated from induced cultures confirmed DNA sequence of *CBF12.2* coding region and thus, enrichment of mutation(s) during culture growth as a reason for the larger than expected proteins on SDS - PAGE seemed unlikely. Both soluble and purified fractions of TrxHisS - CBFs showed slow migration (Figs. 4.10, 4.11), and thus, components of host cells were unlikely the cause of the slow migration on SDS - PAGE gels. Also a C - terminal truncated variant of TrxHisS - CBF12.1, CBF12.2, CBF15.0 and CBF17.0 showed slow migration (Fig. 4.18; Table 4.5). Thus, based on the gel migration patterns of full - length TrxHisS - CBF and truncated versions of TrxHisS - CBF, the slow migration is associated with the first 22 - 30 kDa of TrxHisS - CBF including the AP2 domain. Although proteins are assumed to migrate according to molecular weight in SDS - PAGE gels, sometimes unusual mobility is observed and molecular mass estimates become inaccurate (Griffith, 1972).

4.4.9 Analysis of TrxHisS - CBFs properties

Similar to this study, DREB2A transcription factor from *Pennisetum glaucum* also displayed unexpected slow migration in SDS - PAGE gels and a low pI of 5.55 was suggested to cause the delay (Agarwal *et al.*, 2007). An analysis of purified TrxHisS - CBF12.2 by 2D - gel electrophoresis revealed a single major spot migrating as a 53 kDa protein (predicted value 43 kDa) and a pI of 5.9 (Fig. 4.19). The determined pI agreed well with the predicted value of 5.97 (Table 4.4) and therefore, pI for TrxHisS - CBF12.2 was not modified during production in *E. coli*. A calculation of the predicted pI values for the 15 Norstar TrxHisS - CBFs in the study showed values ranging from 4.74 to 6.80 (Table 4.4). TrxHisS - CBF-17.0 with the lowest pI (pI = 4.47) showed the highest delay in migration (15.4 kDa or + 32 %). On the other hand, TrxHisS - CBF12.1 with highest pI value (pI = 6.8) also showed a high degree of delayed migration (7.87 kDa or + 18 %). The same degree of abnormal migration was observed for TrxHisS - CBF19.2 but its pI was only 5.21. In summary, no correlation between slow migration in SDS - PAGE gels and theoretical pI values could be established for the full - length TrxHisS - CBFs.

For the truncated variants of CBF12.1 and CBF17.0, which became basic (pI = 8 - 9; Table 4.5) when the C - terminal end was removed (constructs TrxHisS - CBF12.1-T4 and TrxHisS - CBF17.0-T4), it cannot be excluded that abnormal migration to some degree was

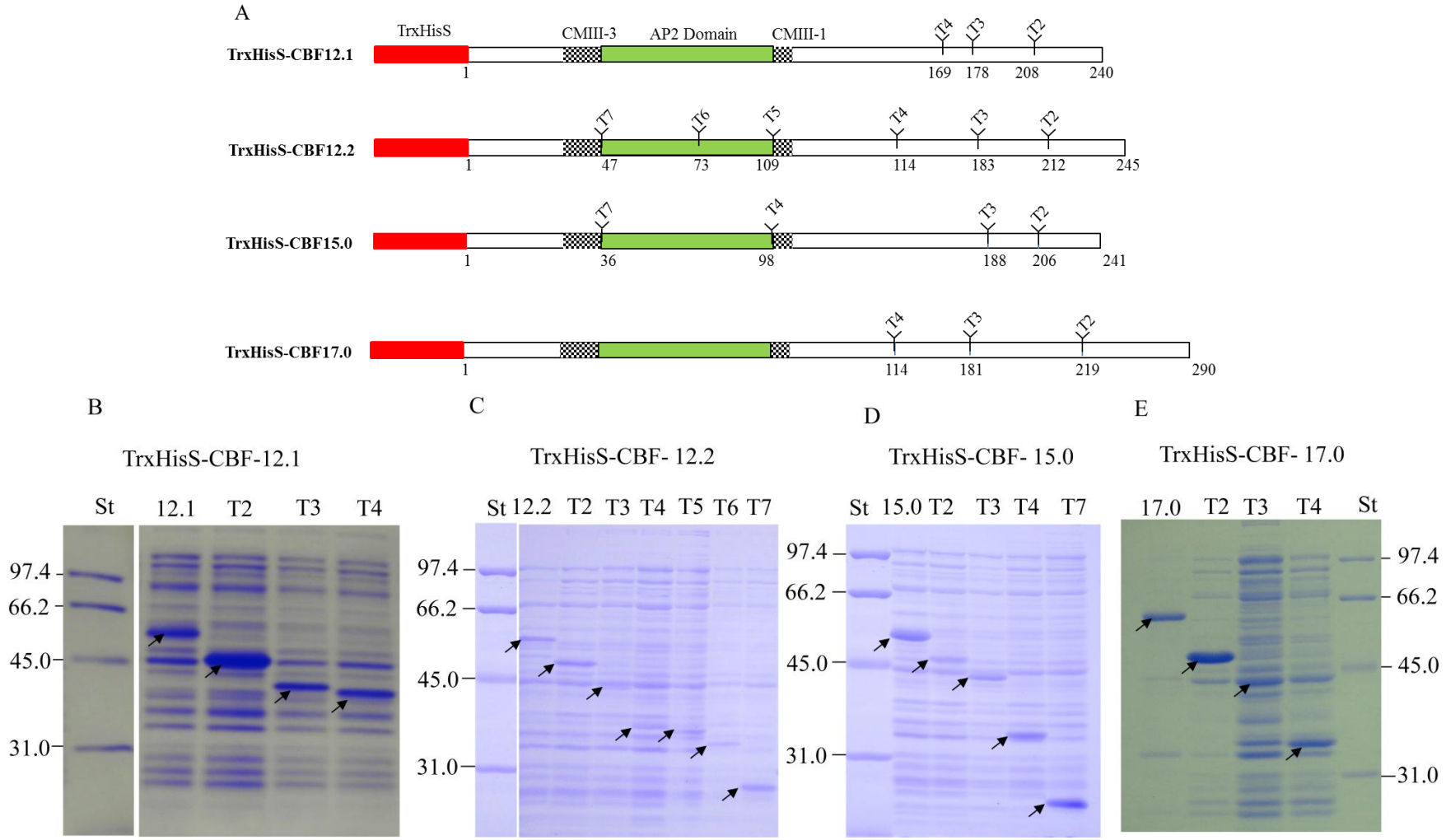


Figure 4.18 Expression of complete and truncated TrxHisS - CBF.

(A) Schematic representation of the complete TrxHisS - CBF12.1, CBF 12.2, CBF15.0, CBF17.0 and the respective truncated polypeptides. TrxHisS - (B) CBF12.1 (C) CBF 12.2 (D) CBF15.0 (E) CBF17.0 : SDS - PAGE gel stained with Coomassie Brillant Blue R250 of complete and the truncated polypeptides in BLR(DE3)plysS host cells induced with 0.5 mM IPTG. Arrows indicates the position of migration of the expressed protein. Molecular weight (kDa) of standard proteins is shown.

Table 4.4 Characteristics of CBFs expressed in *E. coli* .

TrxHisS-CBF	Amino acid	Predicted MW (kDa)	pI	Observed MW (kDa)*	Difference kDa	Difference %
2.1	378	41.6	5.15	50	8.4	20.0
2.2	383	42.2	5.26	46	4.8	11.5
4.0	380	41.2	6	46	4.7	11.5
9.0	427	45.8	6.39	45	-0.8	-1.7
12.1	398	42.1	6.8	50	7.8	18.6
12.2	403	42.8	5.97	53	10.1	23.7
14.1	370	40.3	6	48	7.6	19.0
14.2	372	40.6	6	48	7.4	18.3
15.0	399	42.4	5.52	48	5.5	13.2
17.0	448	47.6	4.74	60	12.4	26.2
19.1	392	42.4	5.11	51	8.6	20.3
19.2	392	42.1	5.21	50	7.8	18.6
20.0	375	40.9	6.57	46	5.1	12.5
21.0	355	38.6	6.28	45	6.3	16.5
22.2	433	46.3	6.38	45	-1.3	-2.7
TrxHisS-tag:	158	17.07	5.64	-	-	-

*The error in assessment of protein migration on gel was estimated to be 1 - 4 kDa.

Table 4.5 Characteristics of TrxHisS - CBF and truncated polypeptides expressed in *E. coli* .

TrxHisS - CBF-	Amino acid	Predicted size kDa	pI	Observed size kDa*	Difference kDa	Difference %
2.2	383	42.19	5.15	50	8	19
2.2-T2	281	31.09	6.96	39	8	26
12.1	398	42.13	6.80	50	8	19
12.1-T2	366	38.84	8.04	46	7	18
12.1-T3	336	35.74	9.07	41	5	14
12.1-T4	327	34.87	9.07	40	5	14
12.2	403	42.82	5.97	55	12	28
12.2-T2	370	39.35	6.28	48	9	23
12.2-T3	341	36.11	6.89	43	7	19
12.2-T4	272	30.22	7.94	38	8	27
12.2-T5	267	28.98	7.95	36	7	24
12.2-T6	231	25.09	7.54	33	8	32
12.2-T7	205	22.11	6.32	29	7	32
15.0	399	42.41	5.52	48	6	13
15.0-T2	364	38.87	6.15	45	6	15
15.0-T3	346	36.44	6.81	42	6	17
15.0-T4	256	28.71	7.88	35	6	21
15.0-T7	194	22.17	6.32	26	4	17
17.0	448	47.58	4.74	60	12	26
17.0-T2	377	40.08	6.4	47	7	17
17.0-T3	339	36.10	7.87	43	7	19
17.0-T4	272	29.29	8.09	35	6	20

*The error in assessment of protein migration on gel was estimated to be 1-4 kDa.

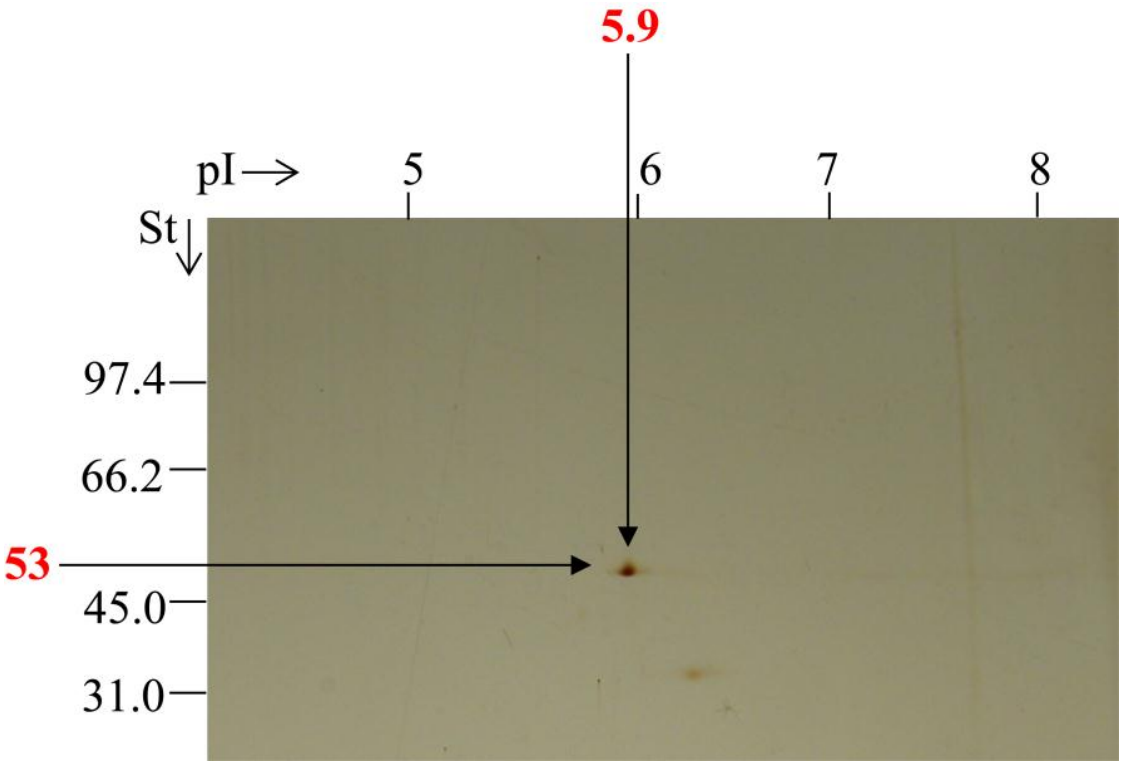


Figure 4.19 Two - dimensional gel analysis of TrxHisS - CBF12.2.

TrxHisS - CBF12.2 purified under native conditions was electrophoresed on pH 3 - 11 gradient IPG strip in first dimension and on 12 % SDS - PAGE in second dimension, followed by silver staining of gel. Apparent molecular weight and pI for TrxHisS - CBF12.2 are indicated. Molecular weight (kDa) of standard proteins is shown on the left.

caused by excessive positive charges on the polypeptides. Highly basic polypeptides like histone proteins are known to migrate slowy in SDS - PAGE gels (Tung and Knight, 1972, Banker and Cotman, 1972, Hamana, 1981).

Post - translational modifications altering the molecular mass or structure may lead to differences in apparent molecular mass as determined from SDS - PAGE and predicted value (Adamson *et al.*, 1992). These modifications may involve protein glycosylation as seen for human erythrocyte membrane proteins (Bretscher, 1971), protein phosphorylation as reported for *Tau* protein (Hosoi *et al.*,1995) or binding of metal ions like calcium, copper, iron etc. (Furthmayr and Timpl, 1970; Takahashi *et al.*, 2000). To determine if any post - translational modification altered the properties of TrxHisS - CBF12.2, the slouble protein was analyzed for glycosylation, phosphorylation and metal chelation. To test for glycosylation, SDS - PAGE separated TrxHisS - CBF12.1, CBF12.2, CBF17.0 and control protein horseradish peroxidase (HRP) were stained with Schiff's base. Only the highly (16 %) glycosylated HRP showed a pink / magenta band, whereas none of the TrxHisS - CBFs indicated any glycosylation (Fig. 4. 20).

Protein phosphorylation as a cause for TrxHisS - CBF12.2 slow migration was excluded by a protein dephosphorylation experiment, where no altered protein mobility was shown upon CIP treatment (Fig. 4.21). The possibility metal ions were bound to TrxHisS - CBFs was tested by incubating TrxHisS - CBF12.2 with imidazole, EGTA, and EDTA, respectively, followed by SDS-PAGE analysis. None of the metal chelators affected TrxHisS - CBF12.2 migration (Fig 4.22); thus metal ions were not likely to induce conformational changes that would affect protein migration. A non - globular protein conformation caused by incomplete denaturation can also have a drastic effect on protein migration in SDS - PAGE. To test for this possibility, the effect on TrxHisS - CBF migration upon treatment with different denaturing conditions was tested. Heat - denaturation in TrxHisCBF12.2 and TrxHisS - CBF17.0 in loading dye with different quantities of SDS (1, 5, and 10 %) followed by 12 % SDS - PAGE analysis did not show any difference in protein migration (Fig. 4.23). Nor was any change observed upon incubation with 50 mM urea, and 98 ^{o}C for 10 to 60 min (Fig. 4.24) or incubation with increasing the urea concentration up to 4 M (Fig. 4.25). Thus, neither heat nor urea alone or in combination had any effect on TrxHisS - CBF12.2 and TrxHisS - CBF17.0 gel migration.

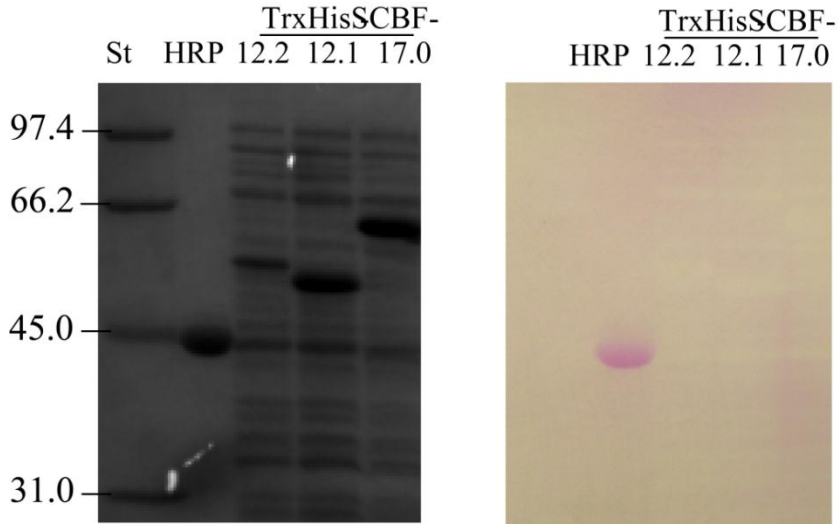


Figure 4.20 Test of protein glycosylation.

(A) 12 % SDS - PAGE gel stained with Coomassie Brilliant Blue R250. Lane 1: Horseradish peroxidase (HRP), positive control, Numbers above each lane indicate the CBF. (B) 12 % SDS - PAGE gel stained with Schiff's base. Loading done in same order as in (A). Molecular weight (kDa) of standard proteins is shown on the left.

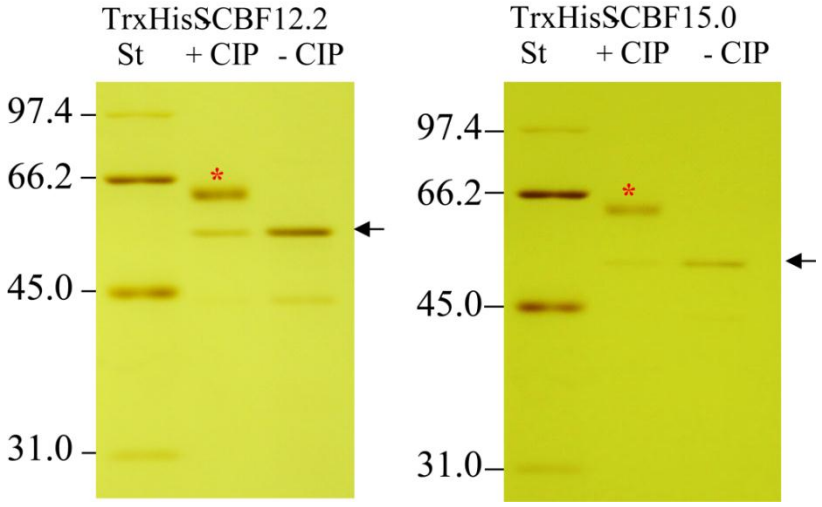


Figure 4.21 Test of protein phosphorylation.

3 µg TrxHisS - (A) CBF12.2 and (B) CBF15.0 was treated with 3 units of CIP for 1 h at 37 ^{o}C in 20 µl reaction. Arrow indicates the position of migration of the expressed protein. Astreix indicated the migration of CIP. Molecular weight (kDa) of standard proteins is shown on the left.

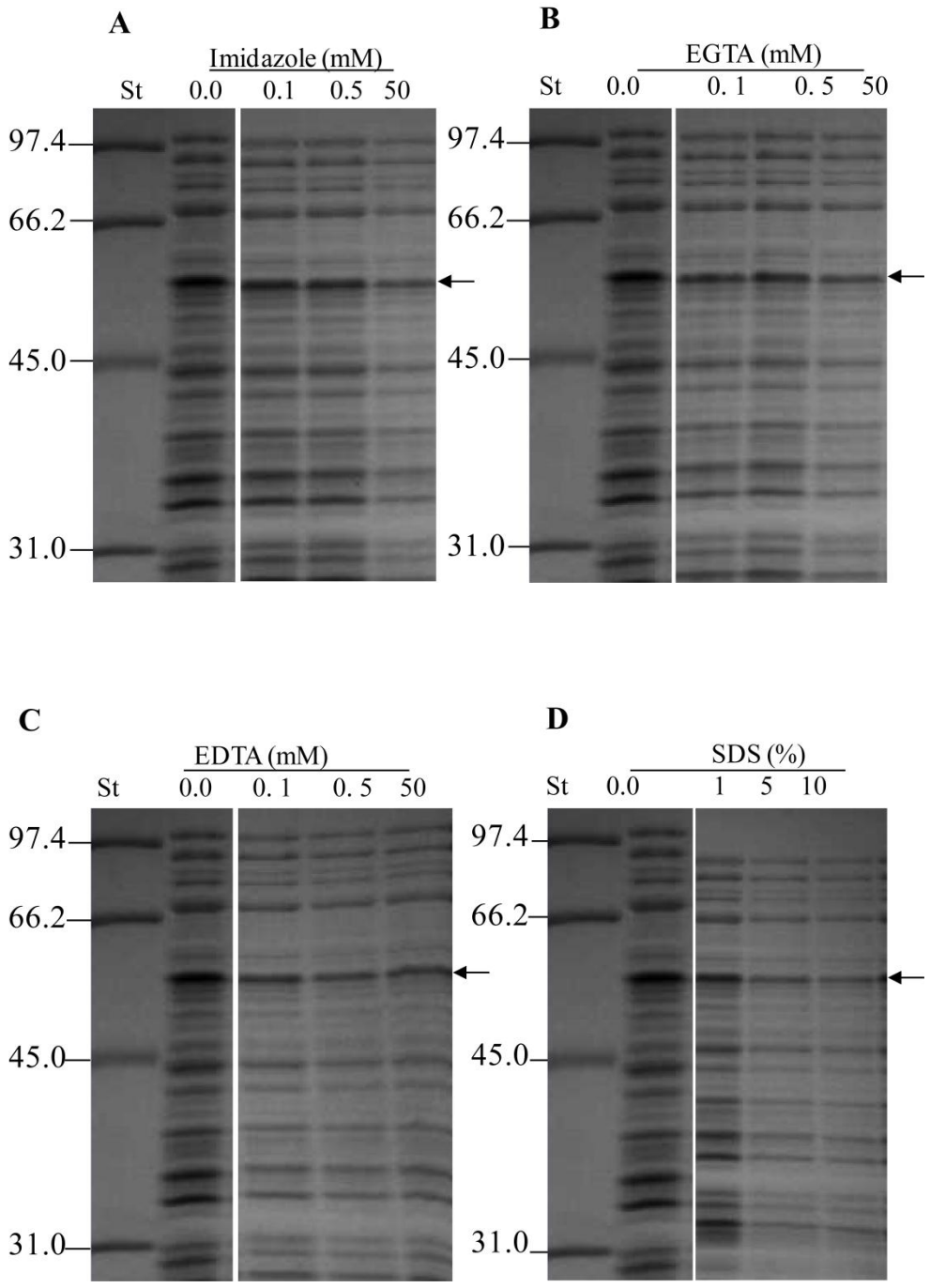


Figure 4.22 Effect of metal chelators on migration of TrxHisS - CBF12.2 on SDS - PAGE gel.

Analysis of TrxHisS - CBF12.2 containing increasing amount of (A) Imidazole, (B) EGTA, (C) EDTA and (D) SDS. Arrow indicates the position of migration of the expressed protein. Molecular weight (kDa) of standard proteins is shown on the left.

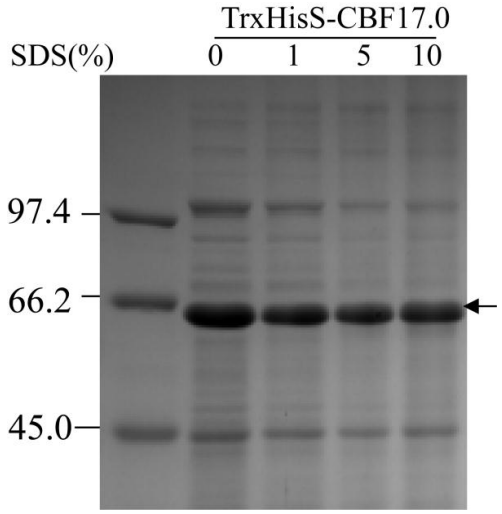


Figure 4.23 Effect of SDS concentrations on TrxHisS - CBF17.0 migration on SDS - PAGE gel.

Samples of TrxHisS - CBF17.0 boiled for 5 min in 1x SDS - PAGE sample buffer containing 0, 1, 5 or 10 % SDS. Arrow indicates migration of fusion protein. Molecular weight (kDa) of standard proteins is shown on left side.

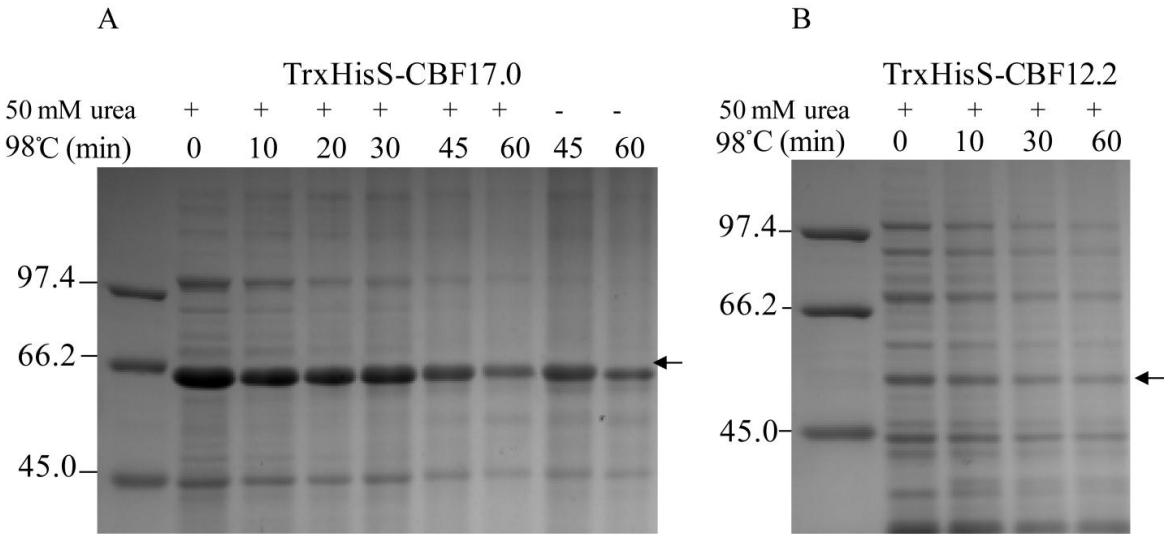


Figure 4.24 Effect of urea and heat treatments on TrxHisS - CBF migration during SDS - PAGE.

Samples of soluble cell fraction of (A) TrxHisS - CBF17.0 incubated in presence of 50 mM urea at 98 ^{o}C for 0, 10, 20, 30, 45, 60 min and for 45, 60 min without urea and (B) TrxHisS - CBF12.2 incubated at 98 ^{o}C in presence of 50 mM urea for 0, 10, 30, 60 min. Arrow indicates migration the fusion protein. Molecular weight (kDa) of standard proteins are shown on the left.

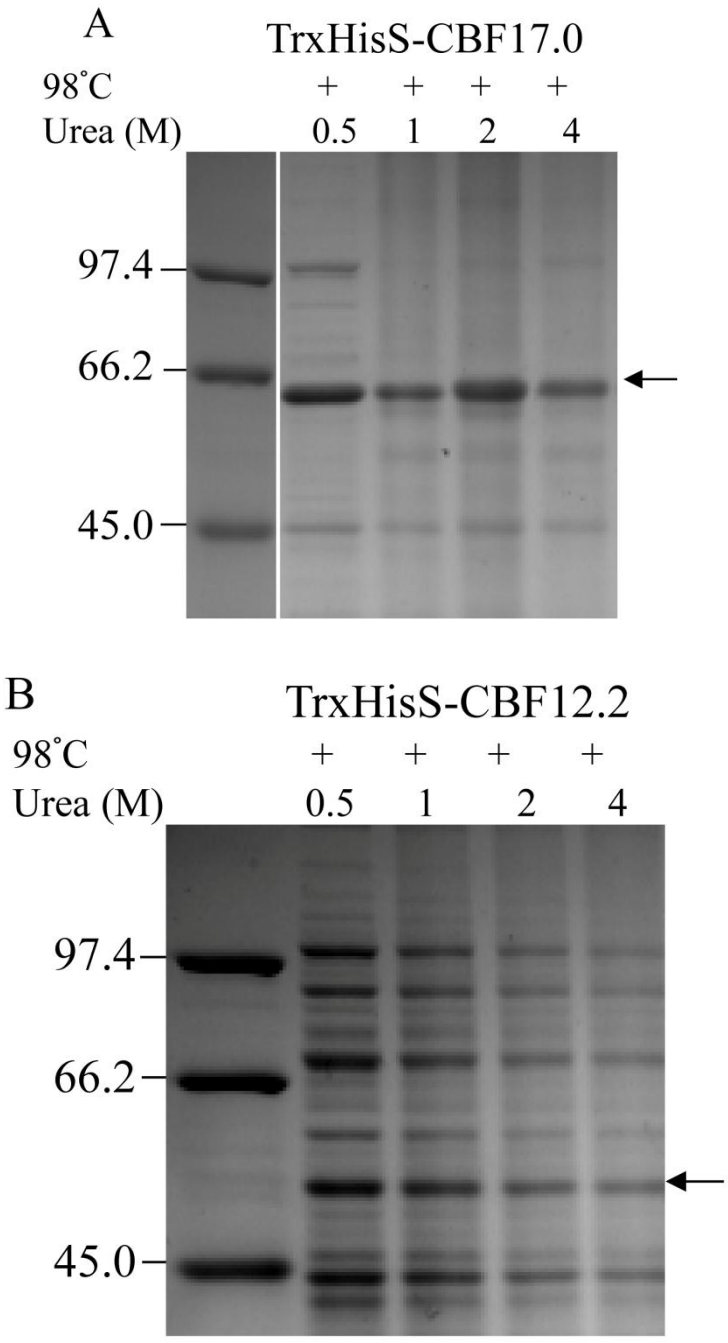


Figure 4.25 Analysis of protein unfolding with increasing amount of urea.

SDS - PAGE with soluble fraction of (A) TrxHisS - CBF17.0 and (B) TrxHisS - CBF12.2, incubated at 98 ^{o}C with 0.5, 1, 2, 4M urea. Arrow indicates the position of migration of the expressed protein. Molecular weight (kDa) of standard proteins is shown on the left.

4.4.10 Structural analysis of TrxHisS - CBF

A structural analysis of purified TrxHisS - CBF12.2, TrxHisS - CBF15.0, TrxHisS - CBF17.0 full-length fusions and truncated variants TrxHisS - CBF12.2-T7, TrxHisS - CBF12.1-T2, and TrxHisS was undertaken to determine degree of homogeneity and secondary structures. The size distribution profiles of purified proteins determined by Dynamic Light Scattering (DLS) showed polydispersity values lower than 18 % (Table 4.6), which suggested all proteins tested including the TrxHisS tag were monomeric. Based on the determined hydro - dynanic radius and calculated sedimentation cofficent ratios (Smax / S), the full - length TrxHisS - CBFs were predicted to be highly (TrxHisS - CBF17.0 and TrxHisS - CBF12.2) or moderately (TrsHisS-CBF15.0) elongated (Table 4.6). The TrxHisS tag was globular. From this analysis it was interesting to note that the two highly elongated fusion proteins also show significantly reduced migration on SDS-PAGE (Fig. 4.17; Table 4.6). Migration of the moderately elongated TrxHisS - CBF15.0 shows 13.2 % difference from expected value, whereas the globular TrxHisS tag migrates according to molecular mass. An elongated structure causing aberrant gel structure agrees with migration of rod - like proteins like collagen α - chains (Furthmayr and Timpl, 1970). A comparison of full - length and truncated variants of TrxHisS - CBF12.2 revealed a more elongated structure for the full - length protein. This suggests that the C - terminal end of TrxHisS - CBF12.2 partially contributes in to the elongated structure of the complete protein.

To study the effect of purification conditions on secondary structure, the far - UV CD spectra of TrxHisS - CBF12.2 protein purified under native and denaturing conditions were compared (Fig. 4.26). The secondary structure calculated using CDNN CD spectra deconvolution software (Böhm *et al.*,1992) showed that the helical content of protein was 15 % when protein was purified under native conditions, but decreased to 10 % for protein purified under denaturing conditions (Table 4.7). Beta sheet content was found to be higher in protein purified under denaturing conditions (41 %) than in protein purified under native conditions (35 %), while the random coli content was quiet similar 33 % under native conditions and 32 % under denaturing conditions, rest were beta turns (21 %) for both samples purified under native or denaturing conditions. This suggests that presence of urea did not denature the CBF12.2 and considerable amount of secondary structure was retained. The amount of secondary structure in complete and C - terminal truncated protein was studied (Fig. 4.26) and it was observed that the truncated

protein too retained most of the secondary structure (Table 4.7), indicating that CBF12.2 structure was mainly due to AP2 domain and/or the amino terminal region was not affected drastically by presence of the C-terminal sequence. Effect of presence of urea during purification was tested for CBF17.0 (Fig. 4.26) and it was observed that it retained the secondary structure with 26 % helical, 15 % β - sheet and 42 % random coli content and 17 % β - turn (Table 4.7). The temperature dependence of the [Θ] value at 220 nm for TrxHisS - CBF12.2 protein is shown in Fig.4.18. The negative molar ellipticity at 220 nm was minimal (- 3815) at 5 ^{o}C, increased by about 750 (- 3051) as the temperature was increased from 5 ^{o}C to 95 ^{o}C. Similar temperature dependence of ellipticity was observed in spectra for TrxHisS - CBF12.1. These results strongly suggest that the protein is stable to thermal denaturation in the range between 5 ^{o}C and 95 ^{o}C (Fig. 4.27).

4.5. DISCUSSION

An evaluation of *E. coli* host strains and conditions for production of native TrxHisS-tagged Norstar CBFs, established a protocol for production of all CBFs, except CBF17. The critical steps of this protocol included growth at 28 °C, two hour induction with 0.5 mm IPTG, production of soluble extract using BugBuster Master Mix and purification using recharged Ni-MAC columns. TrxHisS - CBF17, which was not recovered in the *E. coli* soluble extract, could be purified from cells under denaturing conditions.

Previous studies of DREB / CBF production in *E. coli* used BL21(DE3) host cells, a growth temperatue of 37 ^{o}C and four to six hours of induction with 1 mM IPTG (Stockinger *et al.*,1997; Skinner *et al.*, 2005; Agarwal *et al.*, 2006). Yield of expressed proteins in total cell extracts were not clearly stated in the published protocols but when tested using pTrxHisS - CBF constructs in this study, recombinant protein yields were very low (data not shown). In the previous reports, pET101/D-TOPO (Invitrogen, Carlsbad, CA, USA) or pGEX vector (GE Healthcare, Piscataway, NJ) vectors were used for CBF expression; however these vectors were not tested in this study. Thus, it cannot be excluded that choice of vector may have some effect on protein expression levels and soluble protein yields.

Table 4.6 Molecular mass and conformation of purified TrxHisS - CBF proteins.

Fusion protein	Molecular mass			Protein conformation					
	Predicted (kDa)	Observed* (kDa)	Difference (%)	Radius (nm)	Polydispersity (%)	S	Smax	Smax/S	Predicted shape
TrxHisS	17	17	-	1.29	12.8	3.133	2.39	0.76	Globular
TrxHisS-17.0	47	63	32.6	5.35	13.5	2.089	4.70	2.25	Highly elongated
TrxHisS-15.0	42	48	13.2	3.89	10.4	2.567	4.36	1.7	Moderately elongated
TrxHisS-12.2	43	55	24.8	5.13	13.6	1.993	4.43	2.2	Highly elongated
TrxHisS-12.2-T7	22	28	26.6	315	16.2	2.468	2.83	1.7	Moderately elongated
TrxHisS-12.1-T2	38	46	18.5	3.58	14.6	2.524	4.04	1.61	Moderately elongated

* Calculated from migration on SDS-PAGE.

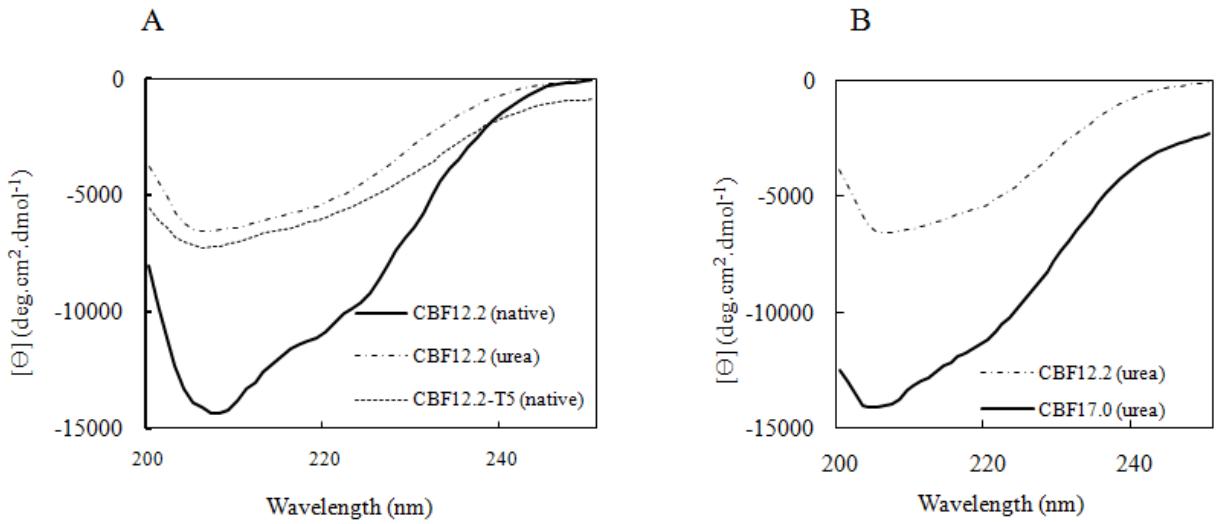


Figure 4.26 Far - UV CD spectra of TrxHisS - CBF12.2, - CBF12.2-T5 and - CBF17.0.
The spectra of native (TrxHisS - CBF12.2 and TrxHisS - CBF12.2-T5) and urea - denatured (TrxHisS - CBF12.2 and TrxHisS - CBF17.0) proteins were generated at 4 ^{o}C.

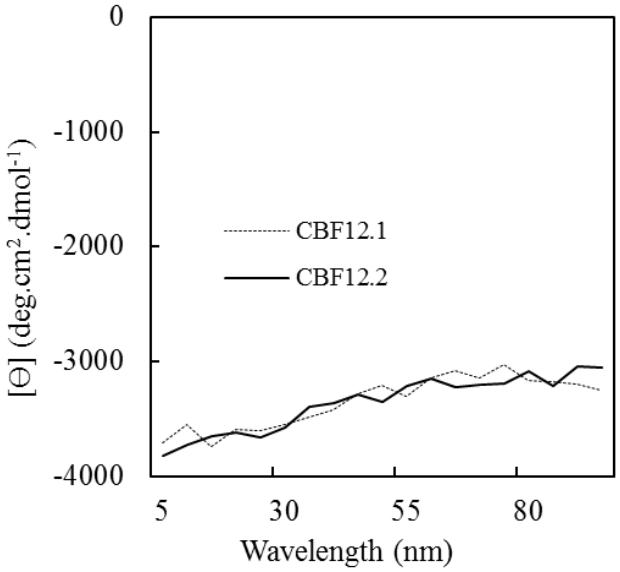


Figure 4.27 Effect of temperature on TrxHisS - CBF12.1 and TrxHisS - CBF12.2 secondary structures.
CD spectra were obtained at constant 220 nm and at 5 ^{o}C temperature intervals ranging from 5 ^{o}C to 95 ^{o}C.

Table 4.7. Secondary structure content of purified TrxHisS - CBF proteins.

Structure predicted *	TrxHisS - CBF12.2		TrxHisS - CBF12.2-T5	TrxHisS - CBF17.0
	Native	Urea - denatured	Native	Urea - denatured
Helix (%)	15	10	14	27
Antiparallel (%)	30	37	31	11
Parallel (%)	05	04	04	04
Beta-Turn (%)	20	21	20	17
Random Coil (%)	33	30	34	42
Total Sum	103	102	103	101

*Data obtained from far - UV spectra

Comparing of TrxHisS - CBF production in BL21(DE3)pLysS host with three other closely related *E. coli* type B host strains (Table 4.1), the least TrxHisS - CBF production was obtained for the BL21(DE3)pLysS strain (Fig. 4.6). The only reported genetic difference between BL21(DE3)LysS and the more successful BRL21(DE3)pLysS strain is a *srl* - *recA* deletion for the latter rendering it *RecA*-. Thus, the supposedly higher plasmid stability in BRL21(DE3)pLysS may have contributed to higher protein synthesis in this strain as compared to BL21(DE3)LysS. However, the two other promising strains, Tuner(DE3)pLysS and Origami B (DE3)pLysS, are both RecA+, but instead have a lacZY deletion not present for BL21(DE3)pLysS or BRL(DE3)pLysS. As the lacZY mutation allows a more uniform IPTG induction, this may have allowed higher TrxHisS - CBF12.2 production in these two strains compared to BL21(DE3)LysS.

During the optimization experiments, it became obvious that a lower growth rate e.g. growth temperature at 28 °C was superior to 37 °C for production of soluble TrxHisS - CBF. Also TrxHisS - CBF solubility was favored by a relatively low IPTG concentration (0.5 mM) and short induction time (two hours). The improved yield at low growth temperature and weak induction of TrxHisS - CBF expression is likely due to reduced protein concentration in *E. coli* cells. High protein levels in *E. coli* cells are known to increase the risk of protein aggregation reducing yield of soluble recombinant proteins (Koptzki *et al.*, 1989; Schirano and Shibata, 1990; Kiefhaber *et al.*, 1991; Sawyer *et al.*, 1994).

For TrxHisS - CBF17, purified protein could only be obtained under denaturing conditions suggested complete sequestration of fusion protein into inclusion bodies (Fig 4.12). A significant amount of inclusion bodies were likely produced by TrxHisS - CBF12.2 as well, since protein yield upon purification under denaturing conditions was three - fold higher as compared to native conditions. Most of the native TrxHisS - CBFs purified showed similar yields as TrxHisS - CBF12.2, thus partial sequestration of overproduced protein within inclusion bodies may have been a general trend.

Since the TrxHisS - CBF17.0 variants, TrxHisS - CBF17.0-T2, -T3 and -T4, were also completely sequestered in inclusion bodies, but not TrxHisS, suggesting that the first 113 amino acids of CBF17.0 as the cause of aggregation. This region includes the complete AP2 domain and 11 amino acids beyond. As only three cysteine residues reside within this region of TrxHisS - CBF17.0-T4 argues against formation of improper cysteine linkages. Misfolding of proteins can

also be caused by hydrophobic regions on protein resulting in "hydrophobic collapse", especially for proteins that are naturally dependent on chaperones for folding (Dill, 1985). As shown in Chapter 3, the C - terminal portion of CBF17.0 is not more hydrophobic compared to CBF15.0 (Fig. 3.7). The distribution of "fast" and "slow" codons along the expressed transcript can also affect protein folding. Generally, a stretch of "slow" codons are present within the first 30 - 50 codons to slow down translation rate and "fast" codons are present in regions that are aggregation - prone (see review by Gatti - Lafranconi *et al.*, 2011). Additional studies are needed to identify residues or regions triggering CBF17.0 aggregation.

Although TrxHisS - CBFs were purified to near - homogenity, small polypeptides were observed in our initial purifications. This occurred even in presence of protease inhibitor cocktail and following the Ni - MAC column purification according to the manufacturer's instructions. Premature variants / forms or degradation products of fusion protein sometimes occur during protein expression in *E. coli* as exemplified by the *Arabidopsis* MYB2, which encodes transcription factor induced by dehydration (Urao *et al.*,1993). The small polypeptides co - purifying with TrxHisS - CBF12.2 were confirmed by mass-spectromerty analysis to be C - terminal truncated products of the full-length fusion protein. As the truncated proteins were not present in proteins accumulated under denaturing conditions, proteolytic cleavage must have occurred post - harvest. Interestingly, this co - purification of small polypeptides was strongly reduced when the Ni - MAC columns were recharged with freshly prepared $NiSO_4$. The cause of protein cleavage has not been identified but may be due to a batch of "bad" Ni - MAC columns.

Among the purified CBFs, CBF2.1, 9.0, 20.0, 21.0 and 22.0 were found to be unstable upon purification. Usually the purified proteins became fully degraded within six hours at 4 ^{o}C and presence of proteinase inhibitor cocktail did not reduce the degradation rate. Surprisingly, the labile CBFs were stable up to five days in *E. coli* soluble extracts stored at 4 ^{o}C or several weeks at - 20 ^{o}C in 10 % glycerol. This suggested component(s) present in *E. coli* may provide some protection against proteolysis of the labile CBFs. A drastic reduction of this protective mechanism during Ni - column chromatography may explain the gradual degradation of TrxHisS - CBF after purification. It could also be that the protease inhibitor cocktail used during cell lysis is not completely effective against all *E. coli* proteases. Proteases escaping inactivation may have some affinity to Ni^{2+} ions and become enriched during purification of fusion proteins, but this

hypothesis would also imply that the CBF isoforms differ in sensitivity for this protease. Lastly, it cannot be excluded that the labile TrxHisS - CBFs display autoproteolytic activity under certain conditions. Self-degradation has been demonstrated for many proteins, for example the zucchini yellow mosaic virus helper component proteinase (ZYMV HC - Pro) is autoproteolytic when expressed as recombinant protein in *E. coli* (Boonrod *et al.*,2011).

AP2 family of transcription factors from non - plant species e.g. those found in protochordates, insects, nematodes, mouse and humans, form either heterodimers or homodimers (reviewed in Eckert *et al.*, 2005). The recombinant TrxHisS - CBFs were found to be monomeric as confirmed by dynamic light scattering (DSL) analysis (Table 4.8). Also, the resultant sedimentation coefficient ratio calculated from the radius obtained by light scattering, clearly indicated that the wheat TrxHisS - CBF proteins studied are elongated, as previously observed for *Arabidopsis* CBF1 (Kanaya *et al.*,1999). Fusion protein elongation is not due to the attached tags, since the TrxHisS protein was predicted to be globular (Table 4.8). As the TrxHisS - CBF12.2 complete protein was found to be highly elongated, but C - terminal truncated version TrxHisS - CBF12.2-T7 was moderately elongated, thus suggested the source of the elongation resides within the C - terminal region of the protein.

In contrast to the five unstable CBFs mentioned above, Norstar CBF12.2 and CBF17.0 produced in this study were found to be extremely resistant to heat and urea treatment (Figs. 4.25, 4.26). This property was further confirmed for TrxHisS - CBF12.2 and TrxHisS - CBF17.0 by calculation of secondary structure using far - UV (200 - 260 nm) CD spectra. The results clearly show that the protein largely retains its secondary structure when heated up to 98 ^{o}C or when purified in presence of urea (Fig. 4.27). A slight shift in the molar ellipticity (- 3815 at 5 ^{o}C, increased by about 750 to - 3051 for TrxHisS - CBF12.2) is noted for the proteins tested, but it does not become zero as expected for a completely denatured protein.

The elongated structure and resistance to conformational change under denaturing conditions as revealed by DLS is a reasonable explanation to slow migration for e.g. TrxHisS - CBF12.2 on SDS - PAGE gels. Similar to TrxHisS - CBF12.2, the thermostable triosephosphate isomerase of *Pyrococcus furiosus* is highly recalcitrant to unfolding and migrates slower than expected on SDS - PAGE gels (Mukherjee *et al.*, 2005). Rod - like proteins, like collagen α - chains and collagen peptides, also show slow gel migration (Furthmayr and Timpl, 1970). As

protein migration on SDS - PAGE gel is due to movement of negatively charged particle, i.e. SDS - coated polypeptide, an elongated protein will have a reduced amount of SDS - coverage, thus slower migration towards positive electrode.

Few plant storage protein and structural protein exhibit such stability demonstrated by some of the purified CBFs in this study. In general, heat stable proteins are mostly found in thermophiles that function effectively at high temperatures or psychrophiles that colonize only in low temperature environments (Feller and Gerday, 2003). However, some heat - stable proteins are also found in mesophiles. One of the few examples of heat - stable proteins in plants are dehydrins (DHN), which are members of late embryogenesis abundant (LEA) D11 family of proteins. Like many CBF genes, the DHN genes are induced by environmental stresses like low temperature and dehydration and the proteins are stable to heat and remain soluble at high temperature (Ceccardi *et al.*, 1994; Houde *et al.*, 1995; Campbell and Close, 1997). Relatively few transcription factors have been identified as heat and urea stable. One of the first reports of heat and urea stable transcription factor was a factor purified from *Drosophila melanogaster* third instar larval nuclei extracts (Jack, 1990). Other example is enolase, a multifunctional protein found in wide range of organisms including bacteria, yeast, plants and humans (reviewed by Pancholi, 2001). Enolase acts as a transcriptional repressor of cold - responsive genes in *Arabidopsis* (Lee *et al.*, 2002). Similar to enolase, the stability of some CBFs may be of high importance for cold acclimation or for the maintenance of frost tolerance during winter in winter wheat and other winter cereals. To understand the molecular mechanism of CBF participation in cold tolerance, 15 recombinant CBFs described in chapter 4 were analyzed for their DNA binding activity to selected motifs in the promoter region of *wcs120*, a cold responsive gene.

CHAPTER 5

ANALYSIS OF CBF INTERACTION WITH TARGET DNA SEQUENCE

5.1 Abstract

Winter wheat low temperature (LT) tolerance and winter survival is associated with a large cluster of C - repeat binding factor (CBF) genes at *Fr-A2* on chromosome 5A. To evaluate the functional properties for 15 CBF genes carried by *Fr-A2* of cold - hardy cv Norstar, the encoded proteins were selected for a functional assessment. The analysis of recombinant CBF produced in *E. coli* revealed *in vitro* binding to dehydration-responsive element / C - repeat element (CRT / DRE) motif (CCGAC) for 10 of the 15 CBFs tested. The CRT interaction was obtained irrespective of protein being purified under native or denaturing conditions (6 M urea). Within hours of exposure to LT (4 °C), native CBFs increased its affinity to CRT motif, which could be related to alterations in their secondary structures. The cold - induced changes in CBF12.1 provided a memory of cold - experience allowing high and immediate affinity to CRT element upon subsequent cold exposure. CBF12.2 produced by cold - sensitive winter wheat cv. Cappelle - Desprez was found to be a non - functional protein caused by a R → Q substitution for a highly conserved residue within the AP2 domain. The DNA binding assays of truncated CBF proteins demonstrated second signature motif (DSAWR) and remaining C - terminal region were dispensable for CBF12.2 and CBF15.0 interaction with CRT. In contrast, CBF12.1 and CBF17.0 required a considerable portion of the C - terminal domain to be active. Analysis of various DNA target sequences showed some CBFs had preference for core GGCCGAC motif, whereas others preferred the TGCCGAC motif. Several of the CBFs encoded from *Fr-A2* were highly stable and showed increased activity under LT and denaturing conditions, which may be important factors for high winter survival in cv Norstar.

5.2 Introduction

The CBF regulon is present in freezing - tolerant cereals such as winter wheat and rye (Choi *et al.*, 2002; Li *et al.*, 2011) and moderately cold tolerant plants like *Arabidopsis* and *Brachypodium* (Thomashow, 2001; Li *et al.*, 2012). It is also found in the chilling - sensitive tomato and rice (Jaglo *et al.*, 2001). Rice and tomato CBF regulon encompasses about 10 genes (Zhang *et al.*, 2004; Hsieh *et al.*, 2002; Dubouzet *et al.*, 2003), whereas *Arabidopsis* has a larger

CBF regulon of more than 100 genes (Maruyama *et al.*, 2004; Vogel *et al.*, 2005) accounting for about 12 % of the *Arabidopsis* COR genes (Fowler and Thomashow, 2002; Vogel *et al.*, 2005). Hundreds of genes are included in the CBF regulon of cold - sensitive *Solanum tuberosum* and a similar number exist for the more cold - hardy *Solanum commersonii* (Carvallo *et al.*, 2011); thus the size of the CBF regulon is not a determinant for freezing tolerance in general. A closer relationship between level of frost resistance and the overall magnitude of expression from the CBF regulon exists for *Arabidopsis* ecotypes (Hannah *et al.*, 2006). The size of the CBF regulon in winter wheat has not been determined; however it is expected to be very large based on number of CBF genes and changes to the transcriptome during cold acclimation (Gulick *et al.*, 2005; Monroy *et al.*, 2007; Ganeshan *et al*., 2011; Laudencia - Chingcuanco *et al.*, 2011).

CBF genes are rapidly induced in *Arabidopsis* upon cold exposure and reach maximal accumulation after three hours, followed by steady decline to minimal levels within 24 h (Zarka *et al.*, 2003). The more cold - hardy winter wheat exhibits a similar rapid and transient expression of CBF genes in response to LT, and display a second up-regulation when the cold period is extended for two to three weeks (Kume *et al*., 2005). Expression levels of CBF genes and downstream CBF regulon show a relatively good correlation to LT tolerance in both winter and spring wheat cultivars (Kobayashi *et al*., 2004). For example, the cold - hardy winter cultivar 'Mironovskaya 808' expresses a 2.5-fold higher level of both CBF and CBF regulon transcripts when compared to the more tender spring cultivar 'Chinese Spring' (Kume *et al.*, 2005). 'Mironovskaya 808' is also able to maintain a high CBF expression during the second up - regulation, whereas Chinese Spring shows a continuous decrease in cold responsive (COR) gene expression at this phase (Kume *et al.*, 2005). Thus, some of the differences in long - term cold - tolerance in wheat can be related to differences in CBF expression levels and profiles during the later stages of cold acclimation.

In addition to cold acclimation, a second phase hardening, also called subzero acclimation occurs at freezing, but non - lethal temperatures (Herman *et al.*, 2006). At this stage, additional freezing tolerance is acquired in wheat and involves a largely different set of factors than during cold acclimation (Herman *et al.*, 2006; Skinner, 2009). Some of the wheat CBF genes are up-regulated in wheat crown tissues during sub - zero acclimation (Skinner, 2009).

Plant CBFs are characterized by a highly conserved central DNA binding AP2/ERF domain (Stockinger *et al*., 1997) made up of a three - stranded β - sheet structure followed by an α - helical domain based on analysis of *Arabidopsis* AP2 domain protein ERF1 (Allen *et al.*, 1998). The AP2 domain is flanked by CBF - specific signature motifs CMIII - 3 (PKK / RPAGRxKFxETRHP) and CMIII - 1 (DSAWR; Jaglo *et al*., 2001). CMIII - 3 is required for transcriptional activity and suggested to determine DNA binding specificity for *Arabidopsis* CBF1 (Canella *et al.*, 2010). Despite similarity to nuclear localisation signals (NLS), CMIII-3 motif is not needed for nuclear import of *Arabidopsis* CBF1; rather, the import signal resides within AP2 domain but precise residues involved have not been identified (Canella *et al.*, 2010). No role for the conserved CMIII - 1 has been postulated to date, whereas regulatory functions have been suggested by the acidic C - terminal regions of CBFs. Multiple hydrophobic regions at the C - terminal end of *Arabidopsis* CBF1 are associated with transcriptional activation (Wang *et al.*, 2005).

CBFs regulate COR genes by binding to their dehydration-responsive / C - repeat element (DRE/CRT) DNA regulatory element (CCGAC; Stockinger *et al.*, 1997; Gilmour *et al.*, 1998; Liu *et al.*, 1998). These elements are two - fold overrepresented in COR promoters as compared to standard promoters (Maruyama *et al.*, 2012). The CBF - mediated regulation of COR genes is ABA - independent and different from ABA-dependent induction where an ABA-responsive element (ABRE; ACGTGT) in the COR promoter is recognized by ABRE - binding bZIP transcription factors (Stockinger *et al.*, 1997; Gilmour *et al.*, 1998; Liu *et al.*, 1998; Haake *et al.*, 2002; Chinnusamy *et al.*, 2003). Besides DRE and ABRE elements, COR promoters often carry the Evening Element (ATATCT) motif, recognized by CIRCADIAN CLOCK - ASSOCIATED (CCA1) and LATE ELONGATED HYPOCOTYL (LHY) proteins (Maruyama *et al.*, 2012).

The consensus sequence determined for *Arabidopsis* CBF AP2-DNA binding targets is T/ACCGACA/T (Yamaguchi - Shinozaki and Shinozaki, 1994; Baker *et al*., 1994), but preference for CCGAC flanking nucleotides differs between CBF variants. For example, *Arabidopsis* DREB1A / CBF3 bind preferably to A/GCCGACNT (Maruyama *et al.*, 2004), but can also bind weakly to CCCGAC, TCCGAC and ATCGAC motifs (Sakuma *et al*., 2002). The barley HvCBF1 and HvCBF2 have TTGCCGACAT (Xue, 2002) and GTCGACAT (Xue, 2003), respectively, as optimal DNA targets. AP2 proteins which are distantly related to CBFs do not

recognize CCGAC motifs, but interact with other GC - rich motifs. For example, a palindromic DNA sequences $GCCN_3GGC$ is common in viral enhancers recognized by AP2 domain proteins (Hilger - Eversheim *et al.*, 2000) and the SV40 enhancer element carries a CCCCAGGC motif that is bound by enhancer - binding protein AP - 2 (Mitchell *et al.*, 1987).

One of the COR genes induced by wheat CBFs is *wcs120*, which carries two copies of CCGAC core motif in its promoter (Ouellet *et al*., 1998). WCS120 is a 50 kDa hydrophilic protein rich in glycine and threonine residues (Houde *et al*., 1992a, b) and accumulates in both cytoplasm and nucleoplasm of cold - acclimated wheat crowns where it is presumed to have a cryoprotective role (Sarhan *et al*., 1997). The protein is unique in remaining soluble upon boiling of plant tissue, which has made it a useful marker for LT tolerance in Gramineae species (Houde *et al.*, 1992a, b, 1995). The *wcs120* promoter fragment with its two CRT elements has also been utilized for functional studies of CBFs (Ouellet *et al*., 1998).

A large cluster of CBF genes underlie the major locus, *Fr-A2*, responsible for cold – hardiness in diploid and hexaploid wheat (Vágújfalvi *et al*., 2005; Båga *et al*., 2007). Many of the CBF genes at the *Fr-A2* are cold - inducible (Badawi *et al*., 2007) but the properties of the individual CBF proteins are largely unknown. In this study, 15 CBFs encoded from the *Fr-A2* of cold hardy cv Norstar were selected for functional assessment using the *wcs120* DNA fragment as probe for DNA binding. A few CBFs produced in less cold - hardy Winter Manitou and Cappelle - Desprez were also included in the study. The CBFs were produced as recombinant proteins in *E. coli* and when analyzed by an *in vitro* DNA binding assay revealed large differences in DNA binding efficiencies. Several CBFs were extremely stable and showed binding properties that were responsive to environmental changes and consistent with efficient control of CBF regulon under LT conditions.

5.3 Material and Methods

5.3.1 Production of recombinant TrxHisS - CBF

The assembly of pTrxHisS - CBF, pTrxHisS and expression vectors producing C-terminal truncated versions of TrxHisS - CBF were described in Chapter 4. Using the same strategy, additional 3'- truncated versions of TrxHisS - CBF and full - length TrxHisS- expression cassettes were assembled for this study. Oligonucleotides for PCR amplification of CBF DNA fragments were synthesized by Sigma Genosys (Sigma - Aldrich, Oakville, ON, Canada) and are

listed (Appendix Table 4.1). Production of recombinant proteins in *E. coli* BRL(DE3)pLysS cells, extraction of soluble proteins and purification was as described in Chapter 4. For experiment described in Fig. 5.11, cell harvest to protein purification was done entirely at room temperature. Fourteen of the CBF fusion proteins were purified under native conditions from soluble *E. coli* BRL(DE3)pLysS cell extracts. TrxHisS - CBF17, which exclusively forms inclusion bodies in *E. coli* (see Chapter 4), was prepared from 6 M urea - soluble *E. coli* extract. Amount of TrxHisS - CBF present in soluble extracts was determined using S - Tag Rapid Assay Kit (EMD Millipore, San Diego, CA, USA) and TrxHisS tag was removed from TrxHisS - CBF12.2 using Tag•off EK Cleavage / Capture Kit (EMD Millipore San Diego, CA, USA) according to manufacturer's instructions.

5.3.2 Analysis of protein sequences

Protein and DNA sequences were analyzed using the DNASTAR Lasergene 7.1 software (DNASTAR Inc. Madison, WI, USA). Search for PEST sequences was done using web - based program 'epestfind' at ExpASy server (http://emboss.bioinformatics.nl/cgi-in/emboss/epestfind), where PEST scores higher than 5.0 indicated possible PEST segment.

5.3.3 Electrophoretic Mobility Shift Assay (EMSA)

Double - stranded DNA fragments used as probes in EMSA are listed in Tables 5.3 and 5.4 and were generated by annealing complementary oligonucleotides synthesized by Sigma Genosys (Sigma-Aldrich, Oakville, ON, Canada). The 21 - bp *wcs120* fragment contains one CRT / DRE motif (CCGAC) and corresponds to 698 - 697 nucleotides of *wcs120* promoter carried by *T. aestivum* cv. Fredrick (accession AF031235; Vazquez - Tello *et al.*, 1998; Ouellet *et al.*, 1998). Probes (5.0 pmol) were end-labelled using [γ-^{32}P] ATP (3,000 Ci / mmol; PerkinElmer, Waltham, MA, USA) and T4 DNA Polynucleotide Kinase (Invitrogen, Carlsbad, CA, USA) and purified using Bio - spin 30 columns (Bio - Rad, Hercules, CA, USA). Quantification of incorporated isotope was done using a Packard 2200CA Liquid Scintillation Analyzer.

EMSA reactions (20 µL) consisted of 10 mM Tris-HCl pH 7.5, 20 mM KCl, 4 % glycerol, 0.2 % Triton - X100, 20 mM DTT, 2.8 µg sheared salmon sperm DNA, and various amounts of protein and end - labelled *wcs120* probe. Purified TrxHisS - CBF fusion protein (0.1 -

3.0 µg) or 6 - 8 µg soluble *E. coli* cell extract containing 0.1 - 3.0 µg TrxHisS - CBF were used in the assays. Amount of [^{32}P] end - labelled probe varied from 3.1 to 8.0 fmol with specific activity ranging from 3,600 - 1,400 cpm / fmol. Competitive binding assays were performed with increasing molar excess of unlabelled probe fragment (300 - 2,000 fmol) included in the reactions. Reactions were performed for 15 min at room temperature or at 4 ^{o}C. Samples were mixed with sample buffer to 1x concentration (0.25 % bromophenol blue, 0.25 % xylene cyanol and 40 % sucrose) and loaded onto pre - run (30 min, 200V) 6 % non - denaturing acrylamide: bis-acrylamide (30 : 0.8) gel. The electrophoresis was done using the vertical electrophoresis system V16 - 2 (Life Technologies Corp., Carlsbad, CA, USA), 1x Tris-glycine buffer (0.25 M Tris - HCl, 1.92 M glycine, pH 8.6). Electrophoresis was done at constant 200 V for 2 h at 4 ^{o}C or 20 ^{o}C. The gels were dried for 40 min at 85 °C using gel dryer Model 583 (Bio - Rad, Hercules, CA, USA), exposed to Kodak BioMax MR X - ray film and developed using AFP imaging system (AFP Imaging Corp., Elmsford, NY, USA).

5.3.4 Circular dichroism spectroscopy (CD)

Far - UV CD spectra (200 - 260 nm) of proteins (0.3 mg / mL in 20 mM sodium phosphate buffer, pH 8.0) were recorded using a PiStar-180 spectrometer (Applied Photophysics, Surrey, UK) calibrated at 290.5 nm with (1S) - (+) -10 - camphorsulfonic acid. The spectra were collected at 4 ^{o}C or 22 ^{o}C using a 1.0 mm pathlength quartz cuvette and entrance/exit slits set at 6 nm. The mean residual ellipticity [Θ], expressed in units of deg cm^{2} dmol^{-1} was calculated using an average amino acid molecular weight of 110. The secondary structures were predicted using CDNN Deconvolution software (V2.1) (Böhm *et al.*, 1992).

Table 5.1 CBF proteins analyzed

Designation	CBF		
	Source	Residues	C-terminal truncation
TrxHisS - CBF2.2	Norstar *Fr-A2*	225	0
TrxHisS - CBF4.0	Norstar *Fr-A2*	222	0
TrxHisS - CBF12.1	Norstar *Fr-A2*	240	0
TrxHisS - CBF12.1-T2	Norstar *Fr-A2*	208	32
TrxHisS - CBF12.1-T3	Norstar *Fr-A2*	178	62
TrxHisS - CBF12.1-T4	Norstar *Fr-A2*	169	71
TrxHisS - CBF12.2	Norstar *Fr-A2*	245	0
TrxHisS - CBF12.2-T2	Norstar *Fr-A2*	212	33
TrxHisS - CBF12.2-T3	Norstar *Fr-A2*	182	63
TrxHisS - CBF12.2-T4	Norstar *Fr-A2*	121	124
TrxHisS - CBF12.2-T5	Norstar *Fr-A2*	109	136
TrxHisS - CBF12.2-T6	Norstar *Fr-A2*	73	172
TrxHisS - CBF12.2-T7	Norstar *Fr-A2*	47	198
CBF12.2	Norstar *Fr-A2*	245	0
TrxHisS - CBF12.2a^{Cap}	Cappelle - Desprez *Fr*	245	0
TrxHisS - CBF12.2b^{Cap}	Cappelle - Desprez *Fr*	245	0
TrxHisS - CBF12.2^{wM}	Winter Manitou *Fr*	245	0
TrxHisS - CBF14.1	Norstar *Fr-A2*	212	0

Table 5.1 cont.

Designation	CBF		
	Source	Residues	C-terminal truncation
TrxHisS - CBF15.0	Norstar *Fr-A2*	241	0
TrxHisS - CBF15.0-T2	Norstar *Fr-A2*	206	35
TrxHisS - CBF15.0-T3	Norstar *Fr-A2*	185	56
TrxHisS - CBF15.0-T4	Norstar *Fr-A2*	107	134
TrxHisS - CBF15.0-T7	Norstar *Fr-A2*	47	194
TrxHisS - CBF17.0	Norstar *Fr-A2*	290	0
TrxHisS - CBF17.0-T2	Norstar *Fr-A2*	219	71
TrxHisS - CBF17.0-T3	Norstar *Fr-A2*	181	109
TrxHisS - CBF17.0-T4	Norstar *Fr-A2*	113	177
TrxHisS - CBF19.1	Norstar *Fr-A2*	234	0
TrxHisS - CBF19.2	Norstar *Fr-A2*	234	0
TrxHisS	Vector encoded	158	0

5.4 Results

5.4.1 Analysis of CBF binding to CRT motif

A total of 15 Norstar CBF proteins encoded from *Fr-A2* were produced in *E. coli* BLR(DE3)pLysS cells as fusions to thioredoxin - histidine - S - protein (TrxHisS) tag and were denoted TrxHisS - CBF (Table 5.1). As shown in Chapter 4, nine of the fusions could be purified in stable form from *E. coli* soluble lysate, whereas TrxHisS - CBF2.1, CBF9.0, CBF20.0, CBF21.0 and CBF22.0 became unstable upon purification. However, these five fusions were relatively stable in the presence of soluble *E. coli* extract. The TrxHisS - CBF17.0 fusion was insoluble in *E. coli* and could only be extracted by 6 M urea (Chapter 4); thus TrxHisS - CBF17.0 prepared under denaturing conditions was used in this study.

The DNA binding ability of the 10 stable fusion proteins was analyzed by EMSA using a 21 - bp [^{32}P] end - labeled *wcs120* fragment containing one CRT motif (CCGAC) as probe (Table 5.3). The initial assays performed with 10 fmol probe (1000 cpm / fmol) and 5.0 µg soluble *E. coli* soluble extract containing 100 ng TrxHisS - CBF12.2 revealed one major and two minor rapid migrating protein / DNA complexes (Fig. 5.1, lane 1). These interactions were not produced by 100 ng (2.4 pmol) TrxHisS tag or *E. coli* cell extract (Fig. 5.1, lanes 4 and 5) confirming the three complexes were specific for CBF12.2. Two of the complexes could be distinguished when reactions were performed with 50 ng (2.0 pmol) TrxHisS-free CBF12.2 (Fig. 5.1 lane 3), whereas EMSA performed with 25 ng TrxHisS - CBF12.2 purified under denaturing conditions produced only one complex (Fig. 5.1, lane 2). The complexes formed with 2.4 pmol purified TrxHisS - CBF12.2 was slightly stronger than complex formed with 2.0 pmol Tag - free CBF12.2 (Fig. 5.1, lanes 1 and 3), indicating no significant effect from TrxHisS tag on TrxHisS - CBF12.2 binding to DNA target. Thus, TrxHisS - CBF fusions present in *E. coli* soluble extract were used in the some of the following EMSA.

5.4.2 Affinity to CRT motif varies for recombinant CBF proteins

To evaluate if the different Norstar CBF genes carried by *Fr-A2* encode any functional difference with regard to *in vitro* DNA binding, the 10 stable TrxHisS - CBF fusions identified in Chapter 4 were analyzed by EMSA. The protein binding analysis of the TrxHisS - CBF fusions revealed considerable differences in profiles and intensities for complexes formed with 2.5 fmol [^{32}P] end - labeled *wcs120* probe (Fig. 5.2). TrxHisS - CBF19.1 and TrxHisS - CBF19.2 formed

two weak complexes of equal intensities and TrxHisS - CBF15.0 showed a major and a minor gel - retarded band. Only one complex could be distinguished for the remaining TrxHisS - CBF fusions including TrxHisS - CBF12.2, which showed the highest affinity for probe. Surprisingly, a strong complex was also obtained for the urea - purified TrxHisS - CBF17.0, suggesting the denaturing conditions (6 M urea) did not abolish protein function.

Five of the tested TrxHisS - CBF fusions (CBF2.1, CBF9.0, CBF20.0, CBF21.0 and CBF22.0) were unable to form complexes with probe, even when 3.0 µg fusion protein present in *E. coli* soluble extract was analyzed by EMSA (data not shown). The CBF's signature sequences and AP2 domain for the non - binding CBFs and the functional CBFs were aligned to find reasons for differences in DNA binding ability (Fig. 5.3). All the critical residues for AP2 function proposed from three - dimensional and mutational studies of *Arabidopsis* ERF1 and CBF1 (Allen *et al.,* 1998; Canella *et al.,* 2010) were present. Thus, the amino acid sequence of the DNA binding domain did not reveal any obvious reason for inactivity for the five CBFs found non - functional by EMSA.

Competition binding experiments were done by including increasing molar excess of unlabeled *wcs120* fragment in the reactions and analyzing the effect on interactions between each TrxHisS - CBF fusion and *wcs120* probe. In these assays of the 10 stable TrxHisS - CBF fusions, the intensity of gel - retarded bands weakened as the amount of competitor in reactions increased, which confirmed all 10 TrxHisS - CBF /DNA interactions were specific (Fig. 5.4). The strongest complex was demonstrated for TrxHisS - CBF12.2, for which complex with 100 ng protein in reaction was eliminated when 800 - 1,000 - fold molar excess of competitor fragment was included (Fig. 5.3). For the weakest binder, TrxHisS - CBF14.1, only a 100-fold excess competitor was required to eliminate complex formation with 3.0 µg protein in the reaction (Fig. 5.4). Based on the amount TrxHisS - CBF protein present in the assays and molar excess competitor fragment estimated to eliminate 50 % of complex formation, the binding strength of the different fusions was ranked as follows: TrxHisS - CBF12.2 > TrxHisS - CBF15.0 > TrxHisS - CBF17.0 > TrxHisS - CBF2.2, CBF4.0, CBF12.1, CBF14.2, CBF19.1, CBF19.2 > TrxHisS - CBF14.2.

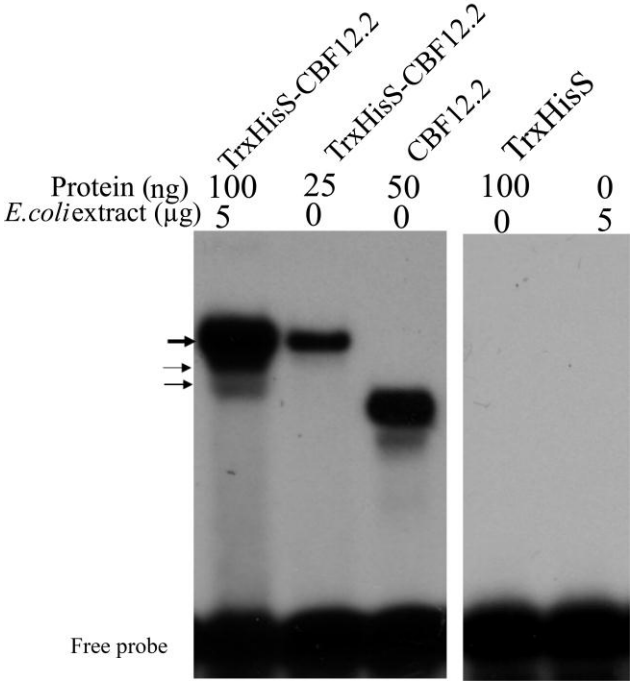


Figure 5.1 Effect of *E. coli* proteins and TrxHisS - tag on TrxHisS - CBF12.2 binding to *wcs120* probe. Autoradiograph of EMSA performed at 4 °C with 10 fmol [^{32}P] end - labeled *wcs120* fragment and proteins indicated. All proteins were extracted at 4 °C. Migrations of distinguished protein / DNA complexes are indicated by arrows.

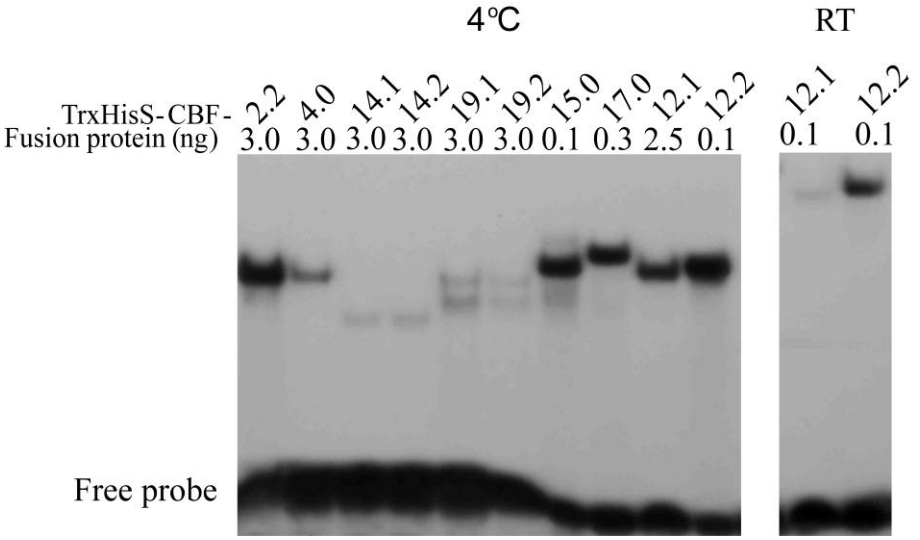
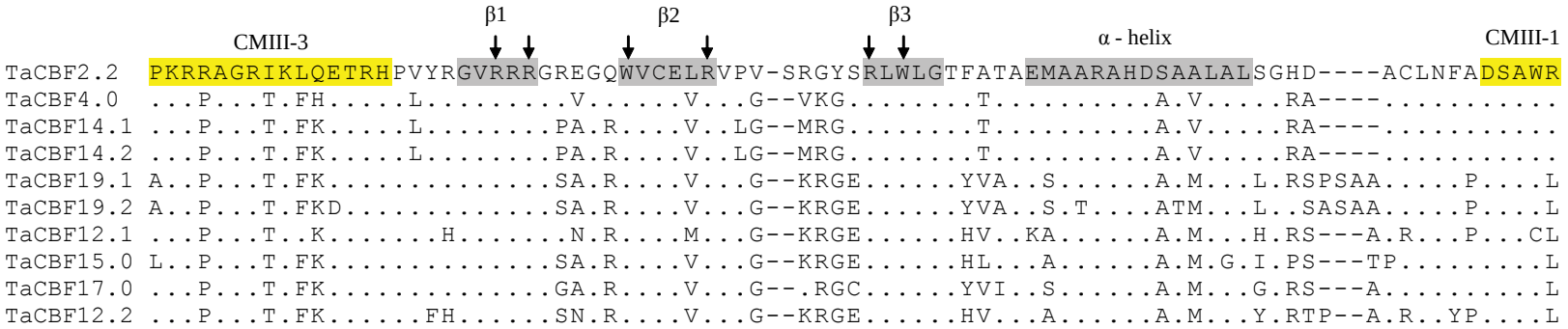


Figure 5.2 Binding of purified TrxHisS - CBF fusions to *wcs120* CRT motif. Autoradiograph of EMSA performed at 4 ^{o}C and RT with 2.5 fmol [^{32}P] end - labeled *wcs120* fragment and various TrxHisS - CBF proteins as indicated.

A



B

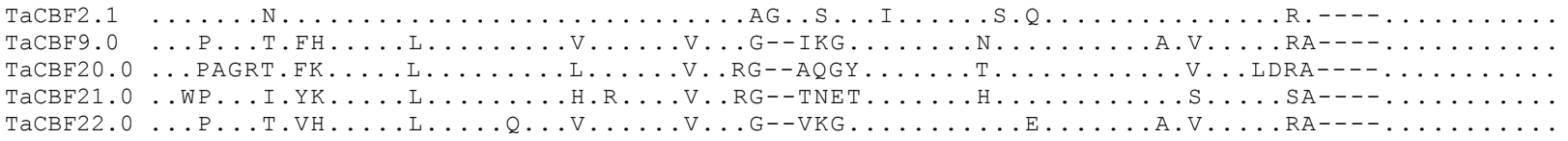


Figure 5.3 Amino acid sequences of AP2 domain and flanking signature motifs of recombinant CBF proteins. Peptide sequences were from CBFs forming complex with CRT motif in EMSA (A) and CBFs with no interaction with CRT motif in EMSA (B). Identical residues (.) and gaps (-) are indicated.

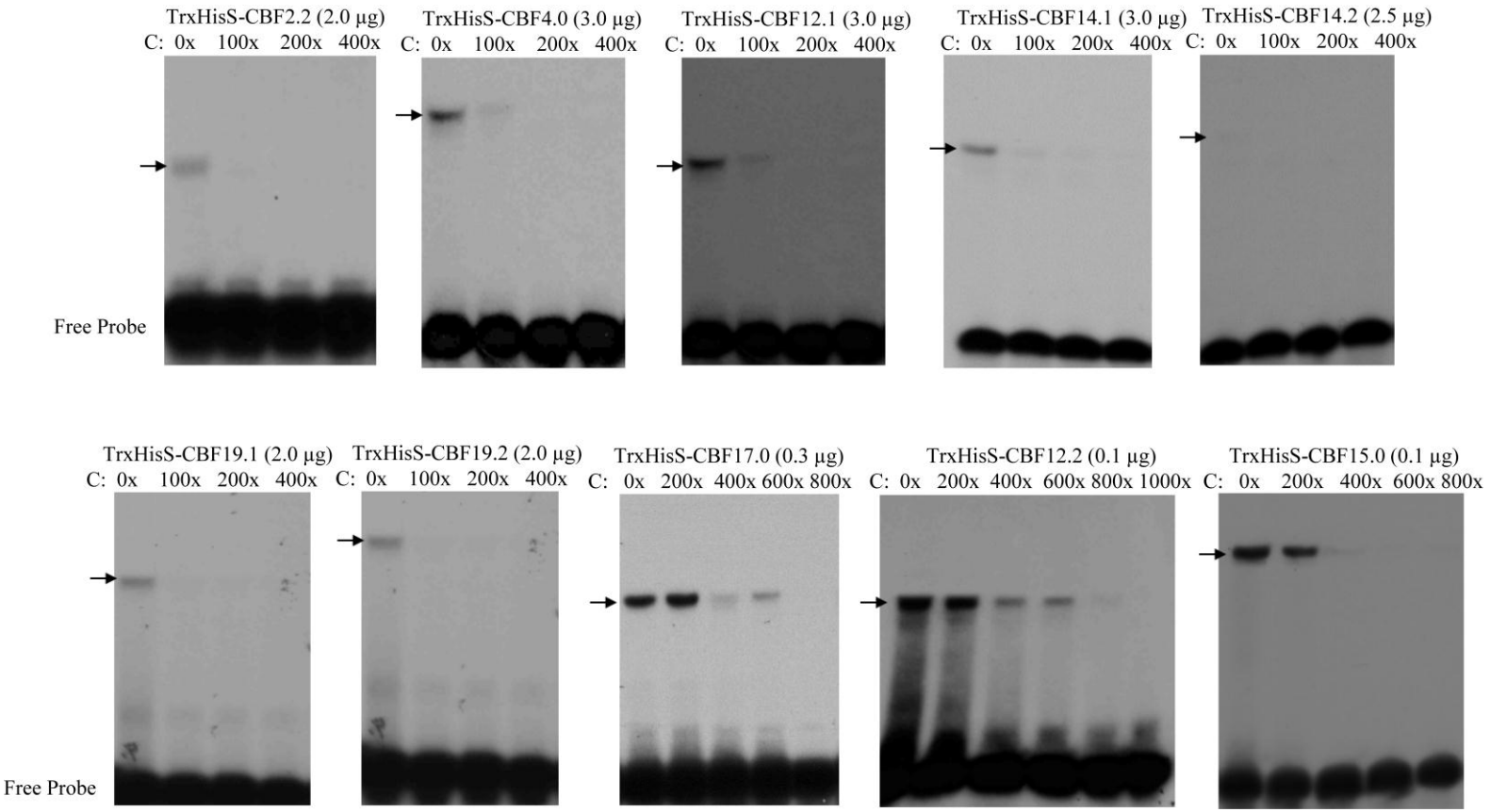


Figure 5.4 Competition assay involving TrxHisS - CBF and CRT motif. Autoradiogram of EMSA performed at 4 °C with various TrxHisS - CBF fusions and 2.5 - 5.0 fmol [^{32}P] end - labeled *wcs120* fragment. Amount protein and of molar excess of competitor (C) included in reactions analysed are indicated. Migration of protein - DNA complexes is indicated by arrows.

The difference in binding strength between the strong TrxHisS - CBF12.2 and the weak TrxHisS - CBF14.2 was estimated to be > 200 - fold.

5.4.3 Allelic variance for *CBF2* and *CBF12* affecting functional properties

These two CBF2 variants, TrxHisS - CBF2.1 and TrxHisS - CBF2.2, belong to the same sub - group (IIId and IVa, respectively) as demonstrated by cluster analysis in Chapter 3, but differ in stability (Chapter 4) and binding to probe *in vitro* (Fig. 5.2). Alignment of the CBF2.1 and CBF2.2 amino acid sequences revealed 81 % identity and 85 % similarity (Fig. 5.5). Within the AP2 DNA binding domain, an insertion of a glycine residue (G70) between second and third β - sheet is observed for CBF2.1 along with several conserved and semi - conserved substitutions.

The amino - terminal signature motif (CMIII - 3) flanking the AP2 - DNA binding domain of CBF2.1 has an asparagine residue (N40) in place of an isoleucine residue (I40) in CBF2.2. Main sequence differences were located in the C - terminal domain, where CBF2.1 carried a six amino acid insertion (PADEST) and a two - amino acid (PT) deletion when aligned to CBF2.2. The PADEST insertion carries the four PEST residues associated with protein instability (Rechsteiner and Rogers, 1996) and / or phosphorylation (Hunter, 2007). CBF2.1 was searched for PEST segments, characterized by Pro, Glu(Asp), Ser, Thr residues flanked by Arg or Lys residues, by epestfind algorithm. A possible PEST motif was found for the 150 - 179 sequence of CBF2.1, which includes the PADEST peptide (PESTfind score of 10.69). Also CBF20.0 showed a potential PEST sequences at the C - terminal end with a PESTfind score of 7.98. Whether any of the potential PEST sequences cause protein instability or protein inability to bind the CRT motif needs to be investigated further.

Similar to the CBF2 isoforms, the two Norstar CBF12 variants also showed differences in DNA binding, where CBF12.2 was estimated to have > 100 - fold higher affinity for probe than CBF12.1 (Fig. 5.2; 5.4). Highest variation was seen in the carboxyl - terminal ends, where multiple non - conserved replacements and deletions of one and two residues were seen for CBF12.1 (Fig. 5.6). Within the AP2 - DNA binding domain, substitution of serine (S57) in CBF12.2 with arginine (R57) in CBF12.1 and insertion of proline between α - helix and second signature motif CMIII - 1 in CBF12.2 are notable differences, whereas other substitutions were conserved. The amino - terminal signature motif (CMIII - 3) sequence, which influences DNA binding strength of *Arabidopsis* CBF1 (Canella *et al.*, 2010), carried a leucine (L40) residue in

CBF12.1 in place of the more conserved phenylalanine (F40) in CBF12.2. The carboxyl - terminal signature motif (CMIII - 1) sequence showed presence of cysteine in CBF12.1 in place of the more conserved tryptophan in CBF12.2.

A comparison of the DNA binding abilities for CBF12.2 variants encoded by Norstar to those produced by winter Manitou (CBF12.2^{wM}) and Cappelle - Desprez (CBF12.2a^{Cap,} CBF12.2b^{Cap}) revealed differences in CRT binding (Fig. 5.7). TrxHisS - CBF12.2, TrxHisS - CBF12.2^{wM}, and TrxHisS - CBF12.2a^{Cap} showed strong binding to probe, but Cappelle - Desprez variant TrxHisS - CBF12.2b^{Cap} did not display any binding at all. An amino acid sequence alignment of the four CBF12.2 sequences revealed differences at only two residues (Fig. 5.7). A glutamine residue (Q60) occurs in non - functional CBF12.2b^{Cap} in place of arginine (R60) in functional CBF12.2 variants of Norstar, winter Manitou and Cappelle - Desprez. Second difference to Norstar CBF12.2 is histidine residue (H236) in non - functional CBF12.2b^{Cap}, but this residue is also found in functional CBF12.2^{wM} and CBF12a^{Cap}. Thus, it was concluded that the inactivity of CBF12.2a^{Cap} is due to a R → Q amino acid substitution within the AP2 - DNA binding domain. The presence of the inactive CBF12.2 allele may contribute towards low freezing tolerance in Cappelle - Desprez.

5.4.4 Effect of C - terminal region on CBF interaction with CRT element

To study the effect of the C - terminal domain on CBF DNA binding ability, various truncated versions of the TrxHisS - CBF12.1, CBF12.2, CBF15.0 and CBF17.0 were analyzed for binding to 2.5 fmol [^{32}P] end - labeled *wcs120* fragment. EMSA performed with TrxHisS - CBF12.2 and its truncated variants revealed no significant effect on DNA binding capacity with removal of 32, 63 or 124 residues from the C - terminal end (TrxHisS - CBF12.2-T2, -T3 and – T4; Fig. 5.8). Elimination of additional 12 residues including the second signature motif resulted in approximately two - fold reduction in affinity for probe (TrxHisS - CBF12.2-T5). As expected, removal of half or the entire AP2 domain resulted in non - functional proteins (TrxHisS - CBF12.2-T6, -T7).

For TrxHisS - CBF15.0, protein binding was not influenced by the last 35 amino acids (TrxHisS - CBF15.0-T2), but further reductions of 21 residues resulted in an approximate two -

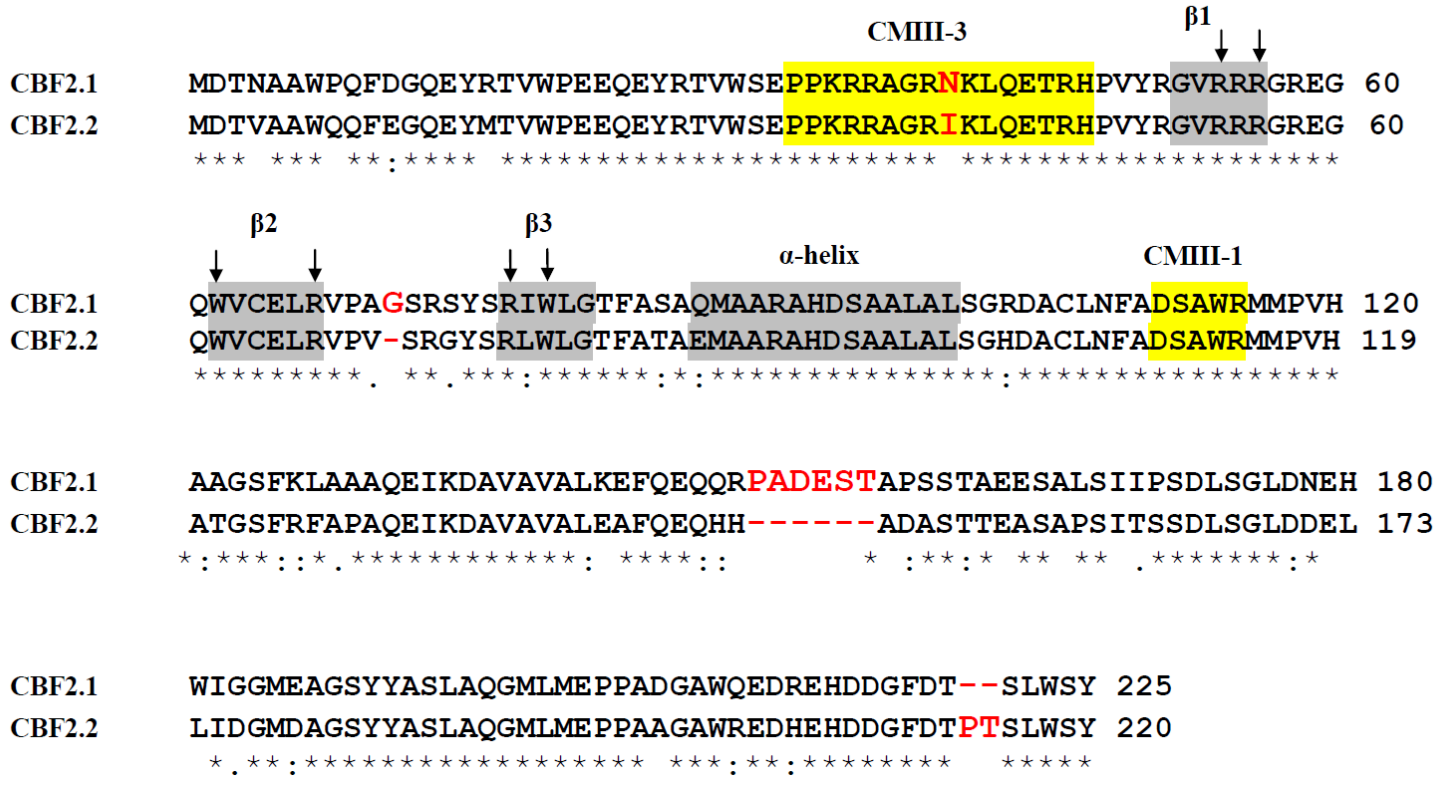


Figure 5.5 Alignment of Norstar CBF2.1 and CBF2.2 sequences. Identical residues (*), conserved substitutions (:), and semi - conserved substitutions (.) are indicated. Location of signature motifs (CMIII - 3 and CMIII - 1) flanking AP2 DNA binding domain are shown by yellow areas. Amino acids forming the three β - sheets and single α - helix (grey areas) of AP2 domain are shadowed. The six conserved residues in contact with target sequence (Allen *et al.*, 1998) are indicated by vertical arrows. Notable sequence differences between the two isoforms are indicated by red residues.

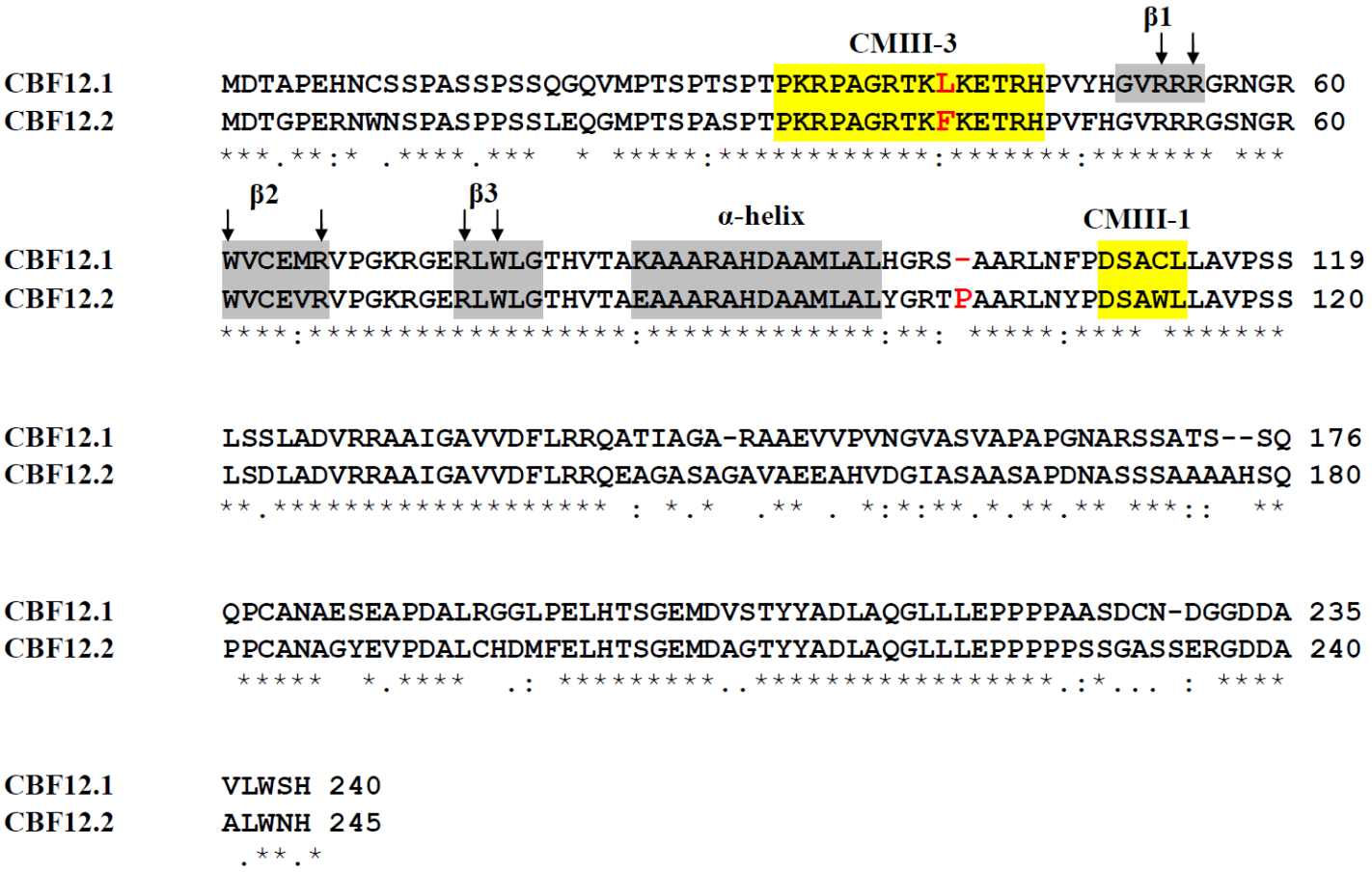


Figure 5.6 Alignment of Norstar CBF12.1 and CBF12.2 sequences. Identical residues (*), conserved substitutions (:), and semi - conserved substitutions (.) are indicated. Location of signature motifs (CMIII - 3 and CMIII - 1) flanking AP2 DNA binding domain are shown by yellow areas. Amino acids forming the three β - sheets and single α - helix (grey areas) within AP2 domain are shadowed. The six conserved residues in contact with target sequence (Allen *et al.*, 1998) are indicated by vertical arrows. Notable sequence differences within AP2 domain are indicated by red residues.

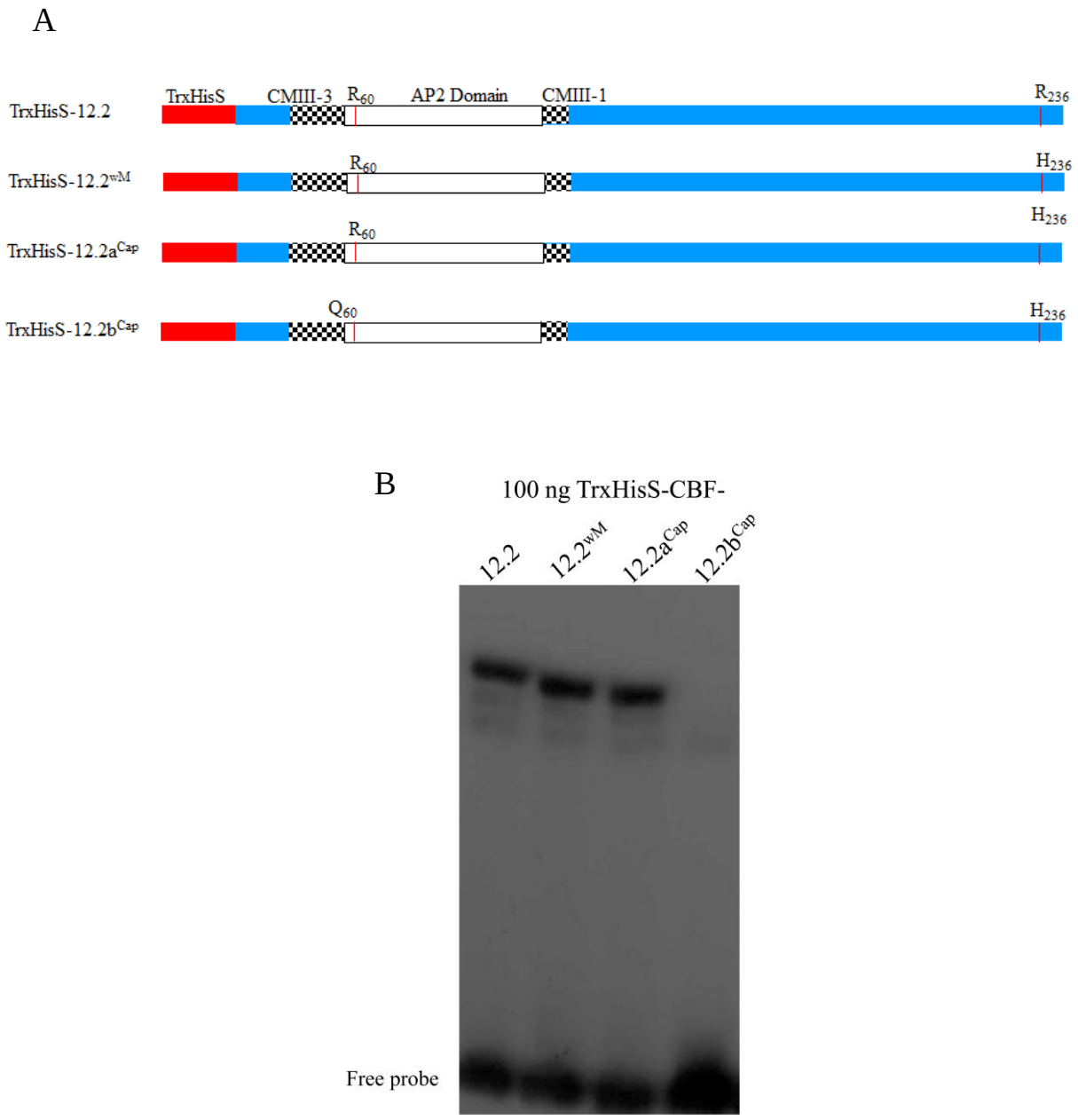


Figure 5.7 Analysis of CRT binding by TrxHisS - CBF12.2 isoforms produced in winter wheat genotypes. (A) Schematic illustration of CBF12.2 variants produced in Norstar, winter Manitou and Cappelle - Desprez. Location of the two variant residues are indicated. (B) Autoradiograph of EMSA performed at 4 ^{o}C with 100 ng purified TrxHisS - CBF12.2 and 2.5 fmol [^{32}P] end - labeled *wcs120* fragment. The ThxHisS-CBF fusions analyzed were CBF12.2 of cold - hardy Norstar (lane1), CBF12.2^{wM} of moderately cold-hardy winter Manitou (lane 2), CBF12.2a^{Cap} and CBF12.2b^{Cap} from moderately cold - hardy Cappelle - Desprez (lanes 3 and 4).

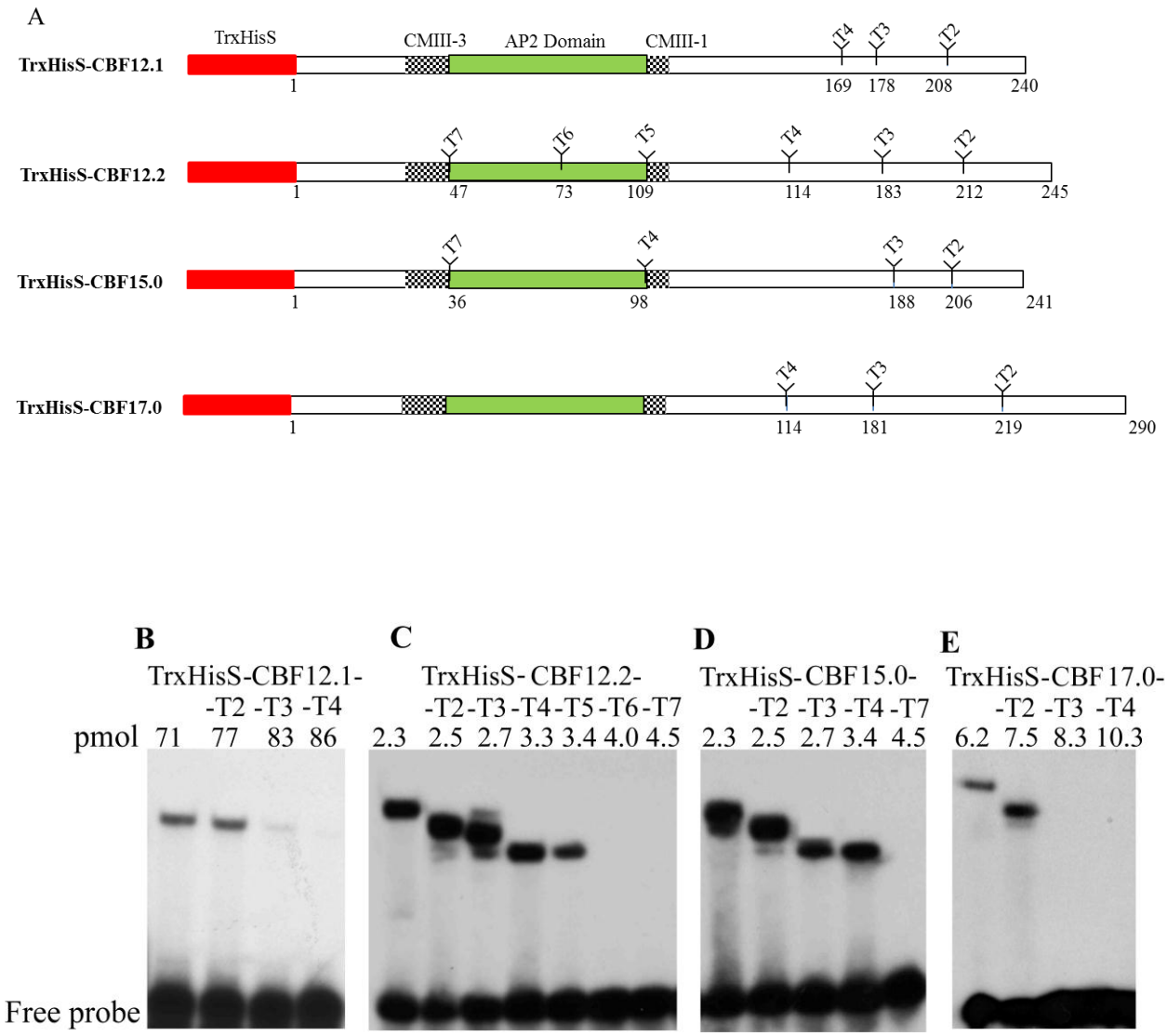


Figure 5.8 Effect of CBF C - terminal region on binding to CRT motif. (A) Schematic representation of full - length TrxHisS - CBF12.1, TrxHisS - CBF12.2, TrxHisS - CBF15.0 and TrxHisS - CBF17.0. For each full - length construct, the end - points of C - terminal truncated TrxHisS - CBF variants are indicated. (B) Autoradiogram of EMSA performed with 2.5 - 3.0 fmol [^{32}P] end - labelled *wcs120* probe and TrxHisS - CBF proteins indicated.

fold reduced level of protein / DNA complexes (TrxHisS - CBF15.0-T3) suggesting the 343 - 364 region of CBF15.0 had a positive role in binding to CRT. Like TrxHisS - CBF12.2, the protein retained about half of its activity with the C - terminal and second signature motif removed (TrxHisS - CBF15.0-T4). For TrxHisS - CBF12.1 and CBF17.0, the C - terminal domain appeared to have a more important role for DNA binding than for TrxHisS - CBF15 and - CBF12.2. Although the last 32 amino acids of CBF12.1 did not affect DNA binding (TrxHisS - CBF12.1-T2), a further truncation of 30 amino acids (residues 337 - 366) reduced DNA binding about five-fold (TrxHisS - CBF12.1-T3). When 71 amino acids of CBF12.1 end were missing, but second signature motif maintained, the truncated protein demonstrated no interaction with DNA (TrxHisS - CBF12.1-T4).

Thus, the 169 - 178 sequences had a positive effect on CBF12.1 activity. TrxHisS - CBF17.0-T2 with 71 residues removed from C - terminal end showed an increase in DNA binding as compared to the full - length construct, suggesting this portion of the protein had a negative effect on DNA binding. Further removal of 38 or 106 amino acids resulted in non - functional proteins (TrxHisS - CBF17.0-T3, -T4). Thus, the C - terminal region of CBF17.0 appeared to contain domains with both positive and negative effects on protein function.

5.4.5 Effect of presence urea on TrxHisS - CBF - DNA interaction

The structural analysis performed on TrxHisS - CBF17.0 in Chapter 4 showed the secondary structure of the protein was very resistant to denaturation by 6 M urea, and purification of TrxHisS - CBF17.0 using urea did not render the protein inactive (Fig. 5.2). To explore if this resilience displayed by CBF17.0 was shared by other Norstar CBFs, the DNA binding affinities for four TrxHisS - CBF fusions prepared under native and denaturing conditions were compared by EMSA. Unexpectedly, the analysis demonstrated an about five - fold increase in DNA binding for urea - extracted TrxHisS - CBF2.2, CBF12.1 and CBF15.0 as compared to corresponding protein extracted in native form (Fig. 5.9). In this assay, CBF2.2 and CBF12.1 with relatively weak affinity to probe in native form became highly active upon urea treatment. However, TrxHisS - CBF12.2 showed strong activity irrespective of purification method.

The effect of urea present during the EMSA was tested for the relatively weak binder TrxHisS - CBF12.1. EMSA performed in the presence of increasing concentrations of urea (0 to

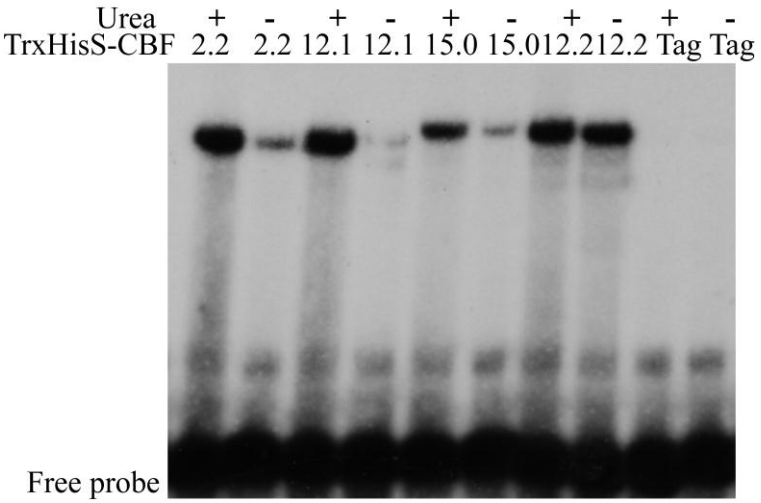


Figure 5.9 Effect of TrxHisS - CBF purification methods on DNA binding capacity. Autoradiograph of EMSA perfomed with 3.0 fmol [^{32}P] end - labeled *wcs120* probe and 100 ng TrxHisS - CBF fusion or TrxHisS (tag) purified under denaturing (6 M urea) or native conditions.

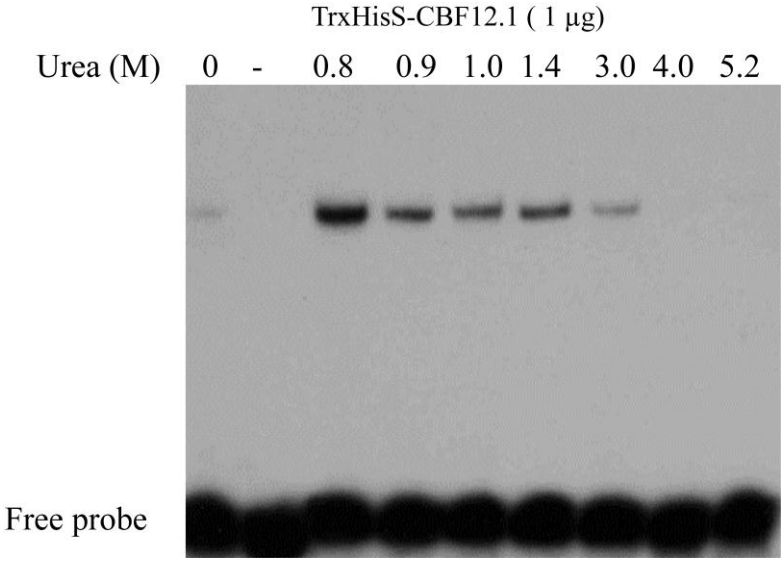


Figure 5.10 Effect of urea on TrxHisS - CBF12.1 DNA binding to CRT motif. Autoradiograph of EMSA performed with 2.5 fmol [^{32}P] end - labelled *wcs120*, 1.0 µg native TrxHisS - CBF12.1 and increasing amount of urea (0 - 5.2 M) as indicated. No protein sample, only probe (-).

5.2 M) revealed higher binding to *wcs120* probe in the presence of 0.8 M urea as compared to absence of urea (Fig. 5.10). Subsequent higher concentrations of urea caused a reduction in formation of protein / DNA complexes, which became non - detectable at 4.0 M urea. To explore if denaturing conditions could transform the five unstable and non - binding CBFs into functional proteins, EMSA was performed with urea - purified TrxHisS - CBF2.1, CBF9.0, CBF20.0, CBF21.0, CBF22.0, respectively. Despite using high concentrations of fusion protein (5 µg), binding to wild type *wcs120* fragment could not be demonstrated for any of the tested fusion proteins (data not shown).

5.4.6 Effect of temperature on CBF DNA binding affinity

The effect of exposure to freezing, cold, warm and hot temperatures on TrxHisS - CBF12.1 and TrxHisS - CBF12.2 functional properties at 4 °C was tested in the following experiment. Samples of purified TrxHisS - CBF12.1 and TrxHisS - CBF12.2 incubated for 15 min at various temperatures (-80, -20, 3, 6, 9, 12, 15, 22 and 100 ^{o}C) revealed no differences in affinity to *wcs120* probe regardless of protein exposure prior to assay (Fig. 5.11). Also samples boiled for 5 min followed by 15 min incubation on ice and 15 min at - 80 °C before analysis did not show any effect on functional properties for both CBF12 variants. In addition to denaturing conditions, the CBF12.1 and CBF12.2 were also very resilient to extreme temperatures and drastic temperature fluctuations.

The barley HvCBF2 binds *in vitro* to the CRT element in temperature - dependent manner where higher binding is obtained at 0 °C as compared to 25 °C (Xue, 2003). To test if interactions between Norstar CBFs and CRT element vary with temperature, the 10 TrxHisS - CBF fusions purified at room temperature or 4 °C were tested for binding to CRT element. In these experiments, the temperature was kept constant at either 4 °C or room temperature during the binding reactions and gel separation. EMSA performed with purified TrxHisS - CBF protein and 2.5 fmol [^{32}P] end - labeled *wcs120* fragment revealed a reduced affinity for CRT motif at RT as compared to cold - exposed samples analyzed at 4 ^{o}C (Fig. 5.12). The largest differences in binding strength were noted for TrxHisS - CBF12.1, and CBF17.0, for which CRT binding was very low at 20 °C and > 5 - fold stronger at the low temperature. TrxHisS - CBF14.1, CBF14.2, and CBF2.2 appeared to be less controlled by temperature as CRT binding was about two - fold lower for the 20 °C samples as compared to those assayed at 4 °C.

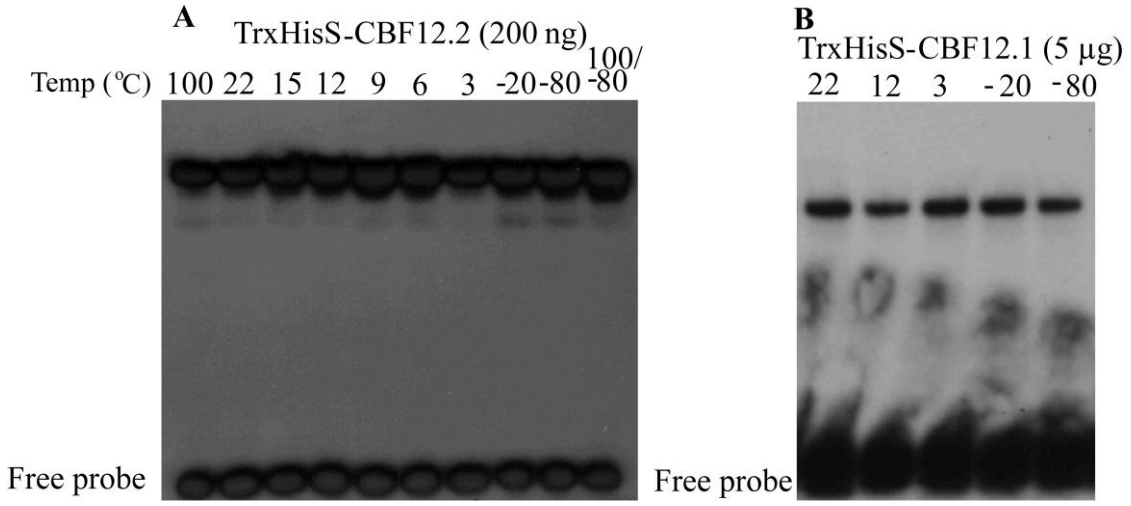


Figure 5.11 Effect of temperature exposure on TrxHisS - CBF12 CRT binding. Autoradiograph of EMSA with 3.0 fmol [^{32}P] end - labeled *wcs120* fragment and 200 ng TrxHisS - CBF12.2 (A) and 5.0 µg TrxHisS - CBF12.1 (B) exposed to different temperatures (100 to - 80 ^{o}C, 15 min). Sample 100 / - 80 was incubated at 100 °C 5 min, ice 15 min, and - 80 °C for 15 min before assayed.

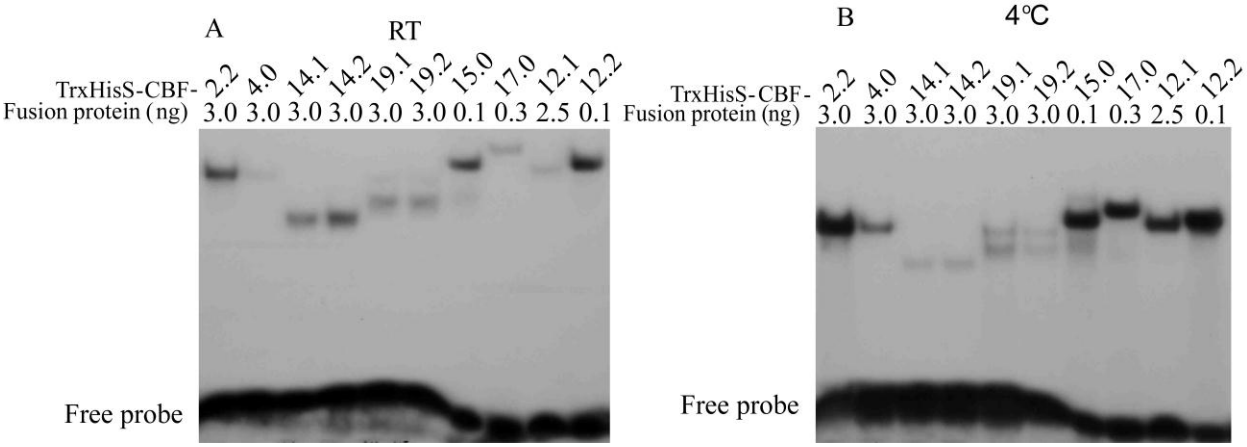


Figure 5.12 Analysis of Norstar CBF proteins binding to CRT motif at different temperatures. Autoradiograph of EMSA performed with 2.5 fmol [^{32}P] end - labeled *wcs120* fragment and various TrxHisS - CBF fusions at RT (A) and 4 ^{o}C (B). Proteins analyzed at RT were not exposed to LT prior to analysis.

5.4.7 Effect of temperature cycling on CBF function

To further study the temperature dependence on CRT binding, a set of experiments were performed as outlined in Fig. 5.13. In these tests, the two TrxHisS - CBF12 variants were assayed upon passage through a cycle of warm-cool-warm temperatures to mimic alterations in temperature occurring naturally during cold acclimation in the field. TrxHisS - CBF12.1 and TrxHisS - CBF12.2 were selected for analysis as these two closely related proteins show different affinity for CRT (Fig. 5.2), are differentially affected by urea (Fig. 5.9A) and temperature (Fig. 5.12). Initial experiment was done with TrxHisS - CBF12.1 and TrxHisS - CBF12.2 purified in native form at room temperature, and thus lacking LT experience. The proteins were incubated at 4 °C for various times (0.5, 1, 2, 3, 4, and 16 h) as illustrated in Fig. 5.13. Subsequently, the 0.5 to 4 h LT - treated samples were analyzed by 15 min gel binding assay followed by gel separation of products performed entirely at 4 °C. Under these conditions, EMSA performed with 3.0 fmol end - labeled *wcs120* fragment and 3.0 µg TrxHisS - CBF12.1 or 100 ng TrxHisS - CBF12.2 showed a constant increase in binding affinities to probe with time spent at 4 ^{o}C (Figs. 5.14A and 5.15A), suggesting cold slowly alters TrxHisS - CBF protein conformation to facilitate stronger binding to CRT motif.

Upon maximizing protein activity by 4 ^{o}C incubation for 16 h, the proteins were exposed to warm temperature (RT) for various time periods (2, 3, and 4 h) and thereafter analyzed by EMSA conducted entirely at 4 °C (Figs. 5.14B and 5.15B) or RT (Figs. 5.14C and 5.15C). From this analysis it was observed that TrxHisS - CBF12.2 maintained strong interaction with probe at both 4 °C and RT (Figs. 5.15B and 5.15C), and thus did not lose its cold - induced strong binding state after being exposed to RT for up to 4 h. In contrast, TrxHisS - CBF12.1 rapidly lost its cold-induced strong binding state when transitioned from 4 °C to 20 °C (Figs. 5.14B and 5.14C). However, the CBF12.1 fusion did not revert back to its original non - induced state, but rather could bind immediately and strongly to CRT element when exposed to a second round of cold provided during the 4 °C EMSA assay (compare Fig. 5.14B and 5.14C). These data suggested the first - time experience of cold provided CBF12.1 with a built - in memory of cold, which improved protein response in the following cold exposure.

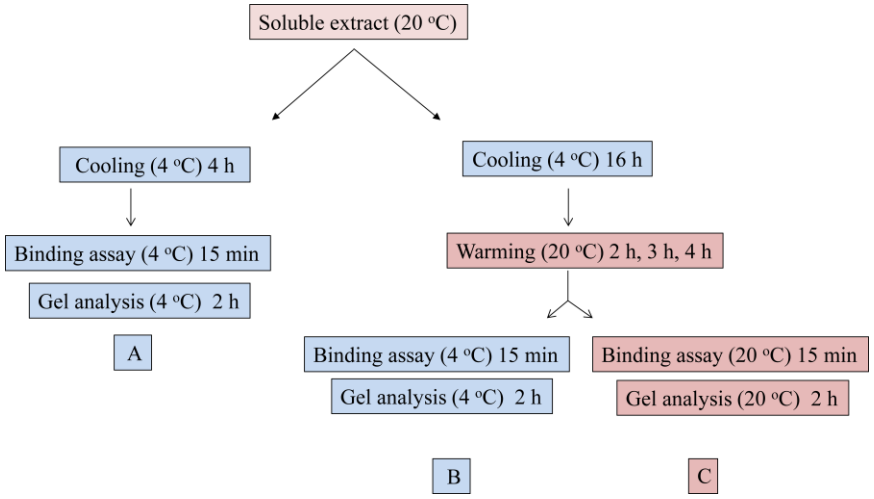


Figure 5.13 Schematic outline of experiments presented in Fig. 5.15

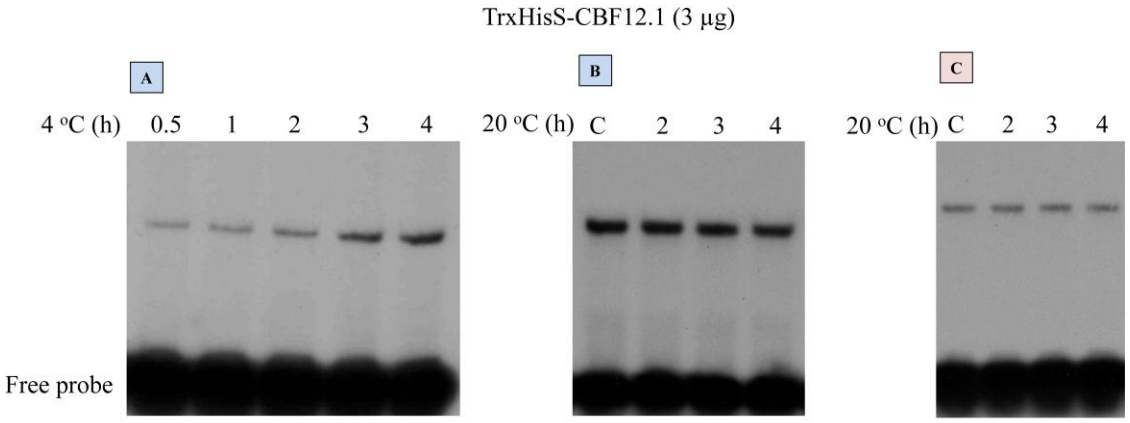


Figure 5.14 Analysis of effect of warm - coo l - warm incubation temperature on TrxHisS - CBF12.1 DNA binding ability. Autoradiograph of EMSA with 3.0 fmol [^{32}P] end - labeled *wcs120* probe and 5.0 µg *E. coli* soluble extract containing 3.0 µg TrxHisS - CBF12.1.

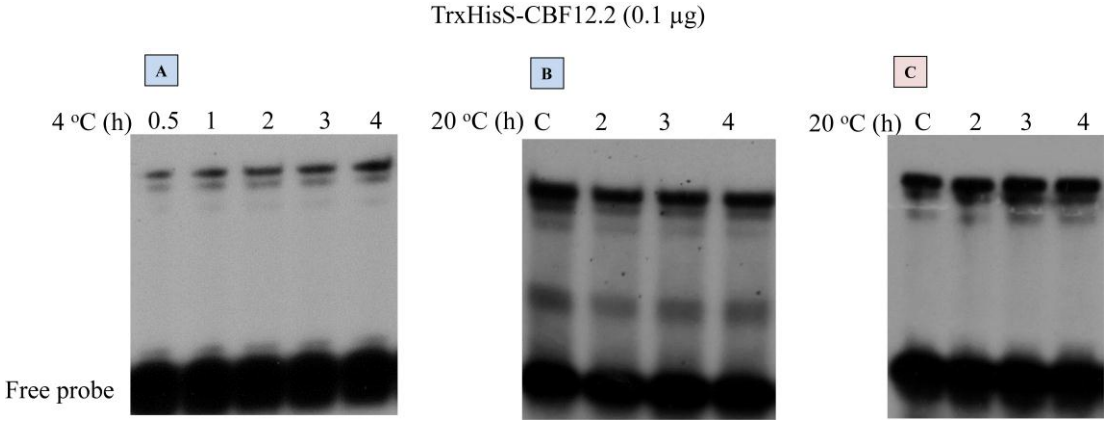


Figure 5.15 Analysis of effect of warm - cool - warm incubation temperature on TrxHisS - CBF12.2 DNA binding ability. Autoradiograph of EMSA with 3.0 fmol [^{32}P] end - labeled *wcs120* probe and 5.0 µg *E. coli* soluble extract containing 0.1 µg TrxHisS - CBF12.

5.4.8 Effect of temperature on secondary structure of CBF12.1 and CBF12.2

The increased binding to CRT element demonstrated for cold - acclimated TrxHisS - CBF12.1 and TrxHisS - CBF12.2 were likely due to conformational changes induced by the lower temperature. A comparison of CD scans obtained at 4 °C for two preparations of TrxHisS - CBF12.2 showed very similar values except for the α - helix content which varied from 15.1 to 17.5. The α - helix and anti-parallel β - sheet content of TrxHisS - CBF12.2 was almost identical to its truncated variant, TrxHisS - CBF12.2-T5, which encompasses the CBF12.2 N - terminal region with the first signature motif and AP2 domain but not second signature motif (Fig. 5.8; Table 5.1). A comparison of the two TrxHisS - CBF12.2 4 °C spectra to TrxHisS - CBF12.1 spectrum revealed almost no difference for parallel β - sheets, β - turns or random coil structures (Table 5.2). The α - helix value was similar (15.8 %) to that of the two TrxHisS - CBF12.2 preparations and parallel β - sheets content was slightly higher for TrxHisS - CBF12.1 (33.6 % versus 30.9 %; Table 5.2). Since both fusions analyzed were cold - acclimated, the 4 °C spectra represented secondary structures in proteins with high affinity to CRT motif. Upon shift to 22 °C, TrxHisS - CBF12.2 showed only a slight increase in anti - parallel β - sheet content (30.9 to 32.2 %; Table 5.2), which appeared consistent with the protein maintaining its high DNA binding ability at the warmer temperature. The 4 to 22 °C shift for TrxHisS - CBF12.1 resulted in an increase in α - helix content from 15.8 to 17.5 % and a simultaneous decrease in anti - parallell β - sheets from 33.6 to 30.9 %. As CBF binding to DNA target is mediated by residues on the anti - parallel β - sheets of AP2 domain (Allen *et al.*, 1998) and strengthened by the alpha - helical configuration of signature motif CMIII - 3 (Canella *et al.*, 2010), the changes in α - helix and anti - parallel β - sheets observed for CBF12.1 fusion may have involved these secondary structures. Thus, a reduction in contact points between protein and DNA may underlie the lower affinity to CRT motif obtained upon transition to warm temperature.

5.4.9 Identification of optimal target sequence for Norstar CBFs

To more precisely specify the optimal target sequence for Norstar CBF variants, a set of double - stranded probes (Table 5.3) were tested by EMSA. Mutated *wcs120* fragments (MUT1 - MUT27) included in the analysis are all derivatives of the 21 - nt *wcs120* fragment with systematic substitutions of one or two nucleotides.

q

Table 5.2 Effect of temperature on secondary structure content in TrxHisS - CBF12.1 and TrxHisS - CBF12.2.

Secondary structure	CBF12.1		CBF12.2 (prep. 1)		CBF12.2 (prep. 2)*	TrxHisS - CBF12.2-T5*
	4 °C	22 °C	4 °C	22 °C	4 °C	4 °C
Helix (%)	15.8	17.7	17.5	17.3	15.1	13.6
Anti-parallel β-sheet (%)	33.6	28.9	30.9	32.2	30.3	30.8
Parallel β-sheet (%)	5.0	5.3	5.3	5.3	4.7	4.4
β-turn (%)	19.1	19.5	19.7	19.9	20.0	20.0
Random coil (%)	31.9	32.9	32.9	32.5	33.1	34.2
Total sum:	105.4	104.4	106.3	107.2	103.2	103.1

* Data from Chapter 4.

The mutations did not involve the central GC dinucleotide which is regarded as the first recognized nucleotides by AP2 domain proteins. The binding assays included 2.5 - 3.0 fmol of each [^{32}P] end - labeled probe and an optimal amount (3.0 µg - 100 ng) of each TrxHisS - CBF fusion needed for detection of complex.

Analysis of probes MUT1 and MUT2 with base substitutions within the core motif ($T^{-2}G^{-1}C^{1}C^{2}G^{3}A^{4}C^{5}$) revealed no binding to any of the CBF fusions when C^{1} nucleotide was replaced with a purine (A/G) (Table. 5.3). If C^{1} was substituted with pyrimidine residue T (MUT2), protein binding was not affected. Similarly, substitution of A^{4} with other purine G resulted in reduction in binding affinity for the strong binders TrxHisS - CBF12.2 and CBF15.0, whereas binding to the weaker TrxHisS - CBFs was not affected. However, if the A^{4} was substituted with a pyrimidine base, it resulted in no detectable binding for all CBFs tested. Interestingly, if the C^{1} and A^{4} substitutions were combined (MUT8), the binding affinity was increased two - fold for all the TrxHisS - CBFs (Fig. 5.16). The increased binding could be explained by the core sequence $G^{-2}G^{-1}C^{1}C^{2}G^{3}A^{4}C^{5}A$ appearing on the reverse strand of the probe fragment, but with -2 residue altered. MUT 25 has the the same core sequence on forward strand as MUT8 and showed same binding strength to protein fusions as MUT8. Similarly MUT24 is the same core sequence as MUT8 on reverse strand but with different 3' flanking nucleotide base pair. No difference in the binding strength of MUT8, MUT24, and MUT25 was observed, indicating that 3' flanking base pair was not affecting the binding. Binding experiments with consensus sequences for AP2 domain proteins from vertebrate and invertebrates (Table 5.4) resulted in no detectable interaction with the CBF protein, indicating that the CBF - CRT interaction is specific for plant proteins only.

5.4.10 Importance of the CRT 5' flanking base pair in DNA - binding

Substituting conserved 5' flanking base pair G^{-1} with A ($T^{-2}A^{-1}C^{1}C^{2}G^{3}A^{4}C^{5}$) reduced binding ability of the CBF proteins to target DNA (MUT4), but did not abolish it. As already mentioned, substitution of C^{1} with A resulted in complete loss of DNA binding, however, both these substitutions together (MUT5) resulted in no detectable interaction of DNA - CBF proteins. Substitution with any other base at these positions also resulted in no binding between CBF protein and the DNA fragment, which indicates that these two positions play important role in the

Table 5.3 Double-stranded DNA fragments and their efficiency as TrxHisS - CBF targets.

Designation	DNA sequence (5'→ 3')	Efficiency*
WCS120	G C C A C C T **G C C G A** C C A C T G A T C	++
MUT2	- - - - - - - **-** T **- - -** - - - - - - - - -	++
MUT4	- - - - - - - A - - - - - - - - - - - - -	+
MUT8	- - - - - - - - T - - G - - - - - - - - -	++++
MUT11	- - - - - - - - - - - G - - - - - - - - -	+
MUT1	- - - - - - - - A - - - - - - - - - - - -	-
MUT3	- - - - - - - - G - - - - - - - - - - - -	-
MUT5	- - - - - - - A A - - - - - - - - - - - -	-
MUT6	- - - - - - - A T - - - - - - - - - - - -	-
MUT7	- - - - - - - A G - - - - - - - - - - - -	-
MUT9	- - - - - - - - T - - C - - - - - - - - -	-
MUT10	- - - - - - - - T - - T - - - - - - - - -	-
MUT12	- - - - - - - - - - - C - - - - - - - - -	-
MUT13	- - - - - - - - - - - T - - - - - - - - -	-
MUT14	- - - - - - - C - - - - - - - - - - - - -	-
MUT15	- - - - - - - T - - - - - - - - - - - - -	-
MUT16	- - - - - - - - A - - G - - - - - - - - -	-
MUT17	- - - - - - - - G - - G - - - - - - - - -	-
MUT24	- - - - - - - - G C C G A C C T C T G A T C	++++
MUT25	- - - - - - - - G C C G A C C G G T G A T C	++++
MUT26	A T T T C A - - G C C T A C C T G C T T T T	-
MUT27	A T T T C A - - G C C G T C C T G C T T T T	-
		Y =T/C;U
Preferred:	G c c a c c **Y U Y Y U U Y** c a c t g a t c	=A/G

* As determined by EMSA in this study.

Table 5.4 Various AP2 domain target sequences and their efficacy as TrxHisS - CBF targets.

Designation	DNA sequence (5'→ 3')	Efficiency*	Reference
MUT18	A C C T G T G T G C A C A T G C A C T G A	no binding	De Silva *et al.*, 2008
MUT19	A C C C T A T G C A T G C A T C A C A C T	no binding	De Silva *et al.*, 2008
MUT20	A C C G C A G C C A G G G G C C G G C A C	no binding	Eckert *et al.*, 2005
MUT22	A C C A A G T C C C C A G G C T C C C A C	no binding	Eckert *et al.*, 2005

* As determined by EMSA in this study.

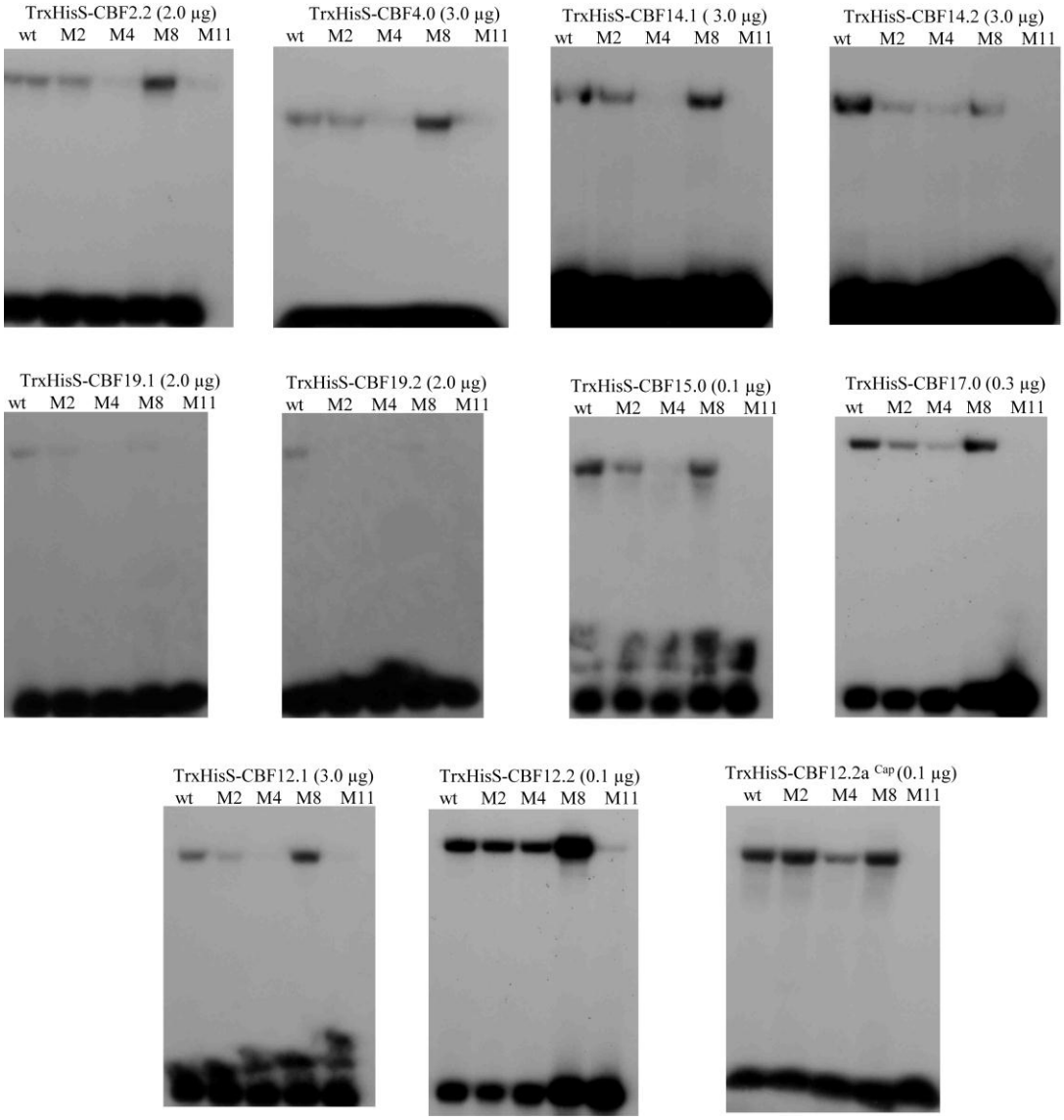


Figure 5.16 Analysis of target sequences for TrxHisS - CBF fusions. Autoradiograph of EMSA performed with 2.5 - 3.0 fmol [^{32}P] end - labeled probe and purified TrxHisS - CBF (0.1 - 3.0 µg). Probes analyzed were *wcs120* and its derivatives MUT2 (M2), MUT4 (M4), MUT8 (M8) and MUT11 (M11) presented in Table 5.3.

specific interaction with the CBF proteins. The base substitution revealed that G^{-1} was the preferred base at the conserved 5' flanking base pair ($T^{-2}G^{-1}C^{1}C^{2}G^{3}A^{4}C^{5}$) and A was alternative base with lower binding affinity. The substitution with the pyrimidine base C or T at this position resulted in no detectable binding affinity. The importance of this position's effect on the binding ability of CBFs was further confirmed by observing a lack of detectable binding between MUT14 ($T^{-2}C^{-1}C^{1}C^{2}G^{3}A^{4}C^{5}$) or MUT15 ($T^{-2}T^{-1}C^{1}C^{2}G^{3}A^{4}C^{5}$). These results clearly show that even with the intact core motif, the change of purine to pyrimidine at the 5' flanking base pair has the most crucial effect. From this analysis it was concluded GGCCGAC was the optimal DNA recognition motif for all Norstar CBFs.

5.5 Discussion

5.5.1 CRT binding activity *in vitro* was demonstrated for 10 Norstar CBFs

The analysis of Norstar CBF proteins encoded from *Fr-A2* demonstrated DNA binding activity to CRT motif by 10 of the 15 purified recombinant TrxHisS - CBF fusions. For TrxHisS - CBF2.1, CBF9.0, CBF20, CBF21 and CBF22, no binding could be demonstrated to *wcs120* probe although the protein binding assays were performed in the presence of *E. coli* soluble extract which had shown stabilizing effect on the fusions. The sequences of the first signature motif and AP2 domain did not give any reason to believe the non - binding CBFs were all pseudo - proteins. Rather, the conditions for protein production or DNA binding assays were likely not conducive for these five CBFs to show activity *in vitro*.

5.5.2 Norstar CBF binding to different DNA fragments

The possibility the five non - active CBFs recognized a DNA target that was slightly different from the CRT CCGAC core sequence was explored by testing 21 different MUT probes. These probes covered several possible substitutions of the CRT core motif (Table 5.3), but none was recognized by the five non - functional CBFs. Also five different binding sites determined for AP2 proteins, but with little similarity to CCGAC motif (Table 5.4), were found non - functional as DNA targets for any of the 15 CBFs studied. The fact that binding of the functional CBFs was not restricted to only one fragment, but rather tolerated some modification

of the CCGAC core and flanking sequences argued against a CCGAC - like sequence being the true target sequence for the non - binding CBFs.

For the functional CBFs, five different target motifs were recognized by one or several CBFs (Fig. 5.16) and the 5' flanking bases of probes were found to be essential for protein / DNA interactions. The preferred target sequence for CBF14.2, CBF19.1 and CBF19.2 was TG**CCGAC** carried by *wcs120* probe. CBF2.2, CBF4.0, CBF12.1, CBF12.2, CBF14.1, CBF17.0 preferred the GG**CCGAC** motif carried by MUT8 probe (Table 5.3), whereas CBF15.0 did not show any preference between TG**CCGAC** or GG**CCGAC** (Fig. 5.16). The *wcs120* probe variants TGTCGACC, TA**CCGAC**C, TGCCGGCCA motifs were also recognized by several CBFs, but with lower affinity than the preferred targets.

Most preferred binding site for HvCBF2 DNA is YGTCGACAT (Xue, 2003), but the core of this element (TCGAC) was a very weak binding element when tested with Norstar CBFs. Similarly, *Arabidopsis* DREB1A exhibits weak binding with TCCGAC and ATCGAC (Sakuma *et al*., 2002), but wheat CBFs did not show any detectable binding to these sequences. Thus, the CBF preferred motifs appeared to be species specific, and this difference may have implications on genes or regulons targeted by the CBFs during cold acclimation as suggested by Skinner *et al.*, (2005). The finding that the two closely related CBF14.1 and CBF14.2 showed different preference for target sequence was unexpected. The two CBF14 variants do not differ for AP2 domain or signature sequences (Appendix Table 4.1), and only show eight amino acid differences. One difference is for the fourth residue (Ala / Val variation) at the amino terminus and the remainder substitutions are found within the last 62 residues of the protein. Since analysis of the truncated CBF17.0 and CBF12.1 suggested C - terminal sequences could affect DNA binding (Fig. 5.8), the C - terminal end of CBF14 may also have this role affecting DNA binding specificity.

In contrast to CBF17.0 and CBF12.1, the CBF12.2 and CBF15.0 were able to bind the target sequence even when the entire C - terminal region and secondary signature motif DSAWL were missing (Fig. 5.8). Thus, the C - terminal region did not seem to have equal importance for all CBFs, at least not in the *in vitro* assays. The role for the well conserved secondary signature

motif has not been demonstrated previously, but as shown here, it is dispensable for CBF12.2 and CBF15.0 function, although protein activity was reduced about two - fold when motif was removed. Therefore, the DSAWL motif may have a minor role in stabilizing the AP2 domain.

The Norstar CBFs carry multiple clusters of hydrophobic residues in their C - terminal regions (Chapter 3). These clusters constitute trans - activation domains in *Arabidopsis* CBF1 *in vivo* (Wang *et al.*, 2005). Regulatory regions are also located within C - terminal region of *Arabidopsis* AP2 domain protein, DREB2A, which like the CBFs binds to CRT / DRE motif (Liu *et al* 1998; Sakuma *et al.*, 2006). Removal of the DREB2 repressor domain located after the AP2 domain increases binding to CRT element 30 - fold. If such regulatory domains exist on the non-binding Norstar CBFs and negatively affecting DNA binding ability *in vitro* remains to be demonstrated by additional experiments.

5.5.3 Allelic variants of Norstar *CBF2* and *CBF12* encode proteins with functional differences

Norstar *Fr-A2* carries two *CBF2* alleles and two *CBF12* alleles, for which the extra copies were likely derived from gene duplication events within the locus (Båga *et al.*, unpublished). CBF2.1 and CBF2.2 showed differences in DNA binding activity where CBF2.1 was inactive whereas binding could be demonstrated for CBF2.2 (Fig 5.2). An explanation for the functional differences between the CBF2 variants could not be precisely identified from alignment of the two amino acid sequences (Fig. 5.5). The fact CBF2.1 was unstable upon purification from *E. coli* extract, but not CBF2.2, suggested CBF2.1 had an instability element that was absent for CBF2.2. Thus, one possible reason for CBF2.1 inactivity could be the presence of a PADEST insertion in C - terminal region as this element added the amino acids Pro, Glu(Asp), Ser and Thr, which are associated with protein instability when flanked by Arg(Lys) residues (Rechsteiner and Rogers, 1996). Analysis of CBF2.1 and CBF2.2 by 'epestfind' software further supported the hypothesis that the protein segment carrying the PADEST insertion was a potential instability element, but no instability was suggested for CBF2.2. However, further studies are needed to confirm if indicated PEST segment on CBF2.1 confers instability and / or inactivity.

Despite high sequence similarity, EMSA showed CBF12.1 had lower affinity to CRT element than CBF12.2 (Fig. 5.3). As CBF12.2 was not dependent on second signature motif or sequences beyond for DNA binding, the difference in activity between CBF12.1 and CBF12.2 could be related to F41L substitution within first signature motif or extra proline residue (P102) positioned after α - helix of CBF12.2 AP2 domain (Fig. 5.6). Of the two possibilities, the substitution seems more likely as it changes a highly conserved residue. The preferred residue at this position is Phe or Tyr, which both confer higher DNA binding activity in *Arabidopsis* CBF1 (Canella *et al.*, 2010). As replacement of an aromatic side chain amino acid, Phe40, to aliphatic side chain amino acid, Leu40, as in CBF12.1, may change protein structure causing reduced binding of CBF 12.1. As for the CBF2 variants, further studies are needed to confirm this hypothesis.

5.5.4 CBF12 gene of cold - sensitive Cappelle - Desprez encodes an inactive CBF

CBF12.2^{Cap} of cold - sensitive Cappelle - Desprez was found to be inactive in contrast to the near identical CBF12.2 produced by cold - hardy Norstar (two amino acid difference) and CBF12.2^{wM} of cold - sensitive winter Manitou (one amino acid difference), which both interacted strongly with CRT element. Due to the one amino acid differences of winter Manitou CBF12.2^{wM}, the CBF12.2^{Cap} inactivity could be pinpointed to R60Q substitution within AP2 domain. Both arginine and lysine are highly favored (122 out of 129 studied) at this position in ERF / AP2 domain proteins (Sakuma *et al.*, 2002). This residue in *Arabidopsis* ERF1 forms hydrophobic interactions through its side-chain with other residues within AP2 as determined by three - dimensional studies (Allen *et al.*, 1998). Thus, the R → Q substitution may alter AP2 structure resulting in loss of contact with DNA. The extent by which this *CBF12.2b*^{Cap} mutation contributes to lower frost resistance in Cappelle - Desprez is difficult to predict as the cultivar has the potential to produce a functional CBF12.2 from the *CBF12.2a*^{Cap} allele.

5.5.5 Norstar produces CBFs with exceptional stability

The Norstar CBFs DNA activities were found to be extremely resistant to high (i.e. boiling) as well as low (i.e. freezing at - 80 ^{o}C) temperatures (Fig. 5.11). No loss of secondary structure within the temperature range from 4 - 95 ^{o}C could be demonstrated by CD spectra, which confirmed the recombinant CBF protein does not exhibit a normal thermal denaturation.

The unusual stability demonstrated for CBF12.1 and CBF12.2 may be specific to wheat CBFs, as *Arabidopsis* CBF1 show a complete loss of secondary structure between 40 - 60 ^{o}C (Kanaya *et al*., 1999). The Norstar CBFs were extremely resistant to urea and were able to interact with DNA target in the presence up to 3 M urea. Low concentrations of urea (e.g. 0.8 M urea) had a positive effect on interaction between CBF12.1 and its target (Fig. 5.10). A high concentration of urea (6 M) was not able to fully disrupt the tertiary and secondary structure of protein as shown in Chapter 4.

5.5.6 Cold experience inceases DNA binding activity and memory of cold

All the 10 CBFs from Norstar bound to CRT motif under both warm and cold conditions (Fig. 5.12), in agreement with barley HvCBF3 and HvCBF4 - subgroup members (Skinner *et al*., 2005). However, there were large differences in activities for the functional Norstar CBFs depending on temperature, previous exposure to cold or denaturing conditions. For CBF12 fusions never exposed to cold, the initial affinity for CRT motif at low temperature (4 °C) was relatively weak, but the proteins became very active depending on time (h) in the cold. The CD spectra of CBF12.1 revealed an increase in α - helix and lower anti - parallell β - sheet content upon shift to from cold to warm temperature, which could be correlated with protein losing some of its affinity for CRT motif. A comparison of two CD spectra for CBF12.1 and CBF12.2 also suggested that secondary structures of the two proteins are not identical, which may underlie their different reaction to warm temperature after being cold - exposed.

Cycling of CBF12.1 from warm - cool - warm temperatures appeared to involve four activity stages. In the initial warm stage CBF12.1 is weakly active, in the second cooling stage the protein becomes highly active, the third stage at warm temperature the protein become less active but different from first stage, in the fourth stage at second cold exposure, the protein is able to quickly switch to highly active state. This response to temperature and acquisition of cold memory may be important during cold acclimation in the fall when daily temperatures switch from 10 - 20 °C during day - time to 0 - 10 °C at night - time.

A high CBF expression at non - acclimating conditions by ectopic expression of CBF genes has several negative effects of overall fitness of *Arabidopsis* and other plants (Gilmour *et*

al., 2004). Thus, a mechanism that reduces CBF activity during the day when photosynthesis is fully active may be important for energy conservation during the cold acclimation period. However, the down-regulation of activity did not seem to be general as demonstrated by cold - acclimated CBF12.2.

Freeze - thaw cycling has been associated with development of high cold - tolerance in birch (*Betula pendula*), which has a CBF regulon and like many artic and temperate trees can achieve very high freezing tolerance during winter (Welling and Palva, 2008). During freeze - thaw cycles, the expression level of BpCBF genes is increased during thawing adding additional cold - tolerance to the plant. This might be the underlying reason for increase in freezing tolerance of winter wheat cultivars during sub - zero acclimation.

CHAPTER 6

GENERAL DISCUSSION

6.1 Background

One - fifth of earth's land surface is covered by grasses which belong to the family Poaceae and include some of the most important agricultural crops such as wheat, rice, maize, barley and other minor cereals and forage grasses. Poaceae family originated about 70 - 80 million years ago (Mya), and about 50 Mya diverged in to Panicoideae and BEP (Bambusoideae, Ehrhartoideae and Pooideae sub-families) clade which 46 - 41 Mya diverged in to the three respective sub-families (Fig 6.1; Gaut 2002; Stromberg 2005; Sandve *et al*., 2008). The early Pooideae evolution (45 - 35 Mya) also coincides with the global climate change towards a cooler climate also referred to as the Coenozoic 'greenhouse to icehouse transition' (Zachos *et al*., 2001; 2008). The global climate cooling was associated with large scale extinctions and shifts in global distribution of fauna and flora (Ivany *et al*., 2000; Seiffert 2007). Genome duplication and polyploidization has been suggested as a strategy for evolution and rapid diversification of angiosperms (Soltis *et al*., 2008). Recently available complete genome sequences of some plants such as *Arabidopsis*, rice and selected genes in some other crops such as wheat, maize have been used to predict the whole genome duplication events and their role in evolution and survival under adverse conditions (Figure 6.1). The current consensus is that species which underwent genome duplication were better adapted to changing environmental conditions and increased gene content contributed to plants success in survival during the Cretacious - Tertiary extinction event (Fawcett *et al*., 2009). Grass family is adapted to cool environments, the origin of which has been traced to the Eocene - Oligocene boundary about ~33.5 - 26 Mya. Genome duplication was also used by the ancestors of modern day Pooideae members that aided in their survival and adaptation to rapid cooling during the Eocene - Oligocene boundary cool climates. Recent studies have shown that during the Eocene - Oligocene boundary cool climate saw rapid expansion in three gene families, Fructosyl transferase (FT), C - repeat binding factors (CBF) and Ice recrystallisation inhibition (IRI) protein family, which are associated with cold tolerance in plants (Sandve *et al*., 2008; Sandve and Fjellhelm, 2010).

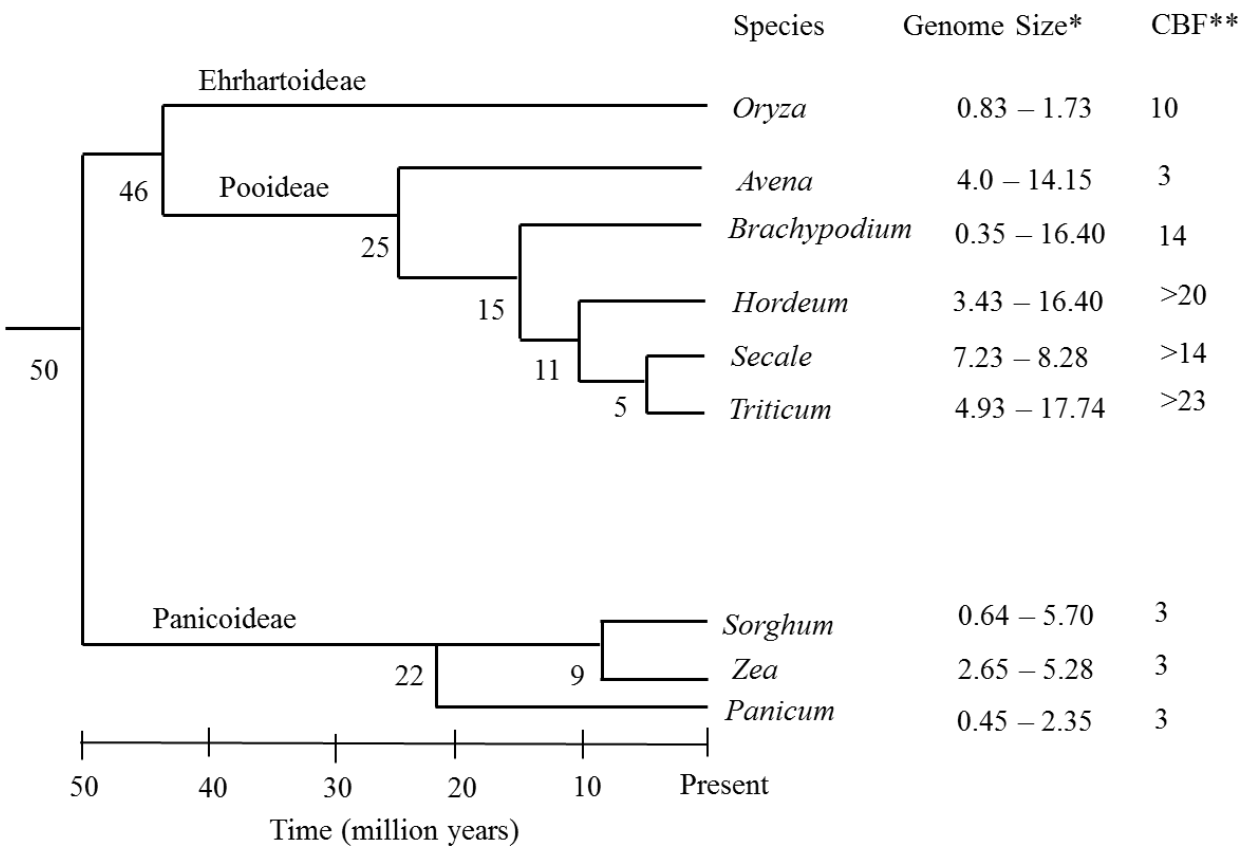


Figure 6.1 Evolution of wheat with respect to other economically important species within each subfamily.

Divegence time at each node is in million years (Bossolini *et al*., 2007; Bortiri *et al*., 2008).
*Genome size is the range of haploid DNA content, in picograms, for diploid and polyploid species collected from Angiosperm C - value database (http://www.rbgkew.org.uk/cval/).
** number of CBF genes in each species.

Among the three protein families which were shown to have rapidly expanded in response to the Eocene - Oligocene boundary cooling period, the CBF gene family is of special interest as a transcription factor it can regulate the expression of several cold responsive genes (Thomashaw 1999). In the present day cereals comparing mostly warm season non cold tolerant crop rice to *Brachypodium*, less cold tolerant barley and diploid wheat to winter wheat and most cold hardy crop rye, there has been a gradual expansion in the number of CBF gene family members (Fig. 3.1). Although in rye a few CBF genes have been characterized, but there may be several more which will be identified as rye genome is sequenced. The increase in CBF members is likely due to point mutations that can modulate their function and could be of recent occurrence to help plants adapt to their changing environment.

6.2 CBF polypeptide sequence based cluster analyses reveal functional specialization

Cluster analyses of winter wheat cv Norstar CBF polypeptides and their constituents clustered the peptides into five groups / subgroups (IIIa, IVa, IVb, IVc and IVd) out of the 10 previously identified groups for CBF genes (Badawi *et al*., 2007). Members of these five groups have been associated with higher LT tolerance capacity (Badawi *et al*., 2007; Skinner *et al*., 2005; Miller *et al*., 2006). Further analysis based on functional domains of CBFs show that mainly groups / subgroups remain the same but individual members are changed. This suggests that functional difference in CBF can account for differences among the present day genotypes.

Bioinformatic analysis of AP2 - DNA binding domain and flanking signature sequence amino acid sequence between cold tolerant rye and wheat show that there are difference in the residues contacting with DNA. Also, rye has at least two copies of each group IV member (Fig. 3.2). This could explain higher LT tolerance of rye. The amino acid sequence differences also support the suggestion that such point mutations can modulate their function and help plants to adapt to changing climates.

Another difference observed in our study is that barley signature sequence that was considered representative of cereals (Skinner *et al*., 2005) is not conserved for Norstar CBF. Since these signature sequences are used to classify CBFs in different groups, we suggest the CMIII - 3 motifs from wheat can be defined for subgroup III as xKRPAGRTK(F/L)K(E/D)TRHP

and CMIII - 1 as DSA(W/C)L. Similarly, for subgroup IV the motifs can be defined as PK(R/W)(P/R)AGRxKxxETRHP and DSAWR.

6.3 CBF12 isoforms, CBF15.0 and CBF17.0 show extreme stability

To identify individual Norstar CBF function, a set of 15 *CBF* from cold hardy winter wheat cv Norstar were expressed in *E. coli* to produce TrxHisS - CBF fusion proteins. An interesting observation that although all the genes from one specific cultivar, but some of the fusion proteins behaved differently. All TrxHisS - CBF fusion proteins except TrxHisS - CBF17.0 could be recovered in the cell soluble phase. Ten of the TrxHisS - CBF fusion proteins were very stable to heat and strong protein denaturing conditions. Most of the CBFs were found to extremely stable, both as *E. coli* cell extracts and in the purified form. However, a few CBFs, namely CBF 2.1, 9.0, 20.0, 21.0, and 22.0, were found to be stable as cell extract but unstable under purified form. SDS - PAGE analyses show that all CBFs have anomalous migration on the gel, most likely due to incomplete denaturation, indicating its heat stability. This was confirmed by CD spectra analysis, where it was observed that 12.2 retained secondary structure both on heating as well as in presence of urea. Several COR proteins like LEA, dehydrins and enolase are known to be heat stable (Ceccardi *et al*., 1994; Houde *et al*., 1995; Campbell and Close, 1997; Pancholi, 2001; Lee *et al*., 2002) but none of them act as transcription factors. Not only were the proteins found to be extremely stable in presence of urea but DNA - binding activity of CBFs also increased at least two fold in the presence of urea. The only other transcription factor resistant to denaturation by heat and urea was reported from third instar larvae nuclei of *Drosophila melanogaster* (Jack, 1990). This protein retains and increases its DNA - binding ability, similar to CBFs observed in the study. In addition, CBFs are seen to be stable to cold denaturation.

Taken together, these results indicate that CBF stability, especially CBF12.1, 12.2, 15.0 and CBF17.0, might be an important property for higher LT tolerance of Norstar. The higher threshold induction temperature of Norstar could be due to the presence of these stable CBFs as well as longer duration of up - regulation of COR genes, resulting in accumulation of LT tolerance, both in fall and early spring season.

6.4 CBF C - terminal region show variable role for DNA binding

All the CBFs were found to be stable even in the absence of a C - terminal region. However, difference in function of CBFs was observed due to presence / absence of a C - terminal region. It has been suggested that hydrophobic residues of the C - terminal region of CBFs might affect the conformation of the protein and in turn its DNA - binding abilty. In the case of CBF12.2 and CBF15.0, polypeptide containing only an AP2 - DNA binding domain was sufficient to produce a complex with DNA, whereas for CBF12.1 and CBF17.0 polypeptide a second signature motif along with part of the C - terminal were required for detectable binding. CD - spectra analysis of intact CBF12.2 and truncated polypeptide containing region up to the AP2 - DNA binding domain showed no significant difference in the secondary structure confirming that there was no conformation change due to C - terminal region, and that it plays a role in functional property of CBF12.2. However, it is most likely that the C - terminal is essential for proper functioning of CBF15.0 and some other CBFs (i.e. CBF12.1 and17.0). Since, plants have evolved to utilize their resources / machinery and energy most economically, there has to be a reason for retained region in proteins. Another function attributed to CBF C - terminal region is that it may act as transcriptional activation region (Sakuma *et al*., 2006). Therefore where it is not required for the functional activity of individual CBF e.g. CBF12.2 and CBF15.0, it may be interacting with some other gene. The complete complex formed due to the interaction of CBF AP2 - domain with CRT motif and C - terminal with other DNA sequence could be regulating some other, yet to be identified regulon. Another possibility is that the conformation change in CBF due to interaction with CRT could enable the C - terminal to interact with other DNA sequence to activate a different regulon.

6.5 Consensus sequence for CRT motif

CBFs from *Fr-A2* locus of Norstar bind with varying strength to the wild type oligonucleotide sequence based on *wcs120* promoter region. In addition, our experiments show that 5’ base pair flanking the CRT core motif (CCGAC) has an important effect on the protein-DNA binding. Presence of CRT motif alone does not ensure CBF - DNA binding, as observed with MUT14 - 15 oligonucleotide sequences. For example when the CRT motif was preced by a non-preferred 5’ flanking base pair, no detectable binding was observed. The most preferred base

pair was ‘GG’ (MUT8, 24, 25) resulting in very strong binding, that is at least two - fold greater than the wild type present in most of the CBFs. The most preferred CRT sequence is C/t CGA/g C, where lower case indicates the alternate base resulting in lower affinity and any other sequence results in no - detectable binding. Therefore, we suggest that instead of five base pair as core CRT motif (CCGAC) for CBF - DNA interaction, we should consider seven base pairs (xxC/t CGA/g C). Since presence of CBFs and the known CBF - regulon members can explain up to 50 % difference in LT tolerance in Norstar, we suggest these differences in the binding motif preferred by CBFs might indicate that there are more members of the existing regulon or existence of a different regulon regulated by these CBFs. The present members of CBF regulon have been identified on the basis of 5 bp CRT core motif, but a 7 bp core motif might lead us to identify different regulon activated by the CBFs.

6.6 CBF12.1 function is regulated by temperature through structural changes

Another important factor contributing to difference in LT tolerance is presence of CBF influenced by temperature fluctuation. This study, for the first time reports that between the CBF12 isoforms from Norstar, CBF12.1 has the property to modulate its DNA binding ability with exposure to LT whereas CBF12.2 remains a consistent strong binder of the DNA fragments tested. The HvCBF2 was unable to bind to CRT / DRE motif at 30 ^{o}C but there was increase in its DNA binding activity with lowering of temperature (Xue, 2003). Similar influence of temperature on DNA binding ability of CBF12.1 was observed in this study. However, unlike HvCBF2, Norstar CBF12.1 showed strong binding to CRT motif both at extremely high (100 ^{o}C) and low temperatures (- 80 ^{o}C). In addition, DNA binding ability at 4 ^{o}C of CBF12.1 increased with pre-incubating protein alone through a cycle of cold - warm - cold (4 ^{o}C - 22 ^{o}C - 4 ^{o}C) temperature, simulating the time and temperature for acclimation - deacclimation - reacclimation process in nature.

Norstar CBF12 isoforms show continuous increase in DNA - binding ability with pre-incubation of protein at 4 ^{o}C. The above mentioned two observations suggest that conformation change due to temperature modulate Norstar CBFs from a less active to a more active state. These observations were further supported by studying the secondary structure of CBF 12.1 with

CD spectra and increase in amount of beta - sheets present at 4 ^{o}C was more than at 22 ^{o}C. The HvCBF2 changed from inactive state at 30 ^{o}C to active state at 2 ^{o}C (Xue, 2003). Taken together, these results suggest that some CBFs are stable (e.g. CBF12.2) to temperature fluctuation while other CBFs (e.g. CBF12.1) respond more to these fluctuations and this change is species specific. Therefore, results suggest that *in vivo*, the stable CBFs might provide the basal LT tolerance and other CBFs prepare the plant for better LT tolerance after first few spell of cold temperatures in autumn season. Also, another possibility could be that higher number of genes encoding for stable CBF might be the reason for superior threshold induction temperature leading to higher LT tolerance in some cultivars.

6.7 Candidate genes for high LT tolerance in Norstar

Even though there is a hierarchical, and to some extent decisive, relationship between the transcriptome and the proteome response, gene expression levels and protein levels in many cases are not highly correlated, as exemplified by comparing protein abundance with transcript abundance in *Arabidopsis* leaf and root (Mooney *et al*., 2006). Therefore, it would be inappropriate to make functional conclusions based mainly on data from transcript or proteomic profiling. In this study, difference in the function of isoforms of CBFs from *Fr-A2* locus was observed. Among the two isoforms of CBF2, one binds to the CRT motif whereas its variant does not show any detectable protein - DNA interaction. Similarly, CBF12 isoforms exhibited 200 - fold difference in DNA binding strength. These differences in protein function of CBFs indicate that presence or up - regulation of *cbf* alone cannot determine or ensure the level of LT tolerance. In fact, unless specific protein function data is available, conclusions from such studies may be misleading.

TmCBF12 and *TmCBF15* have been associated with higher LT tolerance in *T. momcoccum* accession G3116 and a five amino acid deletion in AP2 - DNA binding domain was identified as the underlying reason for the difference in level of LT tolerance between the two accessions (G3116 and DV92) studied. This study narrows down that region for difference (i.e. five amino acids) to a single amino acid, resulting in complete loss of DNA binding activity between a hardy winter wheat cv Norstar and less cold - hardy Cappelle - Desprez. Presence of

R_{60} in Norstar ensures presence of active protein whereas replacement of R_{60} with Q_{60} results in inactive CBF12 in the less cold - hardy cultivar. The non - functional CBF12 allele in Cappelle - Desprez may contribute to its lower frost tolerance similar to difference between to the two above mentioned *T. monococcum* accessions. This study improves the understanding of CBF function between spring and winter wheat. Further, CBF15.0 was found to be stable protein and capable of strong DNA binding ability. Therefore, combination of stable CBFs, namely CBF12.2, 15.0 and 17.0 and CBF 12.1 that can modulate its activity in response to temperature, appear to be the candidate genes for high LT tolerance of Norstar.

In conclusion, a major event of CBF gene family expansion during the Eocene - Oligocene boundary cool climate allowed the progenitors of wheat to survive. The complex gene families with several members provided adequate genetic buffer for mutations to occur which in some cases change protein functionality. The change in protein functionality gave those genotypes an advantage to adapt to their niche climates. Several such examples were discovered in this work and it stresses the need that to develop strategies of plant adaptation to environment we need to better understand protein functionality.

6.8 Conclusions and novel scientific findings

- Difference in clusters members were based on individual functional domain amino acid sequences of Norstar CBFs.
- The Norstar Group III signature motif was redefined: CMIII - 3 motifs as xKRPAGRTK(F/L)K(E/D)TRHP and CMIII - 1 as DSA(W/C)L. Similarly, for subgroup IV the motifs can be defined as PK(R/W)(P/R)AGRxKxxETRHP and DSAWR.
- Optimized conditions for production of recombinant CBFs in *E. coli* were achieved.
- CBFs were found to be recalcitrant to denaturation, both by temperature and urea.
- CBFs retained DNA - binding activity in extreme conditions, both temperature and in the presence of urea
- A difference in the effect of C - terminal region on DNA - binding activity of CBFs.
- Importance and preference of 5’ base pair and the CRT motif was observed.

- CBF12.1 DNA - binding activity was regulated by temperature by bringing conformational changes in the protein structure.
- CBF12 and CBF15.0 were identified as candidate genes resulting in higher LT tolerance of Norstar.
- A difference in the DNA-binding function of CBFs was observed, proving that the hypothesis that '*The function of one or several CBFs encoded from Fr-A2 has an effect on low temperature tolerance in winter wheat'* can be accepted.

6.9 Future studies

Differences in the DNA binding strength observed for the CBFs tested suggested that although they were present at the same locus, *in vitro* studies suggested that individual CBF contributed differently to LT tolerance. This study focused on a winter hardy cv Norstar, but general validation of these results should be done by studying CBF gene arrangements and their functions in very cold - hardy genotypes of cereals and grasses present in high latitude regions of Europe and Russia. This is the first study of CBF protein structure and its interaction with temperature however the results need to be expanded with CBF that are responsive to and non - responsive to cold. Another important line of research is to study the interaction of CBF with other abiotic stress such as drought and heat, which also cause water limiting conditions. It would be very interesting to discover if CBF could also participate in the cross talk of biochemical / genetic pathways, to determine plants response to multiple abiotic stresses likely to happen as a result of climate change.

REFERENCES

Achard, P., Gong, F., Cheminant, S., Alioua, M., Hedden, P., & Genschik, P. (2008). The cold - inducible *CBF1* factor - dependent signaling pathway modulates the accumulation of the growth - repressing DELLA proteins via its effect on gibberellin metabolism. *Plant Cell*, *20*, 2117 - 2129.

Adams, W. W., Demmig - Adams, B., Rosenstiel, T. N., Brightwell, A. K., & Ebbert, V. (2002). Photosynthesis and photoprotection in overwintering plants. *Plant Biology*, *4*, 545 - 557.

Adams, W., & Demmigadams, B. (1994). Carotenoid composition and down - regulation of photosystem - II in three conifer species during the winter. *Physiologia Plantarum*, *92*, 451 - 458.

Adamson, P., Marshall, C., HALL, A., & Tilbrook, P. (1992). Post translational modifications of P21(rho)proteins. *Journal of Biological Chemistry*, *267*, 20033 - 20038.

Agarwal, M., Hao, Y., Kapoor, A., Dong, C., Fujii, H., & Zheng, X., (2006). A R2R3 type MYB transcription factor is involved in the cold regulation of *CBF* genes and in acquired freezing tolerance. *Journal of Biological Chemistry*, *281*, 37636 - 37645.

Agarwal, P., Agarwal, P. K., Nair, S., Sopory, S. K., & Reddy, M. K. (2007). Stress - inducible *DREB2A* transcription factor from *Pennisetum glaucum* is a phosphoprotein and its phosphorylation negatively regulates its DNA - binding activity. *Molecular Genetics and Genomics*, *277*, 189 - 198.

Aguilar, P. S., Hernandez - Arriaga, A. M., Cybulski, L. E., Erazo, A. C., & de Mendoza, D. (2001). Molecular basis of thermosensing: A two - component signal transduction thermometer in *Bacillus subtilis*. *EMBO Journal*, *20*, 1681 - 1691.

Ahmad, M., & Cashmore, A. R. (1993). *HY4* gene of *A. thaliana* encodes a protein with characteristics of a blue - light photoreceptor. *Nature*, *366*, 162 - 166.

Akar, T., Francia, E., Tondelli, A., Rizza, F., Stanca, A. M., & Pecchioni, N. (2009). Marker - assisted characterization of frost tolerance in barley (*Hordeum vulgare* L.). *Plant Breeding*, *128*, 381 - 386.

Ala, P., Chong, P., Ananthanarayanan, V., Chan, N., & Yang, D. (1993). Synthesis and characterization of a fragment of an ice nucleation protein. *Biochemistry and Cell Biology*, *71*, 236 - 240.

Allagulova, C., Gimalov, F., Shakirova, F., & Vakhitov, V. (2003). The plant dehydrins: Structure and putative functions. *Biochemistry - Moscow*, *68*, 945 - 951.

Allen, M. D., Yamasaki, K., Ohme - Takagi, M., Tateno, M., & Suzuki, M. (1998). A novel mode of DNA recognition by a beta - sheet revealed by the solution structure of the GCC - box binding domain in complex with DNA. *EMBO Journal*, *17*, 5484 - 5496.

Alm, V., Busso, C. S., Ergon, A., Rudi, H., Larsen, A., Humphreys, M. W., & Rognli, O. A. (2011). QTL analyses and comparative genetic mapping of frost tolerance, winter survival and drought tolerance in meadow fescue (*Festuca pratensis* Huds.). *Theoretical and Applied Genetics*, *123*, 369 - 382.

Alonso, A., Queiroz, C. S., & Magalhaes, A. C. (1997). Chilling stress leads to increased cell membrane rigidity in roots of coffee (*Coffea arabica* L) seedlings. *Biochimica Et Biophysica Acta - Biomembranes*, 1323, 75 - 84.

Alonso - Blanco, C., Gomez - Mena, C., Llorente, F., Koornneef, M., Salinas, J., & Martinez - Zapater, J. (2005). Genetic and molecular analyses of natural variation indicate *CBF2* as a candidate gene for underlying a freezing tolerance quantitative trait locus in *Arabidopsis*. *Plant Physiology*, *139*, 1304 - 1312.

Amid, A., Lytovchenko, A., Fernie, A. R., Warren, G., & Thorlby, G. J. (2012). The sensitive to freezing3 mutation of *Arabidopsis thaliana* is a cold - sensitive allele of homomeric acetyl - CoA carboxylase that results in cold - induced cuticle deficiencies. *Journal of Experimental Botany*, *63*, 5289 - 5299.

Antikainen, M., & Griffith, M. (1997). Antifreeze protein accumulation in freezing - tolerant cereals. *Physiologia Plantarum*, *99*, 423 - 432.

Artuso, A., Guidi, L., Soldatini, G. F., Pardossi, A., & Tognoni, F. (2000). The influence of chilling on photosynthesis and activities of some enzymes of sucrose metabolism in *Lycopersicon esculentum* Mill. *Acta Physiologiae Plantarum*, *22*, 95 - 101.

Athmer, B., Perovic, D., Himmelbach, A., Szucs, A., Vashegyi, I., & Kocsy, G. (2010). Regulation of gene expression by chromosome 5A during cold hardening in wheat. *Molecular Genetics and Genomics*, *283*, 351 - 363.

Badawi, M., Danyluk, J., Boucho, B., Houde, M., & Sarhan, F. (2007). The *CBF* gene family in hexaploid wheat and its relationship to the phylogenetic complexity of cereal CBFs. *Molecular Genetics and Genomics*, *277*, 533 - 554.

Badawi, M., Reddy, Y. V., Agharbaoui, Z., Tominaga, Y., Danyluk, J., Sarhan, F., & Houde, M. (2008). Structure and functional analysis of wheat *ICE* (inducer of *CBF* expression) genes. *Plant Cell Physiology*, *49*, 1237 - 1249.

Båga, M., Chodaparambil, S. V., Limin, A. E., Pecar, M., Fowler, D. B., & Chibbar, R. N. (2007). Identification of quantitative trait loci and associated candidate genes for low - temperature tolerance in cold - hardy winter wheat. *Functional and Integrative Genomics*, *7*, 53 - 68.

Båga, M., Fowler, D. B., & Chibbar, R. N. (2009). Identification of genomic regions determining the phenological development leading to floral transition in wheat (*Triticum aestivum* L.). *Journal of Experimental Botany*, *60*, 3575 - 3585.

Baker, S. S., Wilhelm, K. S., & Thomashow, M. F. (1994). The 5' - region of *Arabidopsis thaliana cor15a* has *cis* - acting elements that confer cold - regulated, drought - regulated and ABA - regulated gene - expression. *Plant Molecular Biology*, *24*, 701 - 713.

Ballare, C. L. (1999). Keeping up with the neighbours: Phytochrome sensing and other signaling mechanisms. *Trends in Plant Science*, *4*, 97 - 102.

Banker, G. A., & Cotman, C. W. (1972). Measurement of free electrophoretic mobility and retardation coefficient of protein sodium dodecyl sulfate complexes by gel - electrophoresis

- method to validate molecular - weight estimates. *Journal of Biological Chemistry*, *247*, 5856 - 5861.

Battaglia, M., Olvera - Carrillo, Y., Garciarrubio, A., Campos, F., & Covarrubias, A. A. (2008). The enigmatic LEA proteins and other hydrophilins. *Plant Physiology*, *148*, 6 - 24.

Baurle, I., & Dean, C. (2006). The timing of developmental transitions in plants. *Cell*, *125*, 655 - 664.

Bayles, D. O., & Wilkinson, B. J. (2000). Osmoprotectants and cryoprotectants for *Listeria* monocytogenes. *Letters in Applied Microbiology*, *30*, 23 - 27.

Becker, L. A., Evans, S. N., Hutkins, R. W., & Benson, A. K. (2000). Role of sigma(B) in adaptation of *Listeria* monocytogenes to growth at low temperature. *Journal of Bacteriology*, *182*, 7083 - 7087.

Benedict, C., Skinner, J. S., Meng, R., Chang, Y., Bhalerao, R., Huner, N. P., Finn, C. E., Chen, T. H., & Hurry, V. (2006). The *CBF1* - dependent low temperature signalling pathway, regulon and increase in freeze tolerance are conserved in populus spp. *Plant, Cell and Environment*, *29*, 1259 - 1272.

Bennett, M. D., Bhandol, P., & Leitch, I. J. (2000). Nuclear DNA amounts in angiosperms and their modern uses - 807 new estimates. *Annals of Botany*, *86*, 859 - 909.

Blanc, G., & Wolfe, K. H. (2004). Functional divergence of duplicated genes formed by polyploidy during *Arabidopsis* evolution. *Plant Cell*, *16*, 1679 - 1691.

Böhm, G., Muhr, R., & Jaenicke, R. (1992). Quantitative analysis of protein far UV circular dichroism spectra by neural networks. *Protein Engineering*, *5*, 191 - 195.

Boonrod, K., Fuellgrabe, M. W., Krczal, G., & Wassenegger, M. (2011). Analysis of the autoproteolytic activity of the recombinant helper component proteinase from zucchini yellow mosaic virus. *Biological Chemistry*, *392*, 937 - 945.

Bortiri, E., Coleman - Derr, D., Lazo, G. R., Anderson, O. D., & Gu, Y. Q. (2008). The complete chloroplast genome sequence of *Brachypodium distachyon*: Sequence comparison and phylogenetic analysis of eight grass plastomes. *BMC Research Notes*, *1*, 61 - 68.

Boss, P. K., Bastow, R. M., Mylne, J. S., & Dean, C. (2004). Multiple pathways in the decision to flower: Enabling, promoting, and resetting. *Plant Cell*, *16*, S18 - S31.

Bossolini, E., Wicker, T., Knobel, P. A., & Keller, B. (2007). Comparison of orthologous loci from small grass genomes *Brachypodium* and rice: Implications for wheat genomics and grass genome annotation. *Plant Journal*, *49*, 704 - 717.

Bouche, N., Scharlat, A., Snedden, W., Bouchez, D., & Fromm, H. (2002). A novel family of calmodulin - binding transcription activators in multicellular organisms. *Journal of Biological Chemistry*, *277*, 21851 - 21861.

Bowen, B., Steinberg, J., Laemmli, U. K., & Weintraub, H. (1980). The detection of DNA - binding proteins by protein blotting. *Nucleic Acids Research*, *8*, 1 - 20.

Bray, E. A. (1993). Molecular responses to water - deficit. *Plant Physiology*, *103*, 1035 - 1040.

Bretsche.M S. (1971). Major human erythrocyte glycoprotein spans cell membrane. *Nature - New Biology*, *231*, 229 - 232.

Brower, A., DeSalle, R., & Vogler, A. (1996). Gene trees, species trees, and systematics: A cladistic perspective. *Annual Review of Ecology and Systematics*, *27*, 423 - 450.

Browne, J., Tunnacliffe, A., & Burnell, A. (2002). Anhydrobiosis - plant desiccation gene found in a nematode. *Nature*, *416*, 38 - 38.

Brule - Babel, A. L., & Fowler, D. B. (1988). Genetic control of cold hardiness and vernalization requirement in winter wheat. *Crop Science*, *28*, 879–884.

Byrne, M. E., Barley, R., Curtis, M., Arroyo, J. M., Dunham, M., Hudson, A., & Martienssen, R. A. (2000). Asymmetric leaves1 mediates leaf patterning and stem cell function in *Arabidopsis*. *Nature*, *408*, 967 - 971.

Callebaut, I., Labesse, G., Durand, P., Poupon, A., Canard, L., Chomilier, J., Henrissat, B., & Mornon, J. P. (1997). Deciphering protein sequence information through hydrophobic cluster analysis (HCA): Current status and perspectives. *Cellular and Molecular Life Sciences*, *53*, 621 - 645.

Campbell, S. A., & Close, T. J. (1997). Dehydrins: Genes, proteins, and associations with phenotypic traits. *New Phytologist*, *137*, 61 - 74.

Campoli, C., Matus - Cadiz, M. A., Pozniak, C. J., Cattivelli, L., & Fowler, D. B. (2009). Comparative expression of *cbf* genes in the Triticeae under different acclimation induction temperatures. *Molecular Genetics and Genomics*, *282*, 141 - 152.

Canella, D., Gilmour, S. J., Kuhn, L. A., & Thomashow, M. F. (2010). DNA binding by the arabidopsis CBF1 transcription factor requires the PKKP / RAGRxKFxETRHP signature sequence. *Biochimica Et Biophysica Acta - Gene Regulatory Mechanisms*, *1799*, 454 - 462.

Canella, D., Gilmour, S. J., Kuhn, L. A., & Thomashow, M. F. (2010). DNA binding by the *Arabidopsis* CBF1 transcription factor requires the PKKP / RAGRxKFxETRHP signature sequence. *Biochimica Et Biophysica Acta - Gene Regulatory Mechanisms*, *1799*, 454 - 462.

Carvallo, M. A., Pino, M., Jeknic, Z., Zou, C., Doherty, C. J., Shiu, S., Chen, T. H. H., & Thomashow, M.F. (2011). A comparison of the low temperature transcriptomes and CBF regulons of three plant species that differ in freezing tolerance: *Solanum commersonii*, *Solanum tuberosum*, and *Arabidopsis thaliana*. *Journal of Experimental Botany*, *62*, 3807 - 3819.

Cattivelli, L., Baldi, P., Crosatti, C., Di Fonzo, N., Faccioli, P., Grossi, M., Mastrangelo, A. M., Pecchioni, N., & Stanca, A. M. (2002). Chromosome regions and stress - related sequences involved in resistance to abiotic stress in Triticeae. *Plant Molecular Biology*, *48*, 649 - 665.

Ceccardi, T. L., Meyer, N. C., & Close, T. J. (1994). Purification of a maize dehydrin. *Protein Expression and Purification*, *5*, 266 - 269.

Cerdan, P. D., & Chory, J. (2003). Regulation of flowering time by light quality. *Nature*, *423*, 881 - 885.

Chauvin, L. P., Houde, M., & Sarhan, F. (1993). A leaf - specific gene stimulated by light during wheat acclimation tolow temperature. *Plant Molecular Biology*, *23*, 255 - 265.

Chen, T. H. H., & Gusta, L. V. (1983). Abscisic acid - induced freezing resistance in cultured plant - cells. *Plant Physiology*, *73*, 71 - 75.

Cheng, S. H., Willmann, M. R., Chen, H. C., & Sheen, J. (2002). Calcium signaling through protein kinases: The *Arabidopsis* calcium - dependent protein kinase gene family. *Plant Physiology*, *129*, 469 - 485.

Chew, Y. H., & Halliday, K. J. (2010). A stress - free walk from *Arabidopsis* to crops. *Current Opinon in Biotechnology*, *22*, 1 - 6.

Chinnusamy, V., Ohta, M., Kanrar, S., Lee, B. H., Hong, X. H., Agarwal, M., & Zhu, J. K. (2003). ICE1: A regulator of cold - induced transcriptome and freezing tolerance in *Arabidopsis*. *Genes and Development*, *17*, 1043 - 1054.

Chinnusamy, V., Zhu, J., & Zhu, J.K. (2006). Gene regulation during cold acclimation in plants. *Physiologia Plantarum*, 126, 52–61.

Choi, D. W., Rodriguez, E. M., & Close, T. J. (2002). Barley *Cbf3* gene identification, expression pattern, and map location. *Plant Physiology*, *129*, 1781 - 1787.

Choi, D. W., Zhu, B., & Close, T. J. (1999). The barley (*Hordeum vulgare* L.) dehydrin multigene family: Sequences, allele types, chromosome assignments, and expression characteristics of 11 *Dhn* genes of cv dicktoo. *Theoretical and Applied Genetics*, *98*, 1234 - 1247.

Choi, H. I., Hong, J. H., Ha, J. O., Kang, J. Y., & Kim, S. Y. (2000). ABFs, a family of ABA - responsive element binding factors. *Journal of Biological Chemistry*, *275*, 1723 - 1730.

Choulet, F., Wicker, T., Rustenholz, C., Paux, E., Salse, J., Leroy, P., Schlub S., Le Paslier, M. C., Magdelenat, G., Gonthier, C., Couloux, A., Budak, H., Breen, J., Pumphrey, M., Liu, S., Kong, X., Jia, J., Gut, M., Brunel, D., Anderson, J. A., Gill, B. S., Appels, R., Keller, B., & Feuillet, C. (2010). Megabase level sequencing reveals contrasted organization and evolution patterns of the wheat gene and transposable element spaces. *Plant Cell*, *22*, 1686 - 1701.

Clapham, D. E. (2003). TRP channels as cellular sensors. *Nature*, *426*(6966), 517 - 524.

Clapham, D. E. (2007). Calcium signaling. *Cell*, *131*, 1047 - 1058.

Close, T. J. (1996). Dehydrins: Emergence of a biochemical role of a family of plant dehydration proteins. *Physiologia Plantarum*, *97*, 795 - 803.

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

Colasanti, J., & Sundaresan, V. (2000). 'Florigen' enters the molecular age: Long - distance signals that cause plants to flower. *Trends in Biochemical Sciences*, *25*, 236 - 240.

Collins, N. C., Tardieu, F., & Tuberosa, R. (2008). Quantitative trait loci and crop performance under abiotic stress: Where do we stand? *Plant Physiology*, *147*, 469 - 486.

Conerly, M. L., Teves, S. S., Diolaiti, D., Ulrich, M., Eisenman, R. N., & Henikoff, S. (2010). Changes in H2A.Z occupancy and DNA methylation during B - cell lymphomagenesis. *Genome Research*, *20*, 1383 - 1390.

Cook, D., Fowler, S., Fiehn, O., & Thomashow, M. F. (2004). A prominent role for the CBF cold response pathway in configuring the low - temperature metabolome of *Arabidopsis*. *Proceedings of the National Academy of Sciences of the United States of America*, *101*, 15243 - 15248.

Crowe, J. H., Crowe, L. M., & Chapman, D. (1984). Preservation of membranes in anhydrobiotic organisms - the role of trehalose. *Science, 223*, 701 - 703.

Dahal, K., Kane, K., Gadapati, W., Webb, E., Savitch, L. V., Singh, J., Sharma, P., Sarhan, F., Longstaffe, F. J., Grodzinski, B., & Hüner, P. A. (2012). The effects of phenotypic plasticity on photosynthetic performance in winter rye, winter wheat and *Brassica napus*. *Physiologia Plantarum, 144*, 169 - 188.

Danyluk, J. (2003). *TaVRT - 1*, a putative transcription factor associated with vegetative to reproductive transition in cereals. *Plant Physiology, 132*, 1849 - 1860.

Danyluk, J., Houde, M., Rassart, E., & Sarhan, F. (1994). Differential expression of a gene encoding an acidic dehydrin in chilling sensitive and freezing tolerant gramineae species. *FEBS Letters, 344*, 20 - 24.

Danyluk, J., Perron, A., Houde, M., Limin, A., Fowler, B., Benhamou, N., & Sarhan, F. (1998). Accumulation of an acidic dehydrin in the vicinity of the plasma membrane during cold acclimation of wheat. *Plant Cell, 10*, 623 - 638.

Davey, M. P., Woodward, F. I., & Quick, W. P. (2009). Intraspecfic variation in cold - temperature metabolic phenotypes of *Arabidopsis lyrata* ssp petraea. *Metabolomics, 5*, 138 - 149.

De Bodt, S., Maere, S., & Van de Peer, Y. (2005). Genome duplication and the origin of angiosperms. *Trends in Ecology & Evolution, 20*, 591 - 597.

DeFalco, T. A., Bender, K. W., & Snedden, W. A. (2010). Breaking the code: Ca^{2+} sensors in plant signalling. *Biochemical Journal, 425*, 27 - 40.

DePauw, R.M. (2011). History of Canada western hard red spring wheat improvement in Canada. *In*: Chibbar, R. N & Dexter, J. E (Eds), *Wheat Science Dynamics: Challenges and Opportunities*, AACC International Inc. Minnesota, USA, pp. 47 - 58.

Dhillon, T., Pearce, S. P., Stockinger, E. J., Distelfeld, A., Li, C., Knox, A. K., Vashegyi, I., Vágújfalvi, A, Galiba, G., & Dubcovsky, J. (2010). Regulation of freezing tolerance and flowering in temperate cereals: The *VRN - 1* connection. *Plant Physiology, 153*, 1846 - 1858.

Dietz, K., Vogel, M. O., & Viehhauser, A. (2010). AP2 / EREBP transcription factors are part of gene regulatory networks and integrate metabolic, hormonal and environmental signals in stress acclimation and retrograde signalling. *Protoplasma, 245*, 3 - 14.

Dill, K. (1985). Theory for the folding and stability of globular - proteins. *Biochemistry, 24*, 1501 - 1509.

Dinh, T. T., Girke, T., Liu, X., Yant, L., Schmid, M., & Chen, X. (2012). The floral homeotic protein APETALA2 recognizes and acts through an AT - rich sequence element. *Development, 139*, 1978 - 1986.

Distelfeld, A., Tranquilli, G., Li, C., Yan, L., & Dubcovsky, J. (2009). Genetic and molecular characterization of the *VRN2* loci in tetraploid wheat. *Plant Physiology, 149*, 245 - 257.

Doherty, C. J., Van Buskirk, H. A., Myers, S. J., & Thomashow, M. F. (2009). Roles for *Arabidopsis CAMTA* transcription factors in cold - regulated gene expression and freezing tolerance. *Plant Cell, 21*, 972 - 984.

Dong, C. H., Agarwal, M., Zhang, Y. Y., Xie, Q., & Zhu, J. K. (2006). The negative regulator of plant cold responses, HOS1, is a RING E3 ligase that mediates the ubiquitination and degradation of ICE1. *Proceedings of the National Academy of Sciences of the United States of America*, *103*, 8281 - 8286.

Dong, M. A., Farre, E. M., & Thomashow, M. F. (2011). Circadian clock - associated 1 and late elongated hypocotyl regulate expression of the C - repeat binding factor (CBF) pathway in *Arabidopsis*. *Proceedings of the National Academy of Sciences of the United States of America*, *108*, 7241 - 7246.

Dörffling, K., Dörffling, H., & Luck, E. (2009). Improved frost tolerance and winter hardiness in proline overaccumulating winter wheat mutants obtained by *in vitro* selection is associated with increased carbohydrate, soluble protein and abscisic acid (ABA) levels. *Euphytica*, *165*, 545 - 556.

Doxey, A. C., Yaish, M. W., Griffith, M., & McConkey, B. J. (2006). Ordered surface carbons distinguish antifreeze proteins and their ice - binding regions. *Nature Biotechnology*, *24*, 852 - 855.

Drobak, B. K., Dewey, R. E., & Boss, W. F. (1999). Phosphoinositide kinases and the synthesis of polyphosphoinositides in higher plant cells. *International Review of Cytology - a Survey of Cell Biology*, *189*, 95 - 130.

Dubcovsky, J., & Dvorak, J. (2007). Genome plasticity a key factor in the success of polyploid wheat under domestication. *Science*, *316*, 1862 - 1866.

Dubcovsky, J., Lijavetzky, D., Appendino, L., & Tranquilli, G. (1998). Comparative RFLP mapping of *Triticum monococcum* genes controlling vernalization requirement. *Theoretical and Applied Genetics*, *97*, 968 - 975.

Dubcovsky, J., Luo, M. C., & Dvorak, J. (1995). Linkage relationships among stress - induced genes in wheat. *Theoretical and Applied Genetics*, *91*, 795 - 801.

Dubouzet, J. G., Sakuma, Y., Ito, Y., Kasuga, M., Dubouzet, E. G., Miura, S., Seki, M, Shinozaki, K., & Yamaguchi - Shinozaki K. (2003). *OsDREB* genes in rice, *Oryza sativa* L., encode transcription activators that function in drought - , high - salt - and cold - responsive gene expression. *Plant Journal*, *33*, 751 - 763.

Dure, L. (2001). Occurrence of a repeating 11 - mer amino acid sequence motif in diverse organisms. *Protein and Peptide Letters*, *8*, 115 - 122.

Dvorak, J., Luo, M. C., Yang, Z. L., & Zhang, H. B. (1998). The structure of the *Aegilops tauschii* genepool and the evolution of hexaploid wheat. *Theoretical and Applied Genetics*, *97*, 657 - 670.

Eckert, D., Buhl, S., Weber, S., Jager, R., & Schorle, H. (2005). The *AP - 2* family of transcription factors. *Genome Biology*, *6*, 246 - 251.

Elbein, A. D., Pan, Y. T., Pastuszak, I., & Carroll, D. (2003). New insights on trehalose: A multifunctional molecule. *Glycobiology*, *13*, 17R - 27R.

Ensminger, I., Busch, F., & Hüner, N. P. A. (2006). Photostasis and cold acclimation: Sensing low temperature through photosynthesis. *Physiologia Plantarum*, *126*, 28 - 44.

Eriksson, S., Hurme, R., & Rhen, M. (2002). Low - temperature sensors in bacteria. *Philosophical Transactions of the Royal Society of London Series B - Biological Sciences, 357*), 87 - 893.

FAO Satistical Year Book. (2012). http: / / www.fao.org / docrep / 015 / i2490e / i2490e00.htm.

Fawcett, J. A., Maere, S., & Van de Peer, Y. (2009). Plants with double genomes might have had a better chance to survive the cretaceous - tertiary extinction event. *Proceedings of the National Academy of Sciences of the United States of America*, *106*, 5737 - 5742.

Feldman, M. (2000) Origin of cultivated wheat. *In*: Bonjean, P. A., & Angus J.W. (Eds), *The world wheat book*, Springer Verlag. France, UK. pp. 3 - 56

Feldman, M., Lupton, F. G. H., & Miller, T. E. (1995). Wheats. *In*: Smartt, J., & Simmonds, N. W. (Eds), *Evolution of Crop Plants*, Longman Group Ltd., London. UK. pp. 184 - 192

Feng, W. G., & Williams, T. (2003). Cloning and characterization of the mouse AP - 2 epsilon gene: A novel family member expressed in the developing olfactory bulb. *Molecular and Cellular Neuroscience*, *24*, 460 - 475.

Flood, R. G., & Halloran, G. M. (1984). Basic development rate in spring wheat. *Agronomy Journal*, *76*, 260 - 264.

Fourrier, N., Bédard, J., Lopez - Juez, E., Barbrook, A., Bowyer, J., Jarvis, P., Warren, G., & Thorlby, G. (2008). A role for SENSITIVE TO FREEZING2 in protecting chloroplasts against freeze - induced damage in *Arabidopsis*. *Plant Journal*, *55*, 734 - 745.

Fowler, D. B. (2002). Winter wheat production manual.

Fowler, D. B. (2008). Cold acclimation threshold induction temperatures in cereals. *Crop Science*, *48*, 1147 - 1154.

Fowler, D. B. (2012). Wheat production in the high winter sress climate of the great plains of North America - An experiment in crop adaptation. *Crop Science*, *52*, 11 – 20.

Fowler, D. B., & Carles, R. J. (1979). Growth, development, and cold tolerance of fall - acclimated cereal - grains. *Crop Science*, *19*, 915 - 922.

Fowler, D. B., & Gusta, L. V. (1979). Selection for winterhardiness in wheat .1. identification of genotypic variability. *Crop Science*, *19*, 769 - 772.

Fowler, D. B., & Limin, A. E. (2004). Interactions among factors regulating phenological development and acclimation rate determine low - temperature tolerance in wheat. *Annals of Botany*, *94*, 717 - 724.

Fowler, D. B., Breton, G., Limin, A. E., Mahfoozi, S., & Sarhan, F. (2001). Photoperiod and temperature interactions regulate low - temperature - induced gene expression in barley. *Plant Physiology*, *127*, 1676 - 1681

Fowler, D. B., Chauvin, L. P., Limin, A. E., & Sarhan, F. (1996a). The regulatory role of vernalization in the expression of low - temperature - induced genes in wheat and rye. *Theoretical and Applied Genetics*, *93*, 554 - 559.

Fowler, D. B., Gusta, L. V., & Tyler, N. J. (1981). Selection for winter - hardiness in wheat .3. screening methods. *Crop Science*, *21*, 896 - 901

Fowler, D. B., Limin, A. E., & Ritchie, J. T. (1999). Low - temperature tolerance in cereals: Model and genetic interpretation. *Crop Science*, *39*, 626 - 633.

Fowler, D. B., Limin, A. E., Wang, S. Y., & Ward, R. W. (1996b). Relationship between low - temperature tolerance and vernalization response in wheat and rye. *Canadian Journal of Plant Science*, *76*, 37 - 42.

Fowler, S. & Thomashow, M. F. (2002). *Arabidopsis* transcriptome profiling indicates that multiple regulatory pathways are activated during cold acclimation in addition to the CBF cold response pathway. *The Plant Cell*, *14*, 1675 - 1690.

Fowler, S. G., Cook, D., & Thomashow, M. F. (2005). Low temperature induction of *Arabidopsis* CBF1, 2, and 3 is gated by the circadian clock. *Plant Physiology*, *137*, 961 - 968.

Francia, E., Barabaschi, D., Tondelli, A., Laido, G., Rizza, F., Stanca, A. M., Busconi, M., Foger, C., Stockinger, E. J., Pecchioni, N. (2007). Fine mapping of a *HvCBF* gene cluster at the frost resistance locus *Fr - H2* in barley. *Theoretical and Applied Genetics*, *115*, 1083 - 1091.

Francia, E., Rizza, F., Cattivelli, L., Stanca, A. M., Galiba, G., Tóth, B., Hayes, P. M., Skinner, J. S., & Pecchioni, N. (2004). Two loci on chromosome 5H determine low - temperature tolerance in a 'Nure' (winter) x 'Tremois' (spring) barley map. *Theoretical and Applied Genetics*, *108*, 670 - 680.

Franklin, K. A. (2009). Light and temperature signal crosstalk in plant development. *Current Opinion in Plant Biology*, *12*, 63 - 68.

Franklin, K. A., & Whitelam, G. C. (2007). Light - quality regulation of freezing tolerance in *Arabidopsis thaliana*. *Nature Genetics*, *39*, 1410 - 1413.

Fricano, A., Rizza, F., Faccioli, P., Pagani, D., Pavan, P., Stella, A., Rossini, L., Piffanelli, P., & Cattivelli, L. (2009). Genetic variants of *HvCbf14* are statistically associated with frost tolerance in a European germplasm collection of *Hordeum vulgare*. *Theoretical and Applied Genetics*, *119*, 1335 - 1348.

Fu, D. L., Szucs, P., Yan, L. L., Helguera, M., Skinner, J. S., von Zitzewitz, J., Hayes, P. M., & Dubcovsky, J. (2005). Large deletions within the first intron in *VRN - 1* are associated with spring growth habit in barley and wheat. *Molecular Genetics and Genomics*, *273*, 54 - 56.

Fuhrmann, M., Oertel, W., & Hegemann, P. (1999). A synthetic gene coding for the green fluorescent protein (GFP) is a versatile reporter in *Chlamydomonas reinhardtii*. *Plant Journal*, *19*, 353 - 361.

Fujita, M., Fujita, Y., Noutoshi, Y., Takahashi, F., Narusaka, Y., Yamaguchi - Shinozaki, K., & Shinozaki, K. (2006). Crosstalk between abiotic and biotic stress responses: A current view from the points of convergence in the stress signaling networks. *Current Opinon in Plant Biology*, *9*, 436 - 442.

Fukaki, H., Fujisawa, H., & Tasaka, M. (1996). Gravitropic response of inflorescence stems in *Arabidopsis thaliana*. *Plant Physiology*, *110*, 933 - 943.

Furthmay, H., & Timpl, R. (1971). Characterization of collagen peptides by sodium dodecylsulfate - polyacrylamide electrophoresis. *Analytical Biochemistry*, *41*, 510 - 516.

Gaboriaud, C., Bissery, V., Benchetrit, T., & Mornon, J. (1987). Hydrophobic cluster - analysis - an efficient new way to compare and analyze amino - acid - sequences. *FEBS Letters*, *224*, 149 - 155.

Galiba, G. (2002). Mapping of genes regulating abiotic stress tolerance in cereals. *Acta Agronomica Hungarica*, *50*, 235 - 247.

Galiba, G., Quarrie, S. A., Sutka, J., Morgounov, A., & Snape, J. W. (1995). RFLP mapping of the vernalization (*Vrn1*) and frost - resistance (*Fr1*) genes on chromosome 5A of wheat. *Theoretical and Applied Genetics*, *90*, 1174 - 1179.

Galiba, G., Vágújfalvi, A., Li, C., Soltesz, A., & Dubcovsky, J. (2009). Regulatory genes involved in the determination of frost tolerance in temperate cereals. *Plant Science*, *176*, 12 - 19.

Ganeshan, S., Denesik, T., Fowler, D. B., & Chibbar, R. N. (2009). Quantitative expression analysis of selected low temperature - induced genes in autumn - seeded wheat (*Triticum aestivum* L.) reflects changes in soil temperature. *Environmental and Experimental Botany*, *66*, 46 - 53.

Ganeshan, S., Sharma, P., Young, L., Kumar, A., Fowler, D. B., & Chibbar, R. N. (2011). Contrasting cDNA - AFLP profiles between crown and leaf tissues of cold - acclimated wheat plants indicate differing regulatory circuitries for low temperature tolerance. *Plant Molecular Biology*, *75*, 379 - 398.

Ganeshan, S., Vitamvas, P., Fowler, D. B., & Chibbar, R. N. (2008). Quantitative expression analysis of selected COR genes reveals their differential expression in leaf and crown tissues of wheat (*Triticum aestivum* L.) during an extended low temperature acclimation regimen. *Journal of Experimental Botany*, *59*, 379 - 398.

Garay - Arroyo, A., Colmenero - Flores, J., Garciarrubio, A., & Covarrubias, A. (2000). Highly hydrophilic proteins in prokaryotes and eukaryotes are common during conditions of water deficit. *Journal of Biological Chemistry*, *275*, 5668 - 5674.

Gatti - Lafranconi, P., Natalello, A., Ami, D., Doglia, S. M., & Lotti, M. (2011). Concepts and tools to exploit the potential of bacterial inclusion bodies in protein science and biotechnology. *FEBS Journal*, *278*, 2408 - 2418.

Gaut, B. S. (2002). Evolutionary dynamics of grass genomes. *New Phytologist*, 154, 15 - 28.

Gery, C., Zuther, E., Schulz, E., Legoupi, J., Chauveau, A., McKhann, H., Hincha, D. K., & Téoulé, E. (2011). Natural variation in the freezing tolerance of *Arabidopsis thaliana*: Effects of RNAi - induced CBF depletion and QTL localisation vary among accessions. *Plant Science*, *180*, 12 - 23.

Gilmour, S. J., Fowler, S. G., & Thomashow, M. F. (2004). *Arabidopsis* transcriptional activators CBF1, CBF2, and CBF3 have matching functional activities. *Plant Molecular Biology*, *54*, 767 - 781.

Gilmour, S. J., Hajela, R. K., & Thomashow, M. F. (1988). Cold - acclimation in *Arabidopsis thaliana*. *Plant Physiology*, *87*, 745 - 750.

Gilmour, S. J., Sebolt, A. M., Salazar, M. P., Everard, J. D., & Thomashow, M. F. (2000). Overexpression of the *Arabidopsis CBF3* transcriptional activator mimics multiple biochemical changes associated with cold acclimation. *Plant Physiology*, *124*, 1854 - 1865.

Gilmour, S. J., Zarka, D. G., Stockinger, E. J., Salazar, M. P., Houghton, J. M., & Thomashow, M. F. (1998). Low temperature regulation of the *Arabidopsis* CBF family of AP2 transcriptional activators as an early step in cold - induced COR gene expression. *Plant Journal*, *16*, 433 - 442.

Goldman, N., & Yang, Z. H. (1994). Codon - based model of nucleotide substitution for protein - coding DNA - sequences. *Molecular Biology and Evolution*, *11*, 725 - 736.

Gong, Z. Z., Lee, H., Xiong, L. M., Jagendorf, A., Stevenson, B., & Zhu, J. K. (2002). RNA helicase - like protein as an early regulator of transcription factors for plant chilling and freezing tolerance. *Proceedings of the National Academy of Sciences of the United States of America*, *99*, 11507 - 11512.

Gorsuch, P. A., Pandey, S., & Atkin, O. K. (2010b). Temporal heterogeneity of cold acclimation phenotypes in *Arabidopsis* leaves. *Plant Cell and Environment*, *33*, 1488 - 1498.

Gorsuch, P. A., Pandey, S., & Atkin, O. K. (2010c). Thermal de - acclimation: How permanent are leaf phenotypes when cold - acclimated plants experience warming? *Plant Cell and Environment*, *33*, 1124 - 1137.

Gorsuch, P. A., Sargeant, A. W., Penfield, S. D., Quick, W. P., & Atkin, O. K. (2010a). Systemic low temperature signaling in *Arabidopsis*. *Plant Cell Physiology*, *51*, 1488 - 1498.

Goulas, E., Schubert, M., Kieselbach, T., Kleczkowski, L. A., Gardestrom, P., Schröder, W., & Hurry, V. (2006). The chloroplast lumen and stromal proteomes of *Arabidopsis thaliana* show differential sensitivity to short - and long - term exposure to low temperature. *Plant Journal*, *47*, 720 - 734.

Gray, G. R., & Heath, D. (2005). A global reorganization of the metabolome in *Arabidops* during cold acclimation is revealed by metabolic fingerprinting. *Physiologia Plantarum*, *124*, 236 - 248.

Greenup, A. G., Sasani, S., Oliver, S. N., Walford, S. A., Millar, A. A., & Trevaskis, B. (2011). Transcriptome analysis of the vernalization response in barley (*Hordeum vulgare*) seedlings. *PLoS One*, *6*(3), e17900 - e17911.

Griffith, I. P. (1972). Effect of cross - links on mobility of proteins in dodecyl sulfate polyacrylamide gels. *Biochemical Journal*, *126*, 553 - 560.

Griffith, M., & Mcintyre, H. C. H. (1993). The interrelationship of growth and frost tolerance in winter rye. *Physiologia Plantarum*, *87*, 335 - 344.

Griffith, M., & Yaish, M. W. F. (2004). Antifreeze proteins in overwintering plants: A tale of two activities. *Trends in Plant Science*, *9*, 399 - 405.

Gulick, P., Drouin, S., Yu, Z., Danyluk, J., Poisson, G., Monroy, A., & Sarhan, F. (2005). Transcriptome comparison of winter and spring wheat responding to low temperature. *Genome*, *48*, 913 - 923.

Gulig, P. A., Danbara, H., Guiney, D. G., Lax, A. J., Norel, F., & Rhen, M. (1993). Molecular analysis of spv virulence genes of the salmonella virulence plasmids. *Molecular Microbiology*, *7*, 825 - 830.

Gundersen, G. G., & Cook, T. A. (1999). Microtubules and signal transduction. *Current Opinion in Cell Biology*, *11*, 81 - 94.

Guo, J., Duan, R., Hu, X., Li, K., & Fu, S. (2010). *Isopentenyl transferase* gene (*ipt*) downstream transcriptionally fused with gene expression improves the growth of transgenic plants. *Transgenic Research*, *19*, 197 - 209.

Guo, Y., Xiong, L., Ishitani, M., & Zhu, J. K. (2002). An *Arabidopsis* mutation in translation elongation factor 2 causes superinduction of *CBF / DREB1* transcription factor genes but blocks the induction of their downstream targets under low temperatures. *Proceedings of the National Academy of Sciences of the United States of America*, *99*, 7786 - 7791.

Gusta, L. V., & Fowler, D. B. (1976). Effects of temperature on dehardening and rehardening of winter cereals. *Canadian Journal of Plant Science*, *56*, 673 - 678.

Gusta, L. V., & Wisniewski, M. (2013). Understanding plant cold hardiness: An opinion. *Physiologia Plantarum*, *147*, 4 - 14.

Gustafsson, C., Govindarajan, S., & Minshull, J. (2004). Codon bias and heterologous protein expression. *Trends in Biotechnology*, *22*, 346 - 353.

Guy, C., Kaplan, F., Kopka, J., Selbig, J., & Hincha, D. K. (2008). Metabolomics of temperature stress. *Physiologia Plantarum*, *132*, 220 - 235.

Haake, V., Cook, D., Riechmann, J. L., Pineda, O., Thomashow, M. F., & Zhang, J. Z. (2002). Transcription factor *CBF4* is a regulator of drought adaptation in *Arabidopsis*. *Plant Physiology*, *130*, 639 - 648.

Hamana, K. (1981). Gel - electrophoresis and gel chromatography of calf thymus - histones in the presence of a cationic surfactant. *Journal of Biochemistry*, *90*, 1591 - 1596.

Hanin, M., Brini, F., Ebel, C., Toda, Y., Takeda, S., & Masmoudi, K. (2011). Plant dehydrins and stress tolerance: Versatile proteins for complex mechanisms. *Plant Signaling and Behavior*, *6*, 1503 - 1509.

Hannah, M. A., Wiese, D., Freund, S., Fiehn, O., Heyer, A. G., & Hincha, D. K. (2006). Natural genetic variation of freezing tolerance in *Arabidopsis*. *Plant Physiology*, *142*, 98 - 112.

Hao, D., Yamasaki, K., Sarai, A., & Ohme - Takagi, M. (2002). Determinants in the sequence specific binding of two plant transcription factors, CBF1 and NtERF2, to the DRE and GCC motifs. *Biochemistry*, *41*, 4202 - 4208.

Harrison, M. J., Lawton, M. A., Lamb, C. J., & Dixon, R. A. (1991). Characterization of a nuclear - protein that binds to three elements within the silencer region of a bean chalcone synthase gene promoter. *Proceedings of the National Academy of Sciences of the United States of America*, *88*, 2515 - 2519.

Havaux, M., & Kloppstech, K. (2001). The protective functions of carotenoid and flavonoid pigments against excess visible radiation at chilling temperature investigated in *Arabidopsis* npq and tt mutants. *Planta*, *213*, 953 - 966.

Hawkins, G., Deng, Z., Kubik, T., & Johnson - Flanagan, A. (2002). Characterization of freezing tolerance and vernalization in *vern* - , a spring - type *Brassica napus* line derived from a winter cross. *Planta*, 216, 220 - 226.

Hayama, R., & Coupland, G. (2003). Shedding light on the circadian clock and the photoperiodic control of flowering. *Current Opinion in Plant Biology*, *6*, 13 - 19.

Heino, P., Sandman, G., Lang, V., Nordin, K., & Palva, E. T. (1990). Abscisic - acid deficiency prevents development of freezing tolerance in *Arabidopsis thaliana* (L) heynh. *Theoretical and Applied Genetics*, *79*, 801 - 806.

Henderson, I. R., & Dean, C. (2004). Control of *Arabidopsis* flowering: The chill before the bloom. *Development*, *131*, 3829 - 3838.

Herman, E. M., Rotter, K., Premakumar, R., Elwinger, G., Bae, R., Ehler - King, L., Chen, & Livingston III, D. P. (2006). Additional freeze hardiness in wheat acquired by exposure to - 3 degrees C is associated with extensive physiological, morphological, and molecular changes. *Journal of Experimental Botany*, *57*, 3601 - 3618.

Hilger - Eversheim, K., Moser, M., Schorle, H., & Buettner, R. (2000). Regulatory roles of AP – 2 transcription factors in vertebrate development, apoptosis and cell - cycle control. *Gene*, *260*, 1 - 12.

Hodges, C., Bintu, L., Lubkowska, L., Kashlev, M., & Bustamante, C. (2009). Nucleosomal fluctuations govern the transcription dynamics of RNA polymerase II. *Science*, *325*, 626–628.

Hoekstra, F. A., & Golovina, E. A. (2002). The role of amphiphiles. *Comparative Biochemistry and Physiology. Part A, Molecular and Integrative Physiology*, *131*, 527 - 533.

Holder, M., & Lewis, P. O. (2003). Phylogeny estimation: Traditional and bayesian approaches. *Nature Reviews Genetics*, *4*, 275 - 284.

Hon, W. C., Griffith, M., Mlynarz, A., Kwok, Y. C., & Yang, D. S. C. (1995). Antifreeze proteins in winter rye are similar to pathogenesis - related proteins. *Plant Physiology*, *109*, 879 - 889.

Hoogendoorn, J. (1985). A reciprocal F1 monosomic analysis of the genetic - control of time of ear emergence, number of leaves and number of spikelets in wheat (*Triticum aestivum* L). *Euphytica*, *34*, 545 - 558.

Hosoi, T., Uchiyama, M., Okumura, E., Saito, T., Ishiguro, K., Uchida, T., Okuyama, A., Kishimoto, T., & Hisanaga, S. (1995). Evidence for cdk5 as a major activity phosphorylating tau protein in porcine brain extract. *Journal of Biochemistry*, *117*, 741 - 749.

Houde, M., Daniel, C., Lachapelle, M., Allard, F., Laliberte, S., & Sarhan, F. (1995). Immunolocalization of freezing - tolerance - associated proteins in the cytoplasm and nucleoplasm of wheat crown tissues. *Plant Journal*, *8*, 583 - 593.

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

Houde, M., Dhindsa, R. S., & Sarhan, F. (1992b). A molecular marker to select for freezing tolerance in gramineae. *Molecular and General Genetics*, *234*, 43 - 48.

Hrabak, E. M., Chan, C. W., Gribskov, M., Harper, J. F., Choi, J. H., Halford, N., Kudla, J., Luan, S., Nimmo, H. G., Sussman, M. R., Thomas, M., Walker - Simmons, K., Zhu, J. K., & Harmon, A. C. (2003). The *Arabidopsis* CDPK - SnRK superfamily of protein kinases. *Plant Physiology*. *132*, 666 - 680.
http: / / www.usask.ca / agriculture / plantsci / winter_cereals / index.php

Huang, G. T., Ma, S. L., Bai, L. P., Zhang, L., Ma, H., Jia, P., Liu, J., Zhong, M., & Guo, Z. F. (2012). Signal transduction during cold, salt, and drought stresses in plants. *Molecular Biology Reports 3*, *39*, 969 - 987.

Hughes, M. A., & Dunn, M. A. (1996). The molecular biology of plant acclimation to low temperature. *Journal of Experimental Botany*, *47*, 291 - 305.

Huner, N. P. A., Oquist, G., Hurry, V. M., Krol, M., Falk, S., & Griffith, M. (1993). Photosynthesis, photoinhibition and low - temperature acclimation in cold tolerant plants. *Photosynthesis Research*, *37*, 19 - 39.

Hunter, T. (2007). The age of crosstalk: Phosphorylation, ubiquitination, and beyond. *Molecular Cell*, *28*, 730 - 738.

Iordachescu, M., & Imai, R. (2008). Trehalose biosynthesis in response to abiotic stresses. *Journal of Integrative Plant Biology*, *50*, 1223 - 1229.

Ishikawa, M. (1984). Deep supercooling in most tissues of wintering *Sasa senanensis* and its mechanism in leaf blade tissues. *Plant Physiology*, *75*, 196 - 202.

Ivanov, A. G., Hendrickson, L., Krol, M., Selstam, E., Öquist, G., Hurry, V., & Huner, P.A. (2006). Digalactosyl - diacylglycerol deficiency impairs the capacity for photosynthetic intersystem electron transport and state transitions in *Arabidopsis thaliana* due to photosystem I acceptor side limitations. *Plant and Cell Physiology*, *47*, 1146 - 1157.

Ivanov, A. G., Sane, P. V., Zeinalov, Y., Malmberg, G., Gardeström, P., Huner, N. P. A., & Öquist, G. (2001). Photosynthetic electron transport adjustments in overwintering scots pine *(Pinus sylvestris* L.). *Planta*, *213*, 575 - 585.

Ivany, L. C., Patterson, W. P., & Lohmann, K. C. (2000). Cooler winters as a possible cause of mass extinctions at the eocene / oligocene boundary. *Nature*, *407*, 887 - 890.

Jack, R. S. (1990). An unusually stable DNA - binding protein can locate its specific binding - site in the presence of high - concentrations of urea. *Biochemical and Biophysical Research Communications*, *169*, 840 - 845.

Jaglo, K. R., Kleff, S., Amundsen, K. L., Zhang, X., Haake, V., Zhang, J. Z., Deits, T., & Thomashow, M. F. (2001). Components of the *Arabidopsis* C - Repeat / Dehydration - responsive element binding factor cold - response pathway are conserved in *Brassica napus* and other plant species. *Plant Physiology*, *127*, 910 - 917.

Jaglo - Ottosen, K. R., Gilmour, S. J., Zarka, D. G., Schabenberger, O., & Thomashow, M. F. (1998). *Arabidopsis* CBF1 overexpression induces COR genes and enhances freezing tolerance. *Science*, *280*, 104 - 106.

Janská, A., Maršík, P., Zelenková, S., & Ovesná, J. (2010). Cold stress and acclimation - what is important for metabolic adjustment? *Plant Biology*, *12*, 395 - 405.

Jeon, J., Kim, N. Y., Kim, S., Kang, N. Y., Novák, O., Ku, S., Cho, C., Lee, D. J., Lee, E., Strnad, M & Kim, J. (2010). A subset of cytokinin two - component signaling system plays a role in cold temperature stress response in *Arabidopsis*. *Journal of Biological Chemistry*, *285*, 23371 - 23386.

Jiang, C., Iu, B., & Singh, J. (1996). Requirement of a CCGAC cis - acting element for cold induction of the BN115 gene from winter *Brassica napus*. *Plant Molecular Biology*, *30*, 679 - 684.

Jiang, W., Jin, Y., Lee, J., Lee, K., Piao, R., Han, L., Shin, J., Jin, R., Cao, T, Pan, H., Du, X., & Koh, H. (2011). Quantitative trait loci for cold tolerance of rice recombinant inbred lines in low temperature environments. *Molecules and Cells*, *32*, 579 - 587.

Jin, H., & Martin, C. (1999) Multifunctionality and diversity within the plant MYB - gene family. *Plant Molecular Biology*, *41*, 577 - 585.

Jofuku, K. D., Denboer, B. G. W., Vanmontagu, M., & Okamuro, J. K. (1994). Control of *Arabidopsis* flower and seed development by the homeotic gene *Apetala2*. *Plant Cell*, *6*, 1211 - 1225.

Jurczyk, B., Rapacz, M., Budzisz, K., Barcik, W., & Sasal, M. (2012). The effects of cold, light and time of day during low - temperature shift on the expression of *CBF6*, *FpCor14b* and *LOS2* in *Festuca pratensis*. *Plant Science*, *183*, 143 - 148.

Kagale, S., Divi, U. K., Krochko, J. E., Keller, W. A., & Krishna, P. (2007). Brassinosteroid confers tolerance in *Arabidopsis thaliana* and *Brassica napus* to a range of abiotic stresses. *Planta*, *225*, 353 - 364.

Kagaya, Y., Ohmiya, K., & Hattori, T. (1999). RAV1, a novel DNA - binding protein, binds to bipartite recognition sequence through two distinct DNA - binding domains uniquely found in higher plants. *Nucleic Acids Research*, *27*, 470 - 478.

Kalberer, S. R., Wisniewski, M., & Arora, R. (2006). Deacclimation and reacclimation of cold – hardy plants: Current understanding and emerging concepts. *Plant Science*, *171*, 3 - 16.

Kanaya, E., Nakajima, N., Morikawa, K., Okada, K., & Shimura, Y. (1999). Characterization of the transcriptional activator CBF1 from *Arabidopsis thaliana* evidence for cold denaturation in regions outside of the DNA binding domain. *Journal of Biological Chemistry*, *274*, 16068 - 16076.

Kandror, O., DeLeon, A., & Goldberg, A. L. (2002). Trehalose synthesis is induced upon exposure of *Escherichia coli* to cold and is essential for viability at low temperatures. *Proceedings of the National Academy of Sciences of the United States of America*, *99*, 9727 - 9732.

Kane, J. F. (1995). Effects of rare codon clusters on high - level expression of heterologous proteins in *Escherichia coli*. *Current Opinion in Biotechnology*, *6*, 494 - 500.

Kang, H., Kim, J., Kim, B., Jeong, H., Choi, S. H., Kim, E. K., Lee, H. Y., & Lim, P. O. (2011). Overexpression of *FTL1* / *DDF1*, an AP2 transcription factor, enhances tolerance to cold, drought, and heat stresses in *Arabidopsis thaliana*. *Plant Science*, *180*, 634 - 641.

Kang, J. Y., Choi, H. I., Im, M. Y., & Kim, S. Y. (2002). *Arabidopsis* basic leucine zipper proteins that mediate stress - responsive abscisic acid signaling. *Plant Cell*, *14*, 343 - 357.

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

Kaplan, F., Kopka, J., Haskell, D. W., Zhao, W., Schiller, K. C., Gatzke, N., Sung, D. Y., & Guy, C. L. (2004). Exploring the temperature - stress metabolome of *Arabidopsis*. *Plant Physiology*, *136*, 4159 - 4168.

Kaplan, F., Kopka, J., Sung, D. Y., Zhao, W., Popp, M., Porat, R., & Guy, C. L. (2007). Transcript and metabolite profiling during cold acclimation of *Arabidopsis* reveals an intricate relationship of cold - regulated gene expression with modifications in metabolite content. *Plant Journal*, *50*, 967 - 981.

Karimzadeh, G., Mahfoozi, S., Ghanati, F., & Javadian, N. (2010). Cold - induced changes of enzymes, proline, carbohydrates, and chlorophyll in wheat. *Russian Journal of Plant Physiology*, *57*, 540 - 547.

Kariola, T., Brader, G., Helenius, E., Li, J., Heino, P., & Palva, E. T. (2006). EARLY RESPONSIVE TO DEHYDRATION 15, a negative regulator of abscisic acid responses in *Arabidopsis*. *Plant Physiology*, *142*, 1559 - 1573.

Karlson, D., & Imai, R. (2003). Conservation of the cold shock domain protein family in plants. *Plant Physiology*, *131*, 12 - 15.

Karlson, D., Nakaminami, K., Toyomasu, T., & Imai, R. (2002). A cold - regulated nucleic acid - binding protein of winter wheat shares a domain with bacterial cold shock proteins. *Journal of Biological Chemistry*, *277*, 35248 - 35256.

Kasuga, J., Hashidoko, Y., Nishioka, A., Yoshiba, M., Arakawa, K., & Fujikawa, S. (2008). Deep supercooling xylem parenchyma cells of katsura tree (*Cercidiphyllum japonicum*) contain flavonol glycosides exhibiting high anti - ice nucleation activity. *Plant Cell and Environment*, *31*, 1335 - 1348.

Kelly, S. M., Jess, T. J., & Price, N. C. (2005). How to study proteins by circular dichroism. *Biochimica et Biophysica Acta*, *1751*, 119 - 139.

Kidokoro, S., Maruyama, K., Nakashima, K., Imura, Y., Narusaka, Y., Shinwari, Z. K., Osakabe, Y., Fujita, Y., Mizoi, J., Shinzaki, K., & Yamaguchi - Shinozaki, K. (2009). The phytochrome - interacting factor PIF7 negatively regulates DREB1 expression under circadian control in *Arabidopsis*. *Plant Physiology*, *151*, 2046 - 2057.

Kiefhaber, T., Rudolph, R., Kohler, H. H., & Buchner, J. (1991). Protein aggregation *in vitro* and *in vivo*: a quantitative model of the kinetic competition between folding and aggregation. *Biotechnology*, *9*, 825 - 829.

Kim, J. B., Kang, J. Y., & Kim, S. Y. (2004). Over - expression of a transcription factor regulating ABA - responsive gene expression confers multiple stress tolerance. *Plant Biotechnology Journal*, *2*, 459 - 466.

Kim, S., Kang, J. Y., Cho, D. I., Park, J. H., & Kim, S. Y. (2004). ABF2, an ABRE - binding bZIP factor, is an essential component of glucose signaling and its overexpression affects multiple stress tolerance. *Plant Journal*, *40*, 75 - 87.

Kingsley, P. D., & Palis, J. (1994). Grp2 proteins contain both CCHC zinc fingers and a cold shock domain. *Plant Cell*, *6*, 1522 - 1523.

Klumpp, S., Dong, J., & Hwa, T. (2012). On ribosome load, codon bias and protein abundance. *Plos One*, *7*, e48542 - e48554.

Knox, A. K., Dhillon, T., Cheng, H., Tondelli, A., Pecchioni, N., & Stockinger, E. J. (2010). CBF gene copy number variation at *frost resistance - 2* is associated with levels of freezing tolerance in temperate - climate cereals. *Theoretical and Applied Genetics*, *121*, 221 - 35.

Knox, A. K., Li, C., Vágújfalvi, A., Galiba, G., Stockinger, E. J., & Dubcovsky, J. (2008). Identification of candidate CBF genes for the frost tolerance locus *Fr - A^m2* in *Triticum monococcum*. *Plant Molecular Biology*, *67*, 257 - 270.

Ko, R., Smith, L. T., & Smith, G. M. (1994). Glycine betaine confers enhanced osmotolerance and cryotolerance on listeria - monocytogenes. *Journal of Bacteriology*, *176*, 426 - 431.

Kobayashi, F., Takumi, S., Nakata, M., Ohno, R., Nakamura, T., & Nakamura, C. (2004). Comparative study of the expression profiles of the *Cor* / *Lea* gene family in two wheat cultivars with contrasting levels of freezing tolerance. *Physiologia Plantarum*, *120*, 585 - 594.

Kočova, M., Holá, D., Wilhelmová, N., & Rothová, O. (2009). The influence of low - temperature on the photochemical activity of chloroplasts and activity of antioxidant enzymes in maize leaves. *Biologia Plantarum*, *53*, 475 - 483.

Kocsy, G., Athmer, B., Perovic, D., Himmelbach, A., Szücs, A., Vashegyi, I., Schweizer, P., Galiba, G., & Stein, N. (2010). Regulation of gene expression by chromosome 5A during cold hardening in wheat. *Molecular Genetics and Genomics*, *283*, 351 - 363.

Koike, M., Kato, H., & Imai, R. (2008). Diversification of the barley and wheat *blt101* / *wpi6* promoters by the xumet element without affecting stress responsiveness. *Molecular Genetics and Genomics*, *280*, 41 - 47.

Koornneef, M., Jorna, M. L., Derswan, D. L. C. B., & Karssen, C. M. (1982). The isolation of abscisic - acid (ABA) deficient mutants by selection of induced revertants in non - germinating gibberellin sensitive lines of *Arabidopsis thaliana* (L) heynh. *Theoretical and Applied Genetics*, *61*, 385 - 393.

Kopetzki, E., Schumacher, G., & Buckel, P. (1989). Control of formation of active soluble or inactive insoluble bakers - yeast alpha - glucosidase - pi in *Escherichia coli* by induction and growth - conditions. *Molecular and General Genetics*, *216*, 149 - 155.

Kosová, K., Prášil, I. T., Vitámvás, P., Dobrev, P., Motyka, V., Floková, K., Novák, O, Turečková, V., Rolčik, J., Pešek, B., Trávničková, A., Gaudinová, A., Galiba, G., Janda, T.,

Vlasáková, E., Prášilová, P., & Vanková, R. (2012). Complex phytohormone responses during the cold acclimation of two wheat cultivars differing in cold tolerance, winter Samanta and spring Sandra. *Journal of Plant Physiology*, *169*, 567 - 576.

Kranz, H.D., Denekamp, M., Greco, R., Jin, H., Leyva, A., Meissner, R., Petroni, K., Urzainqui, A., Bevan, M., Martin, C., Smeekens, S., Tonelli, C., Paz - Ares, J., & Weisshaar, B. (1998). Towards functional characterisation of members of the R2R3 - MYB gene family from *Arabidopsis thaliana*. *Plant Journal*, *16*, 263 - 276.

Krell, P. J. (1996). Passage effect of virus infection in insect cells. *Cytotechnology*, *20*, 125 - 137.

Kreps, J. A., Wu, Y. J., Chang, H. S., Zhu, T., Wang, X., & Harper, J. F. (2002). Transcriptome changes for *Arabidopsis* in response to salt, osmotic, and cold stress. *Plant Physiology*, *130*, 2129 - 2141.

Krivosheeva, A., Tao, D. L., Ottander, C., Wingsle, G., Dube, S. L., & Öquist, G. (1996). Cold acclimation and photoinhibition of photosynthesis in scots pine. *Planta*, *200*, 296 - 305.

Krizek, B. A. (2003). AINTEGUMENTA utilizes a mode of DNA recognition distinct from that used by proteins containing a single AP2 domain. *Nucleic Acids Research*, *31*, 1859 - 1868.

Krol, M., Hurry, V.M., Maxwell, D.P., Malek, L., Ivanov, A. G. & Huner, N. P. A. (2002). Low growth temperature inhibition of photosynthesis in cotyledons of jack pine seedlings (*Pinus banksianna*) is due to impaired chloroplast development. *Candian Journal of Botany*, *80*, 1042 – 1051.

Kudla, G., Murray, A. W., Tollervey, D., & Plotkin, J. B. (2009). Coding - sequence determinants of gene expression in *Escherichia coli*. *Science*, *324*, 255 - 258.

Kumar, S. V., & Wigge, P. A. (2010). H2A.Z - containing nucleosomes mediate the thermosensory response in *Arabidopsis*. *Cell*, *140*, 136 - 147.

Kume, S., Kobayashi, F., Ishibashi, M., Ohno, R., Nakamura, C., & Takumi, S. (2005). Differential and coordinated expression of *cbf* and *cor* / *lea* genes during long - term cold acclimation in two wheat cultivars showing distinct levels of freezing tolerance. *Genes and Genetic Systems*, *80*, 185 - 197.

Kurkela, S., & Franck, M. (1990). Cloning and characterization of a cold - inducible and ABA - inducible *Arabidopsis* gene. *Plant Molecular Biology*, *15*, 137 - 144.

Kurland, C., & Gallant, J. (1996). Errors of heterologous protein expression. *Current Opinion in Biotechnology*, *7*, 489 - 493.

Lai, E. C. (2003). RNA sensors and riboswitches: Self - regulating messages. *Current Biology*, *13*, R285 - R291.

Lang, V., Mantyla, E., Welin, B., Sundberg, B., & Palva, E. T. (1994). Alterations in water status, endogenous abscisic - acid content, and expression of *Rab18* gene during the development of freezing tolerance in *Arabidopsis thaliana*. *Plant Physiology*, *104*, 1341 - 1349.

Larcher, W. (2003). Plant under stress. *In*: Larcher,W. (Ed) *Physiological plant ecology*, Springer Verlag, Berlin, Germany, pp 345 - 348

Latorre, R., Brauchi, S., Madrid, R., & Orio, P. (2011). A cool channel in cold transduction. *Physiology*, *26*, 273 - 285.

Laudencia - Chingcuanco, D., Ganeshan, S., You, F., Fowler, B., Chibbar, R., & Anderson, O. (2011). Genome - wide gene expression analysis supports a developmental model of low temperature tolerance gene regulation in wheat (*Triticum aestivum* L.). *BMC Genomics*, *12*, 299 - 317.

Law, C. N., Sutka, J., & Worland, A. J. (1978). Genetic study of day - length response in wheat. *Heredity*, *41*, 185 - 191.

Lee, H. J., Xiong, L. M., Gong, Z. Z., Ishitani, M., Stevenson, B., & Zhu, J. K. (2001). The *Arabidopsis* HOS1 gene negatively regulates cold signal transduction and encodes a RING finger protein that displays cold - regulated nucleo - cytoplasmic partitioning. *Genes and Development*, *15*, 912 - 924.

Lee, H., Guo, Y., Ohta, M., Xiong, L. M., Stevenson, B., & Zhu, J. K. (2002). LOS2, a genetic locus required for cold - responsive gene transcription encodes a bi - functional enolase. *EMBO Journal*, *21*, 2692 - 2702.

Lemeslevarloot, L., Henrissat, B., Gaboriaud, C., Bissery, V., Morgat, A., & Mornon, J. P. (1990). Hydrophobic cluster - analysis: procedures to derive structural and functional information from 2 - D - representation of protein sequences. *Biochimie*, *72*, 555 - 574.

Li, C., Rudi, H., Stockinger, E. J., Cheng, H., Cao, M., Fox, S. E., Mockler, T. C., Westereng, B., Fjellheim, S., Rognli, O. A., & Sandve, S. R. (2012). Comparative analyses reveal potential uses of *Brachypodium distachyon* as a model for cold stress responses in temperate grasses. *BMC Plant Biology*, *12*, 65 - 79.

Li, Q., & Yang, H. (2007). Cryptochrome signaling in plants. *Photochemistry and Photobiology*, *83*, 94 - 101.

Li, W. L., Zhang, P., Fellers, J. P., Friebe, B., & Gill, B. S. (2004). Sequence composition, organization, and evolution of the core triticeae genome. *Plant Journal*, *40*, 500 - 511.

Li, Y., Böck, A., Haseneyer, G., Korzun, V., Wilde, P., & Schön, C. C., Ankerst, D. P., & baure, E. (2011). Association analysis of frost tolerance in rye using candidate genes and phenotypic data from controlled, semi - controlled, and field phenotyping platforms. *BMC Plant Biology*, *11*, 146 - 160.

Licausi, F., Giorgi, F. M., Zenoni, S., Osti, F., Pezzotti, M., & Perata, P. (2010). Genomic and transcriptomic analysis of the AP2 / ERF superfamily in *Vitis vinifera*. *BMC Genomics*, *11*, 719 - 724.

Limin, A. E., & Fowler, D. B. (2002). Developmental traits affecting low - temperature tolerance response in near - isogenic lines for the vernalization locus *vrn - A1* in wheat (*Triticum aestivum* L. em thell). *Annals of Botany*, *89*, 579 - 585.

Limin, A. E., & Fowler, D. B. (2006). Low - temperature tolerance and genetic potential in wheat (*Triticum aestivum* L.): Response to photoperiod, vernalization, and plant development. *Planta*, *224*, 360 - 366.

Limin, A. E., Houde, M., Chauvin, L. P., Fowler, D. B., & Sarhan, F. (1995). Expression of the cold - induced wheat gene *wcs120* and its homologs in related species and interspecific combinations. *Genome*, *38*, 1023 - 1031.

Lin, Y., Hwang, S., Hsu, P., Chiang, Y., Huang, C., Wang, C., & Lin, T. (2008). Molecular population genetics and gene expression analysis of duplicated CBF genes of *Arabidopsis thaliana*. *BMC Plant Biology*, *8*, 111 - 117.

Liu, Q., Kasuga, M., Sakuma, Y., Abe, H., Miura, S., Yamaguchi - Shinozaki, K., & Shinozaki, K. (1998). Two transcription factors, DREB1 and DREB2, with an EREBP / AP2 DNA binding domain separate two cellular signal transduction pathways in drought - and low - temperature - responsive gene expression, respectively, in *Arabidopsis*. *Plant Cell*, *10*, 1391 - 1406.

Livingston, D. P. (1996). The second phase of cold hardening: Freezing tolerance and fructan isomer changes in winter cereal crowns. *Crop Science*, *36*, 1568 - 1573.

Llorente, F., Oliveros, J. C., Martinez - Zapater, J. M., & Salinas, J. (2000). A freezing - sensitive mutant of *Arabidopsis*, *frs1*, is a new *aba3* allele. *Planta*, *211*, 648 - 655.

Loukoianov, A., Yan, L., Blechl, A., Sanchez, A., & Dubcovsky, J. (2005). Regulation of VRN - 1 vernalization genes in normal and transgenic polyploid wheat. *Plant Physiology*, *138*, 2364 - 2373.

Maeda, K., & Imae, Y. (1979). Thermosensory transduction in *Escherichia coli* - inhibition of the thermoresponse by L - serine. *Proceedings of the National Academy of Sciences of the United States of America*, *76*, 91 - 95.

Magnani, E., Sjolander, K., & Hake, S. (2004). From endonucleases to transcription factors: Evolution of the AP2 DNA binding domain in plants. *Plant Cell*, *16*, 2265 - 2277.

Magome, H., Yamaguchi, S., Hanada, A., Kamiya, Y., & Oda, K. (2004). Dwarf and delayed - flowering 1, a novel *Arabidopsis* mutant deficient in gibberellin biosynthesis because of overexpression of a putative AP2 transcription factor. *Plant Journal*, *37*, 720 - 729.

Magome, H., Yamaguchi, S., Hanada, A., Kamiya, Y., & Oda, K. (2008). The DDF1 transcriptional activator upregulates expression of a gibberellin - deactivating gene, GA2ox7, under high - salinity stress in *Arabidopsis*. *Plant Journal*, *56*, 613 - 626.

Mahajan, S., & Tuteja, N. (2005). Cold, salinity and drought stresses: An overview. *Archives of Biochemistry and Biophysics*, *444*, 139 - 158.

Mahfoozi, S., Limin, A. E., & Fowler, D. B. (2001). Influence of vernalization and photoperiod responses on cold hardiness in winter cereals. *Crop Science*, *41*, 1006 - 1011.

Mahfoozi, S., Limin, A. E., Hayes, P. M., Hucl, P., & Fowler, D. B. (2000). Influence of photoperiod response on the expression of cold hardiness in wheat and barley. *Canadian Journal of Plant Science*, *80*, 721 - 724.

Manning, G., Plowman, G., Hunter, T., & Sudarsanam, S. (2002). Evolution of protein kinase signaling from yeast to man. *Trends in Biochemical Sciences*, *27*, 514 - 520.

Mantyla, E., Lang, V., & Palva, E. (1995). Role of abscisic - acid in drought - induced freezing tolerance, cold - acclimation, and accumulation of LT178 and RAB18 proteins in *Arabidopsis thaliana*. *Plant Physiology*, *107*, 141 - 148.

Mao, D., & Chen, C. (2012). Colinearity and similar expression pattern of rice DREB1s reveal their functional conservation in the cold - responsive pathway. *PloS One*, *7*, e47275 - e47275.

Marston, F. A. O. (1986). The purification of eukaryotic polypeptides synthesized in *Escherichia - coli*. *Biochemical Journal*, *240*, 1 - 12.

Maruyama, K., Sakuma, Y., Kasuga, M., Ito, Y., Seki, M., Goda, H., Shimada, Y., Yoshida, S., Shinozki, K & Yamaguchi - Shinozaki, K. (2004). Identification of cold - inducible downstream genes of the *Arabidopsis* DREB1A / CBF3 transcriptional factor using two microarray systems. *Plant Journal*, *38*, 982 - 993.

Maruyama, K., Todaka, D., Mizoi, J., Yoshida, T., Kidokoro, S., Matsukura, S., Takasaki, H., Sakurai, T., Yamamoto, Y. Y., Yoshiwara, K., Kojima, M., Sakakibara, H., Shinozaki, K., & Yamaguchi - Shinozaki, K. (2012). Identification of cis - acting promoter elements in cold - and dehydration - induced transcriptional pathways in Arabidopsis, rice, and soybean. *DNA Research*, *19*, 37 - 49.

Mazzucotelli, E., Tartari, A., Guerra, D., Cattivelli, L., & Forlani, G. (2006). Metabolism of gamma - aminobutyric acid during cold acclimation and freezing and its relationship to frost tolerance in barley and wheat. *Journal of Experimental Botany 57, 3755 - 3766.*

McKemy, D. D., Neuhausser, W. M., & Julius, D. (2002). Identification of a cold receptor reveals a general role for TRP channels in thermosensation. *Nature*, *416*, 52 - 58.

Médigue, C., Krin, E., Pascal, G., Barbe, V., Bernsel, A., Bertin, P., Cheung, F., Cruveiller, S, D'Amico, S., Duilio, A., fang, G., Feller, G., Ho, C., Mangenot, S., Marino, G., Nilsson, J., Parrilli, E., Rocha, E. P., Rouy, Z., Sekowska, A., Tutino, M. L., Vallenet, D., von Heijne, G., & Danchin, A. (2005). Coping with cold: The genome of the versatile marine antarctica bacterium *Pseudoalteromonas haloplanktis* TAC125. *Genome Research*, *15*, 1325 - 1335.

Medina, J., Bargues, M., Terol, J., Perez - Alonso, M., & Salinas, J. (1999). The *Arabidopsis CBF* gene family is composed of three genes encoding AP2 domain - containing proteins whose expression is regulated by low temperature but not by abscisic acid or dehydration. *Plant Physiology*, *119*, 463 - 469.

Medina, J., Catala, R., & Salinas, J. (2011). The CBFs: Three *Arabidopsis* transcription factors to cold acclimate. *Plant Science*, *180*, 3 - 11.

Meyer, K., Keil, M., & Naldrett, M. J. (1999). A leucine - rich repeat protein of carrot that exhibits antifreeze activity. *FEBS Letters*, *447*, 171 - 178.

Miller, A. K., Galiba, G., & Dubcovsky, J. (2006). A cluster of 11 CBF transcription factors is located at the frost tolerance locus *Fr - A^m2* in *Triticum monococcum*. *Molecular Genetics and Genomics*, *275*, 193 - 203.

Mitchell, P. J., Wang, C., & Tjian, R. (1987). Positive and negative regulation of transcription in vitro enhancer - binding protein AP - 2 is inhibited by SV40 T antigen. *Cell*, *50*, 847 - 861.

Miura, H., & Worland, A. J. (1994). Genetic - control of vernalization, day - length response, and earliness per se by homoeologous group - 3 chromosomes in wheat. *Plant Breeding*, *113*, 160 - 169.

Miura, K., Jin, J. B., Lee, J., Yoo, C. Y., Stirm, V., Miura, T., *et al*. (2007). SIZ1 - mediated sumoylation of ICE1 controls CBF3 / DREB1A expression and freezing tolerance in *Arabidopsis*. *Plant Cell*, *19*, 1403 - 1414.

Mizoguchi, T., Gotoh, Y., Nishida, E., Yamaguchi - Shinozaki, K., Hayashida, N., Iwasaki, T., Kamada, H., & Shinozaki, K. (1994). Characterization of two cDNAs that encode map kinase homologs in *Arabidopsis thaliana* and analysis of the possible role of auxin in activating such kinase - activities in cultured - cells. *Plant Journal*, *5*, 111 - 122.

Mizoi, J., Shinozaki, K., & Yamaguchi - Shinozaki, K. (2012). AP2 / ERF family transcription factors in plant abiotic stress responses. *Biochimica et Biophysica Acta - Gene Regulatory Mechanisms*, *1819*, 86 - 96.

Mizuno, T., & Imae, Y. (1984). Conditional inversion of the thermoresponse in *Escherichia coli*. *Journal of Bacteriology*, *159*, 360 - 367.

Moellering, E. R., Muthan, B., & Benning, C. (2010). Freezing tolerance in plants requires lipid remodeling at the outer chloroplast membrane. *Science*, *330*, 226 - 228.

Mohseni, S., Che, H., Djillali, Z., Dumont, E., Nankeu, J., & Danyluk, J. (2012). Wheat CBF gene family: Identification of polymorphisms in the CBF coding sequence. *Genome*, *55*, 865 - 881.

Monroy, A. F., Dryanova, A., Malette, B., Oren, D. H., Farajalla, M. R., Liu, W., Danyluk, J., Ubayasena, L. W., Kane, K., Scoles, G. J., Sarhan, F., & Gulick, P. J. (2007). Regulatory gene candidates and gene expression analysis of cold acclimation in winter and spring wheat. *Plant Molecular Biology*, *64*, 409 - 423.

Monroy, A. F., Labbe, E., & Dhindsa, R. S. (1997). Low temperature perception in plants: Effects of cold on protein phosphorylation in cell - free extracts. *FEBS Letters*, *410*, 206 - 209.

Mooney, B. P., Miernyk, J. A., Michael Greenlief, C., & Thelen, J. J. (2006). Using quantitative proteomics of *Arabidopsis* roots and leaves to predict metabolic activity. *Physiologia Plantarum*, *128*, 237 - 250.

Morran, S., Eini, O., Pyvovarenko, T., Parent, B., Singh, R., Ismagul, A., Eliby, S., Shirley, N., Langridge, P., & Lopato, S. (2010). Improvement of stress tolerance of wheat and barley by modulation of expression of DREB / CBF factors. *Plant Biotechnology Journal*, *9*, 230 - 249.

Mouradov, A., Cremer, F., & Coupland, G. (2002). Control of flowering time: Interacting pathways as a basis for diversity. *Plant Cell*, *14*, S111 - S130.

Mukherjee, S., Sharma, S., Kumar, S., & Guptasarma, P. (2005). Slow irreversible unfolding of *Pyrococcus furiosus* triosephosphate isomerase: Separation and quantitation of conformers through a novel electrophoretic approach. *Analytical Biochemistry*, *347*, 49 - 59.

Muller, J., Boller, T., & Wiemken, A. (1995). Trehalose and trehalase in plants: Recent developments. *Plant Science*, *112*, 1 - 9.

Murai, K., Miyamae, M., Kato, H., Takumi, S., & Ogihara, Y. (2003). WAP1, a wheat APETALA1 homolog, plays a central role in the phase transition from vegetative to reproductive growth. *Plant and Cell Physiology*, *44*, 1255 - 1265.

Murray, M. B., Cape, J. N., & Fowler, D. (1989). Quantification of frost damage in plant - tissues by rates of electrolyte leakage. *New Phytologist*, *113*, 307 - 311.

Naegele, T., Kandel, B. A., Frana, S., Meissner, M., & Heyer, A. G. (2011). A systems biology approach for the analysis of carbohydrate dynamics during acclimation to low temperature in *Arabidopsis thaliana*. *FEBS Journal*, *278*, 506 - 518.

Nagy, F., & Schafer, E. (2002). Phytochromes control photomorphogenesis by differentially regulated, interacting signaling pathways in higher plants. *Annual Review of Plant Biology*, *53*, 329 - 355.

Nakaminami, K., Karlson, D. T., & Imai, R. (2006). Functional conservation of cold shock domains in bacteria and higher plants. *Proceedings of the National Academy of Sciences of the United States of America*, *103*, 10122 - 10127.

Nakano, T., Suzuki, K., Fujimura, T., & Shinshi, H. (2006). Genome - wide analysis of the ERF gene family in *Arabidopsis* and rice. *Plant Physiology*, *140*, 411 - 432.

Nara, T., Lee, L., & Imae, Y. (1991). Thermosensing ability of Trg and Tap chemoreceptors in *Escherichia coli*. *Journal of Bacteriology*, *173*, 1120 - 1124.

Narberhaus, F., Waldminghaus, T., & Chowdhury, S. (2006). RNA thermometers. *FEMS Microbiology Reviews*, *30*, 3 - 16.

NDong, C., Danyluk, J., Wilson, K. E., Pocock, T., Huner, N. P. A., & Sarhan, F. (2002). Cold - regulated cereal chloroplast late embryogenesis abundant - like proteins. Molecular characterization and functional analyses. *Plant Physiology*, *129*, 1368 - 1381.

Nishizawa, A., Yabuta, Y., & Shigeoka, S. (2008). Galactinol and raffinose constitute a novel function to protect plants from oxidative damage. *Plant Physiology*, *147*, 1251 - 1263.

Notredame, C., Higgins, D. G., & Heringa, J. (2000). T - coffee: A novel method for fast and accurate multiple sequence alignment. *Journal of Molecular Biology*, *302*, 205 - 217.

Novillo, F., Medina, J., & Salinas, J. (2007). *Arabidopsis* CBF1 and CBF3 have a different function than CBF2 in cold acclimation and define different gene classes in the CBF regulon. *Proceedings of the National Academy of Sciences of the United States of America*, *104*, 21002 - 21007.

Novillo, F., Medina, J., Rodriguez - Franco, M., Neuhaus, G., & Salinas, J. (2012). Genetic analysis reveals a complex regulatory network modulating CBF gene expression and *Arabidopsis* response to abiotic stress. *Journal of Experimental Botany*, *63*, 293 - 304.

Oh, S., Kwon, C., Choi, D., Song, S. I., & Kim, J. (2007). Expression of barley *HvCBF4* enhances tolerance to abiotic stress in transgenic rice. *Plant Biotechnology Journal*, *5*, 646 - 656.

Ohme - Takagi, M., & Shinshi, H. (1995). Ethylene - inducible DNA - binding proteins that

interact with an ethylene - responsive element. *Plant Cell*, *7*, 173 - 182.

Ohno, R., Takumi, S., & Nakamura, C. (2001). Expression of a cold - responsive *Lt - Cor* gene and development of freezing tolerance during cold acclimation in wheat (*Triticum aestivum* L.). *Journal of Experimental Botany*, *52*, 2367 - 2374.

Ohno, R., Takumi, S., & Nakamura, C. (2003). Kinetics of transcript and protein accumulation of a low molecular weight wheat LEA D - 11 dehydrin in response to low temperature. *Journal of Plant Physiology*, *160*, 193 - 200.

Ohta, T. (1990). How gene families evolve. *Theoretical Population Biology*, *37*, 213 - 219.

Ohta, T. (2000). Evolution of gene families. *Gene*, *259*, 45 - 52.

Olien, C. R. (1984). An adaptive response of rye to freezing. *Crop Science*, *24*, 51 - 54.

Oono, Y., Seki, M., Satou, M., Iida, K., Akiyama, K., Sakurai, T., Fujita, M., Yamaguchi - Shinozaki, K., & Shinozaki, K. (2006). Monitoring expression profiles of *Arabidopsis* genes during cold acclimation and deacclimation using DNA microarrays. *Functional and Integrative Genomics*, *6*, 212 - 234.

Oquist, G., & Huner, N. P. A. (2003). Photosynthesis of overwintering evergreen plants. *Annual Review of Plant Biology*, *54*, 329 - 355.

Orvar, B. L., Sangwan, V., Omann, F., & Dhindsa, R. S. (2000). Early steps in cold sensing by plant cells: The role of actin cytoskeleton and membrane fluidity. *Plant Journal*, *23*, 785 - 794.

Ottander, C., Campbell, D., & Oquist, G. (1995). Seasonal - changes in photosystem - II organization and pigment composition in *Pinus sylvestris*. *Planta*, *197*, 176 - 183.

Ouellet, F., Vazquez - Tello, A., & Sarhan, F. (1998). The wheat *wcs120* promoter is cold - inducible in both monocotyledonous and dicotyledonous species. *FEBS Letters*, 42, 324 - 328.

Palta, J. P., Levitt, J., & Stadelmann, E. J. (1977). Freezing injury in onion bulb cells. 2. Post thawing injury or recovery. *Plant Physiology*, *60*, 398 - 401.

Pancholi, V. (2001). Multifunctional alpha - enolase: Its role in diseases. *Cellular and Molecular Life Sciences*, *58*, 902 - 920.

Pavangadkar, K., Thomashow, M. F., & Triezenberg, S. J. (2010). Histone dynamics and roles of histone acetyltransferases during cold - induced gene regulation in *Arabidopsis*. *Plant Molecular Biology*, *74*, 183 - 200.

Penfield, S. (2008). Temperature perception and signal transduction in plants. *New Phytologist*, *179*, 615 - 628.

Pocock, T. H., Hurry, V., Savitch, L. V., & Huner, N. P. A. (2001). Susceptibility to low - temperature photoinhibition and the acquisition of freezing tolerance in winter and spring wheat: The role of growth temperature and irradiance. *Physiologia Plantarum*, *113*, 499 - 506.

Porcar - Castell, A., Juurola, E., Nikinmaa, E., Berninger, F., Ensminger, I., & Hari, P. (2008).

Seasonal acclimation of photosystem II in *Pinus sylvestris*. I. Estimating the rate constants of sustained thermal energy dissipation and photochemistry. *Tree Physiology*, *28*, 1475 - 1482.

Rabbani, M. A., Maruyama, K., Abe, H., Khan, M. A., Katsura, K., Ito, Y., Yoshiwara, K., Seki, M., shinozaki, K., & Yamaguch - Shinozaki, K. (2003). Monitoring expression profiles of rice genes under cold, drought, and high - salinity stresses and abscisic acid application using cDNA microarray and RNA get - blot analyses. *Plant Physiology*, *133*, 1755 - 1767.

Rapacz, M. (2002). Regulation of frost resistance during cold de - acclimation and re - acclimation in oilseed rape. A possible role of PSII redox state. *Physiologia Plantarum*, *115*, 236 - 243.

Ratnayaka, I., Båga, M., Fowler, D. B., & Chibbar, R. N. (2005). Construction and characterization of a BAC library of a cold - tolerant hexaploid wheat cultivar. *Crop Science*, *45*, 1571 - 1577.

Rechsteiner, M., & Rogers, S. W. (1996). PEST sequences and regulation by proteolysis. *Trends in Biochemical Sciences*, *21*, 267 - 271.

Reid, G., & Flonta, M. L. (2001). Physiology: cold current in thermoreceptive neurons. *Nature*, *413*, 480 - 480.

Rhzetaky, A., & Nei, M. (1993). Theoretical foundation of the minimum - evolution method of phylogenetic inference. *Molecular Biology and Evolution*, *10*, 1073 - 1095.

Riechmann, J. L., & Meyerowitz, E. M. (1998). The AP2 / EREBP family of plant transcription factors. *Biological Chemistry*, *379*, 633 - 646.

Riederer, M., & Schreiber, L. (2001). Protecting against water loss: Analysis of the barrier properties of plant cuticles. *Journal of Experimental Botany*, *52*, 2023 - 2032.

Rinalducci, S., Egidi, M. G., Mahfoozi, S., Godehkahriz, S. J., & Zolla, L. (2011). The influence of temperature on plant development in a vernalization - requiring winter wheat: A 2 - DE based proteomic investigation. *Journal of Proteomics*, *74*, 643 - 659.

Robertson, M. J., Brooking, I. R., & Ritchie, J. T. (1996). Temperature response of vernalization in wheat: Modelling the effect on the final number of mainstem leaves. *Annals of Botany*, *78*, 371 - 381.

Rolland, F., Baena - Gonzalez, E., & Sheen, J. (2006). Sugar sensing and signaling in plants: Conserved and novel mechanisms. *Annual Review of Plant Biology*, *57*, 675 - 709.

Romero, I., Fuertes, A., Benito, M., Malpica, J., Leyva, A., & Paz - Ares, J. (1998). More than 80R2R3 - MYB regulatory genes in the genome of *Arabidopsis thaliana*. *Plant Journal*, *14*, 273 - 284.

Rosinski, J. A., & Atchley, W. R. (1998). Molecular evolution of the MYB family of transcription factors: Evidence for polyphyletic origin. *Journal of Molecular Evolution*, *46*, 74 - 83.

Rowbury, R. J. (2003). Temperature effects on biological systems: Introduction. *Science Progress*, *86*, 1 - 8.

Rozen, S., & Skaletsky, H. (2000). Primer3 on the WWW for general users and for biologist programmers. *Methods in Molecular Biology, 132*, 365 - 86.

Ruelland, E., Vaultier, M., Zachowski, A., & Hurry, V. (2009). Cold signalling and cold acclimation in plants. *In*: J. D. Kader M. (Ed.), *Advances in botanical research,* Academic Press, Elsevier, San Diego, CA, USA. pp. 35 - 150.

Saitou, N., & Nei, M. (1987). The neighbor - joining method - a new method for reconstructing phylogenetic trees. *Molecular Biology and Evolution, 4*, 406 - 425.

Sakai, A. and W. Larcher. 1987. The freezing process in plants. *In*: Billings, W. D., Golley, F., Lange, O. L., Olson, J. S., & Remmert, H. (Ed.) *Frost Survival of Plants. Responses and Adaptation to Freezing Stress*. Springer Verlag, Berlin, Germany. pp. 21 - 38.

Sakuma, Y., Liu, Q., Dubouzet, J. G., Abe, H., Shinozaki, K., & Yamaguchi - Shinozaki, K. (2002). DNA - binding specificity of the ERF / AP2 domain of *Arabidopsis* DREBs, transcription factors involved in dehydration - and cold - inducible gene expression. *Biochemical and Biophysical Research Communications, 290*, 998 - 1009.

Sakuma, Y., Maruyama, K., Osakabe, Y., Qin, F., Seki, M., Shinozaki, K., Yamaguchi – Shinozaki, K. (2006). Functional analysis of an *Arabidopsis* transcription factor, DREB2A, involved in drought-responsive gene expression. *Plant Cell*, *18*, 1292 – 1309.

Sandve, S. R., & Fjellheim, S. (2010). Did gene family expansions during the eocene - oligocene boundary climate cooling play a role in Pooideae adaptation to cool climates? *Molecular Ecology, 19*, 2075 - 2088.

Sandve, S. R., Rudi, H., Asp, T., & Rognli, O. A. (2008). Tracking the evolution of a cold stress associated gene family in cold tolerant grasses. *BMC Evolutionary Biology, 8*, 245 - 260.

Sappl, P., Heazlewood, J., & Millar, A. (2004). Untangling multi - gene families in plants by integrating proteomics into functional genomics. *Phytochemistry, 65*, 1517 - 1530.

Sasaki, K., Kim, M., & Imai, R. (2007). *Arabidopsis* COLD SHOCK DOMAIN PROTEIN2 is a RNA chaperone that is regulated by cold and developmental signals. *Biochemical and Biophysical Research Communications, 364*, 633 - 638.

Savitch, L. V., Leonardos, E. D., Krol, M., Jansson, S., Grodzinski, B., Huner, N. P. A., & Öquist G. (2002). Two different strategies for light utilization in photosynthesis in relation to growth and cold acclimation. *Plant Cell and Environment, 25*, 761 - 771.

Sawyer, J. R., Schlom, J., & Kashmiri, S. V. S. (1994). The effects of induction conditions on production of a soluble antitumor sFv in *Escherichia coli*. *Protein Engineering, 7*, 1401 - 1406.

Scarth, R., & Law, C. N. (1983). The location of the photoperiod gene, *Ppd2* and an additional genetic - factor for ear - emergence time on chromosome - 2B of wheat. *Heredity, 51*, 607 - 619.

Scheidt, H., Pampel, A., Nissler, L., Gebhardt, R., & Huster, D. (2004). Investigation of the membrane localization and distribution of flavonoids by high - resolution magic angle spinning NMR spectroscopy. *Biochimica Et Biophysica Acta - Biomembranes, 1663*, 97 - 107.

Seiffert, E. R. (2007). Evolution and extinction of afro - arabian primates near the eocene - oligocene boundary. *Folia Primatologica*, 78, 314 - 327.

Seki, M., Ishida, J., Narusaka, M., Fujita, M., Nanjo, T., Umezawa, T., Kamiya, A., Nakajima, M., Enju, A., Sakurai, T., Satou, M., Akiyama, K., Yamaguchi - Shinozaki, K., Caninci, P., Kawai, J., Hayashizaki, Y., & Shinozaki, K. (2002a). Monitoring the expression pattern of around 7,000 *Arabidopsis* genes under ABA treatments using a full - length cDNA microarray. *Functional and Integrative Genomics*, *2*, 282 - 291.

Seki, M., Kamei, A., Yamaguchi - Shinozaki, K., & Shinozaki, K. (2003). Molecular responses to drought, salinity and frost: Common and different paths for plant protection. *Current Opinion in Biotechnology*, *14*, 194 - 199.

Seki, M., Narusaka, M., Abe, H., Kasuga, M., Yamaguchi - Shinozaki, K., Carninci, P., Hayashizaki, Y., & Shinozaki, K. (2001). Monitoring the expression pattern of 1300 *Arabidopsis* genes under drought and cold stresses by using a full - length cDNA microarray. *Plant Cell*, *13*, 61 - 72.

Seki, M., Narusaka, M., Ishida, J., Nanjo, T., Fujita, M., Oono, Y., Kamiya, A., Nakajima, M., Enju, A. Sakurai, T., Satou, M., Akiyama, K., Taji, T., Yamaguchi - Shinozaki, K., Carninci, P., Kawai, J., Hayashizaki, Y., and Shinozaki, K. (2002b). Monitoring the expression profiles of 7000 *Arabidopsis* genes under drought, cold and high - salinity stresses using a full - length cDNA microarray. *Plant Journal*, *31*, 279 - 292.

Shi, Y., Tian, S., Hou, L., Huang, X., Zhang, X., Guo, H., & Yang, S. (2012). Ethylene signaling negatively regulates freezing tolerance by repressing expression of CBF and type - A ARR genes in *Arabidopsis*. *Plant Cell*, *24*, 2578 - 2595.

Shibasaki, K., Uemura, M., Tsurumi, S., & Rahman, A. (2009). Auxin response in *Arabidopsis* under cold stress: Underlying molecular mechanisms. *Plant Cell*, *21*, 3823 - 3838.

Shigyo, M., & Ito, M. (2004). Analysis of gymnosperm two - AP2 - domain - containing genes. *Development Genes and Evolution*, *214*, 105 - 114.

Shigyo, M., Hasebe, M., & Ito, M. (2006). Molecular evolution of the AP2 subfamily. *Gene*, *366*, 256 - 265.

Shinozaki, K., & Yamaguchi - Shinozaki, K. (2000). Molecular responses to dehydration and low temperature: Differences and cross - talk between two stress signaling pathways. *Current Opinion in Plant Biology*, *3*, 217 - 223.

Shinozaki, K., & Yamaguchi - Shinozaki, K. (2000). Molecular responses to dehydration and low temperature: Differences and cross - talk between two stress signaling pathways. *Current Opinion in Plant Biology*, *3*, 217 - 223.

Shitsukawa, N., Ikari, C., Shimada, S., Kitagawa, S., Sakamoto, K., Saito, H., Ryuto, H., Fukunishi, N., Abe, T., Takumi, S., Nasuda, S., & Murai, K. (2007). The einkorn wheat (*Triticum monococcum*) mutant, maintained vegetative phase, is caused by a deletion in the *VRN1* gene. *Genes and Genetic Systems*, *82*, 167 - 170.

Simpson, G. G., & Dean, C. (2002). Flowering - *Arabidopsis*, the rosetta stone of flowering time? *Science, 296*, 285 - 289.

Singh, B., Lee, C. B., Park, J. W., & Sohng, J. K. (2012). The amino acid sequences in the C - terminal region of glucose - 1 - phosphate thymidylyltransferases determine their soluble expression in *Escherichia coli*. *Protein Engineering, 25*, 179 - 187.

Single, W.V. (1964). Studies on frost injury to wheat. II. Ice formation within the plant. *Australian Journal of Agricultural Research, 15*, 869–875.

Skinner, D. Z. (2009). Post - acclimation transcriptome adjustment is a major factor in freezing tolerance of winter wheat. *Functional and Integrative Genomics, 9*, 513 - 523.

Skinner, D. Z., & Bellinger, B. S. (2010). Exposure to subfreezing temperature and a freeze - thaw cycle affect freezing tolerance of winter wheat in saturated soil. *Plant and Soil, 332*, 289 - 297.

Skinner, D. Z., & Garland - Campbell, K. A. (2008). The relationship of LT_{50} to prolonged freezing survival in winter wheat. *Canadian Journal of Plant Science, 88*, 885 - 889.

Skinner, J. S., von Zitzewitz, J., Szucs, P., Marquez - Cedillo, L., Filichkin, T., Amundsen, K., Stockinger, E. J., Thomashow, M. F., Chen, T. H., & Hayes, P. M. (2005). Structural, functional, and phylogenetic characterization of a large CBF gene family in barley. *Plant Molecular Biology, 59*, 533 - 551.

Smallwood, M., & Bowles, D. J. (2002). Plants in a cold climate. *Philosophical Transactions of the Royal Society of London Series B - Biological Sciences, 357*, 831 - 846.

Smith, D. B., & Flavell, R. B. (1975). Characterization of wheat genome by renaturation kinetics. *Chromosoma, 50*, 223 - 242.

Snape, J. W., Sarma, R., Quarrie, S. A., Fish, L., Galiba, G., & Sutka, J. (2001). Mapping genes for flowering time and frost tolerance in cereals using precise genetic stocks. *Euphytica, 120*, 309 - 315.

Snape, J. W., Semikhodskii, A., Fish, L., Sarma, R. N., Quarrie, S. A., Galiba, G., & Sutka, J (1997). Mapping frost tolerance loci in wheat and comparative mapping with other cereals. *Acta Agronomica Hungarica, 45*, 265 - 270.

Solomon, A., Salomon, R., Paperna, I., & Glazer, I. (2000). Desiccation stress of entomopathogenic nematodes induces the accumulation of a novel heat - stable protein. *Parasitology, 121*, 409 - 416.

Soltis, D. E., Bell, C. D., Kim, S., & Soltis, P. S. (2008). Origin and early evolution of angiosperms. *Annals of the New York Academy of Sciences, 1133*, 3 - 25.

Statistics Canada. (2011) http: / / www.agr.gc.ca / pol / mad - dam / index_e.php

Steponkus, P. L., Uemura, M., Balsamo, R. A., Arvinte, T., & Lynch, D. V. (1988). Transformation of the cryobehavior of rye protoplasts by modification of the plasma - membrane lipid - composition. *Proceedings of the National Academy of Sciences of the United States of America, 85*, 9026 - 9030.

Stockinger, E. J., Gilmour, S. J., & Thomashow, M. F. (1997). *Arabidopsis thaliana* CBF1 encodes an AP2 domain - containing transcriptional activator that binds to the C - repeat /

DRE, a cis - acting DNA regulatory element that stimulates transcription in response to low temperature and water deficit. *Proceedings of the National Academy of Sciences of the United States of America*, *94*, 1035 - 1040.

Stockinger, E. J., Mao, Y. P., Regier, M. K., Triezenberg, S. J., & Thomashow, M. F. (2001). Transcriptional adaptor and histone acetyltransferase proteins in *Arabidopsis* and their interactions with *CBF1*, a transcriptional activator involved in cold - regulated gene expression. *Nucleic Acids Research*, *29*, 1524 - 1533.

Stockinger, E. J., Skinner, J. S., Gardner, K. G., Francia, E., & Pecchioni, N. (2007). Expression levels of barley *cbf* genes at the frost resistance - H2 locus are dependent upon alleles *at Fr - H1* and *Fr - H2*. *Plant Journal*, *51*, 308 - 321.

Stracke, R., Werber, M., & Weisshaar, B. (2001). The R2R3 - MYB gene family in *Arabidopsis thaliana*. *Current Opinion in Plant Biology*, *4*, 447 - 456.

Stromberg, C. A. E. (2005). Decoupled taxonomic radiation and ecological expansion of open - habitat grasses in the cenozoic of north america. *Proceedings of the National Academy of Sciences of the United States of America*, 102, 11980 - 11984.

Sukumaran, N. P., & Weiser, C. J. (1972). An excised leaflet test for evaluating potato frost tolerance. *HortScience*, *7*, 564 - 567.

Sutka, J. (1994). Genetic - control of frost tolerance in wheat (*Triticum aestivum* L). *Euphytica*, *77*, 277 - 282.

Sutka, J., & Snape, J. W. (1989). Location of a gene for frost - resistance on chromosome - 5A of wheat. *Euphytica*, *42*, 41 - 44.

Sutton, F., Chen, D. G., Ge, X., & Kenefick, D. (2009). *Cbf* genes of the *Fr - A2* allele are differentially regulated between long - term cold acclimated crown tissue of freeze - resistant and - susceptible, winter wheat mutant lines. *BMC Plant Biology*, *9*, 34 - 43.

Suzuki, I., Los, D. A., Kanesaki, Y., Mikami, K., & Murata, N. (2000). The pathway for perception and transduction of low - temperature signals in synechocystis. *EMBO Journal*, *19*, 1327 - 1334.

Suzuki, N., Koussevitzky, S., Mittler, R., & Miller, G. (2012). ROS and redox signalling in the response of plants to abiotic stress. *Plant Cell and Environment*, *35*, 259 - 270.

Svensson, J. T., Crosatti, C., Campoli, C., Bassi, R., Stanca, A. M., Close, T. J., & Cattivelli, L. (2006). Transcriptome analysis of cold acclimation in barley *Albina* and *Xantha* mutants. *Plant Physiology*, *141*, 257 - 270.

Swamy, K. H. S., & Goldberg, A. L. (1981). *Escherichia coli* contains eight soluble proteolytic activities, one being atp dependent. *Nature*, *292*, 652 - 654.

Takahashi, R., Yasuda, S. (1971) Genetics of earliness and growth habit in barley. *In*: Nilan, R. A. (Ed) *Barley genetics II* (Proceedings of the Second International Barley Genetics Symposium,). Washington State University Press, Pullman, USA. pp 388–408.

Takahashi, S., Katagiri, T., Yamaguchi - Shinozaki, K., & Shinozaki, K. (2000). An *Arabidopsis* gene encoding a Ca2+ - binding protein is induced by abscisic acid during dehydration. *Plant and Cell Physiology*, *41*, 898 - 903.

Takumi, S., Koike, A., Nakata, M., Kume, S., Ohno, R., & Nakamura, C. (2003). Cold - specific and light - stimulated expression of a wheat (*Triticum aestivum* L.) *cor* gene *Wcor15* encoding a chloroplast - targeted protein. *Journal of Experimental Botany*, *54*, 2265 - 2274.

Talbert, L., Smith, L., & Blake, M. (1998). More than one origin of hexaploid wheat is indicated by sequence comparison of low - copy DNA. *Genome*, *41*, 402 - 407.

Tamura, K., & Yamada, T. (2007). A perennial ryegrass CBF gene cluster is located in a region predicted by conserved synteny between Poaceae species. *Theoretical and Applied Genetics*, *114*, 273 - 283.

Tamura, K., Dudley, J., Nei, M., & Kumar, S. (2007). MEGA4: Molecular evolutionary genetics analysis (MEGA) software version 4.0. *Molecular Biology and Evolution*, *24*, 1596 - 1599.

Tanaka, A. (2007) Photosynthetic activity in winter needles of the evergreen tree *Taxus cuspidata* at low temperatures. *Tree Physiology 27* 641 – 648 .

Thambirajah, A. A., Dryhurst, D., Ishibashi, T., Li, A., Maffey, A. H., & Ausio, J. (2006). H2A.Z stabilizes chromatin in a way that is dependent on core histone acetylation. *Journal of Biological Chemistry*, *281*, 20036 - 20044.

Thomashow, M. F. (1998). Role of cold - responsive genes in plant freezing tolerance. *Plant Physiology*, *118*, 1 - 7.

Thomashow, M. F. (1999). Plant cold acclimation: Freezing tolerance genes and regulatory mechanisms. *Annual Review of Plant Physiology and Plant Molecular Biology*, *50*, 571 - 599.

Thomashow, M. F., Gilmour, S. J., Stockinger, E. J., Jaglo - Ottosen, K. R., & Zarka, D. G. (2001). Role of the *Arabidopsis* CBF transcriptional activators in cold acclimation. *Physiologia Plantarum*, *112*, 171 - 175.

Thorlby, G., Fourrier, N., & Warren, G. (2004). The sensitive to *freezing2* gene, required for freezing tolerance in *Arabidopsis thaliana*, encodes a beta - glucosidase. *Plant Cell*, *16*, 2192 - 2203.

Tian, Y., Zhang, H., Pan, X., Chen, X., Zhang, Z., Lu, X., & Huang, R. (2011). Overexpression of ethylene response factor TERF2 confers cold tolerance in rice seedlings. *Transgenic Research*, *20*, 857 - 866.

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

Tóth, B., Galiba, G., Feher, E., Sutka, J., & Snape, J. W. (2003). Mapping genes affecting flowering time and frost resistance on chromosome 5B of wheat. *Theoretical and Applied Genetics*, *107*, 509 - 514.

Tremblay, K., Ouellet, F., Fournier, J., Danyluk, J., & Sarhan, F. (2005). Molecular characterization and origin of novel bipartite cold - regulated ice recrystallization inhihition proteins from cereals. *Plant and Cell Physiology*, *46*, 884 - 891.

Trethowan, R. M., & Mujeeb - Kazi, A. (2008). Novel germplasm resources for improving environmental stress tolerance of hexaploid wheat. *Crop Science*, *48*, 1255 - 1265.

Trevaskis, B., Bagnall, D. J., Ellis, M. H., Peacock, W. J., & Dennis, E. S. (2003). MADS box genes control vernalization - induced flowering in cereals. *Proceedings of the National Academy of Sciences of the United States of America*, *100*, 13099 - 13104.

Trevaskis, B., Hemming, M. N., Dennis, E. S., & Peacock, W. J. (2007). The molecular basis of vernalization - induced flowering in cereals. *Trends in Plant Science*, *12*, 1360 - 1385.

Trevaskis, B., Hemming, M. N., Peacock, W. J., & Dennis, E. S. (2006). *HvVRN2* responds to daylength, whereas *HvVRN1* is regulated by vernalization and developmental status. *Plant Physiology*, *140*, 1397 - 1405.

Tsuda, K., Tsvetanov, S., Takumi, S., Mori, N., Atanassov, A., & Nakamura, C. (2000). New members of a cold - responsive group - 3 Lea / Rab - related cor gene family from common wheat (*Triticum aestivum* L.). *Genes and Genetic Systems*, *75*, 179 - 188.

Tsvetanov, S., Ohno, R., Tsuda, K., Takumi, S., Mori, N., Atanassov, A., & Nakamura, C. (2000). A cold - responsive wheat (*Triticum aestivum* L.) gene *wcor14* identified in a winter - hardy cultivar 'Mironovska 808'. *Genes and Genetic Systems*, *75*, 49 - 57.

Tung, J. S., & Knight, C. A. (1972). Relative importance of some factors affecting electrophoretic migration of proteins in sodium dodecyl sulfate - polyacrylamide gels. *Analytical Biochemistry*, *48*, 153 - 163.

Uemura, M., & Yoshida, S. (1984). Involvement of plasma - membrane alterations in cold - acclimation of winter rye seedlings (*Secale cereale* L cv puma). *Plant Physiology*, *75*, 818 - 826.

Uemura, M., Joseph, R., & Steponkus, P. (1995). Cold - acclimation of *Arabidopsis thaliana* - effect on plasma - membrane lipid - composition and freeze - induced lesions. *Plant Physiology*, *109*, 15 - 30.

Uno, Y., Furihata, T., Abe, H., Yoshida, R., Shinozaki, K., & Yamaguchi - Shinozaki, K. (2000). *Arabidopsis* basic leucine zipper transcription factors involved in an abscisic acid - dependent signal transduction pathway under drought and high - salinity conditions. *Proceedings of the National Academy of Sciences of the United States of America*, *97*, 11632 - 11637.

Urao, T., Yakubov, B., Yamaguchi - Shinozaki, K., & Shinozaki, K. (1998). Stress - responsive expression of genes for two - component response regulator - like proteins in *Arabidopsis thaliana*. *FEBS Letters*, *427*, 175 - 178.

Urao, T., Yamaguchishinozaki, K., Urao, S., & Shinozaki, K. (1993). An *Arabidopsis myb* homolog is induced by dehydration stress and its gene - product binds to the conserved MYB recognition sequence. *Plant Cell*, *5*, 1529 - 1539.

Usadel, B., Bläesing, O. E., Gibon, Y., Poree, F., Höehne, M., Güenter, M., Trethewey, R., Kamlage, B., Poorter, H., & Stitt, M. (2008). Multilevel genomic analysis of the response of transcripts, enzyme activities and metabolites in *Arabidopsis* rosettes to a progressive

decrease of temperature in the non - freezing range. *Plant Cell and Environment*, *31*, 518 - 547.

Vágújfalvi, A., Aprile, A., Miller, A., Dubcovsky, J., Delugu, G., Galiba, G., & Cattivelli, L. (2005). The expression of several *cbf* genes at the *Fr - A2* locus is linked to frost resistance in wheat. *Molecular Genetics and Genomics*, *274*, 506 - 514.

Vágújfalvi, A., Crosatti, C., Galiba, G., Dubcovsky, J., & Cattivelli, L. (2000). Two loci on wheat chromosome 5A regulate the differential cold - dependent expression of the *cor14b* gene in frost - tolerant and frost - sensitive genotypes. *Molecular and General Genetics*, *263*, 194 - 200.

Vágújfalvi, A., Galiba, G., Cattivelli, L., & Dubcovsky, J. (2003). The cold - regulated transcriptional activator *Cbf3* is linked to the frost - tolerance locus *Fr - A2* on wheat chromosome 5A. *Molecular Genetics and Genomics*, *269*, 60 - 67.

Vahala, T., Oxelman, B., & von Arnold, S. (2001). Two APETALA2 - like genes of picea abies are differentially expressed during development. *Journal of Experimental Botany*, *52*, 1111 - 1115.

Valluru, R., & Van den Ende, W. (2008). Plant fructans in stress environments: Emerging concepts and future prospects. *Journal of Experimental Botany*, *59*, 2905 - 2916.

Valluru, R., Lammens, W., Claupein, W., & Van den Ende, W. (2008). Freezing tolerance by vesicle - mediated fructan transport. *Trends in Plant Science*, *13*, 1360 - 1385.

Van Eldik, L. J., Grossman, A. R., Iverson, D. B., & Watterson, D. M. (1980). Isolation and characterization of calmodulin from spinach *Spinacia oleracea* leaves and *in vitro* translation mixtures. *Proceedings of the National Academy of Sciences of the United States of America*, *77*, 1912 - 1916.

Vargas, W. A., Pontis, H. G., & Salerno, G. L. (2007). Differential expression of alkaline and neutral invertases in response to environmental stresses: Characterization of an alkaline isoform as a stress - response enzyme in wheat leaves. *Planta*, *226*, 1535 - 1545.

Verhoeven, A., Osmolak, A., Morales, P., & Crow, J. (2009). Seasonal changes in abundance and phosphorylation status of photosynthetic proteins in eastern white pine and balsam fir. *Tree Physiology*, *29*, 361 - 374.

Vítámvás, P., Prášil, I. T., Kosová, K., Planchon, S., & Renaut, J. (2012). Analysis of proteome and frost tolerance in chromosome 5A and 5B reciprocal substitution lines between two winter wheats during long - term cold acclimation. *Proteomics*, *12*, 68 - 85.

Vogel, J. T., Zarka, D. G., Van Buskirk, H. A., Fowler, S. G., & Thomashow, M. F. (2005). Roles of the CBF2 and ZAT12 transcription factors in configuring the low temperature transcriptome of *Arabidopsis*. *Plant Journal*, *41*, 195 - 211.

Vogg, G., Heim, R., Hansen, J., Schafer, C., & Beck, E. (1998). Frost hardening and photosynthetic performance of scots pine (*Pinus Sylvestris* L.) needles. I. seasonal changes in the photosynthetic apparatus and its function. *Planta*, *204*, 193 - 200.

Wang, D. G., Harper, J. F., & Gribskov, M. (2003). Systematic trans - genomic comparison of protein kinases between *Arabidopsis* and *Saccharomyces cerevisiae*. *Plant Physiology*, *132*, 2152 - 2165.

Wang, H. V., Vaupel, K., Buettner, R., Bosserhoff, A. K., & Moser, M. (2004). Identification and embryonic expression of a new AP - 2 transcription factor, AP - 2 epsilon. *Developmental Dynamics*, *231*, 128 - 135.

Wang, H., Datla, R., Georges, F., Loewen, M., & Cutler, A. J. (1995). Promoters from *kin1* and *cor6.6*, two homologous *Arabidopsis thaliana* genes: transcriptional regulation and gene - expression induced by low - temperature, ABA, osmoticum and dehydration. *Plant Molecular Biology*, *28*, 605 - 617.

Wang, Z. B., Triezenberg, S. J., Thomashow, M. F., & Stockinger, E. J. (2005). Multiple hydrophobic motifs in *Arabidopsis* CBF1 COOH - terminus provide functional redundancy in trans - activation. *Plant Molecular Biology*, *58*, 543 - 559.

Wanner, L. A., & Junttila, O. (1999). Cold - induced freezing tolerance in *Arabidopsis*. *Plant Physiology*, *120*, 391 - 399.

Warren, G., McKown, R., Marin, A., & Teutonico, R. (1996). Isolation of mutations affecting the development of freezing tolerance in *Arabidopsis thaliana* (L) heynh. *Plant Physiology*, *111*, 1011 - 1019.

Welling, A., & Palva, E. T. (2008). Involvement of CBF transcription factors in winter hardiness in birch. *Plant Physiology*, *147*, 1199 - 1211.

Winfield, M. O., Lu, C., Wilson, I. D., Coghill, J. A., & Edwards, K. J. (2010). Plant responses to cold: Transcriptome analysis of wheat. *Plant Biotechnology Journal*, *8*, 749 - 771.

Wise, M. J. (2003). LEAping to conclusions: A computational reanalysis of late embryogenesis abundant proteins and their possible roles. *BMC Bioinformatics*, *4*, 52 - 70.

Woodcock, S., Mornon, J., & Henrissat, B. (1992). Detection of secondary structure elements in proteins by hydrophobic cluster - analysis. *Protein Engineering*, *5*, 629 - 635.

Wormit, A., Trentmann, O., Feifer, I., Lohr, C., Tjaden, J., Meyer, S., Schmidt, U., Martinoia, E., & Neuhaus, H. E. (2006). Molecular identification and physiological characterization of a novel monosaccharide transporter from *Arabidopsis* involved in vacuolar sugar transport. *Plant Cell*, *18*, 3476 - 3490.

Xin, Z. G., & Browse, J. (1998). eskimo1 mutants of *Arabidopsis* are constitutively freezing - tolerant. *Proceedings of the National Academy of Sciences of the United States of America*, *95*, 7799 - 7804.

Xin, Z., & Browse, J. (2000). Cold comfort farm: The acclimation of plants to freezing temperatures. *Plant Cell and Environment*, *23*, 893 - 902.

Xiong, L. M., Schumaker, K. S., & Zhu, J. K. (2002). Cell signaling during cold, drought, and salt stress. *Plant Cell*, *14*, S165 - S183.

Xue, G. P. (2002). An AP2 domain transcription factor HvCBF1 activates expression of cold - responsive genes in barley through interaction with a (G / a)(C / t)CGAC motif. *Biochimica Et Biophysica Acta - Gene Structure and Expression*, *1577*, 63 - 72.

Xue, G. P. (2003). The DNA - binding activity of an AP2 transcriptional activator *HvCBF2* involved in regulation of low - temperature responsive genes in Barley is modulated by temperature. *Plant Journal*, *33*, 373 - 383.

Yamaguchi - Shinozaki, K., & Shinozaki, K. (1994). A novel cis - acting element in an *Arabidopsis* gene is involved in responsiveness to drought, low - temperature, or high - salt stress. *Plant Cell*, *6*, 251 - 264.

Yamaguchi - Shinozaki, K., & Shinozaki, K. (2006). Transcriptional regulatory networks in cellular responses and tolerance to dehydration and cold stresses. *Annual Review of Plant Biology*, *57*, 781 - 803.

Yan, L. L., Loukoianov, A., Blechl, A., Tranquilli, G., Ramakrishna, W., SanMiguel, P., Bennetzen, J. l., Echenique, V., & Dubcovsky, J. (2004a). The wheat VRN2 gene is a flowering repressor down - regulated by vernalization. *Science*, *303*, 1640 - 1644.

Yan, L., Fu, D., Li, C., Blechl, A., Tranquilli, G., Bonafede, M., Sanchez, A., Valarik, M., Yasuda, S., & Dubcovsky, J. (2006). The wheat and barley vernalization gene VRN3 is an orthologue of FT. *Proceedings of the National Academy of Sciences of the United States of America*, *103*, 19581 - 19586.

Yan, L., Helguera, M., Kato, K., Fukuyama, S., Sherman, J., & Dubcovsky, J. (2004b). Allelic variation at the *VRN - 1* promoter region in polyploid wheat. *Theoritical and Applied Genetics*, *109*, 1677 - 1686.

Yoshida, S., & Uemura, M. (1984). Protein and lipid compositions of isolated plasma - membranes from orchard grass (*Dactylis glomerata* L) and changes during cold - acclimation. *Plant Physiology*, *75*, 31 - 37.

Yoshida, T., Nishida, H., Zhu, J., Nitcher, R., Distelfeld, A., Akashi, Y., Kato, K., & Dubcovsky, J. (2010). Vrn - D4 is a vernalization gene located on the centromeric region of chromosome 5D in hexaploid wheat. *Theoretical and Applied Genetics*, *120*, 543 - 552.

Yu, X. M., Griffith, M., & Wiseman, S. B. (2001). Ethylene induces antifreeze activity in winter rye leaves. *Plant Physiology*, *126*, 1232 - 1240.

Zachos, J. C., Dickens, G. R., & Zeebe, R. E. (2008). An early cenozoic perspective on greenhouse warming and carbon - cycle dynamics. *Nature*, *451*, 279 - 283.

Zachos, J., Pagani, M., Sloan, L., Thomas, E., & Billups, K. (2001). Trends, rhythms, and aberrations in global climate 65 ma to present. *Science*, *292*, 686 - 693.

Zarka, D. G., Vogel, J. T., Cook, D., & Thomashow, M. F. (2003). Cold induction of *Arabidopsis* CBF genes involves multiple ICE (inducer of CBF expression) promoter elements and a cold - regulatory circuit that is desensitized by low temperature. *Plant Physiology*, *133*, 910 - 918.

Zech, A. C., & Pauli, A. W. (1960). Cold resistance in 3 varieties of winter wheat as related to notrogen fractions and total sugar. *Agronomy Journal*, *52*, 334 - 337.

Zhang, J., Wang, Y., Wu, S., Yang, J., Liu, H., & Zhou, Y. (2012). A single nucleotide polymorphism at the vrn - D1 promoter region in common wheat is associated with vernalization response. *Theoretical and Applied Genetics*, *125*, 1697 - 1704.

Zhang, M. I. N., & Willison, J. H. M. (1987). An improved conductivity method for the measurement of frost hardiness. *Canadian Journal of Botany*, *65*, 710 - 715.

Zhang, X., Fowler, S. G., Cheng, H. M., Lou, Y. G., Rhee, S. Y., Stockinger, E. J., & Thomashow, M. F. (2004). Freezing - sensitive tomato has a functional CBF cold response pathway, but a CBF regulon that differs from that of freezing - tolerant *Arabidopsis*. *Plant Journal*, *39*, 905 - 919.

Zhang, Z., & Huang, R. (2010). Enhanced tolerance to freezing in tobacco and tomato overexpressing transcription factor TERF2 / LeERF2 is modulated by ethylene biosynthesis. *Plant Molecular Biology*, *73*, 241 - 249.

Zhao, F., Satoda, M., Licht, J. D., Hayashizaki, Y., & Gelb, B. D. (2001). Cloning and characterization of a novel mouse AP - 2 transcription factor, ap - 2 gamma, with unique DNA binding and transactivation properties. *Journal of Biological Chemistry*, *276*, 40755 - 40760.

Zilberman, D., Coleman - Derr, D., Ballinger, T., & Henikoff, S. (2008). Histone H2A.Z and DNA methylation are mutually antagonistic chromatin marks. *Nature*, *456*, 125–129.

APPENDICES

Table 3.1 Amino acid sequence alignment of selected CBFs of the Poaceae family members. Sequences were aligned using ClustalW2 at EBI ClustalW server, (http://www.ebi.ac.uk/Tools/msa/clustalw2/). Accession numbers are listed in Table 3.1 in chapter 3.

```
OsCBFIa-1G          ---------------------------MDVSAALSSDYSSGTPSP----VAADAD--DGS 27
HvCBFIa-1           ---------------------------MDVG-ALSSDYSSGTPSP----VGADGGNSEGF 28
HvCBF11             ---------------------------MEWACCGSGYSSSGTQSP----AAGDGE----E 25
TaCBFIa-A11         ---------------------------MEWAYSGGGHSSSGTKSP----AAGGRE----E 25
OsCBFIa-1E          ---------------------------MEWAYYGSGYSSSGTPSP----VGGDGD----E 25
HvCBFII-5           ---------------------------MDHC----------------------GVGLYG 10
TmCBFII-5           ---------------------------MDNS----------------------GVVFYG 10
TaCBFII-5           ---------------------------MDQYSYRGG-----------------GDDNGQ 15
TaCBFII-5.2         ---------------------------MDQYSYGGGD----------------GDQAGQ 16
TaCBFII-5.3         ---------------------------MDQYNYGGGVAYYG-------STTGGVGDNGQG 26
OsCBFII-1C          ---------------------------MEYYE--------------------------Q 6
SbCBFII-5           ---------------------------MEYGVADDYGYGYGGYDDQQDLPSSSSVDG--- 30
ZmCBF4              ---------------------------MEYAAVG-YGYGYG-YDERQEPAESADGGGGGD 31
OsCBFI-1F           ---------------------------MDTEDTSSASSSSVSPPS--------------S 19
Bradi4g35580        -------------------------------MD----AASSPSS-------------- 9
Bradi4g35640        -------------------------------MD----AASSPSS-------------- 9
Bradi4g35570        MIDVRTCIRHGRPPASLMHLFKELASCVEIFKANG----YVSSPSS-------------- 42
Bradi1g57970        ---------------------------MD---QWI----SFPPSPS-------------- 12
Bradi1g77120        ---------------------------MDGSCQWM----SFTSSSS-------------- 15
Bradi3g57360        ---------------------------MDGSGQWI----SFPSSSS-------------- 15
Bradi4g35600        -------------------------MAMDG-YDQW----ISSA-SS-------------- 15
Bradi4g35610        -------------------------MAMDG-SDQR----ISSA-SS-------------- 15
Bradi4g35590        -------------------------MDMDG-SDQQ----ITSP-SS-------------- 15
Bradi4g35620        -------------------------MAMDG-SDQW----TSSTTSS------------- 16
Bradi2g60331        -------------------------MAMDG-SDRC----LSSPSSS------------- 16
Bradi2g60340        -------------------------MAMDGFDERI----CSSPSSS------------- 17
TmCBFIIId-17        ---------------------------MDMGSEQW----SSPSTSA-------------- 15
TaCBF17.0           ---------------------------MDMGSEQW----SSPSTSA-------------- 15
```

```
AsCBFIIId-16B      ------------------------MDMSGSEQ-------WSSP---------------- 12
LpCBF2             ------------------------MDMTGSEQ-------WSSPSL-------------- 14
TmCBFIIId-16       ---MPLVQTASGKTIKQCTPQDTKILTLPSQAQPALTLHRPPSTVR-------------- 43
TaCBFIIId-A15      ------------------------MDMTGSDQ------QRSSPSS-------------- 15
TmCBF15            ------------------------MDMTGSDQ------QRSSPSS-------------- 15
HvCBF15B           ------------------------MDMAGSDQ------QRCSPSS-------------- 15
HvCBF15A           --------------------------MAGSDQ------QRCSPSS-------------- 13
AsCBFIIId-16A      ------------------------MDMTGSEQ------WSSSSSS-------------- 15
TaCBF12.2          --------------------------MDTGPERN----WNSPASP-------------- 15
TaCBFIIId-B12      --------------------------MDTGPERN----WNSPASP-------------- 15
HvCBFIIId-12       --------------------------MDTVPERN----WNSPASP-------------- 15
TaCBF12.1          --------------------------MDTAPEHN----CSSPASS-------------- 15
AsCBFIIId-12       --------------------------MDTGPEYN----LT-PTSS-------------- 14
TaCBFIIId-A19      --------------------------MDFGINGW----ISSPSSS-------------- 15
TaCBF19.1          --------------------------MDFGINGW----ISSPSSS-------------- 15
TaCBFIIId-D19      --------------------------MDMGINGW----ISSPSSS-------------- 15
TaCBFIIId-B19      --------------------------MDMGINGW----ISSPSSS-------------- 15
TaCBF19.2          --------------------------MDMAIDSW----ISSPSSS-------------- 15
ZmCBFIIIb-1A       --------------------------MDTAGLVQHATSSSSTSTS---------A---- 20
ZmCBF1             --------------------------MDMGRHQLQLQHAASSSST---------S---- 20
SbCBF1a            --------------------------MDMGRLQQHQHATCSSSTST------SAS---- 23
OsCBFIIIb-1H       --------------------------MDMAGHEVNSSSSSSGAES---------S---- 20
TmCBFIIIb-18       --------------------------MDMS--LEHSSSASSSSTT---------E---- 18
TaCBFIIIc-3.2      --------------------------MDMG---LEVSSSSPSSSS-------------- 16
TmCBF3             --------------------------MDMG---LEVSSSSPSSSS-------------- 16
TaCBFIIIc-D3       --------------------------MDMG---LEVSSSSPSSSS-------------- 16
TaCBFIIIc-3        ------------------------------------SSSPSSSS-------------- 8
HvCBFIIIc-3        --------------------------MDMG---LEVSSSSPSSSP-------------- 16
TaCBFIIIc-B10      --------------------------MDMG---LEVSSSSP------------------ 12
TmCBFIIIc-10       --------------------------MGMG---LEISSSSP------------------ 12
HvCBFIIIc-10A      --------------------------MDMG---LEVSSSPP------------------ 12
HvCBFIIIc-13       ----------------------------MG---MDLCSSSPSSS-------------- 13
TmCBFIIIc-13       ------------------------------MDLSSSSPSSS-------------- 11
Bradi4g35630       --------------------------MDMG---LQLSSSPSSSLS-------------- 16
```

```
FaCBFIIIa-6       --------------------------MCGIKREMSGE-SGLS--CS------GEY--HS 22
LpCBFIIIa-6       --------------------------MCQIKKEMSGE-SGSP--CN------GEY--CS 22
TaCBFIIIa-6       --------------------------MCPIKKEMSGE-SGSP--CS------GEN-FYS 23
TaCBFIIIa-D6      --------------------------MCPIKREMSGE-SGSP--CS------GES-FYS 23
HvCBFIIIa-6       --------------------------MCQIKKEMSGE-SGSP--CS------GENYYYS 24
TaCBFIIIa-6.2     --------------------------MCPIKREMSGE-SGSPSPCS------GEN-FCS 25
Bradi4g35650      --------------------------MCQIKKEMTGE-SGSP--CS------GDS---- 20
OsCBFIIIa-1A      --------------------------MCGIKQEMSGESSGSP----------------C 17
SbCBFIIIa-6       --------------------------MCPIKKEMTGE-SSSP--CS------------S 18
OsCBFIII-1I       --------------------------MFIRMRAASTT---------------------- 11
OsCBFIII-1D       --------------------------MEKNTAASGQLMTSS------------------ 15
OsCBFIII-1J       --------------------------MEKNTTAMGQLMSSSATTA-------------- 19
HvCBFIVa-2A       --------------------------MDTVAAWPQFEEQDYMTVWP-----------EE 22
HvCBF2B           --------------------------MDTVAAWPQFEGQDYMTVWP-----------EE 22
HbCBFIVa-2        --------------------------MDTVAVWQQFDGQEYMTGCP-----------EE 22
TaCBF2.2          --------------------------MDTVAAWQQFEGQEYMTVWP-----------EE 22
TaCBF2.1          --------------------------MDTNAAWPQFDGQEYRTVWP-----------EE 22
TaCBFIVa-2.3      --------------------------MDTIAAWPQFDGQEYRTVWP-----------EE 22
TmCBFIVa-2        --------------------------MDTAGAWPHFEGQEYRTVWP-----------EE 22
TaCBFIVa-2.2      ------------------------------------------------------------
FaCBFIVa-2        --------------------------MDAAVAAS-------LSLQS-----------GE 15
TaCBF14.1         --------------------------MDAADAASPC---------------------- 10
TaCBF14.2         --------------------------MDAADAASPC---------------------- 10
TmCBF14           --------------------------MDAADAASPC---------------------- 10
TaCBFIVc-B14      --------------------------MDAADAASPC---------------------- 10
HvCBFIVc-14       --------------------------MDAADAASPC---------------------- 10
TaCBFIVb-A20      --------------------------MDTAAPGSPR---------------------- 10
TaCBF20.0         --------------------------MDTAAPGSPR---------------------- 10
TaCBFIVb-B20      --------------------------MDTAAPGSPR---------------------- 10
TaCBFIVb-D20      --------------------------MDTAAAGSPR---------------------- 10
ScCBFIVb-20       --------------------------MDAADAGSPR---------------------- 10
TaCBFIVb-21       --------------------------MD-ADAASPS---------------------- 9
TaCBF21.0         --------------------------MD-ADAASPS---------------------- 9
TaCBFIVb-D21      --------------------------MD-ADAASLS---------------------- 9
```

```
TaCBF4.0          --------------------------MDVADAASKSGQH------------------E 14
TaCBFIVd-B4       --------------------------MDVADAASKSGHQ------------------E 14
HvCBFIVd-4A       --------------------------MDVADIASPSGQQK-----------------Q 15
HvCBFIVd-9        ---------------------MSNPIQTDVAGIASPSGQQE-----------------Q 21
ScCBFIVd-9A       --------------------------MDVADIASRSGQQ------------------Q 14
TaCBFIVd-9        --------------------------MDVADIASPSGQQ------------------E 14
TaCBF9.0          --------------------------MDVADIASPSGQQ------------------E 14
TmCBF9            --------------------------MDVADIASPSGQQ------------------E 14
TaCBFIVd-B9       --------------------------MDVADIASRSGQQ------------------E 14
TaCBFIVd-D9       --------------------------MDVADIASPSGQQG-----------------H 15
ScCBFIVd-9B       --------------------------MDVADIASPSGQQ------------------E 14
TaCBF22.0         --------------------------MDVADAASSSGQE------------------Q 14
TaCBFIVd-D22      --------------------------MDVADAASPSGQD------------------Q 14
TmCBFIVd-4        ----------------------------------MPSGQE------------------E 7
TaCBFIVd-B22      --------------------------MDVADAASPSGQE------------------Q 14
OsCBFIV-1B        --------------------------MEVEEAAY------------------------ 8

OsCBFIa-1G        SAYMTVSSAP--------------------------------PKRRAGRTKFKETRHPVF 55
HvCBFIa-1         STYMTVSSAP--------------------------------PKRRAGRTKFKETRHPVY 56
HvCBF11           GSYMTVSSAP--------------------------------PKRRAGRTKVRETRHPVY 53
TaCBFIa-A11       GSYMTVSSAP--------------------------------PKRRAGRTKVRETRHPVY 53
OsCBFIa-1E        DSYMTVSSAP--------------------------------PKRRAGRTKFKETRHPVY 53
HvCBFII-5         -EYATVTSAP--------------------------------PKRPAGRTKFRETRHPVY 37
TmCBFII-5         GAYATVMSAP--------------------------------PKRPAGRTKFRETRHPVY 38
TaCBFII-5         GGYATVTSAP--------------------------------PKRPAGRTKFRETRHPVY 43
TaCBFII-5.2       GGYATVTSAP--------------------------------PKRPAGRTKFRETRHPVY 44
TaCBFII-5.3       GGYATVTSAP--------------------------------PKRPAGRTKFRETRHPVY 54
OsCBFII-1C        EEYATVTSAP--------------------------------PKRPAGRTKFRETRHPVY 34
SbCBFII-5         DEYATVLSAP--------------------------------PKRPAGRTKFRETRHPVY 58
ZmCBF4            DEYATVLSAP--------------------------------PKRPAGRTKFRETRHPVY 59
OsCBFI-1F         PGGGHHHRLP--------------------------------PKRRAGRKKFRETRHPVY 47
Bradi4g35580      -SSSHEQHGAAC---------------PQPAPPR--------PKRPAGRTKFKETRHPVY 45
Bradi4g35640      -SSSHEQHGAAC---------------PQPAPPR--------PKRPAGRTKFKETRHPVY 45
```

```
Bradi4g35570       -SSSHEQHGAAC---------------PPPAPPR--------PKRPAGRTKFKETRHPVY 78
Bradi1g57970       -SQDQQEE--------------------WSPAP--------PKRPAGRTKFKETRHPVY 42
Bradi1g77120       -SSSSSHNGQAG---------------APWPPPP--------PKRPAGRTKFKETRHPVY 51
Bradi3g57360       -SHDHECHG------------------VAWSPKP---------KRLAGRSKFKETRHPVY 47
Bradi4g35600       -SSSHDQYG------------------VVWSPPP---------KRPAGRTKFKETRHPVY 47
Bradi4g35610       -SSSYDQYG------------------VVWSPPP---------KRPAGRTKFKETRHPVY 47
Bradi4g35590       -SSSHDQYG------------------VVWSPPP--------PKRPAGRTKFKETRHPVY 48
Bradi4g35620       -SSSHDQYM------------------VVWSPPP---------KRPAGRTKFKETRHPVY 48
Bradi2g60331       -SSSHDRQEGL----------------AVWS------------KRPAGRTKFKETRHPVY 47
Bradi2g60340       -PSSSSSQEGL----------------AVWS------------KRPAGRTKFKETRHPVY 48
TmCBFIIId-17       -SSRDQHAA-----------------------AP--------PKRPAGRTKFKETRHPVY 43
TaCBF17.0          -SSRDQHAA-----------------------AP--------PKRPAGRTKFKETRHPVY 43
AsCBFIIId-16B      -SSSLEH-----------------GGPAVWTTP---------PKRPAGRTKFKETRHPVY 45
LpCBF2             -SSSSQE-----------------QGVPVWVTP---------PKRPAGRTKFKETRHPVY 47
TmCBFIIId-16       -SSSSQHRPPSAMDMT--------GSDQQWSSSSSPSSTSSHPKRPAGRTKFKETRHPVY 94
TaCBFIIId-A15      -PSSSSH-------------------------------------LKRPAGRTKFKETRHPVY 39
TmCBF15            -PSSSSH-------------------------------------LKRPAGRTKFKETRHPVY 39
HvCBF15B           -PSLSSH-------------------------------------LKRPAGRTKFKETRHPVY 39
HvCBF15A           -PSLSSH-------------------------------------LKRPAGRTKFKETRHPVY 37
AsCBFIIId-16A      -SYSVEHG-----------------GPAVWTTP---------PKRPAGRTKFKETRHPVY 48
TaCBF12.2          -PSSLEQGMP----------------TSPASPTP---------KRPAGRTKFKETRHPVF 49
TaCBFIIId-B12      -PSSLEQGMP----------------TSPASPTP---------KRPAGRTKFKETRHPVF 49
HvCBFIIId-12       -PSSLEQGMP----------------SSPASPTP---------KRPAGRTKFKETRHPVF 49
TaCBF12.1          -PSSQGQVMP----------------TSPTSPTP---------KRPAGRTKLKETRHPVY 49
AsCBFIIId-12       -SSSQELGMAL---------------SSPTSPTP---------KRPAGRTKFKETRHPVY 49
TaCBFIIId-A19      -TSGHELGDA----------------VPVWSPAA---------KRPAGRTKFKETRHPVY 49
TaCBF19.1          -TSGHELGDA----------------VPVWSPAA---------KRPAGRTKFKETRHPVY 49
TaCBFIIId-D19      -TSGHELGEA----------------VPVWSQAA---------KRPAGRTKFKETRHPVY 49
TaCBFIIId-B19      -TSGHELGET----------------VPVWSPAA---------KRPAGRTKFKETRHPVY 49
TaCBF19.2          -TSGHEHGEV----------------VLVWSPAA---------KRPAGRTKFKDTRHPVY 49
ZmCBFIIIb-1A       SSSSSEQQSRKAAWPPSTASSPQQ----------------PPKKRPAGRTKFRETRHPVF 64
ZmCBF1             ASSSSEQDK------PLCCSGP-------------------KKRPAGRTKFRETRHPVF 54
SbCBF1a            STSSSEQTIKAVVWSPSSSSSPQ-----------------PPKKRPAGRTKFRETRHPVY 66
OsCBFIIIb-1H       SSSSGRQQY-----------------------------------KKRPAGRTKFRETRHPVY 47
```

```
TmCBFIIIb-18      RGGTAWPWP---------------------------------PKRPAGRTKFRETRHPVF 45
TaCBFIIIc-3.2     ASSSPEHAAGRAS-----------------------------LAKRPAGRTKFRETRHPVY 48
TmCBF3            ASSSPEHAAGRAS-----------------------------LAKRPAGRTKFRETRHPVY 48
TaCBFIIIc-D3      VSSSPVHAAGRAS-----------------------------LAKRPAGRTKFRETRHPVY 48
TaCBFIIIc-3       VSSSPEHAAGRAS-----------------------------LAKRPAGRTKFRETRHPVY 40
HvCBFIIIc-3       VSSSPEHAARRAS-----------------------------PAKRPAGRTKFRETRHPVY 48
TaCBFIIIc-B10     -SSSS-------------------------------------LAKRPAGRTKFRETRHPVY 35
TmCBFIIIc-10      -SSSNENAL---------------------------------VAKRPAGRTKFRETRHPVY 39
HvCBFIIIc-10A     -SSSNENASGRSS-----------------------------TAKRPAGRTKFRETRHPVY 43
HvCBFIIIc-13      VSSSPEHASG--------------------------------PAKRPVGRTKFRETRHPVY 42
TmCBFIIIc-13      ASSSPEHASGRAS-----------------------------PAKRPAGRTKFRETRHPVY 43
Bradi4g35630      ASSSPEHDAGRSSS-------PAP----------------APAKRPAGRTKFRETRHPVF 53
FaCBFIIIa-6       PSTSPEQQQGH------------------------SQKQTAWMKRPAGRTKFRETRHPVF 58
LpCBFIIIa-6       PSTSSEQKQ---------------------------QTVWTKRPAGRTKFRETRHPVY 53
TaCBFIIIa-6       PSTSREHQQAK----------------------QAAWTSAPAKRPAGRTKFRETRHPVY 60
TaCBFIIIa-D6      PSTSPENQQAR----------------------QAAWTSAPAKRPAGRTKFRETRHPVY 60
HvCBFIIIa-6       PSTSPEHQQAKQ---------------------QAAWTSAPAKRPAGRTKFRETRHPVY 62
TaCBFIIIa-6.2     PSASPERQQAR----------------------QAGWTSAPAKRPAGRTKFRETRHPVY 62
Bradi4g35650      ASSSSQQQQQQ----------------------QTVWTS-PPKRPAGRTKFRETRHPVF 56
OsCBFIIIa-1A      SSASAERQHQT--------------------------VWTAPPKRPAGRTKFRETRHPVF 51
SbCBFIIIa-6       ASTSSEHHQTV--------------------------WTSRPKRPAGRTKFRETRHPVF 51
OsCBFIII-1I       -SSSEPCRRLS-------------------------PPSSKRPAGRTKFHETRHPVF 42
OsCBFIII-1D       -AEAT-PSSP---------------------------------KRPAGRTKFQETRHLVF 40
OsCBFIII-1J       -ATATGPASP---------------------------------KRPAGRTKFQETRHPVF 45
HvCBFIVa-2A       QEYRTVWSEP---------------------------------PKRRAGRIKLQETRHPVY 50
HvCBF2B           QEYRTVWSEP---------------------------------PKRRAGRIKLQETRHPVY 50
HbCBFIVa-2        QEYRTVWSEP---------------------------------PKRRAGRIKLQETRHPVY 50
TaCBF2.2          QEYRTVWSEP---------------------------------PKRRAGRIKLQETRHPVY 50
TaCBF2.1          QEYRTVWSEP---------------------------------PKRRAGRNKLQETRHPVY 50
TaCBFIVa-2.3      QEYRTVWSEP---------------------------------PKRRAGRNKLQETRHPVY 50
TmCBFIVa-2        -EYRTVWSEP---------------------------------PKRRAGRNKLQETRHPVY 49
TaCBFIVa-2.2      ----TVWSEP---------------------------------PKRRAGRNKLQETRHPVY 24
FaCBFIVa-2        QEYRTVWSEP---------------------------------PKPRSGRTKFQETRHPVY 43
TaCBF14.1         DGHRTVWSEP---------------------------------PKRPAGRTKFKETRHPLY 38
```

```
TaCBF14.2          DGHRTVWSEP--------------------------------PKRPAGRTKFKETRHPLY 38
TmCBF14            DGHRTVWSEP--------------------------------PKRPAGRTKFKETRHPLY 38
TaCBFIVc-B14       DGHRTVWSEP--------------------------------PKRPAGRTKFKETRHPLY 38
HvCBFIVc-14        DGHRTVWSEP--------------------------------PKRPAGRTKFKETRHPLY 38
TaCBFIVb-A20       EGHRTVCSEP--------------------------------PKRPAGRTKFKETRHPLY 38
TaCBF20.0          EGHRTVCSEP--------------------------------PKRPAGRTKFKETRHPLY 38
TaCBFIVb-B20       EGHRTVCSEP--------------------------------PKRPAGRTKFKETRHPLY 38
TaCBFIVb-D20       EGHRTVCSEP--------------------------------PKRPAGRTKFRETRHPLY 38
ScCBFIVb-20        FGHRTVCSEP--------------------------------PKRPAGRTKFKETRHPLY 38
TaCBFIVb-21        DQHRTVWTEP--------------------------------PKRPAGRIKYKETRHPLY 37
TaCBF21.0          DQHRTVWTEP--------------------------------PKWPAGRIKYKETRHPLY 37
TaCBFIVb-D21       DQHRTVWTEP--------------------------------PKRPAGRIKYKETRHPLY 37
TaCBF4.0           QGHRTVSSEP--------------------------------PKRPAGRTKFHETRHPLY 42
TaCBFIVd-B4        QGHRTVSSEP--------------------------------PKRPAGRTKFRETRHPLY 42
HvCBFIVd-4A        QGHRTVSSEP--------------------------------PKRPAGRTKFHETRHPLY 43
HvCBFIVd-9         QGHRTVSSEP--------------------------------PKRPAGRTKFHETRHPLY 49
ScCBFIVd-9A        QGHRTVSSEP--------------------------------PKRPAGRTKFHETRHPLY 42
TaCBFIVd-9         QGHRTVSSEP--------------------------------PKRPAGRTKFHETRHPLY 42
TaCBF9.0           QGHRTVSSEP--------------------------------PKRPAGRTKFHETRHPLY 42
TmCBF9             QGHRTVSSEP--------------------------------PKRPAGRTKFHETRHPLY 42
TaCBFIVd-B9        QGHRTVSSEP--------------------------------PKRPAGRTKFHETRHPLY 42
TaCBFIVd-D9        RGHRTVSSEP--------------------------------PKRPAGRTKFHETRHPLY 43
ScCBFIVd-9B        QGHRTVSSEP--------------------------------PKRPAGRTKFHETRHPLY 42
TaCBF22.0          QGHRTVSSEP--------------------------------PKRPAGRTKVHETRHPLY 42
TaCBFIVd-D22       QGHRTVSSEP--------------------------------PKRPAGRTKVHETRHPLY 42
TmCBFIVd-4         QRHRTVRSEP--------------------------------PKRPAGRTKFQETRHPLY 35
TaCBFIVd-B22       QGHRTVSSEP--------------------------------PKRPAGRTKVHETRHPLY 42
OsCBFIV-1B         ---RTVWSEP--------------------------------PKRPAGRTKFRETRHPVY 33
                                                              *   ** * ::*** ::

OsCBFIa-1G         KGVRRRN---PG---RWVCEVREPH-------GKQRIWLGTFETAEMAARAHDVAALALR 102
HvCBFIa-1          KGVRRRN---PG---RWVCEVREPH-------SKQRIWLGTFETAEMAARAHDVAALALR 103
HvCBF11            KGVRSRN---PG---RWVCEVREPQ-------GKQRLWLGTFDTAEMAARAHDVAAMALR 100
TaCBFIa-A11        KGVRSRN---PG---RWVCEVREPH-------GKQRLWLGTFDTAEMAARAHDVAALALR 100
```

```
OsCBFIa-1E          KGVRSRN---PG---RWVCEVREPH-------GKQRIWLGTFETAEMAARAHDVAAMALR 100
HvCBFII-5           RGVRRRGA--AG---RWVCEVREPN-------KKSRIWLGTFATPEAAARAHDVAALALR 85
TmCBFII-5           RGVRRRGA--AG---RWVCEVRQPN-------NKSRIWLGTFASPEAAARAHDVAALALR 86
TaCBFII-5           RGVRRRGA--AG---RWVCEVREPN-------KKSRIWLGTFASPEAAARAHDVAALALR 91
TaCBFII-5.2         RGVRRRGA--AG---RWVCEVREPN-------KKSRIWLGTFASPEAAARAHDVAALALR 92
TaCBFII-5.3         RGVRRRGA--AG---RWVCEVREPN-------KKSRIWLGTFASPEAAARAHDVAALALR 102
OsCBFII-1C          RGVRRRGP--AG---RWVCEVREPN-------KKSRIWLGTFATAEAAARAHDVAALALR 82
SbCBFII-5           RGVRRRGP--AG---RWVCEVREPN-------KKSRIWLGTFATAEAAARAHDVAALALR 106
ZmCBF4              RGVRRRGP--AG---RWVCEVREPN-------KKSRIWLGTFATPEAAARAHDVAALALR 107
OsCBFI-1F           RGVRARAG--GS---RWVCEVREPQ-------AQARIWLGTYPTPEMAARAHDVAAIALR 95
Bradi4g35580        RGVRRRG--AAG---RWVCEVRVPG------KRGQRLWLGTHLTADSAARAHDAAMLALR 94
Bradi4g35640        RGVRRRG--AAG---RWVCEVRVPG------KRGQRLWLGTHLTADSAARAHDAAMLALR 94
Bradi4g35570        RGVRRRG--PAG---RWVCEVRVPG------KRGQRLWLGTHLTADSAARAHDAAMLALR 127
Bradi1g57970        RGVRTRG--AAG---RWVCEIRVPG------KRGKRLWLGTYLTAESAARGHDAAMLMLR 91
Bradi1g77120        HGVRRRG--RAG---RWVCEVRVPGTGSCNKKRGQRLWLGTYFSAECAARAHDAAMLMLR 106
Bradi3g57360        RGVRRRG--ALGGRCRWVCEVRVPG------KHGKRLWLGTHRTAESAGRAHDAAMLTLR 99
Bradi4g35600        RGVRLRG--NAG---RWVCEVRVPG------NRGKRLWLGTHLTAESAARAHDAAMLALH 96
Bradi4g35610        RGVRRRG--SAG---RWVCEVRVPG------MRNKRLWLGTHLTAESAGRAHDAAMLALR 96
Bradi4g35590        RGVRHRG--NAG---RWVCEVRVPG------TSGKRLWLGTHLTAESAARAHDAAMLALH 97
Bradi4g35620        RGVRLRG--TAG---RWVCEVRVPG------MRNKRLWLGTHLTAESAARAHDAAMLALR 97
Bradi2g60331        RGVRSRGAAAAG---RWVCEVRVPG------AHGKRLWLGTHLTAESAGRAHDAAMLALL 98
Bradi2g60340        RGVRHRG--AAG---RWVCEVRVPG------AHGKRLWLGTHLTAESAGRAHDAAMLALL 97
TmCBFIIId-17        RGVRRRG--GAG---RWVCEVRVPG------RRGCRLWLGTYVTAESAARAHDAAMLALG 92
TaCBF17.0           RGVRRRG--GAG---RWVCEVRVPG------RRGCRLWLGTYVIAESAARAHDAAMLALG 92
AsCBFIIId-16B       RGVRRRG--NAG---RWVCEVRVPG------QRGERLWLGTYLTAESAARAHDAAMLGLL 94
LpCBF2              RGVRRRG--NAG---RWVCEVRVPG------QRGERLWLGTYLTAESAARAHDAAMLGLL 96
TmCBFIIId-16        RGVRRRG--NAG---RWVCEVRVPG------QRGERLWLGTYLTADAAARAHDAAMLGLL 143
TaCBFIIId-A15       RGVRRRG--SAG---RWVCEVRVPG------KRGERLWLGTHLTAEAAARAYDAAMLCLI 88
TmCBF15             RGVRRRG--SAG---RWVCEVRVPG------KRGERLWLGTHLTAEAAARAHDAAMLGLI 88
HvCBF15B            RGVRRRG--SAG---RWVCEVRVPG------KRGERLWLGTHLTAEAAARAHDAAMLCLL 88
HvCBF15A            RGVRRRG--SAG---RWVCEVRVPG------KRGERLWLGTHLTAEAAARAHDAAMLCLL 86
AsCBFIIId-16A       RGVRRRG--NAG---RWVCEVRVPG------QRGERLWLGTYLTAESAARAHDAAMLGLH 97
TaCBF12.2           HGVRRRG--SNG---RWVCEVRVPG------KRGERLWLGTHVTAEAAARAHDAAMLALY 98
TaCBFIIId-B12       HGVRRRG--SNG---RWVCEVRVPG------KRGERLWLGTHVTAEAAARAHDAAMLALY 98
```

```
HvCBFIIId-12       HGVRRRG--SNG---RWVCEVRVPG------KRGERLWLGTHVTAEAAARAHDAGMLALY 98
TaCBF12.1          HGVRRRG--RNG---RWVCEMRVPG------KRGERLWLGTHVTAKAAARAHDAAMLALH 98
AsCBFIIId-12       RGVRRRG--SNG---RWVCEVRVPS------KSGERLWLGTHVTAEAAARAHDAAMLAMH 98
TaCBFIIId-A19      RGVRRRG--SAG---RWVCEVRVPG------KRGERLWLGTYVAAESAARAHDAAMLALL 98
TaCBF19.1          RGVRRRG--SAG---RWVCEVRVPG------KRGERLWLGTYVAAESAARAHDAAMLALL 98
TaCBFIIId-D19      RGVRRRG--SAG---RWVCEVRVPG------KRGERLWLGTYVAAESAARAHDAAMLTLL 98
TaCBFIIId-B19      RGVRRRG--SAG---RWVCEVRVPG------KRGERLWLGTYVAAESAARAHDAAMLALL 98
TaCBF19.2          RGVRRRG--SAG---RWVCEVRVPG------KRGERLWLGTYVAAESATRAHDATMLALL 98
ZmCBFIIIb-1A       RGVRRRG--AAG---RWVCEVRVPG------RRGARLWLGTYLAAEAAARAHDAAILALQ 113
ZmCBF1             RGVRRRG--AAG---RWVCEVRVPG------RRGARLWLGTYLGAEAAARAHDAAMLAL- 102
SbCBF1a            RGVRRRG--AAG---RWVCEVRVPG------KRGARLWLGTYLAAESAARAHDAAMLALG 115
OsCBFIIIb-1H       RGVRRRG--GAG---RWVCEVRVPG------KRGARLWLGTYVTAEAAARAHDAAMIALR 96
TmCBFIIIb-18       RGVRRRG--NAG---RWVCEVRVPG------DRGTRLWLGTYFTAEAAARAHDAAMLMLR 94
TaCBFIIIc-3.2      RGVRRRG--NAE---RWVCEVRVPG------KRGARLWLGTYATAEIAARANDAAMLALG 97
TmCBF3             RGVRRRG--NAE---RWVCEVRVPG------KRGARLWLGTYATAEIAARANDAAMLALG 97
TaCBFIIIc-D3       RGVRRRG--NAE---RWVCEVRVPG------KRGARLWLGTYATAEVAARANDAAMLALG 97
TaCBFIIIc-3        RGVRRRG--NAE---RWVCEVRVPG------KRGARLWLGTYDTAELAARANDAAMLALG 89
HvCBFIIIc-3        RGVRRRG--NTE---RWVCEVRVPG------KRGARLWLGTYATAEVAARANDAAMLALG 97
TaCBFIIIc-B10      RGVRRRG--NAQ---RWVCEVRVPG------KRGARLWLGTYATAEIAARANDAAMLALG 84
TmCBFIIIc-10       RGVRRRG--NAE---RWVCEVRVPG------KRGARLWLGTYATAEIAARANDAAMLALG 88
HvCBFIIIc-10A      RGVRRRG--NAE---RWVCEVRVPG------KRGARLWLGTYATAEIAARANDAAMLALG 92
HvCBFIIIc-13       HDVRRRG--NAG---RWVCEVRVPS------KRGARLWLGTYLTAGAAARANDAAMLALG 91
TmCBFIIIc-13       RGVRRRG--NAG---RWVCEVRVPG------KRGSRLWLGTYLTAEAAARANDAAMLALG 92
Bradi4g35630       RGVRRRG--AAC---RWVCEVRVPG------KRGARLWLGTYVTAEAAARAHDAAMLALG 102
FaCBFIIIa-6        RGVRRRG--NAG---RWVCEVRVPG------RRGSRLWVGTFDTAEIAARAHDAAMLALA 107
LpCBFIIIa-6        RGVRRRG--NAG---RWVCEVRVPG------RRGSRLWVGTFDTAEIAARAHDAAMLALA 102
TaCBFIIIa-6        RGVRRRG--NAG---RWVCEVRVPG------RRGSRLWLGTFDTAEAAARANDAAMIALS 109
TaCBFIIIa-D6       RGVRRRG--NAG---RWVCEVRVPG------RRGSRLWLGTFDTAEAAARANDAAMIALS 109
HvCBFIIIa-6        RGVRRRG--NAG---RWVCEVRVPG------RRGSRLWLGTFDTAEAAARANDAAMLALA 111
TaCBFIIIa-6.2      RGVRRRG--NAG---RWVCEVRVPG------RRGSRLWLGTFDTAEAAARANDAVMLMLA 111
Bradi4g35650       RGVRRRG--NAG---RWVCEVRVPG------RRGSRLWLGTFDTAEAAARAHDAAMLALA 105
OsCBFIIIa-1A       RGVRRRG--NAG---RWVCEVRVPG------RRGCRLWLGTFDTAEGAARAHDAAMLAIN 100
SbCBFIIIa-6        RGVRRRG--NAG---RWVCEVRVPG------RRGCRLWLGTFDTADAAARAHDAAMLAIA 100
OsCBFIII-1I        RGVRRRG--RAG---RWVCEVRVPG------RRGCRLWLGTFDAADAAARAHDAAMLALR 91
```

```
OsCBFIII-1D        RGVRWRG--CAG---RWVCKVRVPG------SRGDRFWIGTSDTAEETARTHDAAMLALC 89
OsCBFIII-1J        RGVRRRG--RAG---RWVCEVRVPG------SRGDRLWVGTFDTAEEAARAHDAAMLALC 94
HvCBFIVa-2A        RGVRRRG--KVGQ---WVCELRVP-----VSRGYSRLWLGTFANPEMAARAHDSAALALS 100
HvCBF2B            RGVRRRG--KVGQ---WVCELRVP-----VSRGYSRLWLGTFANPEMAARAHDSAALALS 100
HbCBFIVa-2         RGVRRRG--KVGQ---WVCELRVP-----VSRGYSRLWLGTFANPEMAARAHDSAALALS 100
TaCBF2.2           RGVRRRG--REGQ---WVCELRVP-----VSRGYSRLWLGTFATAEMAARAHDSAALALS 100
TaCBF2.1           RGVRRRG--REGQ---WVCELRVP----AGSRSYSRIWLGTFASAQMAARAHDSAALALS 101
TaCBFIVa-2.3       RGVRRRG--REGQ---WVCELRVP----AGSRSYSRIWLGTFASAQMAARAHDSAALALS 101
TmCBFIVa-2         RGVRRRG--REGQW-VWVCELRVPA---AGSRVYSRIWLGTFADPEMAARAHDSAALALS 103
TaCBFIVa-2.2       RGVRRRG--REGQ---WVCELRVP----AGSRSYSRIWLGTFAGAQMAARAHDSAALALS 75
FaCBFIVa-2         RGVRRRG--RAGQ---WVCEMRVH------GTKGSRLWLGTFDTAEMAARAHDAAALALS 92
TaCBF14.1          RGVRRRG--PAGR---WVCEVRVL------GMRGSRLWLGTFTTAEMAARAHDAAVLALS 87
TaCBF14.2          RGVRRRG--PAGR---WVCEVRVL------GMRGSRLWLGTFTTAEMAARAHDAAVLALS 87
TmCBF14            RGVRRRG--PAGR---WVCEVRVL------GMRGSRLWLGTFTTAEMAARAHDAAVLALS 87
TaCBFIVc-B14       RGVRRRG--PAGR---WVCEVRVL------GMRGSRLWLGTFTTAEMAARAHDAAVLALS 87
HvCBFIVc-14        RGVRRRG--PAGR---WVCEVRVL------GMRGSRLWLGTFTTAEMAARAHDAAVLALS 87
TaCBFIVb-A20       RGVRRRG--RLGQ---WVCEVRVR------GAQGYRLWLGTFTTAEMAARAHDSAVLALL 87
TaCBF20.0          RGVRRRG--RLGQ---WVCEVRVR------GAQGYRLWLGTFTTAEMAARAHDSAVLALL 87
TaCBFIVb-B20       RGVRRRG--RLGQ---WVCEVRVR------GAQGYRLWLGTFTTAEMAARAHDSAVLALL 87
TaCBFIVb-D20       RGVRRRG--RLGQ---WVCEVRVR------GAQGYRLWLGTFTTAEMAARAHDSAVLALL 87
ScCBFIVb-20        RGVRRRG--RLGQ---WVCEVRVR------GAQGYRLWLGTFTTAEMAARAHDSAVLALL 87
TaCBFIVb-21        RGVRRRG--RYGR---WVCEVRVR------GTKETRLWLGTFRTAEMAARAHDSASLALS 86
TaCBF21.0          RGVRRRG--RHGR---WVCEVRVR------GTNETRLWLGTFHTAEMAARAHDSASLALS 86
TaCBFIVb-D21       RGVRRRG--RYGR---WVCEVRVR------GSKETRLWLGTFRTAEMAARAHDSASLALS 86
TaCBF4.0           RGVRRRG--RVGQ---WVCEVRVP------GVKGSRLWLGTFTTAEMAARAHDAAVLALS 91
TaCBFIVd-B4        RGVRRRG--RVGQ---WVCEVRVP------GVKGSRLWLGTFTTAEMAARAHDAAVLALS 91
HvCBFIVd-4A        RGVRRRG--RVGQ---WVCEVRVP------GIKGSRLWLGTFTNPEMAARAHDAAVLALS 92
HvCBFIVd-9         RGVRRRG--RVGQ---WVCEVRVP------GIKGSRLWLGTFNTAEMAARAHDAAALALS 98
ScCBFIVd-9A        RGVRRRG--RVGQ---WVCEVRVP------GIKGSRLWLGTFNTAEMAARAHDAAVLALS 91
TaCBFIVd-9         RGVRRRG--RVGQ---WVCEVRVP------GIKGSRLWLGTFNTAEMAARAHDAAVLALS 91
TaCBF9.0           RGVRRRG--RVGQ---WVCEVRVP------GIKGSRLWLGTFNTAEMAARAHDAAVLALS 91
TmCBF9             RGVRRRG--RVGQ---WVCEVRVP------GIKGSRLWLGTFNTAEMAARAHDAAVLALS 91
TaCBFIVd-B9        RGVRRRG--RVGQ---WVCEVRVP------GIKGSRLWLGTFNTAEMAARAHDAAVLALS 91
TaCBFIVd-D9        RGVRRRG--RVGQ---WVCEVRVP------GIKGSRLWLGTFNTAEMAARAHDAAVLALI 92
```

```
ScCBFIVd-9B         RGVRRRG--RVGQ---WVCEVRVP------GIKGSRLWLGTFNTAEMAARAHDAAVLALS 91
TaCBF22.0           RGVRQRG--RVGQ---WVCEVRVP------GVKGSRLWLGTFATAEMAARAHDAAVLALS 91
TaCBFIVd-D22        RGVRQRG--RVGQ---WVCEVRVP------GVKGSRLWLGTFTTAEMAARAHDAAVLALS 91
TmCBFIVd-4          RGVRQRG--PAGR---WVCEVRVL------GMRGSRLWLGTFVTAEMAARAHDAAVLALS 84
TaCBFIVd-B22        RGVRQRG--RVGQ---WVCEVRVA------GVKGSRLWLGTFTTAEMAARAHDAAVLALS 91
OsCBFIV-1B          RGVRRRGG-RPGAAGRWVCEVRVP------GARGSRLWLGTFATAEAAARAHDAAALALR 86
                    :.** *          ***::*            *:*:**   .  : *  *   : :

OsCBFIa-1G          G------RAACLNFADSPR-RLRVPPI-GASH-----DDIRRAAAEAAEAFR-------P 142
HvCBFIa-1           G------RAACLNFADSPR-RLRVPAV-GASP-----DEIRRAAVEAAEAFL-------P 143
HvCBF11             G------RAACLNFADSPR-RLPVPPQ-GAGH-----DEIRRAAVEAAELFR-------P 140
TaCBFIa-A11         G------RAACLNFADSPR-TLRVPPQ-GAGH-----EEIRRAAVEAAELFR-------P 140
OsCBFIa-1E          G------RAACLNFADSPR-RLRVPPL-GAGH-----EEIRRAAVEAAELFR-------P 140
HvCBFII-5           G------RAACLNFADSAA-LLRVDPATLRTP-----EDIRAAAMALAQAA---C----P 126
TmCBFII-5           G------RAACLNFADSAA-LLAVDPATLRTP-----QDIRAAAITLAQTA---C----P 127
TaCBFII-5           G------RAACLNFADSAA-LLAVDPATLRTP-----DDIRAAAIALAETA---C----P 132
TaCBFII-5.2         G------RAACLNFADSAT-LLAVDPATLRTP-----DDIRAAAIALAETA---C----P 133
TaCBFII-5.3         G------RAACLNFADSAT-LLAVDPATLRTP-----DDIRAAAIALAETA---C----P 143
OsCBFII-1C          G------RGACLNFADSAR-LLRVDPATLATP-----DDIRRAAIELAESCPHDA----A 126
SbCBFII-5           G------RAACLNFADSAR-LLRVDPATLATP-----DDIRRAAIQLAEDS--------S 146
ZmCBF4              G------RAACLNFADSAR-LLQVDPATLATP-----DDIRRAAIQLAD----------A 145
OsCBFI-1F           GE-----RGAELNFPDSPS---TLPRARTASP-----EDIRLAAAQAAELYR-------- 134
Bradi4g35580        GPWCS-AAACLLNFTDSAW-LLAVP---HVLP--NDFSAVRRAALAALADFQRRDQ--VA 145
Bradi4g35640        GPWCS-AAACLLNFTDSAW-LLAVP---HVLP--NDFSAVRRAALAALADFQRRDQ--VA 145
Bradi4g35570        GPFS--ATACHLNFPDSAW-LLAMP---CVLP--SDLAAVRRAALAAVADFQRRE---VA 176
Bradi1g57970        GSYP--VATCLLNFPDSAW-LLDVP---CTLP--ADLGDVRHAALAAVADLQRRE---AA 140
Bradi1g77120        AAP----GARVLNFPDSEW-LLDVPI--MALPAAADLSCVRRASVAAVADFQRREP--AA 157
Bradi3g57360        GPS-----ACPLNFPDSSW-LLDVP---FEVP--EDLPGVRRAALAAVADFQCRE----- 143
Bradi4g35600        GPSAA-AAC-LLNFPDSAW-LLAVTP--P-AL--ADLDDIQRAALAAVADFQRRE----A 144
Bradi4g35610        GPSAA-AKC-LLNFPDSAW-LLAVTP--SSTH--ASLDNIQRAALAAVADFQRQE----A 145
Bradi4g35590        GPSAA-AAARLLNFPDSAW-LLAVTP--S-AL--ADHDEIQRVAIAAVVDFQRRE----A 146
Bradi4g35620        GASAA-AAG-LLNFPDSAW-LLSVTP--P-AL--ADLDGVQRAALAAVADYQRRE----A 145
Bradi2g60331        P--SA-AAC-LLNFPDSA----AATP--P---------ALGCAAVVAVAGFLSRE----A 135
Bradi2g60340        P--SA-AAC-LLNFPDSA----AATP--P---------ALGCAAVVAVAGFLSRE----A 134
```

```
TmCBFIIId-17       GR-----SAACLNFPDSAW-LL--AV--PCAL--ADLADVRRAALAAVAGFQRRE---AA 137
TaCBF17.0          GR-----SAACLNFADSAW-LLAVAV--PSAL--ADLADVRRAALAAVAGFQRRE---AA 139
AsCBFIIId-16B      GRSA---SSACLNFADSAW-RLTVTP--GLS----DLAGVRRAALAAVANFLRRE----A 140
LpCBF2             GHSA---SSTCLNFADSAW-LLDLPP--TLS----DLAHVRRAALAAVAGFLRQE----A 142
TmCBFIIId-16       GR-----SAACLNFADSAW-LLAVPP--ALA----DLAAVRRAALAAVADFQRRH----A 187
TaCBFIIId-A15      GP-----STQCLNFADSAW-LLAVPS--ALP----DFADVRRAALSAVADFQRRE----A 132
TmCBF15            GP-----STPCLNFADSAW-LLAVPS--ALS----DFADVRRAALSAVADFQRRE----A 132
HvCBF15B           DR-----RAPCLNFADSVW-LLAVPS--ALS----DLADVRRAALSAVADFQRRE----A 132
HvCBF15A           DR-----RAPCLNFADSVW-LLAVPS--ALS----DLADVRRAALSAVADFQRRE----A 130
AsCBFIIId-16A      GRS----ASACLNFADSAW-LLDLPS--PLS----DLAAVRRVALAAVVRGQCRK----L 142
TaCBF12.2          GRTP----AARLNYPDSAW-LLAVPS--SLS----DLADVRRAAIGAVVDFLRRQEAGAS 147
TaCBFIIId-B12      GRTP----AARLNYPDSAW-LLAVPS--SLS----DLADVRRAAIGAVVDFLRRQEAGAS 147
HvCBFIIId-12       GRTP----AARLNFPDSAW-LLAVPS--SLS----DMADVRRAAIGAVVDFLRRQETGA- 146
TaCBF12.1          GRS-----AARLNFPDSAC-LLAVPS--SLS----SLADVRRAAIGAVVDFLRRQATIAG 146
AsCBFIIId-12       GHTS----AACLNFPDSAW-LLNVPS--NLS----DLADVRRAAIEAVVEFLRLEAIKD- 146
TaCBFIIId-A19      GRSPS--AAACLNFPDSAW-LLVMPP--RLS----DLADVRRAAIQAVAGFLRP------ 143
TaCBF19.1          GRSPS--AAACLNFPDSAW-LLVMPP--RLS----DLADVRRAAIQAVAGFLRL------ 143
TaCBFIIId-D19      GRSPS--AAACLNFPDSAW-LLVMPP--RLS----DLADVRRAAIQAVVGFLRL------ 143
TaCBFIIId-B19      GRSPC--AAACLNFPDSAW-LLVMPP--RLS----DLADVRRTAIQAVASFLRL------ 143
TaCBF19.2          GHSAS--AAACLNFPDSAW-LLVMPP--WLS----DLADIRRAAIEAVAIFLCL------ 143
ZmCBFIIIb-1A       ----G-RGAGRLNFPDSAR-LLAVPPPSALPG----LDDARRAALEAVAEFQRRSGSGSG 163
ZmCBF1             ----G-RGAACLNFPDSAW-LLAVPPPPALSGG---LDGARRAALEAVAEFQRRR-FGAA 152
SbCBF1a            R---G-GAAGCLNFPDSAW-LLAVPPPSAISG----LDDARRAALEAVAEFQRRFGAAAA 166
OsCBFIIIb-1H       GGAGG-GGAACLNFQDSAW-LLAVPP--AAPSD---LAGVRRAATEAVAGFLQRNKTTNG 149
TmCBFIIIb-18       ----G-RSAACLNFRDSAW-LLSVPP--AFSN----LSDVRRAAVQAVADFLRRPEATGA 142
TaCBFIIIc-3.2      G-----RSAACLNFADSAW-LLAVPP--ALAD----LSDVRRAAVEAVADSQRREAANGS 145
TmCBF3             G-----RSAACLNFADSAW-LLAVPP--ALAD----LGDVRRAAVEAVADFQRREAANGS 145
TaCBFIIIc-D3       G-----RSAACLNFADSAW-LLAVPP--ALSD----LGDVRRAAVEAVADFQRREAANGS 145
TaCBFIIIc-3        G-----RSAACLNFADSAW-LLAVPS--ALSD----LGDVRRAAVEAVANLQRRKAGNGS 137
HvCBFIIIc-3        G-----RSAACLNFADSAW-LLAVPS--ALSD----LADVRRAAVEAVADFQRREAADGS 145
TaCBFIIIc-B10      G-----RSAALLNFPDSAW-LLAVPS--AHSD----LADVRRAAVEAVADLQRREAAGGS 132
TmCBFIIIc-10       G-----RSAARLNFPDSAW-LLAVPS--AHSD----LADVRRAAVEAVADLQRREAAGGS 136
HvCBFIIIc-10A      G-----RSAARLNFSDSAW-LLAVPS--AHSD----LADVRRAAVEAVSDLQRREAAGGS 140
HvCBFIIIc-13       G-----RSARRLNFADSAW-LLAVPF--ALSD----LADVRRTGLQAVANFQRREAASGL 139
```

```
TmCBFIIIc-13       G-----RSARCLNFADSAW-LLAVPS--ALSD----LADVRRAALQAVADFQRWEAANGL 140
Bradi4g35630       ------SSAARLNFPDSAW-LLNVPP--ALAG----LADVRSAAVQAVADFERRETAAID 149
FaCBFIIIa-6        AG------DACLNFADSAE-LLAVPAS--YRN----LAEVRHAVTEAVEDFERR-----Q 149
LpCBFIIIa-6        AG------DVCLNFADSAE-LLDMPASS-YRS----LDEVRHAVTEAVEEFERR-----Q 145
TaCBFIIIa-6        AGG-----AGCLNFADSAE-LLAVPAASSYRS----LDEVRHAVVEAVEDFLRR-----E 154
TaCBFIIIa-D6       AGG-----AGCLNFADSAE-LLAVPAASSYRS----LDEVRHAVVEAVEDLLRR-----E 154
HvCBFIIIa-6        AGG-----AGCLNFADSAE-LLAVPAASSYRS----LDEVRHAVVEAVEDLLRR-----E 156
TaCBFIIIa-6.2      AGG-----AACLNFADSAE-LLSVPVASSYRS----LDEVRHAVVEAVEDLLRR-----E 156
Bradi4g35650       GAG-----AACLNFADSAQ-LLAVPAS--YRS----LDDVRLAVVEAVEDFLRRCEARAE 153
OsCBFIIIa-1A       AGGGGGGGACCLNFADSAW-LLAVPRS--YRT----LADVRHAVAEAVEDFFRR------ 147
SbCBFIIIa-6        GAG------ACLNFADSAW-LLAVPAS--YAS----LAEVRHAVAEAVEEFLLR-----E 142
OsCBFIII-1I        GRA-----AACLNFADSAW-LLAVPPPATLRC----AADVQRAVARALEDFEQRESSSSV 141
OsCBFIII-1D        G------ASASLNFADSAW-LLHVPRAPVVSG---LRPPAARCATRCLQGHRRVPAPGRG 139
OsCBFIII-1J        G------ASASLNFADSAW-LLHVPRAPVASGH-DQLPDVQRAASEAVAEFQRRGS---- 142
HvCBFIVa-2A        G------HDACLNFADSAWRMMPVHAT--GSFRLAPAQEIKDAVAVALEVFQG-QHPADA 151
HvCBF2B            G------HDACLNFADSAWRMMPVHAT--GSFRLAPAQEIKDAVAVALEVFQG-QHPADA 151
HbCBFIVa-2         G------HDACLNFADSAWRMMPVHAT--GSFRFAPAQEIKDAVAVALEVFKG-QHPADA 151
TaCBF2.2           G------HDACLNFADSAWRMMPVHAT--GSFRFAPAQEIKDAVAVALEAFQE-QHHADA 151
TaCBF2.1           G------RDACLNFADSAWRMMPVHAA--GSFKLAAAQEIKDAVAVALKEFQEQQRPADE 153
TaCBFIVa-2.3       G------RDACLNFADSAWRMMPVHAA--GSFKLAAAQEIKDAVAVALKEFQEQQRPADE 153
TmCBFIVa-2         G------RDACLNFADSAWRMMPVHAA--GSFKLAAAQEIKDAVAVALKAFQEQQRPADA 155
TaCBFIVa-2.2       G------RGACLNFADSAWRMMPVHAA--GSFKLAAAQEIKDAVAVALKEFQEQQRPADV 127
FaCBFIVa-2         G------RDACLNFADSAWRMQPVLPAGAGSVCFGGAQEVKDAVAAAVEAFQEEEHHVES 146
TaCBF14.1          G------RAACLNFADSAWRMLPVLAG-----PFSTAKEIKDAVAVAVLAFQRQHPVASM 136
TaCBF14.2          G------RAACLNFADSAWRMLPVLAG-----PFSTAKEIKDAVAVAVLAFQRQHPVASM 136
TmCBF14            G------RAACLNFADSAWRMLPVLAG-----PFSTAKEIKDAVAVAVLAFQRQHPVASM 136
TaCBFIVc-B14       G------RAACLNFADSAWRMLPVLAG-----PFSTAKEIKDAVAVAVLAFQRQHRVASM 136
HvCBFIVc-14        G------RAACLNFADSAWRMLPLLAG-----PFSTAKEIKDAVAVAVLAFQRQHPVAST 136
TaCBFIVb-A20       D------RAACLNFADSAWRMLPVLAA--GSSRFSSAREIKDAVAVAVMEFQRQRPVLST 139
TaCBF20.0          D------RAACLNFADSAWRMLPVLAA--GSSRFSSAREIKDAVAVAVMEFQRQRPVLST 139
TaCBFIVb-B20       D------RAACLNFADSAWRMLPVLAA--GSSRFSSAREIKDAVAIAIVEFQRQRPVVST 139
TaCBFIVb-D20       D------RAACLNFADSAWRMLPVLAA--GSSRFSSAREIKDAVAIAVLEFQRQRPVVST 139
ScCBFIVb-20        D------RAACLNFADSAWRMLPVLAA--GSSRFSSAREIKDAVAVAVVEFQRQRPFVST 139
TaCBFIVb-21        G------SAACLNFADSAWRMLPVLAA--GSSSFSSAREIKDAVAVAVVAFQRQRSVAST 138
```

```
TaCBF21.0          G------SAACLNFADSAWRMLPVLAA--GSSSFSSAREIKDAVAVAVVAFQRQRSIAST 138
TaCBFIVb-D21       G------SAACLNFADSAWRMLPVLAA--GSSSFSSAREIKDAVAVAVVAFQRQRPVAST 138
TaCBF4.0           G------RAACLNFADSAWRMLPVLAA--GSFGFGSAREIKLAVAVAVVAFQQQQIILPV 143
TaCBFIVd-B4        G------RAACLNFADSAWRMLPVLAA--GSFCFGSAREIKLAVAVAVVAFQLQQNIPPA 143
HvCBFIVd-4A        G------RAACLNFADSAWRMRPVLATT-GSFGFSSTREIKLAVAVAVVAFQQQQIILPV 145
HvCBFIVd-9         G------RAACLNFADSAWRMLPVLAA--GSFGFGSAREVKAAVAVAVVAFQRRQ-IIPV 149
ScCBFIVd-9A        G------RAACLNFADSAWRMLPVLAA--GSFGFDSAREVKAAVAVAVVAFQRKQ-IIPV 142
TaCBFIVd-9         G------RAACLNFADSAWRMLPVLAA--GSFGFGSASEIKAAVAVAVVAFQRKQIVLPV 143
TaCBF9.0           G------RAACLNFADSAWRMLPVLAA--GSFGFGSASEIKAAVAVAVVAFQRKQIVLPV 143
TmCBF9             G------RAACLNFADSAWRMLPVLAA--GSFGFGSASEIKAAVAVAVVAFQRKQIVLPV 143
TaCBFIVd-B9        G------RAACLNFADSAWRMLPVLAA--GSFGFDSAREIKAAVAVAVVAFQRKQ-IIPV 142
TaCBFIVd-D9        G------RAACLNFADSAWRMLPVLAA--GSFGFGSASEIKAAVAVAVVAFQRKQ-IIPV 143
ScCBFIVd-9B        C------RAACLNFADSAWRMLPVLAA--GSFGFGSPREIKAAVAVAVIAFQRKQ-IIPV 142
TaCBF22.0          G------RAACLNFADSAWRMLPVLAA--GSFGFGSAREIKAAVAVAVVAFQKEQ-IIPV 142
TaCBFIVd-D22       G------RAACLNFADSAWRMLPVLAP--GSFGFGSAREIKAAVAVAVVAFKKQQ-IIPV 142
TmCBFIVd-4         G------RKACLNFADSAWRMLPVLAA--GSFGFGSAREIKTAVAVAVLAFQRQQIVLPV 136
TaCBFIVd-B22       G------RAACLNFADSAWRMLPVLAA--GSFGFGSAREIKAAVAVAVVAFQKEQ-IIPV 142
OsCBFIV-1B         G------RAACLNFADSAWRMPPVPAS----AALAGARGVRDAVAVAVEAFQRQSAAPSS 136
                             **: **

OsCBFIa-1G         PPDESN---------AATEVAAAASGATN------------------------------ 162
HvCBFIa-1          APDQSN---------APAEEVAAAP--------------------------------- 159
HvCBF11            APGQRN---------AATVAAATAPPVAL---------------------------- 160
TaCBFIa-A11        EPGQRN---------AATTEAPAASPADA---------------------------- 160
OsCBFIa-1E         APGQHNA--------AAEAAAAVAAQATA---------------------------- 161
HvCBFII-5          HDAASSS--------APALKAASAPAPA----------------------------- 146
TmCBFII-5          HDAPRSS----------VSAASAPAPA----------------------------- 144
TaCBFII-5          AAPASSS---------AVAAVASAPAPPM---------------------------- 152
TaCBFII-5.2        AAP--------------VAAEASAPAPAM---------------------------- 148
TaCBFII-5.3        AAPASSS--------SVAAAVASAPAPP----------------------------- 163
OsCBFII-1C         AAAASSS--------AAAVEASAAAAPAM---------------------------- 147
SbCBFII-5          SSTPDAS--------AAAAAVAVASSASVGQATPSSSAY------------------- 177
ZmCBF4             ASQQDET--------AAVAADVVAPS------------------------------- 163
OsCBFI-1F          -RPPPPL--------ALPEDPQEGTSGGG---------------------------- 154
```

```
Bradi4g35580       RGDATVPV-----VDEVASSASALPSYMDEAS---------------------------- 172
Bradi4g35640       RGDATVPV-----VDEVASSASALPSYMDEAS---------------------------- 172
Bradi4g35570       RGDATVPV-----VEDIASSATALPSYMDDAS---------------------------- 203
Bradi1g57970       DGAVNVPD-----IDDAVFSLATTSQPCA------------------------------- 164
Bradi1g77120       NGAAAVLD-----LDEAAVSWATTSSQLARAN---------------------------- 184
Bradi3g57360       -EAATVPV-----VNEST-SWAT-AEPSA------------------------------- 164
Bradi4g35600       ATVPVPVP-----VAASEITSIASMVPVNDAG---------------------------- 171
Bradi4g35610       NNVAVAN------VGAN--VPIASMAPVDNAG---------------------------- 169
Bradi4g35590       ATVQVVNE-----PPIN---PAFAPLPPDNAV---------------------------- 170
Bradi4g35620       ANGAAAVP-----VVNEAVSNEFASSS-DNAV---------------------------- 171
Bradi2g60331       APSIVSVV-----IPVAVVPVAVVPVDHGAAG---------------------------- 162
Bradi2g60340       APSIVSVV-----IPVAVVPVAVVPVDNGAAG---------------------------- 161
TmCBFIIId-17       SGAATVPV-----DEVFDTSSADDAGSWSWATPQPSCAAADGMFEVPAAAL--------A 184
TaCBF17.0          SGAATVPV-----DEVFDTSSADDAGSWSWATPQPSCAAADGVFEVPAAAL--------A 186
AsCBFIIId-16B      AGGAAN------VPADEDTSSASA----DNAGG--------------------------- 163
LpCBF2             DSGAAT------VRADEAAYSASVPSSVDNAGG--------------------------- 169
TmCBFIIId-16       SNSAAT------VPADEETSGASALSSADNASG--------------------------- 214
TaCBFIIId-A15      ASGAATRSLDATVPVDDGTCSQSAQSSMENTGS--------------------------- 165
TmCBF15            ASGAATTSLAATVPVDDGSCSQSAQSSMENTGS--------------------------- 165
HvCBF15B           ASGAATRAQAAAALIDEGTCSQSAQSSMENTGS--------------------------- 165
HvCBF15A           ASGAATRAQAAAALIDEGTCSQSAQSSMENTGS--------------------------- 163
AsCBFIIId-16A      VGDIAA----------------LLCRWDIRGA---------------------------- 158
TaCBF12.2          AGAVAEEA----HVDG-IASAASAPDNASSSAA--------------------------- 175
TaCBFIIId-B12      TGAVTEVA----SIDG-IASAASAPDNASSAAA--------------------------- 175
HvCBFIIId-12       -GAITEVT----SVDG-VASEAYAPGSASSSAA--------------------------- 173
TaCBF12.1          -ARAAEVV----PVNG-VASVAPAPGNARSSAT--------------------------- 173
AsCBFIIId-12       -GAAAVAV----PIDGVVASAALAPSSHADNAS--------------------------- 174
TaCBFIIId-A19      --EAATVV----PDVDEATSPVYLPSPVDN------------------------------ 167
TaCBF19.1          --EAATVV----PDVDEATSPVYLPSPVDN------------------------------ 167
TaCBFIIId-D19      --EAATVV----PDVDEATSPVYLPSPVDN------------------------------ 167
TaCBFIIId-B19      --EAATVV----PDVDEATSPVYLPSLVDN------------------------------ 167
TaCBF19.2          --EAAAVV----PIIDEATSPVYLPSPVDN------------------------------ 167
ZmCBFIIIb-1A       AA-------------DEATSGAS-PPSSSPSLPDV------------------------ 184
ZmCBF1             AA-------------DEATSGTS-PPSSS------------------------------ 167
```

```
SbCBF1a          AAAGGCGSV------DEATSGVSAPPLSTSSLPGIS---------------------- 196
OsCBFIIIb-1H     AS--VAEAM------DEATSGVSAPPPLAN----------------------------- 171
TmCBFIIIb-18     FAG----AA------QEVTSSVTVPSAAA------------------------------ 161
TaCBFIIIc-3.2    LTVTATVT-------EEASCGAPEES-SSES---------------------------- 168
TmCBF3           LTVTATVT-------EEASCGAPEES-SSES---------------------------- 168
TaCBFIIIc-D3     L--TATVT-------EEASCGAPEES-SSES---------------------------- 166
TaCBFIIIc-3      L--TATVT-------EEASCDAPEES-SSES---------------------------- 158
HvCBFIIIc-3      L--AIAVP-------KEASSGAPSLSPSSGS---------------------------- 167
TaCBFIIIc-B10    I--TATATAT---AAEEASCGAPAES-SSES---------------------------- 157
TmCBFIIIc-10     I--TATVN-------EEASCGAPAES-SSES---------------------------- 157
HvCBFIIIc-10A    I--SATVD-------EEASCGAPAES-SSES---------------------------- 161
HvCBFIIIc-13     I--TRTVA--------------------------------------------------- 145
TmCBFIIIc-13     V--TRTAAE-------QAPSSAPAQS-SSES---------------------------- 161
Bradi4g35630     I--AAAAT-------DEATSRVSERT--SSS---------------------------- 169
FaCBFIIIa-6      ELGEKDSL-------SGTSSSTPSSSSLTDDE--------------------------- 174
LpCBFIIIa-6      ALGEEDAL-------SGTESST-----LTDDE--------------------------- 165
TaCBFIIIa-6      AIAEEDAL-------SGTSSSAPSS--LTDDE--------------------------- 177
TaCBFIIIa-D6     AIAEDDAL-------SGTSSSAPSS--LTDDG--------------------------- 177
HvCBFIIIa-6      AHAEDDALS-----VSGTSSSAPSS--ITDDD--------------------------- 181
TaCBFIIIa-6.2    ALAEEDAL-------SGTSSSAPSP--LTDDE--------------------------- 179
Bradi4g35650     EEEEEDAL-------SGASSSLTDN--DTGDE--------------------------- 176
OsCBFIIIa-1A     -RLADDALS-----ATSSSSTTPSTPRTDDDE--------------------------- 173
SbCBFIIIa-6      VVQEDDAL-------SATSSTPPSSPSSSDDG--------------------------- 167
OsCBFIII-1I      FPLAIDVVA-----EDAMSATSEPSAASDDDA--------------------------- 168
OsCBFIII-1D      STATATATS-----GDAASTAPP-SAPVLSAKQC-------------------------- 167
OsCBFIII-1J      TAATATATS-----GDAASTAPPSSSPVLSPND--------------------------- 170
HvCBFIVa-2A      ----------------CTAEES-------------------------------------- 157
HvCBF2B          ----------------CTAEGS------------------------------------- 157
HbCBFIVa-2       ----------------CTAEES-------------------------------------- 157
TaCBF2.2         ----------------STTEAS------------------------------------- 157
TaCBF2.1         STAPS-----------STAEES------------------------------------- 164
TaCBFIVa-2.3     STAPS-----------STAEES------------------------------------- 164
TmCBFIVa-2       SKAPSS--------TDSTSEES------------------------------------- 169
TaCBFIVa-2.2     STAPS-----------STAEES------------------------------------- 138
```

```
FaCBFIVa-2          TAET-----------AKDEES------------------------------------ 156
TaCBF14.1           APLSPA--------RTTDDEKE---------------------------------- 150
TaCBF14.2           APLSPA--------RTTDDEKE---------------------------------- 150
TmCBF14             APLSPA--------RTTDDEKE---------------------------------- 150
TaCBFIVc-B14        APLSPA--------RTTDDEKE---------------------------------- 150
HvCBFIVc-14         APMSPA--------RTAVDEKE---------------------------------- 150
TaCBFIVb-A20        P-------------ETHDGEKD---------------------------------- 148
TaCBF20.0           P-------------ETHDGEKD---------------------------------- 148
TaCBFIVb-B20        S-------------ETHDGEKD---------------------------------- 148
TaCBFIVb-D20        S-------------EMHDGEKD---------------------------------- 148
ScCBFIVb-20         S-------------ETADGEKD---------------------------------- 148
TaCBFIVb-21         ----------------ADGEKD---------------------------------- 144
TaCBF21.0           ----------------ADGEKD---------------------------------- 144
TaCBFIVb-D21        ----------------ADGEKD---------------------------------- 144
TaCBF4.0            ACP-----------TVEAAASP---------------------------------- 154
TaCBFIVd-B4         ACP-----------TVEPTDSP---------------------------------- 154
HvCBFIVd-4A         ACP-----------SPEAPASP---------------------------------- 156
HvCBFIVd-9          AVA-----------VVALQKQQ-VPVAVAVVTLQQKQQQVPVAVAVAALQQQQVPVAVAV 197
ScCBFIVd-9A         AVA-----------VVALQKQQ-VPVAVAVVALQQRQ----------------VPVTVAV 174
TaCBFIVd-9          AVA-----------VVALQQKQ-VPIAVAVVALQQKQ----------------VPVAVAV 175
TaCBF9.0            AVA-----------VVALQQKQ-VPIAVAVVALQQKQ----------------VPVAVAV 175
TmCBF9              AVA-----------VVALQQKQ-VPIAVAVVALQQKQ----------------VPVAVAV 175
TaCBFIVd-B9         AVA-----------VVALQQQQ-VPVAVAVVALQQKQ----------------VPLAVAV 174
TaCBFIVd-D9         AVA-----------VVALQQQQ-VPVAVAVVALQQKQ----------------VPVAVAV 175
ScCBFIVd-9B         AVA-----------VVALQQQQ-VPVAVAVVALKQKQ----------------VPVAVAV 174
TaCBF22.0           AVA-----------VVALQKQQIIPVAVAVVALQKQQ----------------IPVAVAL 175
TaCBFIVd-D22        AVA-----------VVALQKQQIIPVAVAVVALQKQQ----------------IPVAVAL 175
TmCBFIVd-4          ACP-----------AAEPAVAP---------------------------------- 147
TaCBFIVd-B22        AVA-----------VVAIQKQQIIPVAVAVVAIQKQQ----------------IIPSPSWR 176
OsCBFIV-1B          PAETFA--------NDGDEEED---------------------------------- 150


OsCBFIa-1G          ------------------------------------SNAEQFASHPYYEV----MDDGL 181
HvCBFIa-1           ------------------------------------TMQFAGDPYYG-----MDDGM 175
```

```
HvCBF11          ------------------------------------GNAELVADSPYYP------MDGL 177
TaCBFIa-A11      ------------------------------------GNAELVANSPYHL------MDGL 177
OsCBFIa-1E       ------------------------------------ASAELFADFPCYP------MDGL 178
HvCBFII-5        -------------------------------MVMMQEPAAVPYDSYATALY-----GDLT 170
TmCBFII-5        -------------------------------MVITQEAAAAPYDSYA--MY-----GGLA 166
TaCBFII-5        -------------------------------MTMMHESAAVHYDDYPMQYG----YGGIG 177
TaCBFII-5.2      -------------------------------MAMMQEPSAVEYDDYPMQY------GGIG 171
TaCBFII-5.3      -------------------------------MTMMQ------FDDYAMQY------GGIG 180
OsCBFII-1C       -------------------------------MMQYQDDMAATPSSYDYAYY------GNM 170
SbCBFII-5        -------------------------------QAGDDATGAAMYGAEYAAAAM----YGAGM 203
ZmCBF4           -------------------------------QADDVAAAAA------AAAAM----YGGGM 183
OsCBFI-1F        ----------------------------------------ATATSGRPAAVF----VDEDA 171
Bradi4g35580     -----------------SWASSF----QPSEIGDFDVPVVVG--SGMFELD----MSGEM 205
Bradi4g35640     -----------------SWASSF----QPSEIGDFDVPVVVG--SGMFELD----MSGEM 205
Bradi4g35570     -----------------SWAASF----QPCEIGNFDVPV------GMFELD----MAGEM 232
Bradi1g57970     -----------------NNGVLL----G--VFDDFEVPVATG--SGLFELD----VSGDM 195
Bradi1g77120     -----------------NNGGML--------MFDFEVPVAAMGSDGMFELED---ICGET 216
Bradi3g57360     -----------------NNG-------------IFEVPVAMG-SGGMFELD----MSGEM 189
Bradi4g35600     -----------------SWPSFQ----P-CVAGMFDGP-VVMG-SDMFELD----MPDEM 203
Bradi4g35610     -----------------PPPSFR----P-SAAGMFEAP-VAMG-SDMFELD----MPDEM 201
Bradi4g35590     -----------------PWASSQ----PSATTGMFGEP-VAMD-SNMFELD----MTSEM 203
Bradi4g35620     -----------------SWATSQ----ASANNGTSEEP-VVMG-SEMFELG----MPEEM 204
Bradi2g60331     -----------------AWPSFL----GVLFEVPLPVPPMAMG-SGMLELELELDMPSEM 200
Bradi2g60340     -----------------AWPSFL----GVLFEVPLPVPPMAMG-SGMLELELELDMPSEM 199
TmCBFIIId-17     SDMFDFEFDVSWVMDLGSPATSQ----PGCADKVLEVPAAALGGGDMFEFDLELDMSGEM 240
TaCBF17.0        SDMFDFEFDVSWVMDLGSPAASQ----PGCADKVLEVQAAALGGGDMFEFDLELDMSGEM 242
AsCBFIIId-16B    -----------------SSATSQ---PYVDGTFDFEMPAGMG--SDIFELD----MSGEM 197
LpCBF2           -----------------LSATSQ---PYADG--MFELPSALN--SDMFELD----MSGEM 201
TmCBFIIId-16     -----------------SSATSQ---PWAEG--TFEVPSALG--SDMFELD----LSGEM 246
TaCBFIIId-A15    -----------------SWTSSS--LPSGNG--MFEVPATLG--CDMFELD----MSGEM 198
TmCBF15          -----------------SWTSSSSSLPSGDG--MFAVPATLG--CNMFELD----MSGEM 200
HvCBF15B         -----------------SSTSSS--LPSADG--MLEVPATLG--SNMFELD----MSGEM 198
HvCBF15A         -----------------SSTSSS--LPSADG--MLEVPATLG--SNMFELD----MSGEM 196
AsCBFIIId-16A    -----------------GRAGQR----------HVRAGRVRG--NGLRHVL----RGPRG 185
```

```
TaCBF12.2          -----------------AA-HSQ----PPCANAGYEVP--DALCHDMFELH----TSGEM 207
TaCBFIIId-B12      -----------------AS-HSQ----PPCANAEFEVP--DALCHDMFELH----TSGEM 207
HvCBFIIId-12       -----------------SSSHYQ----LPCANAEFVVP--DALCHDMLELH----TSGEM 206
TaCBF12.1          -----------------SS---Q----QPCANAESEAP--DALRGGLPELH----TSGEM 203
AsCBFIIId-12       -----------------PA---A----TSQPSAASEVP--EALGGDMFELH----TSGEM 204
TaCBFIIId-A19      ------------------------------ADEVFQVPTFSPLGSDMFELD----MSGEM 193
TaCBF19.1          ------------------------------ADEVFQVPTFSPLGSDMFELD----MSGEM 193
TaCBFIIId-D19      ------------------------------ADQVFQVPTFSPLGSDMFELD----MSGEM 193
TaCBFIIId-B19      ------------------------------ADEVFQVPTFSPLGSDMFELD----MSGEM 193
TaCBF19.2          ------------------------------AYEVFQVPTFSAQSSDMFELD----MSGEM 193
ZmCBFIIIb-1A       ----------------SAAGSPAAALEHVPVKADEA--VALDLDGDVFGP----DWFGDM 222
ZmCBF1             ----------------SSATKPAPAIERVPVEASET--VALD--GAVFEP----DWFGDM 203
SbCBF1a            ----------------SGSPAPAPELEQVPVKANETA-TALD--GDVFEPA---DWFGDM 234
OsCBFIIIb-1H       ----------------NAGSSETPGPSSIDGTADTAAGAALD----MFEL----DFFGEM 207
TmCBFIIIb-18       ----------------CSVPSSETAQTSGDANFEEPGALSMD----MFDLD---CLFGET 198
TaCBFIIIc-3.2      ----------------DSVGSSETSEPSAD-GEF-EVPVAVD--TDMF-RL---DLFPEL 204
TmCBF3             ----------------DSVGSSETSEPSAD-GEF-EVPVAVD--TDMF-RL---DLFPEL 204
TaCBFIIIc-D3       ----------------DSAGSSETSEPSAD-AEF-EVPVAVD--TDMFSRL---DLFPEM 203
TaCBFIIIc-3        ----------------DSAGSSETSEPSAD-REF-EVPVAVD--TDMF-GL---DLFPEM 194
HvCBFIIIc-3        ----------------DSAGSTGTSEPSAN-GEF-EGPVVMD--SEMF-RL---DLFPEM 203
TaCBFIIIc-B10      ----------------DDAGSSETSKPSAD-GDF-AVPGGMD--IEMFSRL---DLFPEM 194
TmCBFIIIc-10       ----------------DDAGSSETSKPSAR-GDF-ALPGGMD--VEMFSRL---DLFPET 194
HvCBFIIIc-10A      ----------------DGAGSSETSKPSAD-GDL-AVPVGMD--IEMF-RL---DFFPEM 197
HvCBFIIIc-13       ----------------DSADSSETSEPSAD-GDF-ELPVAMD--SDMF-RL---DFLPEM 181
TmCBFIIIc-13       ----------------DSADSSETSEASAD-GEF-EVLATMD--IDMF-RL---DLFPEM 197
Bradi4g35630       ----------------GNAGSLETSEPSADDARFGEVPVAMD--SDMFSRL---DLFREM 208
FaCBFIIIa-6        ----------------EASSQ-------------------ADNSPFELE----VLSDM 193
LpCBFIIIa-6        ----------------ESS-------------------------TPFELD----VLSDM 179
TaCBFIIIa-6        ----------------SSSS--------------------PPEDSPFELD----VLSDM 196
TaCBFIIIa-D6       ----------------SSSSPL------------------PEEDSPFELD----VLSDM 198
HvCBFIIIa-6        ----------------SSSSP-------------------ADEGSPFELD----VLSDM 201
TaCBFIIIa-6.2      ----------------SSSSPL------------------PEEDSPFEQD----VLSEM 200
Bradi4g35650       ----------------MTSSR-------------------SEEDSPFELD----VLSDM 196
OsCBFIIIa-1A       ----------------ESAATDG------------DESSSPASDLAFELD----VLSDM 200
```

```
SbCBFIIIa-6       -----------------STSDGGES--------SESDSSPAATGASPFELD----VFNDM 198
OsCBFIII-1I       -----------------VTSSSS----------TTDADEEASPFELD----VVSDM 193
OsCBFIII-1D       -------------------EFIFLSSLDCWMLMSKLISSSRAKGSLCLRKN---PISFCM 205
OsCBFIII-1J       ----------------------------------DNASSASTPAVAAALDHG---DMFGGM 194
HvCBFIVa-2A       ------------------------------------TTPITSSDLSGLDDEH----WIGGM 178
HvCBF2B           ------------------------------------TTPITSSDLSGLDDEH----WIGGM 178
HbCBFIVa-2        ------------------------------------APSITSSDLSGLDDEH----WISGM 178
TaCBF2.2          ------------------------------------APSITSSDLSGLDDEL----LIDGM 178
TaCBF2.1          ------------------------------------ALSIIPSDLSGLDNEH----WIGGM 185
TaCBFIVa-2.3      ------------------------------------ALSIIPSDLSGLDNEH----WIGGM 185
TmCBFIVa-2        ------------------------------------APSITSNDLSGLDDEH----WIGGM 190
TaCBFIVa-2.2      ------------------------------------ALSIIPSDLSGLDNEH----WIGGM 159
FaCBFIVa-2        ------------------------------------ALSMS-SDLSEHDDER----WIDGM 176
TaCBFIVc-14       ----IDGLPAPSA----------------------LSMSS----ELLNEH----WFGGM 175
TmCBF14           ----IDGSPAPSA----------------------LSMSS----ELLNEH----WFGGM 175
TaCBFIVc-B14      ----IDGSPAPSA----------------------LFMSS----ELLNEH----WFGGM 175
HvCBFIVc-14       ----VDGSPAPSA----------------------LFMSS----ELLNEH----WFGGM 175
TaCBFIVb-A20      ----VQGSPTPSE----------------------LSTSS----DLLDEH----WFGGM 173
TaCBF20.0         ----VQGSPTPSE----------------------LSTSC----DLLDEH----WFGGM 173
TaCBFIVb-B20      ----VQGSPTPSE----------------------LSMSS----DLLDEH----WFGGM 173
TaCBFIVb-D20      ----AQGSPTPSE----------------------LSTSS----DLLDEH----WFGGM 173
ScCBFIVb-20       ----VQGSPRPSE----------------------LSTSS----DLLDEH----WFSGM 173
TaCBFIVb-21       ----VQGSPTPSE----------------------LSTSS----DLLDEH----WFGGT 169
TaCBF21.0         ----VQGSPTPSE----------------------LSTSS----DLLDEH----WFGGT 169
TaCBFIVb-D21      ----VQGSPTPSE----------------------LSTSS----DLLDEH----WFGGT 169
TaCBF4.0          ----------SNS----------------------LFYMSSVDLLELDEEQ----WFGGM 178
TaCBFIVd-B4       ----------SNS----------------------LFYMSSGDLLELDEEH----WFGGM 178
HvCBFIVd-4A       ----------SAA----------------------LFYISSGDLLELDEEQ----WFGGM 180
HvCBFIVd-9        VALQQLQVPVAVAVVALQEQQ-IILPVACLAP---EFYMSSGDLLELDEEQ----WFGGM 249
ScCBFIVd-9A       VALQKLQVPVAVAVVALQKKQ-IILPAACLAP---EFYMSSGDLLELDEEQ----WFGGM 226
TaCBFIVd-9        VALQQLPVPVAVAVVALQQQQ-IILPVACLAP---EFYMSSGDLLELDEEQ----WFGGM 227
TaCBF9.0          VALQQLPVPVAVAVVALQQQQ-IILPVACLAP---EFYMSSGDLLELDEEQ----WFGGM 227
TmCBF9            VALQQLPVPVAVAVVALQQQQ-IILPVACLAP---EFYMSSGDLLELDEEQ----WFGGM 227
TaCBFIVd-B9       VALQQLQVPVAVAVVALQQQQQIILPVACLAP---EFYMSSGDLLELDEEQ----WFGGM 227
```

```
TaCBFIVd-D9        VALQQLQVPVAVAVVALQQQQ-IILPVACLAP---EFYMSSGDLLELDEEQ----WFGGM 227
ScCBFIVd-9B        VALQQLHVPVAVAVVALQQQQ-IILPVACLAP---EFYMSSGDLLELDEEH----WFGGM 226
TaCBF22.0          VALQEQQVPVAVAVVALHRQQVPVACPATSGPGSALFYMSSSDLLELDEEQ----WFGGM 231
TaCBFIVd-D22       VALQEKQIPVAVAVVALHRQQVPVDDPATSGPGSALFYMSSSDLLELDEEQ----WFGGM 231
TmCBFIVd-4         ----------SGA---------------------LFSMSSGDLLELDDEQ----WFGGM 171
TaCBFIVd-B22       SRSSRFQWPSPSWRSRNSRSQSPSPSWRFIGS----RFQSRAREPPARAAL----CFTCR 228
OsCBFIV-1B         -----NKDVLPVA---------------------AAEVFDAGAFELDDGF----RFGGM 179

OsCBFIa-1G         DLGMQG--YLDMAQG----MLIDPPPMAGDP-------AVGSGEDDNDG---------- 217
HvCBFIa-1          DFGMQG--YLDMAQG----MLIAPPPLVG-P-------SATAGDGDDDG---------- 210
HvCBF11            ESEMQG--YLDMAHG----MLIEPPPMAW---------PSTWIEEDYDC---------- 211
TaCBFIa-A11        EFEMQG--YLDMAHG----MLIEPPPMAG---------PSTWIEEDYDC---------- 211
OsCBFIa-1E         EFEMQG--YLDMAQG----MLIEPPPLAG---------QSTWAEEDYDC---------- 212
HvCBFII-5          DLDMHSYYCYDGMSG-GGDWQSISRMDGADE-------D----GIYGAG---------- 207
TmCBFII-5          DLEQHSHCYYDGMSG-SGDWQSISHMNVADE-------D----GGYGAG---------- 203
TaCBFII-5          DLDQDS-YYYDGMSAAGGDWQSGSHMDGADD-------DCNDSGGYGAG---------- 218
TaCBFII-5.2        DFDQHS-YYYDGLSAGGGDWQSSSHMDGADD-------DSNCGGGYGAG---------- 212
TaCBFII-5.3        DLDQHS-YYYDGLSAAGGDWPSGSHMDGADD-------DCNGSGGYGAG---------- 221
OsCBFII-1C         DFDQPS-YYYDGMGG-GGEYQS-WQMDGDDD-------G--GAGGYGGG---------- 207
SbCBFII-5          DFDHSYYYD-GMVGGNE--WQS-AGSSGWHS-------NVDAGDDEGAG---------- 241
ZmCBF4             EFDHSYCYDDGMVSGSSDCWQSGAGAGGWHS-------IVDGDDDDGAS---------- 225
OsCBFI-1F          IFDMPG-LIDDMARG----MMLTPPAIGRSLDDW---AAIDDDDDHYHM---------- 212
Bradi4g35580       --DLG-ACYADLAEG----LLMEPPQTT---------PDTE--ACWGSG---YYYDGG-- 242
Bradi4g35640       --DLG-ACYADLAEG----LLMEPPQTT---------PDTE--ACWGSG---YYYDGG-- 242
Bradi4g35570       --DLG-VYYADLAEG----LLMEPPQMT---------PDTE--ACWEIG---YYSHGG-- 269
Bradi1g57970       --ELG-MYYADLAEG----LLMEPPSPV---------PDAG--ASLESR---DYGHGGS- 233
Bradi1g77120       --DLD-MYYTELAGG----LLMEPP------------PDAG--ACWESR---DAG----- 247
Bradi3g57360       --DVG-MYYADLAEG----LLMEPPQTT--P------DDTG--ACWESG---EYA--EL- 226
Bradi4g35600       --DLG-MYYADLAEG----LLMEPPLP--A-------PDTG--ACWEIG---EYGDGG-- 240
Bradi4g35610       --DLG-MYYTDLAEG----LLMEPPPP--A-------PDAG--ACWESG---DCGNGG-- 238
Bradi4g35590       --DVG-RYYADLAEG----LLMEPPQP--A-------PDTG--ACWESG---DDGD---- 238
Bradi4g35620       --DLG-MYYKDLAEG----LLMEPPPPPPA-------PNTGTGVCWESG---DYGDGG-- 245
Bradi2g60331       --DLG-MHYADLAEG----LLMEPPPD--T-------AEAP---CWESG---DYGHGD-- 236
```

```
Bradi2g60340       --DLG-MHYADLAEG----LLMEPPQD--T-------NEAS---CWESG---DFGDG--- 234
TmCBFIIId-17       --NLVGSYYADFAEG----LLLEPPQP----------ADAT-EARWRNG---DYCGGDGG 280
TaCBF17.0          --DLVGSYYADFAEG----LLLEPPQP----------ADAT-EARWRNG---DYCGGDGG 282
AsCBFIIId-16B      --DQG-TYYAGLAEG----LLLEPPP----------QHAG--ACWDTG---DGG----- 229
LpCBF2             --DLG-TYYAGLAEG----LLLDPPPP----------EHTS--ACWDTG---DGG----- 234
TmCBFIIId-16       --DLG-TYYADLADG----LLLEPPPS----------LDSG--ACWDTG---DGG----- 279
TaCBFIIId-A15      --DLD-TYYAYFAEG----LLLEPPQP----------PVAG--ACWDTE---GGG----- 231
TmCBF15            --DLD-TYYAYFAEG----LLLEPPQP----------PVAG--ACWDTE---GGG----- 233
HvCBF15B           --DLD-TYYAYFAEG----LLLEPPQP----------PAAG--ACWDID---GGG----- 231
HvCBF15A           --DLD-TYYAYFAEG----LLLEPPQP----------PAAG--ACWDTD---GGG----- 229
AsCBFIIId-16A      --GIA-AGAAAAARR----RLLGHRR-----------------RWRS------------ 208
TaCBF12.2          --DAG-TYYADLAQG----LLLEPP-----------PPPS---SGASS---ERGD---- 238
TaCBFIIId-B12      --DAG-TYYADLAQG----LLLEPP-----------PPPS---SGASS---EQGD---- 238
HvCBFIIId-12       --DAG-TYYADLAQG----LLLEPP-----------PPPS---SGASS---EHGD---- 237
TaCBF12.1          --DVS-TYYADLAQG----LLLEPP-----------PPAA---SDCN----DGGD---- 233
AsCBFIIId-12       --GLG-TYYADLAEG----LLLEPP-----------PAAAS--SEHGG---DCGD---- 236
TaCBFIIId-A19      --DLD-AYYAGFAQG----MLLEPP-----------PTPAYWETGECG---DGGA---- 227
TaCBF19.1          --DLD-AYYAGFAQG----MLLEPP-----------PTPAYWETGECG---DGGA---- 227
TaCBFIIId-D19      --DLD-AYYAGFAQG----MLLEPP-----------PTPSYWENGECG---DGGA---- 227
TaCBFIIId-B19      --DLD-AYYAGFAQG----MLLEPP-----------PMPSYWENGECG---DGGA---- 227
TaCBF19.2          --DLD-AYYAGFAQG----MLLEPP-----------PTPTYWENGECG---DGGA---- 227
ZmCBFIIIb-1A       GLELDAYY-ASLAEG----LLVEPPP----------PP----AAWDHGDCCDSGA-AD- 260
ZmCBF1             --DLDLYY-ASLAEG----LLVEPPP----------PPPP--AAWDHGDCCDSG--AD- 240
SbCBF1a            DMELDVYY-ASLAEG----LLVEPPP----------AP-A--AAWDHGDCCDAG--AD- 272
OsCBFIIIb-1H       --DYDTYY-ASLAEG----LLMEPPP----------AA-T--ALWDNG---DEG--AD- 240
TmCBFIIIb-18       --DSDTYYYANLAQG----LLMEPPP----------SMATG-AYWDNGDCADGGAGAD- 239
TaCBFIIIc-3.2      --DLCS-YYASLAEA----LLLDPPA----------PVTTTYAYWDNG---DGGADV-- 241
TmCBF3             --DLCS-YYASLAEA----LLVDPPA----------PVTTTYAYWDNG---DGGADV-- 241
TaCBFIIIc-D3       --DLCS-YYASLAEA----LLVDPPS----------TVAIIDSYWDNG---DDGADV-- 240
TaCBFIIIc-3        --DLCS-YYASLAEA----LLVDPPA----------RVTTTDTYWDNG---DGGADV-- 231
HvCBFIIIc-3        --DLGS-YYMSLAEA----LLMDPPP----------TATIIHAYEDNG---DGGADV-- 240
TaCBFIIIc-B10      --DLGS-YYASLAEA----LLMDPPP----------VATGTGAYWDNG---ECGEAEGA 233
TmCBFIIIc-10       --DLGS-YYASLAEA----LLMDPPP----------VATGTGAYWDDG---EFGEVA-- 231
HvCBFIIIc-10A      --EFGS-YYASLAEA----LLMDPPP----------VANSTGAYWDNG---EFGEVA-- 234
```

```
HvCBFIIIc-13       --DLGS-YYVSLAEA----LLMDPPS-----------TATIIDAHRDNG---DGAKVF-- 218
TmCBFIIIc-13       --DLGS-YYVSLAEA----LLMDPPS-----------TATIIDAYRDNR---DGGADV-- 234
Bradi4g35630       --DLGMYYYTSLAEA----LLVDPPP-----------PETAAGAYWDNGDCVEGGADV-- 249
FaCBFIIIa-6        G---WDLYYSSLAQGM---MLMAPPFL---A------ASAAFGDYGEVN----------L 228
LpCBFIIIa-6        G---WDLYYASLAQG----MLMSSPFL---A------ASAALGDYGEAN----------L 213
TaCBFIIIa-6        G---WDLYYASLAQA----MLMAPPS----S------MAAALGDYGEV----------- 227
TaCBFIIIa-D6       G---WDLYYASLAQA----MLMAPPS----S------MAAALGDYGEV----------- 229
HvCBFIIIa-6        G---WDLYYASLAQG----MLMAPPA----S------LAAALGDYGEAH----------L 234
TaCBFIIIa-6.2      G---WDLYYASLAQA----MLMAPP-----A------AAAALGDYGEAH----------L 232
Bradi4g35650       G---WDLYYASLAQG----MLVEPP-----S------AAAALGDYGEVC----------L 228
OsCBFIIIa-1A       G---WDLYYASLAQG----MLMEPP-------------SAALGDDGDAI----------L 230
SbCBFIIIa-6        S---WDLYYASLAQG----MLVEPP--------------SAVTAFMDEG----------F 227
OsCBFIII-1I        G---WSLYYASLAEG----LLMEPPAS---G------ASSDDDDDAIVDS-------SDI 230
OsCBFIII-1D        V---TNSYTALLLEY----IILQMNSM----------IVLIHELSKYQV----------F 238
OsCBFIII-1J        R---TDLYFASLAQG----LLIEPPPP----------PTTAEGFCDDEG----------C 227
HvCBFIVa-2A        D---AGSYYASLAQG----MLMEPPA-----------AGGWREDDGEHDDG-------FN 213
HvCBF2B            D---AGSYYASLAQG----MLMEPPA-----------AGGWREDDGEHDDG-------FS 213
HbCBFIVa-2         D---AGSYYASLAQG----MLMEPPA-----------AGAWRED-REHDAG-------VD 212
TaCBF2.2           D---AGSYYASLAQG----MLMEPPA-----------AGAWRED-HEHDDG-------FD 212
TaCBF2.1           E---AGSYYASLAQG----MLMEPPA-----------DGA-WREDREHDDG-------FD 219
TaCBFIVa-2.3       E---AGSYYA------------------------------------------------- 192
TmCBFIVa-2         D---AGSYYANLAQG----MLMEPPA-----------AGA-WREDREQDDG-------VD 224
TaCBFIVa-2.2       E---AGSYYASLAQG----MLMEPPA-----------DGASWREDREHDDG-------FD 194
FaCBFIVa-2         D---AGSYYASLAQG----MLVEPPD-----------AGAWRED-GEHG-G-------VE 209
TaCBFIVc-14        D---AGSCYS---E------FMESPD------------TRPWREDFEL----------GG 201
TmCBF14            D---AGSCYS---E------FMESPD------------TRPWREDFEL----------GG 201
TaCBFIVc-B14       D---AGSFYS---EG----LFMESPD------------TRPWREDLEL----------CG 203
HvCBFIVc-14        D---AGSCYS---EG----MFIESPD------------TRPWREDLEL----------GG 203
TaCBFIVb-A20       N---AGSYYASLAQG----MLMEPPA------------ARARSEDGGEY---------SG 205
TaCBF20.0          N---AGSYYASLAQG----MLMEPPA------------ARARSEDGGEY---------SG 205
TaCBFIVb-B20       D---AGSYYASLAQG----MLMEPPS------------ARTWSEDGGEY---------SG 205
TaCBFIVb-D20       D---AGSYYASLAQG----MLMEPPS------------ARTWSEDGGEY---------SA 205
ScCBFIVb-20        D---AGSYYASLAQG----MLMEPPA------------ARAWSEDGGEY---------SG 205
TaCBFIVb-21        D---AGSYYS---PG----MFMEPPE------------RPGNRELGA-----------GE 196
```

```
TaCBF21.0          D---AGSYYS---PG----MFMEPPE------------RPENRQLGA-----------GD 196
TaCBFIVb-D21       N---AGSYYS---PG----MFMESPE------------QPENHELGG-----------GD 196
TaCBF4.0           D---AGSYYESLAQG----MLMAPPD------------DRARREDAEQTG--------VE 211
TaCBFIVd-B4        D---AGSYYDSLAQG----MLVEPPD------------DRARREDAEQTG--------VE 211
HvCBFIVd-4A        D---AGSYYASLAQG----MLVAPPD------------ERARPENREHSG--------VE 213
HvCBFIVd-9         D---AGSYYASLAQG----MLVAPPD------------ERARPEHGEQ----------TG 280
ScCBFIVd-9A        D---AGSYYASLAQG----MLVAPPD------------DRARPENGEQ----------SG 257
TaCBFIVd-9         E---AGSYYASLAQG----MLVAPPD------------ERARPESGEQ----------SG 258
TaCBF9.0           E---AGSYYASLAQG----MLVAPPD------------ERARPESGEQ----------SG 258
TmCBF9             E---AGSYYASLAQG----MLVAPPD------------ERARPESGEQ----------SG 258
TaCBFIVd-B9        E---AGSYYASLAQG----MLVAPPD------------ERAGPEHGEQ----------SG 258
TaCBFIVd-D9        D---AGSYYASLAQG----MLVAPPD------------ERARPEHGEQ----------SG 258
ScCBFIVd-9B        D---AGSYYASLAQG----MLVAPPD------------ERARPENGEQERR-------PD 260
TaCBF22.0          E---AGSYYASLAQG----MLVAPPD------------ERARPEDGEQ----------SG 262
TaCBFIVd-D22       D---AGSYYASLAQG----MLVAPPD------------ERARPEDGEQ----------SG 262
TmCBFIVd-4         V---AGSYYESLAQG----MLVEPPD------------AGAWREDSEHSGV-------AE 205
TaCBFIVd-B22       P---ATCWSSTRSSG----LAAWRPGRTTRAWRRGCSWHRRTKERGRRTAS-------RA 274
OsCBFIV-1B         D---AGSYYASLAQG----LLVEPPA------------AGAWWEDGEL----------AG 210

OsCBFIa-1G         -EVQLWSY------------------------------ 224
HvCBFIa-1          -EVSLWSY------------------------------ 217
HvCBF11            -EISLWNY------------------------------ 218
TaCBFIa-A11        -EVSLWNY------------------------------ 218
OsCBFIa-1E         -EVNLWSY------------------------------ 219
HvCBFII-5          -DVALWSY------------------------------ 214
TmCBFII-5          -DVALWSY------------------------------ 210
TaCBFII-5          -EVPLWSY------------------------------ 225
TaCBFII-5.2        -EVPLWSY------------------------------ 219
TaCBFII-5.3        -EVALWSY------------------------------ 228
OsCBFII-1C         -DVTLWSY------------------------------ 214
SbCBFII-5          -DMSLWSYY----------------------------- 249
ZmCBF4             -DMTLWSY------------------------------ 232
OsCBFI-1F          -DYKLWMD------------------------------ 219
```

```
Bradi4g35580       ADAALWHYGTKHSILL---------------------- 258
Bradi4g35640       ADAALWHYGTKHSILL---------------------- 258
Bradi4g35570       AEATLWNY----------------------------- 277
Bradi1g57970       ADADLWSCY---------------------------- 242
Bradi1g77120       ADADLWSCY---------------------------- 256
Bradi3g57360       WGCEICSLWL--------------------------- 236
Bradi4g35600       TDATLWNY----------------------------- 248
Bradi4g35610       ADASLWSY----------------------------- 246
Bradi4g35590       -DATLWSYRNDLP------------------------ 250
Bradi4g35620       ADAALWSY----------------------------- 253
Bradi2g60331       -AGDLWSY----------------------------- 243
Bradi2g60340       --GDLWSY----------------------------- 240
TmCBFIIId-17       GDAALWSQ----------------------------- 288
TaCBF17.0          GDAAFWSQ----------------------------- 290
AsCBFIIId-16B      ADPALWSY----------------------------- 237
LpCBF2             ADSALWSY----------------------------- 242
TmCBFIIId-16       ADSGLWSY----------------------------- 287
TaCBFIIId-A15      ADAALWSY----------------------------- 239
TmCBF15            ADAALWSY----------------------------- 241
HvCBF15B           ADAALWSY----------------------------- 239
HvCBF15A           ADAALWSY----------------------------- 237
AsCBFIIId-16A      ------------------------------------
TaCBF12.2          -DAALWNH----------------------------- 245
TaCBFIIId-B12      -DAALWNH----------------------------- 245
HvCBFIIId-12       -DAALWNH----------------------------- 244
TaCBF12.1          -DAVLWSH----------------------------- 240
AsCBFIIId-12       -APDMMRRYGNGATEIRLSLAAE-------------- 258
TaCBFIIId-A19      -AAGLWSY----------------------------- 234
TaCBF19.1          -AAGLWSY----------------------------- 234
TaCBFIIId-D19      -AAGLWSY----------------------------- 234
TaCBFIIId-B19      -AAGLWSY----------------------------- 234
TaCBF19.2          -AAGLWSY----------------------------- 234
ZmCBFIIIb-1A       --VALWSYY---------------------------- 267
ZmCBF1             --VALWSY----------------------------- 246
```

```
SbCBF1a           --VALWSY------------------------------ 278
OsCBFIIIb-1H      --IALWSY------------------------------ 246
TmCBFIIIb-18      --VALWSY------------------------------ 245
TaCBFIIIc-3.2     ---ALWSY------------------------------ 246
TmCBF3            ---ALWSY------------------------------ 246
TaCBFIIIc-D3      ---ALWSY------------------------------ 245
TaCBFIIIc-3       ---ALWS------------------------------- 235
HvCBFIIIc-3       ---RLWSYSVDM-------------------------- 249
TaCBFIIIc-B10     TEFALWS------------------------------- 240
TmCBFIIIc-10      TEFALWSL------------------------------ 239
HvCBFIIIc-10A     TEFALWS------------------------------- 241
HvCBFIIIc-13      ---LFWEKTLYSKDFEAVSNPSLNPEKGNIPSRLHVP- 252
TmCBFIIIc-13      ---ALWSY------------------------------ 239
Bradi4g35630      ---ALWSY------------------------------ 254
FaCBFIIIa-6       ADVPLWSYQS---------------------------- 238
LpCBFIIIa-6       ADVPLWSYLS---------------------------- 223
TaCBFIIIa-6       -DVPLWSYQS---------------------------- 236
TaCBFIIIa-D6      -DVPLWSYQS---------------------------- 238
HvCBFIIIa-6       ADVPLWSYQS---------------------------- 244
TaCBFIIIa-6.2     ADVPLWSYQS---------------------------- 242
Bradi4g35650      ADVPLWSYQS---------------------------- 238
OsCBFIIIa-1A      ADVPLWSY------------------------------ 238
SbCBFIIIa-6       ADVPLWSY------------------------------ 235
OsCBFIII-1I       ADVSLWSY------------------------------ 238
OsCBFIII-1D       LLLTMITHHLFQWRR----------------------- 253
OsCBFIII-1J       GGAEMELWS----------------------------- 236
HvCBFIVa-2A       TSASLWSY------------------------------ 221
HvCBF2B           TSTSLWSY------------------------------ 221
HbCBFIVa-2        TSTSLWSY------------------------------ 220
TaCBF2.2          TPTSLWSY------------------------------ 220
TaCBF2.1          T--SLWSY------------------------------ 225
TaCBFIVa-2.3      ------------------------------------
TmCBFIVa-2        T--SLWSYWLDGFG--CVKL------------------ 240
TaCBFIVa-2.2      T--SLWSYQCDQLIKQCKVLESTASARLCFTKYGKNRE 230
```

```
FaCBFIVa-2          T--SLWSYL---------------------------- 216
TaCBF14.1           VETPPWS--YLFD------------------------ 212
TaCBF14.2           VETPPWS--YLFD------------------------ 212
TmCBF14             VETPPWS--YLFD------------------------ 212
TaCBFIVc-B14        VETPPWS--YLFD------------------------ 214
HvCBFIVc-14         VQTPPWS--YLFD------------------------ 214
TaCBFIVb-A20        VQTPLWN--TYPTN----------------------- 217
TaCBF20.0           VQTPLWN--TYPTN----------------------- 217
TaCBFIVb-B20        VYTPLWN------------------------------ 212
TaCBFIVb-D20        VYTPLWN------------------------------ 212
ScCBFIVb-20         VHTPLWN------------------------------ 212
TaCBFIVb-21         VETLLW------------------------------- 202
TaCBF21.0           V-------------------------------------- 197
TaCBFIVb-D21        VKTPLW------------------------------- 202
TaCBFIVd-4          TPTPLWS--YLFD------------------------ 222
TaCBFIVd-B4         TPTPLWS--YLFD------------------------ 222
HvCBFIVd-4A         TPIPLWS--YLFDC----------------------- 225
HvCBFIVd-9          VQTPLWS--CLFD------------------------ 291
ScCBFIVd-9A         VQTPLWS--CLFD------------------------ 268
TaCBFIVd-9          VQTPLWS--CLFD------------------------ 269
TaCBF9.0            VQTPLWS--CLFD------------------------ 269
TmCBF9              VQTPLWS--CLFD------------------------ 269
TaCBFIVd-B9         VQTPLWS--CLFD------------------------ 269
TaCBFIVd-D9         VQTPLWS--CLLD------------------------ 269
ScCBFIVd-9B         AAMELFV--RLI------------------------- 270
TaCBF22.0           VQTPLWSQSHLFN------------------------ 275
TaCBFIVd-D22        VQTPLWSQSHLFN------------------------ 275
TmCBF4.0            TQTPLWS------------------------------ 212
TaCBFIVd-B22        ASRRRYGATCSPNLAV--------------------- 290
OsCBFIV-1B          SDMPLWSY----------------------------- 218
```

Table 3.2 Amino acid sequence alignment of amino terminal and CMIII - 3 of selected CBFs of the Poaceae family members. Sequences were aligned using ClustalW2 at EBI ClustalW server, (http://www.ebi.ac.uk/Tools/msa/clustalw2/). Accession numbers are listed in Table 3.1 in Chapter 3.

```
OsCBFI-1F          ----------------------------------------------------MDTE 4
OsCBFIII-1I        ----------------------------------------------------MFIR 4
HvCBFIVd-9         ------------------------------------------------MSNPIQTDVA 10
TmCBF9             ----------------------------------------------------MDVA 4
TaCBF9.0           ----------------------------------------------------MDVA 4
TaCBFIVd-9         ----------------------------------------------------MDVA 4
ScCBFIVd-9B        ----------------------------------------------------MDVA 4
HvCBFIVd-4A        ----------------------------------------------------MDVA 4
ScCBFIVd-9A        ----------------------------------------------------MDVA 4
TaCBFIVd-B9        ----------------------------------------------------MDVA 4
TaCBFIVd-D9        ----------------------------------------------------MDVA 4
TaCBF22.0          ----------------------------------------------------MDVA 4
TaCBFIVd-B22       ----------------------------------------------------MDVA 4
TaCBFIVd-D22       ----------------------------------------------------MDVA 4
TaCBF4.0           ----------------------------------------------------MDVA 4
TaCBFIVd-B4        ----------------------------------------------------MDVA 4
TmCBFIVd-4         ------------------------------------------------------------
TaCBFIVc-B14       ----------------------------------------------------MDAA 4
TmCBF14            ----------------------------------------------------MDAA 4
HvCBFIVc-14        ----------------------------------------------------MDAA 4
TaCBF14.1          ----------------------------------------------------MDAA 4
TaCBF14.2          ----------------------------------------------------MDAV 4
TaCBFIVb-21        ----------------------------------------------------MDA- 3
TaCBF21.0          ----------------------------------------------------MDA- 3
TaCBFIVb-D21       ----------------------------------------------------MDA- 3
TaCBFIVb-A20       ----------------------------------------------------MDTA 4
TaCBFIVb-B20       ----------------------------------------------------MDTA 4
TaCBF20.0          ----------------------------------------------------MDTA 4
TaCBFIVb-D20       ----------------------------------------------------MDTA 4
```

```
ScCBFIVb-20         ----------------------------------------------------MDAA 4
OsCBFIa-1G          ----------------------------------------------------MDVS 4
HvCBFIa-1           ----------------------------------------------------MDVG 4
HvCBFIa-11          ----------------------------------------------------MEWA 4
HvCBF11             ----------------------------------------------------MEWA 4
TaCBFIa-A11         ----------------------------------------------------MEWA 4
OsCBFIa-1E          ----------------------------------------------------MEWA 4
HvCBFIVa-2A         ----------------------------------------------------MDTV 4
HvCBF2B             ----------------------------------------------------MDTV 4
TaCBF2.2            ----------------------------------------------------MDTV 4
HbCBFIVa-2          ----------------------------------------------------MDTV 4
TaCBFIVa-A2         ----------------------------------------------------MDTN 4
TaCBFIVa-2.3        ----------------------------------------------------MDTI 4
TmCBFIVa-2          ----------------------------------------------------MDTA 4
FaCBFIVa-2          ----------------------------------------------------MDAA 4
OsCBFIV-1B          ----------------------------------------------------MEV- 3
OsCBFII-1C          ----------------------------------------------------MEY- 3
ZmCBF4              ----------------------------------------------------MEYA 4
HvCBFII-5           ----------------------------------------------------MDHC 4
TmCBFII-5           ----------------------------------------------------MDNS 4
TaCBFII-5           ----------------------------------------------------MDQY 4
TaCBFII-5.2         ----------------------------------------------------MDQY 4
TaCBFII-5.3         ----------------------------------------------------MDQY 4
SbCBFII-5           ----------------------------------------------------MEYG 4
FaCBFIIIa-6         ------------------------------------------------MCGIKREMS 9
LpCBFIIIa-6         ------------------------------------------------MCQIKKEMS 9
HvCBFIIIa-6         ------------------------------------------------MCQIKKEMS 9
TaCBFIIIa-6         ------------------------------------------------MCPIKKEMS 9
TaCBFIIIa-6.2       ------------------------------------------------MCPIKREMS 9
TaCBFIIIa-D6        ------------------------------------------------MCPIKREMS 9
SbCBFIIIa-6         ------------------------------------------------MCPIKKEMT 9
Bradi4g35650        ------------------------------------------------MCQIKKEMT 9
OsCBFIIIa-1A        ------------------------------------------------MCGIKQEMS 9
OsCBFIIIb-1H        ----------------------------------------------------MDMA 4
```

```
ZmCBFIIIb-1A        ----------------------------------------------------MDTA 4
SbCBF1a             ----------------------------------------------------MDMG 4
ZmCBF1              ----------------------------------------------------MDMG 4
TaCBFIIIc-3.2       ----------------------------------------------------MDMG 4
TmCBF3              ----------------------------------------------------MDMG 4
TaCBFIIIc-D3        ----------------------------------------------------MDMG 4
TaCBFIIIc-3         ------------------------------------------------------------
HvCBFIIIc-3         ----------------------------------------------------MDMG 4
Bradi4g35630        ----------------------------------------------------MDMG 4
HvCBFIIIc-10A       ----------------------------------------------------MDMG 4
TmCBFIIIc-10        ----------------------------------------------------MGMG 4
TaCBFIIIc-B10       ----------------------------------------------------MDMG 4
HvCBFIIIc-13        ------------------------------------------------------MG 2
TmCBFIIIc-13        ------------------------------------------------------------
TmCBFIIIb-18        ----------------------------------------------------MDMS 4
TaCBF12.2           ----------------------------------------------------MDTG 4
TaCBFIIId-B12       ----------------------------------------------------MDTG 4
HvCBFIIId-12        ----------------------------------------------------MDTV 4
TaCBF12.1           ----------------------------------------------------MDTA 4
AsCBFIIId-12        ----------------------------------------------------MDTG 4
AsCBFIIId-16A       ----------------------------------------------------MDMT 4
AsCBFIIId-16B       ----------------------------------------------------MDMS 4
LpCBF2              ----------------------------------------------------MDMT 4
TmCBFIIId-17        ----------------------------------------------------MDM- 3
TaCBF17.0           ----------------------------------------------------MDM- 3
TaCBF15.0           --------------------------------------------------MDMTG- 5
TmCBF15             --------------------------------------------------MDMTG- 5
TaCBFIIId-A15       --------------------------------------------------MDMTG- 5
TmCBFIIId-16        MPLVQTASGKTIKQCTPQDTKILTLPSQAQPALTLHRPPSTVRSSSSQHRPPSAMDMTG- 59
HvCBF15B            --------------------------------------------------MDMAG- 5
HvCBF15A            ----------------------------------------------------MAG- 3
Bradi2g60331        --------------------------------------------------MAMDG- 5
Bradi2g60340        --------------------------------------------------MAMDGF 6
Bradi4g35600        --------------------------------------------------MAMDG- 5
```

```
Bradi4g35610        ----------------------------------------------------MAMDG- 5
Bradi4g35620        ----------------------------------------------------MAMDG- 5
Bradi4g35590        ----------------------------------------------------MDMDG- 5
Bradi3g57360        ------------------------------------------------------MDG- 3
Bradi1g77120        ------------------------------------------------------MDG- 3
Bradi1g57970        --------------------------------------------------------MD-- 2
TaCBFIIId-A19       ------------------------------------------------------MDFG 4
TaCBF19.1           ------------------------------------------------------MDFG 4
TaCBFIIId-B19       ------------------------------------------------------MDMG 4
TaCBFIIId-D19       ------------------------------------------------------MDMG 4
TaCBF19.2           ------------------------------------------------------MDMA 4
Bradi4g35580        --------------------------------------------------------MD 2
Bradi4g35640        --------------------------------------------------------MD 2
Bradi4g35570        ---------------------------MIDVRTCIRHGRPPASLMHLFKELASCVEIFK 32
OsCBFIII-1D         ------------------------------------------------------MEKN 4
OsCBFIII-1J         ------------------------------------------------------MEKN 4

OsCBFI-1F           D---------TSSASS-----------------SSV-----SPPSSPGGG-------HHH 26
OsCBFIII-1I         ----------MRAAST-----------------TSS-----SEPCR-------------- 18
HvCBFIVd-9          G---------IASPSGQ----------------QEQQ----GHRTVS------------- 28
TmCBF9              D---------IASPSGQ----------------QEQ-----GHRTVS------------- 21
TaCBF9.0            D---------IASPSGQ----------------QEQ-----GHRTVS------------- 21
TaCBFIVd-9          D---------IASPSGQ----------------QEQ-----GHRTVS------------- 21
ScCBFIVd-9B         D---------IASPSGQ----------------QEQ-----GHRTVS------------- 21
HvCBFIVd-4A         D---------IASPSGQ----------------QKQQ----GHRTVS------------- 22
ScCBFIVd-9A         D---------IASRSGQ----------------QQQ-----GHRTVS------------- 21
TaCBFIVd-B9         D---------IASRSGQ----------------QEQ-----GHRTVS------------- 21
TaCBFIVd-D9         D---------IASPSGQ----------------QGHR----GHRTVS------------- 22
TaCBF22.0           D---------AASSSGQ----------------EQQ-----GHRTVS------------- 21
TaCBFIVd-B22        D---------AASPSGQ----------------EQQ-----GHRTVS------------- 21
TaCBFIVd-D22        D---------AASPSGQ----------------DQQ-----GHRTVS------------- 21
TaCBF4.0            D---------AASKSGQ----------------HEQ-----GHRTVS------------- 21
TaCBFIVd-B4         D---------AASKSGH----------------QEQ-----GHRTVS------------- 21
```

```
TmCBFIVd-4          ----------MPSGQ----------------EEQ-----RHRTVR------------ 14
TaCBFIVc-B14        D---------AASPC------------------DG------HRTVW------------ 17
TmCBF14             D---------AASPC------------------DG------HRTVW------------ 17
HvCBFIVc-14         D---------AASPC------------------DG------HRTVW------------ 17
TaCBF14.1           D---------AASPC------------------DG------HRTVW------------ 17
TaCBF14.2           D---------AASPC------------------DG------HRTVW------------ 17
TaCBFIVb-21         D---------AASPS------------------DQ------HRTVW------------ 16
TaCBF21.0           D---------AASPS------------------DQ------HRTVW------------ 16
TaCBFIVb-D21        D---------AASLS------------------DQ------HRTVW------------ 16
TaCBFIVb-A20        A---------PGSPR------------------EG------HRTVC------------ 17
TaCBFIVb-B20        A---------PGSPR------------------EG------HRTVC------------ 17
TaCBF20.0           A---------PGSPR------------------EG------HRTVC------------ 17
TaCBFIVb-D20        A---------AGSPR------------------EG------HRTVC------------ 17
ScCBFIVb-20         D---------AGSPR------------------FG------HRTVC------------ 17
OsCBFIa-1G          A---------ALSSDYS------SGTPSPVAADAD--DGSSAYMTVS------------ 34
HvCBFIa-1           ----------ALSSDYS------SGTPSPVGADGGNSEGFSTYMTVS------------ 35
HvCBFIa-11          C---------CGSGYSS------SGTQSPAAGDGEE----GSYMTVS------------ 32
HvCBF11             C---------CGSGYSS------SGTQSPAAGDGEE----GSYMTVS------------ 32
TaCBFIa-A11         Y---------SGGGHSS------SGTKSPAAGGREE----GSYMTVS------------ 32
OsCBFIa-1E          Y---------YGSGYSS------SGTPSPVGGDGDE----DSYMTVS------------ 32
HvCBFIVa-2A         A---------AWPQF--------EEQDYMTVWPEEQ-----EYRTVW------------ 29
HvCBF2B             A---------AWPQF--------EGQDYMTVWPEEQ-----EYRTVW------------ 29
TaCBF2.2            A---------AWQQF--------EGQEYMTVWPEEQ-----EYRTVW------------ 29
HbCBFIVa-2          A---------VWQQF--------DGQEYMTGCPEEQ-----EYRTVW------------ 29
TaCBFIVa-A2         A---------AWPQF--------DGQEYRTVWPEEQ-----EYRTVW------------ 29
TaCBFIVa-2.3        A---------AWPQF--------DGQEYRTVWPEEQ-----EYRTVW------------ 29
TmCBFIVa-2          G---------AWPHF--------EGQEYRTVWPEE------EYRTVW------------ 28
FaCBFIVa-2          V---------AAS----------------LSLQSGEQ-----EYRTVW------------ 22
OsCBFIV-1B          ----------------------------------EEA-----AYRTVW------------ 12
OsCBFII-1C          ------------------YE--QE-----------------EYATVT------------ 13
ZmCBF4              A---------VGYGYGYGYDERQEPAESADGGGGGD----DEYATVL------------ 38
HvCBFII-5           ----------------------------GVGLYG-----EYATVT------------ 16
TmCBFII-5           ----------------------------GVVFYG----GAYATVM------------ 17
TaCBFII-5           S---------YRGG----------------GDDNGQ----GGYATVT------------ 22
```

```
TaCBFII-5.2         S---------YGGGD---------------GDQAGQ----GGYATVT------------ 23
TaCBFII-5.3         N---------YGGGVAY-----YGSTTGGVGDNGQG----GGYATVT------------ 33
SbCBFII-5           ----------VADDYGYGYGGYDDQQDLPSSSSVDG----DEYATVL------------ 37
FaCBFIIIa-6         G---------E-SGLS------CSGEY--HSPSTSP-----EQQQGH--S-------QKQ 37
LpCBFIIIa-6         G---------E-SGSP------CNGEY--CSPSTSS-----EQKQ--------------Q 32
HvCBFIIIa-6         G---------E-SGSP------CSGENYYYSPSTSP-----EHQQAKQQA-------AWT 41
TaCBFIIIa-6         G---------E-SGSP------CSGEN-FYSPSTSR-----EHQQAK-QA-------AWT 39
TaCBFIIIa-6.2       G---------E-SGSPS----PCSGEN-FCSPSASP-----ERQQAR-QA-------GWT 41
TaCBFIIIa-D6        G---------E-SGSP------CSGES-FYSPSTSP-----ENQQAR-QA-------AWT 39
SbCBFIIIa-6         G---------E-SSSP------CS------SASTSS-----EHHQTV------------W 30
Bradi4g35650        G---------E-SGSP------CSGD----SASSSSQQQ--QQQQTV------------W 35
OsCBFIIIa-1A        G---------ESSGSP------CS------SASAER------QHQTV------------W 30
OsCBFIIIb-1H        G-----HEVNSSSSSS----------------GAESS----SS--------------SSG 25
ZmCBFIIIb-1A        GLVQ--HATSSSSTST----------------SASSS----SSEQQSRKAAWPPSTASSP 42
SbCBF1a             RLQQHQHATCSSSTST----------------SASST----SSSEQTIKAVVWSPSSSSS 44
ZmCBF1              R-HQLQLQHAASSSST----------------SASSS----SEQDK-----------PLC 32
TaCBFIIIc-3.2       L------EVSSSSPSS----------------SSASS----SPEHAAGRA-------S-- 29
TmCBF3              L------EVSSSSPSS----------------SSASS----SPEHAAGRA-------S-- 29
TaCBFIIIc-D3        L------EVSSSSPSS----------------SSVSS----SPVHAAGRA-------S-- 29
TaCBFIIIc-3         ----------SSSPSS----------------SSVSS----SPEHAAGRA-------S-- 21
HvCBFIIIc-3         L------EVSSSSPSS----------------SPVSS----SPEHAARRA-------S-- 29
Bradi4g35630        L------QLSSSPSSS----------------LSASS----SPEHDAGRS-------SSP 31
HvCBFIIIc-10A       L------EVSSSPP--------------------SS----SNENASGRS-------S-- 24
TmCBFIIIc-10        L------EISSSSP--------------------SS----SNENA-----------L-- 20
TaCBFIIIc-B10       L------EVSSSSP--------------------SS----S---------------S-- 16
HvCBFIIIc-13        M------DLCSSSPSS----------------SVSS----SPEHASG----------- 23
TmCBFIIIc-13        M------DLSSSSPSS----------------SASS----SPEHASGRA-------S-- 24
TmCBFIIIb-18        L------EHSSSASS--------------------SS----TTERGGTAW-------P-- 25
TaCBF12.2           P------ERNWNSPAS----------------PPSSL----EQGMP-T----------SP 27
TaCBFIIId-B12       P------ERNWNSPAS----------------PPSSL----EQGMP-T----------SP 27
HvCBFIIId-12        P------ERNWNSPAS----------------PPSSL----EQGMP-S----------SP 27
TaCBF12.1           P------EHNCSSPAS----------------SPSSQ----GQVMP-T----------SP 27
AsCBFIIId-12        P------EYNLT-PTS----------------SSSSQ----ELGMALS----------SP 27
AsCBFIIId-16A       G------SEQWSSSSS----------------SSYSV----EHGGPAV----------WT 28
```

```
AsCBFIIId-16B       G------SEQWSSPSS----------------S---L----EHGGPAV----------WT 25
LpCBF2              G------SEQWSSPSL----------------SSSSQ----EQGVP-V----------WV 27
TmCBFIIId-17        G------SEQWSSPST----------------SASSR----DQHAA-------------- 23
TaCBF17.0           G------SEQWSSPST----------------SASSR----DQHAA-------------- 23
TaCBF15.0           S------DQQRSSPS----------------SPSS-----SSHL--------------- 22
TmCBF15             S------DQQRSSPS----------------SPSS-----SSHL--------------- 22
TaCBFIIId-A15       S------DQQRSSPS----------------SPSS-----SSHL--------------- 22
TmCBFIIId-16        S------DQQWSSSS----------------SPSST----SSHP--------------- 77
HvCBF15B            S------DQQRCSPS----------------SPSL-----SSHL--------------- 22
HvCBF15A            S------DQQRCSPS----------------SPSL-----SSHL--------------- 20
Bradi2g60331        S------DRCLSSPSS---------------SSSSH----DRQEGL------------- 26
Bradi2g60340        D------ERICSSPSS---------------SPSSS----SSQEGL------------- 27
Bradi4g35600        Y------DQWISS-AS---------------SSSSH----DQYGVVW------------ 26
Bradi4g35610        S------DQRISS-AS---------------SSSSY----DQYGVVW------------ 26
Bradi4g35620        S------DQWTSSTTS---------------SSSSH----DQYMVVW------------ 27
Bradi4g35590        S------DQQITS-PS---------------SSSSH----DQYGVVW------------ 26
Bradi3g57360        S------GQWISFPSS---------------SSSHD----HECHGVA----------W 26
Bradi1g77120        S------CQWMSFTSS---------------SSSSS----SSHNGQAGA--------PW 29
Bradi1g57970        --------QWISFPPS---------------PSSQD----QQEE--------------W 20
TaCBFIIId-A19       I------NGWISSPSS---------------STSGH----ELGDAVP----------VW 28
TaCBF19.1           I------NGWISSPSS---------------STSGH----ELGDAVP----------VW 28
TaCBFIIId-B19       I------NGWISSPSS---------------STSGH----ELGETVP----------VW 28
TaCBFIIId-D19       I------NGWISSPSS---------------STSGH----ELGEAVP----------VW 28
TaCBF19.2           I------DSWISSPSS---------------STSGH----EHGEVVL----------VW 28
Bradi4g35580        A---------ASSPSS---------------SSSHE----QHGAACP----------QP 23
Bradi4g35640        A---------ASSPSS---------------SSSHE----QHGAACP----------QP 23
Bradi4g35570        A------NGYVSSPSS---------------SSSHE----QHGAACP----------PP 56
OsCBFIII-1D         T---------AASGQL---------------MTSS-------AEAT------------ 19
OsCBFIII-1J         T---------TAMGQL---------------MSSSA---TTAATATG----------- 24


OsCBFI-1F           RLPP---KRRAGRKKFRETRH-- 44
OsCBFIII-1I         RLSPPSSKRPAGRTKFHETRH-- 39
HvCBFIVd-9          SEPP---KRPAGRTKFHETRH-- 46
```

```
TmCBF9              SEPP---KRPAGRTKFHETRH-- 39
TaCBF9.0            SEPP---KRPAGRTKFHETRH-- 39
TaCBFIVd-9          SEPP---KRPAGRTKFHETRH-- 39
ScCBFIVd-9B         SEPP---KRPAGRTKFHETRH-- 39
HvCBFIVd-4A         SEPP---KRPAGRTKFHETRH-- 40
ScCBFIVd-9A         SEPP---KRPAGRTKFHETRH-- 39
TaCBFIVd-B9         SEPP---KRPAGRTKFHETRH-- 39
TaCBFIVd-D9         SEPP---KRPAGRTKFHETRH-- 40
TaCBF22.0           SEPP---KRPAGRTKVHETRH-- 39
TaCBFIVd-B22        SEPP---KRPAGRTKVHETRH-- 39
TaCBFIVd-D22        SEPP---KRPAGRTKVHETRH-- 39
TaCBF4.0            SEPP---KRPAGRTKFHETRH-- 39
TaCBFIVd-B4         SEPP---KRPAGRTKFRETRH-- 39
TmCBFIVd-4          SEPP---KRPAGRTKFQETRH-- 32
TaCBFIVc-B14        SEPP---KRPAGRTKFKETRH-- 35
TmCBF14             SEPP---KRPAGRTKFKETRH-- 35
HvCBFIVc-14         SEPP---KRPAGRTKFKETRH-- 35
TaCBF14.1           SEPP---KRPAGRTKFKETRH-- 35
TaCBF14.2           SEPP---KRPAGRTKFKETRH-- 35
TaCBFIVb-21         TEPP---KRPAGRIKYKETRH-- 34
TaCBF21.0           TEPP---KWPAGRIKYKETRH-- 34
TaCBFIVb-D21        TEPP---KRPAGRIKYKETRH-- 34
TaCBFIVb-A20        SEPP---KRPAGRTKFKETRH-- 35
TaCBFIVb-B20        SEPP---KRPAGRTKFKETRH-- 35
TaCBF20.0           SEPP---KRPAGRTKFKETRH-- 35
TaCBFIVb-D20        SEPP---KRPAGRTKFRETRH-- 35
ScCBFIVb-20         SEPP---KRPAGRTKFKETRH-- 35
OsCBFIa-1G          SAPP---KRRAGRTKFKETRH-- 52
HvCBFIa-1           SAPP---KRRAGRTKFKETRH-- 53
HvCBFIa-11          SAPP---KRRAGRTKVRETRH-- 50
HvCBF11             SAPP---KRRAGRTKVRETRH-- 50
TaCBFIa-A11         SAPP---KRRAGRTKVRETRH-- 50
OsCBFIa-1E          SAPP---KRRAGRTKFKETRH-- 50
HvCBFIVa-2A         SEPP---KRRAGRIKLQETRH-- 47
```

```
HvCBF2B             SEPP---KRRAGRIKLQETRH-- 47
TaCBF2.2            SEPP---KRRAGRIKLQETRH-- 47
HbCBFIVa-2          SEPP---KRRAGRIKLQETRH-- 47
TaCBFIVa-A2         SEPP---KRRAGRNKLQETRH-- 47
TaCBFIVa-2.3        SEPP---KRRAGRNKLQETRH-- 47
TmCBFIVa-2          SEPP---KRRAGRNKLQETRH-- 46
FaCBFIVa-2          SEPP---KPRSGRTKFQETRH-- 40
OsCBFIV-1B          SEPP---KRPAGRTKFRETRH-- 30
OsCBFII-1C          SAPP---KRPAGRTKFRETRH-- 31
ZmCBF4              SAPP---KRPAGRTKFRETRH-- 56
HvCBFII-5           SAPP---KRPAGRTKFRETRH-- 34
TmCBFII-5           SAPP---KRPAGRTKFRETRH-- 35
TaCBFII-5           SAPP---KRPAGRTKFRETRH-- 40
TaCBFII-5.2         SAPP---KRPAGRTKFRETRH-- 41
TaCBFII-5.3         SAPP---KRPAGRTKFRETRH-- 51
SbCBFII-5           SAPP---KRPAGRTKFRETRH-- 55
FaCBFIIIa-6         TAWM---KRPAGRTKFRETRH-- 55
LpCBFIIIa-6         TVWT---KRPAGRTKFRETRH-- 50
HvCBFIIIa-6         SAPA---KRPAGRTKFRETRH-- 59
TaCBFIIIa-6         SAPA---KRPAGRTKFRETRH-- 57
TaCBFIIIa-6.2       SAPA---KRPAGRTKFRETRH-- 59
TaCBFIIIa-D6        SAPA---KRPAGRTKFRETRH-- 57
SbCBFIIIa-6         TSRP---KRPAGRTKFRETRH-- 48
Bradi4g35650        TSPP---KRPAGRTKFRETRH-- 53
OsCBFIIIa-1A        TAPP---KRPAGRTKFRETRH-- 48
OsCBFIIIb-1H        RQQYK--KRPAGRTKFRETRH-- 44
ZmCBFIIIb-1A        QQPPK--KRPAGRTKFRETRH-- 61
SbCBF1a             PQPPK--KRPAGRTKFRETRH-- 63
ZmCBF1              CSGPK--KRPAGRTKFRETRH-- 51
TaCBFIIIc-3.2       ---LA--KRPAGRTKFRETRH-- 45
TmCBF3              ---LA--KRPAGRTKFRETRH-- 45
TaCBFIIIc-D3        ---LA--KRPAGRTKFRETRH-- 45
TaCBFIIIc-3         ---LA--KRPAGRTKFRETRH-- 37
HvCBFIIIc-3         ---PA--KRPAGRTKFRETRH-- 45
```

```
Bradi4g35630        APAPA--KRPAGRTKFRETRH-- 50
HvCBFIIIc-10A       ---TA--KRPAGRTKFRETRH-- 40
TmCBFIIIc-10        ---VA--KRPAGRTKFRETRH-- 36
TaCBFIIIc-B10       ---LA--KRPAGRTKFRETRH-- 32
HvCBFIIIc-13        ---PA--KRPVGRTKFRETRH-- 39
TmCBFIIIc-13        ---PA--KRPAGRTKFRETRH-- 40
TmCBFIIIb-18        --WPP--KRPAGRTKFRETRH-- 42
TaCBF12.2           ASPTP--KRPAGRTKFKETRH-- 46
TaCBFIIId-B12       ASPTP--KRPAGRTKFKETRH-- 46
HvCBFIIId-12        ASPTP--KRPAGRTKFKETRH-- 46
TaCBF12.1           TSPTP--KRPAGRTKLKETRH-- 46
AsCBFIIId-12        TSPTP--KRPAGRTKFKETRH-- 46
AsCBFIIId-16A       TPP----KRPAGRTKFKETRH-- 45
AsCBFIIId-16B       TPP----KRPAGRTKFKETRH-- 42
LpCBF2              TPP----KRPAGRTKFKETRH-- 44
TmCBFIIId-17        APP----KRPAGRTKFKETRH-- 40
TaCBF17.0           APP----KRPAGRTKFKETRH-- 40
TaCBF15.0           -------KRPAGRTKFKETRH-- 36
TmCBF15             -------KRPAGRTKFKETRH-- 36
TaCBFIIId-A15       -------KRPAGRTKFKETRH-- 36
TmCBFIIId-16        -------KRPAGRTKFKETRH-- 91
HvCBF15B            -------KRPAGRTKFKETRH-- 36
HvCBF15A            -------KRPAGRTKFKETRH-- 34
Bradi2g60331        -AVWS--KRPAGRTKFKETRH-- 44
Bradi2g60340        -AVWS--KRPAGRTKFKETRH-- 45
Bradi4g35600        SPPP---KRPAGRTKFKETRH-- 44
Bradi4g35610        SPPP---KRPAGRTKFKETRH-- 44
Bradi4g35620        SPPP---KRPAGRTKFKETRH-- 45
Bradi4g35590        SPPPP--KRPAGRTKFKETRH-- 45
Bradi3g57360        SPKP---KRLAGRSKFKETRH-- 44
Bradi1g77120        PPPPP--KRPAGRTKFKETRH-- 48
Bradi1g57970        SPAPP--KRPAGRTKFKETRH-- 39
TaCBFIIId-A19       S-PAA--KRPAGRTKFKETRH-- 46
TaCBF19.1           S-PAA--KRPAGRTKFKETRH-- 46
```

```
TaCBFIIId-B19       S-PAA--KRPAGRTKFKETRH-- 46
TaCBFIIId-D19       S-QAA--KRPAGRTKFKETRH-- 46
TaCBF19.2           S-PAA--KRPAGRTKFKDTRH-- 46
Bradi4g35580        APPRP--KRPAGRTKFKETRH-- 42
Bradi4g35640        APPRP--KRPAGRTKFKETRH-- 42
Bradi4g35570        APPRP--KRPAGRTKFKETRH-- 75
OsCBFIII-1D         PSSP---KRPAGRTKFQETRH-- 37
OsCBFIII-1J         PASP---KRPAGRTKFQETRH-- 42
                           *   ** * ::***
```

Table 3.3 Amino acid sequence alignment of AP2 domain of selected CBFs of Poaceae family members . Sequences were aligned using ClustalW2 at EBI ClustalW server, (http://www.ebi.ac.uk/Tools/msa/clustalw2/). Accession numbers are listed in Table 3.1 in Chapter 3.

```
TmCBF9          ---------------PLYRGVRRRG--RVGQ---WVCEVRVP-----GIKG-SRLWLGTF 34
ScCBF2          ---------------PLYRGVRRRG--RVGQ---WVCEVRVP-----GIKG-SRLWLGTF 34
TaCBFIVd-B9     ---------------PLYRGVRRRG--RVGQ---WVCEVRVP-----GIKG-SRLWLGTF 34
TaCBF9.0        ---------------PLYRGVRRRG--RVGQ---WVCEVRVP-----GIKG-SRLWLGTF 34
TaCBFIVd-9      ---------------PLYRGVRRRG--RVGQ---WVCEVRVP-----GIKG-SRLWLGTF 34
ScCBFIVd-9A     ---------------PLYRGVRRRG--RVGQ---WVCEVRVP-----GIKG-SRLWLGTF 34
ScCBFIVd-9B     ---------------PLYRGVRRRG--RVGQ---WVCEVRVP-----GIKG-SRLWLGTF 34
ScCBF3          ---------------PLYRGVRRRG--RVGQ---WVCEVRVP-----GIKG-SRLWLGTF 34
TaCBFIVd-D9     ---------------PLYRGVRRRG--RVGQ---WVCEVRVP-----GIKG-SRLWLGTF 34
HvCBFIVd-9      ---------------PLYRGVRRRG--RVGQ---WVCEVRVP-----GIKG-SRLWLGTF 34
HvCBFIVd-4A     ---------------PLYRGVRRRG--RVGQ---WVCEVRVP-----GIKG-SRLWLGTF 34
TaCBF4.0        ---------------PLYRGVRRRG--RVGQ---WVCEVRVP-----GVKG-SRLWLGTF 34
TaCBFIVd-B4     ---------------PLYRGVRRRG--RVGQ---WVCEVRVP-----GVKG-SRLWLGTF 34
TaCBFIVd-B22    ---------------PLYRGVRQRG--RVGQ---WVCEVRVA-----GVKG-SRLWLGTF 34
TaCBFIVd-D22    ---------------PLYRGVRQRG--RVGQ---WVCEVRVP-----GVKG-SRLWLGTF 34
TaCBF22.0       ---------------PLYRGVRQRG--RVGQ---WVCEVRVP-----GVKG-SRLWLGTF 34
ScCBFIVb-20     ---------------PLYRGVRRRG--RLGQ---WVCEVRVR-----GAQG-YRLWLGTF 34
TaCBFIVb-A20    ---------------PLYRGVRRRG--RLGQ---WVCEVRVR-----GAQG-YRLWLGTF 34
TaCBF20.0       ---------------PLYRGVRRRG--RLGQ---WVCEVRVR-----GAQG-YRLWLGTF 34
CBFIVb-B20      ---------------PLYRGVRRRG--RLGQ---WVCEVRVR-----GAQG-YRLWLGTF 34
TaCBFIVb-D20    ---------------PLYRGVRRRG--RLGQ---WVCEVRVR-----GAQG-YRLWLGTF 34
ScCBF1          ---------------PLYRGVRRRG--RLGQ---WVCEVRVR-----GAQG-YRLWLGTF 34
TaCBFIVb-21     ---------------PLYRGVRRRG--RYGR---WVCEVRVR-----GTKE-TRLWLGTF 34
TaCBFIVb-D21    ---------------PLYRGVRRRG--RYGR---WVCEVRVR-----GSKE-TRLWLGTF 34
TaCBF21.0       ---------------PLYRGVRRRG--RHGR---WVCEVRVR-----GTNE-TRLWLGTF 34
HvCBFIVc-14     ---------------PLYRGVRRRG--PAGR---WVCEVRVL-----GMRG-SRLWLGTF 34
TaCBF14.1       ---------------PLYRGVRRRG--PAGR---WVCEVRVL-----GMRG-SRLWLGTF 34
TaCBFIVc-B14    ---------------PLYRGVRRRG--PAGR---WVCEVRVL-----GMRG-SRLWLGTF 34
TaCBF14.2       ---------------PLYRGVRRRG--PAGR---WVCEVRVL-----GMRG-SRLWLGTF 34
```

```
TmCBF14           ---------------PLYRGVRRRG--PAGR---WVCEVRVL-----GMRG-SRLWLGTF 34
TmCBFIVd-4        ---------------PLYRGVRQRG--PAGR---WVCEVRVL-----GMRG-SRLWLGTF 34
HbCBFIVa-2        ---------------PVYRGVRRRG--KVGQ---WVCELRVP-----VSRGYSRLWLGTF 35
HvCBFIVa-2A       ---------------PVYRGVRRRG--KVGQ---WVCELRVP-----VSRGYSRLWLGTF 35
HvCBF2B           ---------------PVYRGVRRRG--KVGQ---WVCELRVP-----VSRGYSRLWLGTF 35
TaCBF2.2            ---------------PVYRGVRRRG--REGQ---WVCELRVP-----VSRGYSRLWLGTF 35
TaCBF2.1          ---------------PVYRGVRRRG--REGQ---WVCELRVP----AGSRSYSRIWLGTF 36
ScCBFIVa-2B       ---------------PVYRGVRRRG--REGQ---WVCELRVP----AGSRSYSRIWLGTF 36
TaCBFIVa-2.3      ---------------PVYRGVRRRG--REGQ---WVCELRVP----AGSRSYSRIWLGTF 36
ScCBFIVa-2A       ---------------PVYRGVRRRG--REGQ---WVCELRVP----AGSRSYSRIWLGTF 36
TaCBFIVa-2.2      ---------------PVYRGVRRRG--REGQ---WVCELRVP----AGSRSYSRIWLGTF 36
TmCBFIVa-2        ---------------PVYRGVRRRG--REGQW-VWVCELRVPA---AGSRVYSRIWLGTF 39
FaCBFIVa-2        ---------------PVYRGVRRRG--RAGQ---WVCEMRVH-----GTKG-SRLWLGTF 34
OsCBFIII-1D       ---------------LVFRGVRWRG--CAG---RWVCKVRVPG------SRGDRFWIGTS 34
OsCBFIII-1J       ---------------PVFRGVRRRG--RAG---RWVCEVRVPG------SRGDRLWVGTF 34
FaCBFIIIa-6       ---------------PVFRGVRRRG--NAG---RWVCEVRVPG------RRGSRLWVGTF 34
LpCBFIIIa-6       ---------------PVYRGVRRRG--NAG---RWVCEVRVPG------RRGSRLWVGTF 34
OsCBFIIIa-1A      ---------------PVFRGVRRRG--NAG---RWVCEVRVPG------RRGCRLWLGTF 34
SbCBFIIIa-6       ---------------PVFRGVRRRG--NAG---RWVCEVRVPG------RRGCRLWLGTF 34
PvCBFIIIa-6       PKRPAGRTKFRETRHPVFRGVRRRG--NAG---RWVCEVRVPG------RRGCRLWLGTF 49
Bradi4g35650      ---------------PVFRGVRRRG--NAG---RWVCEVRVPG------RRGSRLWLGTF 34
HvCBFIIIa-6       ---------------PVYRGVRRRG--NAG---RWVCEVRVPG------RRGSRLWLGTF 34
TaCBFIIIa-6.2     ---------------PVYRGVRRRG--NAG---RWVCEVRVPG------RRGSRLWLGTF 34
TaCBFIIIa-6       ---------------PVYRGVRRRG--NAG---RWVCEVRVPG------RRGSRLWLGTF 34
TaCBFIIIa-D6      ---------------PVYRGVRRRG--NAG---RWVCEVRVPG------RRGSRLWLGTF 34
ScCBFIIIa-6       GKRPAGRTKFRETRHPVYRGVRRRG--NAG---RWVCEVRVPG------RRGSRLWLGTF 49
OsCBFIII-1I       ---------------PVFRGVRRRG--RAG---RWVCEVRVPG------RRGCRLWLGTF 34
TaCBF12.2         ---------------PVFHGVRRRG--SNG---RWVCEVRVPG------KRGERLWLGTH 34
TaCBFIIId-B12     ---------------PVFHGVRRRG--SNG---RWVCEVRVPG------KRGERLWLGTH 34
HvCBFIIId-12      ---------------PVFHGVRRRG--SNG---RWVCEVRVPG------KRGERLWLGTH 34
ScCBFIIId-12      ---------------PVFHGVRRRG--SNG---RWVCEVRVPG------KRGERLWLGTH 34
TaCBF12.1         ---------------PVYHGVRRRG--RNG---RWVCEMRVPG------KRGERLWLGTH 34
Bradi4g35630      ---------------PVFRGVRRRG--AAC---RWVCEVRVPG------KRGARLWLGTY 34
AsCBFIIId-12      ---------------PVYRGVRRRG--SNG---RWVCEVRVPS------KSGERLWLGTH 34
```

```
TaCBFIIId-A19      ---------------PVYRGVRRRG--SAG---RWVCEVRVPG------KRGERLWLGTY 34
TaCBFIIId-B19      ---------------PVYRGVRRRG--SAG---RWVCEVRVPG------KRGERLWLGTY 34
TaCBF19.1          ---------------PVYRGVRRRG--SAG---RWVCEVRVPG------KRGERLWLGTY 34
ScCBFIIId-19       ---------------PVYRGVRRRG--SAG---RWVCEVRVPG------KRGERLWLGTY 34
TaCBFIIId-D19      ---------------PVYRGVRRRG--SAG---RWVCEVRVPG------KRGERLWLGTY 34
TaCBF19.2          ---------------PVYRGVRRRG--SAG---RWVCEVRVPG------KRGERLWLGTY 34
TmCBFIIId-17       ---------------PVYRGVRRRG--GAG---RWVCEVRVPG------RRGCRLWLGTY 34
TaCBF17.0          ---------------PVYRGVRRRG--GAG---RWVCEVRVPG------RRGCRLWLGTY 34
ZmCBFIIIb-1A       ---------------PVFRGVRRRG--AAG---RWVCEVRVPG------RRGARLWLGTY 34
ZmCBF1             ---------------PVFRGVRRRG--AAG---RWVCEVRVPG------RRGARLWLGTY 34
SbCBF1a            ---------------PVYRGVRRRG--AAG---RWVCEVRVPG------KRGARLWLGTY 34
Bradi4g35580       ---------------PVYRGVRRRG--AAG---RWVCEVRVPG------KRGQRLWLGTH 34
Bradi4g35640       ---------------PVYRGVRRRG--AAG---RWVCEVRVPG------KRGQRLWLGTH 34
Bradi4g35570       ---------------PVYRGVRRRG--PAG---RWVCEVRVPG------KRGQRLWLGTH 34
Bradi4g35590       ---------------PVYRGVRHRG--NAG---RWVCEVRVPG------TSGKRLWLGTH 34
Bradi4g35600       ---------------PVYRGVRLRG--NAG---RWVCEVRVPG------NRGKRLWLGTH 34
Bradi4g35610       ---------------PVYRGVRRRG--SAG---RWVCEVRVPG------MRNKRLWLGTH 34
Bradi4g35620       ---------------PVYRGVRLRG--TAG---RWVCEVRVPG------MRNKRLWLGTH 34
Bradi2g60331       ---------------PVYRGVRSRGAAAAG---RWVCEVRVPG------AHGKRLWLGTH 36
Bradi2g60340       ---------------PVYRGVRHRG--AAG---RWVCEVRVPG------AHGKRLWLGTH 34
Bradi1g57970       ---------------PVYRGVRTRG--AAG---RWVCEIRVPG------KRGKRLWLGTY 34
Bradi3g57360       ---------------PVYRGVRRRG--ALGGRCRWVCEVRVPG------KHGKRLWLGTH 37
Bradi1g77120       ---------------PVYHGVRRRG--RAG---RWVCEVRVPGTGSCNKKRGQRLWLGTY 40
TmCBFIIIb-18       ---------------PVFRGVRRRG--NAG---RWVCEVRVPG------DRGTRLWLGTY 34
OsCBFIIIb-1H       ---------------PVYRGVRRRG--GAG---RWVCEVRVPG------KRGARLWLGTY 34
AsCBFIIId-16B      ---------------PVYRGVRRRG--NAG---RWVCEVRVPG------QRGERLWLGTY 34
LpCBF2             ---------------PVYRGVRRRG--NAG---RWVCEVRVPG------QRGERLWLGTY 34
AsCBFIIId-16A      ---------------PVYRGVRRRG--NAG---RWVCEVRVPG------QRGERLWLGTY 34
TmCBFIIId-16       ---------------PVYRGVRRRG--NAG---RWVCEVRVPG------QRGERLWLGTY 34
TmCBF16            ---------------PVYRGVRRRG--NAG---RWVCEVRVPG------QRGERLWLGTY 34
TaCBF15.0          ---------------PVYRGVRRRG--SAG---RWVCEVRVPG------KRGERLWLGTH 34
TmCBF15            ---------------PVYRGVRRRG--SAG---RWVCEVRVPG------KRGERLWLGTH 34
ScCBFIIId-15       ---------------PVYRGVRRRG--SAG---RWVCEVRVPG------KRGERLWLGTH 34
TaCBFIIId-A15      ---------------PVYRGVRRRG--SAG---RWVCEVRVPG------KRGERLWLGTH 34
```

```
HvCBF15B          ---------------PVYRGVRRRG--SAG---RWVCEVRVPG------KRGERLWLGTH 34
HvCBF15A          ---------------PVYRGVRRRG--SAG---RWVCEVRVPG------KRGERLWLGTH 34
HvCBFIIIc-3       ---------------PVYRGVRRRG--NTE---RWVCEVRVPG------KRGARLWLGTY 34
TaCBFIIIc-D3      ---------------PVYRGVRRRG--NAE---RWVCEVRVPG------KRGARLWLGTY 34
TmCBFIIIc-10      ---------------PVYRGVRRRG--NAE---RWVCEVRVPG------KRGARLWLGTY 34
ScCBFIIIc-10      ----------------VYRGVRRRG--NAE---RWVCEVRVPG------KRGARLWLGTY 33
HvCBFIIIc-10A     ---------------PVYRGVRRRG--NAE---RWVCEVRVPG------KRGARLWLGTY 34
TaCBFIIIc-B10     ---------------PVYRGVRRRG--NAQ---RWVCEVRVPG------KRGARLWLGTY 34
ScCBFIIIc-3B      ---------------PVYRGVRRRG--NAQ---RWVCEVRVPG------KRGARLWLGTY 34
ScCBFIIIc-3A      ---------------PVYRGVRRRG--NTQ---RWVCEVRVPG------KRGARLWLGTY 34
TaCBFIIIc-3.2     ---------------PVYRGVRRRG--NAE---RWVCEVRVPG------KRGARLWLGTY 34
TmCBF3            ---------------PVYRGVRRRG--NAE---RWVCEVRVPG------KRGARLWLGTY 34
TaCBFIIIc-3       ---------------PVYRGVRRRG--NAE---RWVCEVRVPG------KRGARLWLGTY 34
HvCBFIIIc-13      ---------------PVYHDVRRRG--NAG---RWVCEVRVPS------KRGARLWLGTY 34
TmCBFIIIc-13      ---------------PVYRGVRRRG--NAG---RWVCEVRVPG------KRGSRLWLGTY 34
OsCBFIV-1B        ---------------PVYRGVRRRGG-RPGAAGRWVCEVRVPG------ARGSRLWLGTF 38
SbCBFII-5         ---------------PVYRGVRRRG--PAG---RWVCEVREPN-------KKSRIWLGTF 33
PvCBFII-5         ---------------PVYRGVRRRG--PAG---RWVCEVREPN-------KKSRIWLGTF 33
OsCBFII-1C        ---------------PVYRGVRRRG--PAG---RWVCEVREPN-------KKSRIWLGTF 33
TaCBFII-5         ---------------PVYRGVRRRG--AAG---RWVCEVREPN-------KKSRIWLGTF 33
TaCBFII-5.2       ---------------PVYRGVRRRG--AAG---RWVCEVREPN-------KKSRIWLGTF 33
TaCBFII-5.3       ---------------PVYRGVRRRG--AAG---RWVCEVREPN-------KKSRIWLGTF 33
ScCBFII-5         ---------------PAYRGVRRRG--AAG---RWVCEVREPN-------KKSRIWLGTF 33
TmCBFII-5         ---------------PVYRGVRRRG--AAG---RWVCEVRQPN-------NKSRIWLGTF 33
HvCBFII-5         ---------------PVYRGVRRRG--AAG---RWVCEVREPN-------KKSRIWLGTF 33
ZmCBF4            ---------------PVYRGVRRRG--PAG---RWVCEVREPN-------KKSRIWLGTF 33
OsCBFIa-1G        ---------------PVFKGVRRRN---PG---RWVCEVREPH-------GKQRIWLGTF 32
HvCBFIa-1         ---------------PVYKGVRRRN---PG---RWVCEVREPH-------SKQRIWLGTF 32
HvCBFIa-11        ---------------PVYKGVRSRN---PG---RWVCEVREPQ-------GKQRLWLGTF 32
HvCBF11           ---------------PVYKGVRSRN---PG---RWVCEVREPQ-------GKQRLWLGTF 32
OsCBFIa-1E        ---------------PVYKGVRSRN---PG---RWVCEVREPH-------GKQRIWLGTF 32
TaCBFIa-A11       ---------------PVYKGVRSRN---PG---RWVCEVREPH-------GKQRLWLGTF 32
ScCBFIa-11        ---------------PVYKGVRSRN---PG---RWVCEVREPH-------GKQRLWLGTF 32
PvCBfIa-11        ---------------PVYKGVRSRN---PG---RWVCELREPH-------GRHRIWLGTF 32
```

```
OsCBFI-1F          ----------------PVYRGVRARA---GGS--RWVCEVREPQ-------AQARIWLGTY 33
                                    ::.** *          ***::*             *:*:**

TmCBF9             NTAEMAARAHDAAVLALSGR------AACLNFA- 61
ScCBF2             NTAEMAARAHDAAVLALSGR------AACLNFA- 61
TaCBFIVd-B9        NTAEMAARAHDAAVLALSGR------AACLNFA- 61
TaCBF9.0           NTAEMAARAHDAAVLALSGR------AACLNFA- 61
TaCBFIVd-9         NTAEMAARAHDAAVLALSGR------AACLNFA- 61
ScCBFIVd-9A        NTAEMAARAHDAAVLALSGR------AACLNFA- 61
ScCBFIVd-9B        NTAEMAARAHDAAVLALSCR------AACLNFA- 61
ScCBF3             NTAEMAARAHDAAVLALSCR------AACLNFA- 61
TaCBFIVd-D9        NTAEMAARAHDAAVLALIGR------AACLNFA- 61
HvCBFIVd-9         NTAEMAARAHDAAALALSGR------AACLNFA- 61
HvCBFIVd-4A        TNPEMAARAHDAAVLALSGR------AACLNFA- 61
TaCBFIVd-4         TTAEMAARAHDAAVLALSGR------AACLNFA- 61
TaCBFIVd-B4        TTAEMAARAHDAAVLALSGR------AACLNFA- 61
TaCBFIVd-B22       TTAEMAARAHDAAVLALSGR------AACLNFA- 61
TaCBFIVd-D22       TTAEMAARAHDAAVLALSGR------AACLNFA- 61
TaCBF22.0          ATAEMAARAHDAAVLALSGR------AACLNFA- 61
ScCBFIVb-20        TTAEMAARAHDSAVLALLDR------AACLNFA- 61
TaCBFIVb-A20       TTAEMAARAHDSAVLALLDR------AACLNFA- 61
TaCBF20.0          TTAEMAARAHDSAVLALLDR------AACLNFA- 61
CBFIVb-B20         TTAEMAARAHDSAVLALLDR------AACLNFA- 61
TaCBFIVb-D20       TTAEMAARAHDSAVLALLDR------AACLNFA- 61
ScCBF1             TTAEMAARAHDSAVLALLDR------AACLNFA- 61
TaCBFIVb-21        RTAEMAARAHDSASLALSGS------AACLNFA- 61
TaCBFIVb-D21       RTAEMAARAHDSASLALSGS------AACLNFA- 61
TaCBF21.0          HTAEMAARAHDSASLALSGS------AACLNFA- 61
HvCBFIVc-14        TTAEMAARAHDAAVLALSGR------AACLNFA- 61
TaCBFI14.1         TTAEMAARAHDAAVLALSGR------AACLNFA- 61
TaCBFIVc-B14       TTAEMAARAHDAAVLALSGR------AACLNFA- 61
TaCBF14.2          TTAEMAARAHDAAVLALSGR------AACLNFA- 61
TmCBF14            TTAEMAARAHDAAVLALSGR------AACLNFA- 61
TmCBFIVd-4         VTAEMAARAHDAAVLALSGR------KACLNFA- 61
```

```
HbCBFIVa-2          ANPEMAARAHDSAALALSGH------DACLNFA- 62
HvCBFIVa-2A         ANPEMAARAHDSAALALSGH------DACLNFA- 62
HvCBF2B             ANPEMAARAHDSAALALSGH------DACLNFA- 62
TaCBF2.2            ATAEMAARAHDSAALALSGH------DACLNFA- 62
TaCBF2.1            ASAQMAARAHDSAALALSGR------DACLNFA- 63
ScCBFIVa-2B         ASAQMAARAHDSAALALSG--------------- 55
TaCBFIVa-2.3        ASAQMAARAHDSAALALSGR------DACLNFA- 63
ScCBFIVa-2A         ASAQMAARAHDSAALALSGR------DACLNFA- 63
TaCBFIVa-2.2        AGAQMAARAHDSAALALSGR------GACLNFA- 63
TmCBFIVa-2          ADPEMAARAHDSAALALSGR------DACLNFA- 66
FaCBFIVa-2          DTAEMAARAHDAAALALSGR------DACLNFA- 61
OsCBFIII-1D         DTAEETARTHDAAMLALCGAS------ASLNFA- 61
OsCBFIII-1J         DTAEEAARAHDAAMLALCGAS------ASLNFA- 61
FaCBFIIIa-6         DTAEIAARAHDAAMLALAAG------DACLNFA- 61
LpCBFIIIa-6         DTAEIAARAHDAAMLALAAG------DVCLNFA- 61
OsCBFIIIa-1A        DTAEGAARAHDAAMLAINAGGGGGGGACCLNFA- 67
SbCBFIIIa-6         DTADAAARAHDAAMLAIAGAG------ACLNFA- 61
PvCBFIIIa-6         DTAEGAARAHDAAMLAIAGAG------ACLNFA- 76
Bradi4g35650        DTAEAAARAHDAAMLALAGAGA-----ACLNFA- 62
HvCBFIIIa-6         DTAEAAARANDAAMLALAAGG-----AGCLNFA- 62
TaCBFIIIa-6.2       DTAEAAARANDAVMLMLAAGG-----AACLNFA- 62
TaCBFIIIa-6         DTAEAAARANDAAMIALSAGG-----AGCLNFA- 62
TaCBFIIIa-D6        DTAEAAARANDAAMIALSAGG-----AGCLNFA- 62
ScCBFIIIa-6         DTAEAAARANDATMIALNAGG-----AACLNFA- 77
OsCBFIII-1I         DAADAAARAHDAAMLALRGRA-----AACLNFA- 62
TaCBF12.2           VTAEAAARAHDAAMLALYGR----TPAARLNYP- 63
TaCBFIIId-B12       VTAEAAARAHDAAMLALYGR----TPAARLNYP- 63
HvCBFIIId-12        VTAEAAARAHDAGMLALYGR----TPAARLNFP- 63
ScCBFIIId-12        VTAEAAARAHDAAMLALYGR----NPSMRLNFP- 63
TaCBF12.1           VTAKAAARAHDAAMLALHGR----S-AARLNFP- 62
Bradi4g35630        VTAEAAARAHDAAMLALG------SSAARLNFP- 61
AsCBFIIId-12        VTAEAAARAHDAAMLAMHGH----TSAACLNFP- 63
TaCBFIIId-A19       VAAESAARAHDAAMLALLGRSP--SAAACLNFP- 65
TaCBFIIId-B19       VAAESAARAHDAAMLALLGRSP--CAAACLNFP- 65
```

```
TaCBF19.1          VAAESAARAHDAAMLALLGRSP--SAAACLNFP- 65
ScCBFIIId-19       VAAESAARAHDAAMLALLGRSP--SAAACLNFP- 65
TaCBFIIId-D19      VAAESAARAHDAAMLTLLGRSP--SAAACLNFP- 65
TaCBF19.2          VAAESATRAHDATMLALLGHSA--SAAACLNFP- 65
TmCBFIIId-17       VTAESAARAHDAAMLALGGRS-----AACLNFP- 62
TaCBF17.0          VIAESAARAHDAAMLALGGRS-----AACLNF-- 61
ZmCBFIIIb-1A       LAAEAAARAHDAAILALQ-G----RGAGRLNFP- 62
ZmCBF1             LGAEAAARAHDAAMLAL--G----RGAACLNFP- 61
SbCBF1a            LAAESAARAHDAAMLALGRG----GAAGCLNFP- 63
Bradi4g35580       LTADSAARAHDAAMLALRGPWC-SAAACLLNFT- 66
Bradi4g35640       LTADSAARAHDAAMLALRGPWC-SAAACLLNFT- 66
Bradi4g35570       LTADSAARAHDAAMLALRGPFS--ATACHLNFP- 65
Bradi4g35590       LTAESAARAHDAAMLALHGPSA-AAAARLLNFP- 66
Bradi4g35600       LTAESAARAHDAAMLALHGPSA-AAAC-LLNFP- 65
Bradi4g35610       LTAESAGRAHDAAMLALRGPSA-AAKC-LLNFP- 65
Bradi4g35620       LTAESAARAHDAAMLALRGASA-AAAG-LLNFP- 65
Bradi2g60331       LTAESAGRAHDAAMLALLPS---AAAC-LLNFP- 65
Bradi2g60340       LTAESAGRAHDAAMLALLPS---AAAC-LLNFP- 63
Bradi1g57970       LTAESAARGHDAAMLMLRGSYP-VATC-LLNFP- 65
Bradi3g57360       RTAESAGRAHDAAMLTLRGP----SAC-PLNFP- 65
Bradi1g77120       FSAECAARAHDAAMLMLRAA----PGARVLNFP- 69
TmCBFIIIb-18       FTAEAAARAHDAAMLMLRGR-----SAACLNFR- 62
OsCBFIIIb-1H       VTAEAAARAHDAAMIALRGGAG-GGGAACLNFQ- 66
AsCBFIIId-16B      LTAESAARAHDAAMLGLLGR---SASSACLNFA- 64
LpCBF2             LTAESAARAHDAAMLGLLGH---SASSTCLNFA- 64
AsCBFIIId-16A      LTAESAARAHDAAMLGLHGR---SAS-ACLNFA- 63
TmCBFIIId-16       LTADAAARAHDAAMLGLLGR---SAA--CLNFA- 62
TmCBF16            LTADAAARAHDAAMLGLLGR---SAA--CLNFA- 62
TaCBF15.0          LTAEAAARAHDAAMLGLIGP---STP--CLNFA- 62
TmCBF15            LTAEAAARAHDAAMLGLIGP---STP--CLNFA- 62
ScCBFIIId-15       LTAEAAARAHDAAMLGLIGP---STP--CLNFA- 62
TaCBFIIId-A15      LTAEAAARAYDAAMLCLIGP---STQ--CLNFA- 62
HvCBF15B           LTAEAAARAHDAAMLCLLDR---RAP--CLNFA- 62
HvCBF15A           LTAEAAARAHDAAMLCLLDR---RAP--CLNFA- 62
```

```
HvCBFIIIc-3         ATAEVAARANDAAMLALGGR---SAA--CLNFA- 62
TaCBFIIIc-D3        ATAEVAARANDAAMLALGGR---SAA--CLNFA- 62
TmCBFIIIc-10        ATAEIAARANDAAMLALGGR---SAA--RLNFP- 62
ScCBFIIIc-10        ATAEIAARANDAAMLALGGR---SAA--RLNFP- 61
HvCBFIIIc-10A       ATAEIAARANDAAMLALGGR---SAA--RLNFS- 62
TaCBFIIIc-B10       ATAEIAARANDAAMLALGGR---SAA--LLNFP- 62
ScCBFIIIc-3B        ATAEIAAHANDAAMLALGGR---SAA--CLNFA- 62
ScCBFIIIc-3A        ATAEIAARANDAAMLALGGR---SAA--CLNFAL 63
TaCBFIIIc-3.2       ATAEIAARANDAAMLALGGR---SAA--CLNFA- 62
TmCBF3              ATAEIAARANDAAMLALGGR---SAA--CLNFA- 62
TaCBFIIIc-3         DTAELAARANDAAMLALGGR---SAA--CLNFA- 62
HvCBFIIIc-13        LTAGAAARANDAAMLALGGR---SAR--RLNFA- 62
TmCBFIIIc-13        LTAEAAARANDAAMLALGGR---SAR--CLNFA- 62
OsCBFIV-1B          ATAEAAARAHDAAALALRGR------AACLNFA- 65
SbCBFII-5           ATAEAAARAHDVAALALRGR------AACLNFA- 60
PvCBFII-5           ATAEAAARAHDVAALALRGR------AACLNFA- 60
OsCBFII-1C          ATAEAAARAHDVAALALRGR------GACLNFA- 60
TaCBFII-5           ASPEAAARAHDVAALALRGR------AACLNFA- 60
TaCBFII-5.2         ASPEAAARAHDVAALALRGR------AACLNFA- 60
TaCBFII-5.3         ASPEAAARAHDVAALALRGR------AACLNFA- 60
ScCBFII-5           ASPEAAARAHDVAALALRGR------AACLNFA- 60
TmCBFII-5           ASPEAAARAHDVAALALRGR------AACLNFA- 60
HvCBFII-5           ATPEAAARAHDVAALALRGR------AACLNFA- 60
ZmCBF4              ATPEAAARAHDVAALALRGR------AACLNFA- 60
OsCBFIa-1G          ETAEMAARAHDVAALALRGR------AACLNFA- 59
HvCBFIa-1           ETAEMAARAHDVAALALRGR------AACLNFA- 59
HvCBFIa-11          DTAEMAARAHDVAAMALRGR------AACLNFA- 59
HvCBF11             DTAEMAARAHDVAAMALRGR------AACLNFA- 59
OsCBFIa-1E          ETAEMAARAHDVAAMALRGR------AACLNFA- 59
TaCBFIa-A11         DTAEMAARAHDVAALALRGR------AACLNFA- 59
ScCBFIa-11          DTAEMAARAHDVAALALRGR------AACLNFAT 60
PvCBfIa-11          ETAEMAARAHDVAALALRGR------AACLNFA- 59
OsCBFI-1F           PTPEMAARAHDVAAIALRGE-----RGAELNFP- 61
                      .  : :  *   : :
```

Table 3.4 Amino acid sequence alignment of C - terminal region and CMIII - 1 of selected CBFs of the Poaceae family members. Sequences were aligned using ClustalW2 at EBI ClustalW server, (http://www.ebi.ac.uk/Tools/msa/clustalw2/). Accession numbers are listed in Table 3.1 in Chapter 3.

```
OsCBFIa-1G          DSPRRL----RVPPI------GASH-DDIRRAAAEAAEAFR------------------ 30
HvCBFIa-1           DSPRRL----RVPAV------GASP-DEIRRAAVEAAEAFL------------------ 30
HvCBFIa-11          DSPRRL----PVPPQ------GAGH-DEIRRAAVEAAELFR------------------ 30
HvCBF11             DSPRRL----PVPPQ------GAGH-DEIRRAAVEAAELFR------------------ 30
TaCBFIa-A11         DSPRTL----RVPPQ------GAGH-EEIRRAAVEAAELFR------------------ 30
ScCBFIa-11          DSPRTL----RVPPQ------GAGH-DEIRRAAVEAAELFR------------------ 30
OsCBFIa-1E          DSPRRL----RVPPL------GAGH-EEIRRAAVEAAELFR------------------ 30
HvCBFII-5           DSAALL----RVDPA------TLRTPEDIRAAAMALAQAA------------------- 30
TmCBFII-5           DSAALL----AVDPA------TLRTPQDIRAAAITLAQTA------------------- 30
ScCBFII-5           DSATLL----AVDPA------TLRTPDDIRAAAIALAETA------------------- 30
TaCBFII-5           DSAALL----AVDPA------TLRTPDDIRAAAIALAETA------------------- 30
TaCBFII-5.2         DSATLL----AVDPA------TLRTPDDIRAAAIALAETA------------------- 30
TaCBFII-5.3         DSATLL----AVDPA------TLRTPDDIRAAAIALAETA------------------- 30
OsCBFII-1C          DSARLL----RVDPA------TLATPDDIRRAAIELAESCPHD----------------- 33
SbCBFII-5           DSARLL----RVDPA------TLATPDDIRRAAIQLAEDS------------------- 30
ZmCBF4              DSARLL----QVDPA------TLATPDDIRRAAIQLAD---------------------- 28
Bradi2g60331        DSA-------AATP------------PALGCAAVVAVAGFLSREAAPSIVSV-------- 33
Bradi2g60340        DSA-------AATP------------PALGCAAVVAVAGFLSREAAPSIVSV-------- 33
Bradi1g57970        DSAWLLDVP-CTLP-------AD--LGDVRHAALAAVADLQRRE-AADGAVN-------- 41
Bradi1g77120        DSEWLLDVPIMALPA-----AAD--LSCVRRASVAAVADFQRREPAANGAAA-------- 45
Bradi3g57360        DSSWLLDVP-FEVP-------ED--LPGVRRAALAAVADFQCRE----EAAT-------- 38
Bradi4g35580        DSAWLL----AVPHV----LPND--FSAVRRAALAALADFQRRDQVARGDAT-------- 42
Bradi4g35640        DSAWLL----AVPHV----LPND--FSAVRRAALAALADFQRRDQVARGDAT-------- 42
Bradi4g35570        DSAWLL----AMPCV----LPSD--LAAVRRAALAAVADFQRRE-VARGDAT-------- 41
Bradi4g35600        DSAWLL----AVTPP----ALAD--LDDIQRAALAAVADFQRRE-AAT------------ 37
Bradi4g35610        DSAWLL----AVTPSS---THAS--LDNIQRAALAAVADFQRQE-ANN------------ 38
Bradi4g35590        DSAWLL----AVTPS----ALAD--HDEIQRVAIAAVVDFQRRE-AAT------------ 37
Bradi4g35620        DSAWLL----SVTPP----ALAD--LDGVQRAALAAVADYQRRE-AANGAAA-------- 41
TmCBFIIId-17        DSAWLL------AVPC---ALAD--LADVRRAALAAVAGFQRRE-AASGAAT-------- 40
```

```
TaCBF17.0          DSAWLLA----VAVPS---ALAD--LADVRRAALAAVAGFQRRE-AASGAAT-------- 42
AsCBFIIId-16B      DSAWRL----TVTPG-----LSD--LAGVRRAALAAVANFLRRE-AAGGAAN-------- 40
LpCBF2             DSAWLL----DLPPT-----LSD--LAHVRRAALAAVAGFLRQE-ADSGAAT-------- 40
TmCBFIIId-16       DSAWLL----AVPPA-----LAD--LAAVRRAALAAVADFQRRH-ASNSAAT-------- 40
TmCBF16            DSAWLL----AVPPA-----LAD--LAAVRRAALAAVADFQRRH-ASNSAAT-------- 40
TaCBFIIId-A15      DSAWLL----AVPSA-----LPD--FADVRRAALSAVADFQRRE-AASGAATRSLDA--- 45
TmCBF15            DSAWLL----AVPSA-----LSD--FADVRRAALSAVADFQRRE-AASGAATTSLAA--- 45
HvCBF15B           DSVWLL----AVPSA-----LSD--LADVRRAALSAVADFQRRE-AASGAATRAQAA--- 45
HvCBF15A           DSVWLL----AVPSA-----LSD--LADVRRAALSAVADFQRRE-AASGAATRAQAA--- 45
TaCBF12.2          DSAWLL----AVPSS-----LSD--LADVRRAAIGAVVDFLRRQ--EAGASA-------- 39
TaCBFIIId-B12      DSAWLL----AVPSS-----LSD--LADVRRAAIGAVVDFLRRQ--EAGAST-------- 39
HvCBFIIId-12       DSAWLL----AVPSS-----LSD--MADVRRAAIGAVVDFLRRQ--ETGA---------- 37
TaCBF12.1          DSACLL----AVPSS-----LSS--LADVRRAAIGAVVDFLRRQ--ATIAG--------- 38
AsCBFIIId-12       DSAWLL----NVPSN-----LSD--LADVRRAAIEAVVEFLRLE--AIKD---------- 37
TaCBFIIId-A19      DSAWLL----VMPPR-----LSD--LADVRRAAIQAVAGFLRP----------------- 32
TaCBF19.1          DSAWLL----VMPPR-----LSD--LADVRRAAIQAVAGFLRL----------------- 32
TaCBFIIId-D19      DSAWLL----VMPPR-----LSD--LADVRRAAIQAVVGFLRL----------------- 32
TaCBFIIId-B19      DSAWLL----VMPPR-----LSD--LADVRRTAIQAVASFLRL----------------- 32
TaCBF19.2          DSAWLL----VMPPW-----LSD--LADIRRAAIEAVAIFLCL----------------- 32
OsCBFIIIb-1H       DSAWLL----AVPPAA--P--SD--LAGVRRAATEAVAGFLQRNKTTNGASVAEAMD--- 47
TmCBFIIIb-18       DSAWLL----SVPPAF-----SN--LSDVRRAAVQAVADFLRRPEATG--AFAGAAQ--- 44
HvCBFIIIc-13       DSAWLL----AVPFAL-----SD--LADVRRTGLQAVANFQRREAASGLI--TRTVA--- 44
TmCBFIIIc-13       DSAWLL----AVPSAL-----SD--LADVRRAALQAVADFQRWEAANGLV--TRTAAE-- 45
HvCBFIIIc-3        DSAWLL----AVPSAL-----SD--LADVRRAAVEAVADFQRREAADGSL--AIAVPK-- 45
TaCBFIIIc-B10      DSAWLL----AVPSAH-----SD--LADVRRAAVEAVADLQRREAAGGSI--TATATATA 47
TmCBFIIIc-10       DSAWLL----AVPSAH-----SD--LADVRRAAVEAVADLQRREAAGGSI--TATVN--- 44
HvCBFIIIc-10A      DSAWLL----AVPSAH-----SD--LADVRRAAVEAVSDLQRREAAGGSI--SATVD--- 44
TaCBFIIIc-3        DSAWLL----AVPSAL-----SD--LGDVRRAAVEAVANLQRRKAGNGSL--TATVT--- 44
TaCBFIIIc-D3       DSAWLL----AVPPAL-----SD--LGDVRRAAVEAVADFQRREAANGSL--TATVT--- 44
TaCBFIIIc-3.2      DSAWLL----AVPPAL-----AD--LSDVRRAAVEAVADSQRREAANGSLTVTATVT--- 46
TmCBF3             DSAWLL----AVPPAL-----AD--LGDVRRAAVEAVADFQRREAANGSLTVTATVT--- 46
Bradi4g35630       DSAWLL----NVPPAL-----AG--LADVRSAAVQAVADFERRETAAIDI--AAAAT--- 44
ZmCBFIIIb-1A       DSARLL----AVPPPS--ALPG---LDDARRAALEAVAEFQRRSGSGSGAA--------- 42
ZmCBF1             DSAWLL----AVPPPP--ALSGG--LDGARRAALEAVAEFQRRR-FGAAAA--------- 42
```

```
SbCBF1a           DSAWLL----AVPPPS--AISG---LDDARRAALEAVAEFQRRFGAAAAAAAGGCGSV-- 49
OsCBFIII-1J       DSAWLL----HVPRAP--VASGHDQLPDVQRAASEAVAEFQRRGSTAATAT--------- 45
FaCBFIIIa-6       DSAELL----AVPAS-------YRNLAEVRHAVTEAVEDFERR-----QELGEKDSL--- 41
LpCBFIIIa-6       DSAELL----DMPASS------YRSLDEVRHAVTEAVEEFERR-----QALGEEDAL--- 42
TaCBFIIIa-6       DSAELL----AVPAAS-----SYRSLDEVRHAVVEAVEDFLRR-----EAIAEEDAL--- 43
TaCBFIIIa-D6      DSAELL----AVPAAS-----SYRSLDEVRHAVVEAVEDLLRR-----EAIAEDDAL--- 43
TaCBFIIIa-6.2     DSAELL----SVPVAS-----SYRSLDEVRHAVVEAVEDLLRR-----EALAEEDAL--- 43
HvCBFIIIa-6       DSAELL----AVPAAS-----SYRSLDEVRHAVVEAVEDLLRR-----EAHAEDDALS-- 44
Bradi4g35650      DSAQLL----AVPAS-------YRSLDDVRLAVVEAVEDFLRRCEARAEEEEEEDAL--- 46
OsCBFIIIa-1A      DSAWLL----AVPRS-------YRTLADVRHAVAEAVEDFFRR-------RLADDALS-- 40
SbCBFIIIa-6       DSAWLL----AVPAS-------YASLAEVRHAVAEAVEEFLLR-----EVVQEDDAL--- 41
OsCBFIII-1I       DSAWLL----AVPPPA-----TLRCAADVQRAVARALEDFEQRESSSSVFPLAIDVVA-- 49
HvCBFIVa-2A       DSAWRMM---PVHAT--G-SFRLAPAQEIKDAVAVALEVFQG-QHPADA----------- 42
HvCBF2B           DSAWRMM---PVHAT--G-SFRLAPAQEIKDAVAVALEVFQG-QHPADA----------- 42
HbCBFIVa-2        DSAWRMM---PVHAT--G-SFRFAPAQEIKDAVAVALEVFKG-QHPADA----------- 42
TaCBF2.2          DSAWRMM---PVHAT--G-SFRFAPAQEIKDAVAVALEAFQE-QHHADA----------- 42
TaCBFIVa-2.2      DSAWRMM---PVHAA--G-SFKLAAAQEIKDAVAVALKEFQEQQRPADVSTAPS------ 48
TaCBFIVa-2.3      DSAWRMM---PVHAA--G-SFKLAAAQEIKDAVAVALKEFQEQQRPADESTAPS------ 48
TaCBF2.1          DSAWRMM---PVHAA--G-SFKLAAAQEIKDAVAVALKEFQEQQRPADESTAPS------ 48
TmCBFIVa-2        DSAWRMM---PVHAA--G-SFKLAAAQEIKDAVAVALKAFQEQQRPADASKAPSS----- 49
FaCBFIVa-2        DSAWRMQ---PVLPAGAG-SVCFGGAQEVKDAVAAAVEAFQEEEHHVESTAET------- 49
TaCBF14.1         DSAWRML---PVLAG------PFSTAKEIKDAVAVAVLAFQRQHPVASMAPLSPA----- 46
TaCBF14.2         DSAWRML---PVLAG------PFSTAKEIKDAVAVAVLAFQRQHPVASMAPLSPA----- 46
TmCBF14           DSAWRML---PVLAG------PFSTAKEIKDAVAVAVLAFQRQHPVASMAPLSPA----- 46
TaCBFIVc-B14      DSAWRML---PVLAG------PFSTAKEIKDAVAVAVLAFQRQHRVASMAPLSPA----- 46
HvCBFIVc-14       DSAWRML---PLLAG------PFSTAKEIKDAVAVAVLAFQRQHPVASTAPMSPA----- 46
TaCBFIVb-A20      DSAWRML---PVLAA--G-SSRFSSAREIKDAVAVAVMEFQRQRPVLSTP---------- 44
TaCBF20.0         DSAWRML---PVLAA--G-SSRFSSAREIKDAVAVAVMEFQRQRPVLSTP---------- 44
ScCBFIVb-20       DSAWRML---PVLAA--G-SSRFSSAREIKDAVAVAVVEFQRQRPFVSTS---------- 44
TaCBFIVb-B20      DSAWRML---PVLAA--G-SSRFSSAREIKDAVAIAIVEFQRQRPVVSTS---------- 44
TaCBFIVb-D20      DSAWRML---PVLAA--G-SSRFSSAREIKDAVAIAVLEFQRQRPVVSTS---------- 44
TaCBFIVb-21       DSAWRML---PVLAA--G-SSSFSSAREIKDAVAVAVVAFQRQRSVAST----------- 43
TaCBF21.0         DSAWRML---PVLAA--G-SSSFSSAREIKDAVAVAVVAFQRQRSIAST----------- 43
TaCBFIVb-D21      DSAWRML---PVLAA--G-SSSFSSAREIKDAVAVAVVAFQRQRPVAST----------- 43
```

```
TaCBF4.0           DSAWRML---PVLAA--G-SFGFGSAREIKLAVAVAVVAFQQQQIILPVACP-------- 46
TaCBFIVd-B4        DSAWRML---PVLAA--G-SFCFGSAREIKLAVAVAVVAFQLQQNIPPAACP-------- 46
HvCBFIVd-4A        DSAWRMR---PVLATT-G-SFGFSSTREIKLAVAVAVVAFQQQQIILPVACP-------- 47
TmCBFIVd-4         DSAWRML---PVLAA--G-SFGFGSAREIKTAVAVAVLAFQRQQIVLPVACP-------- 46
HvCBFIVd-9         DSAWRML---PVLAA--G-SFGFGSAREVKAAVAVAVVAFQRRQ-IIPVAVA-------- 45
ScCBFIVd-9A        DSAWRML---PVLAA--G-SFGFDSAREVKAAVAVAVVAFQRKQ-IIPVAVA-------- 45
TaCBFIVd-9         DSAWRML---PVLAA--G-SFGFGSASEIKAAVAVAVVAFQRKQIVLPVAVA-------- 46
TmCBF9             DSAWRML---PVLAA--G-SFGFGSASEIKAAVAVAVVAFQRKQIVLPVAVA-------- 46
TaCBF9.0           DSAWRML---PVLAA--G-SFGFGSASEIKAAVAVAVVAFQRKQIVLPVAVA-------- 46
TaCBFIVd-B9        DSAWRML---PVLAA--G-SFGFDSAREIKAAVAVAVVAFQRKQ-IIPVAVA-------- 45
TaCBFIVd-D9        DSAWRML---PVLAA--G-SFGFGSASEIKAAVAVAVVAFQRKQ-IIPVAVA-------- 45
ScCBFIVd-9B        DSAWRML---PVLAA--G-SFGFGSPREIKAAVAVAVIAFQRKQ-IIPVAVA-------- 45
TaCBF22.0          DSAWRML---PVLAA--G-SFGFGSAREIKAAVAVAVVAFQKEQ-IIPVAVA-------- 45
TaCBFIVd-D22       DSAWRML---PVLAP--G-SFGFGSAREIKAAVAVAVVAFKKQQ-IIPVAVA-------- 45
TaCBFIVd-B22       DSAWRML---PVLAA--G-SFGFGSAREIKAAVAVAVVAFQKEQIIPVAVAVVAIQKQ-- 52
OsCBFIV-1B         DSAWRMP---PVPAS-----AALAGARGVRDAVAVAVEAFQRQSAAPSSPAETFAN---- 48
OsCBFI-1F          DSPSTLPR--ARTAS----------PEDIRLAAAQAAELYRRPP---------------- 32
AsCBFIIId-16A      DSAWLLD----LPSP-------LSDLAAVRRVALAAVVRGQCRK---------------- 33
OsCBFIII-1D        DSAWLLH----VPRAP----VVSGLRPPAARCATRCLQGHRRVPAPGRGS---------- 42
                   **

OsCBFIa-1G         -------PPPDES-------NAATEVAAAASG--------------------------- 48
HvCBFIa-1          -------PAPDQS-------NAPAEEVAAAP---------------------------- 47
HvCBFIa-11         -------PAPGQR-------NAATVAAATAPP--------------------------- 48
HvCBF11            -------PAPGQR-------NAATVAAATAPP--------------------------- 48
TaCBFIa-A11        -------PEPGQR-------NAATTEAPAASP--------------------------- 48
ScCBFIa-11         -------PAPGQP-------NAAA-EAPAASP--------------------------- 47
OsCBFIa-1E         -------PAPGQH-------NAAAEAAAAVAAQ--------------------------- 49
HvCBFII-5          ---CPHDAASSSA-------PALKAASAPAPA---------------------------- 52
TmCBFII-5          ---CPHDAPRSS----------VSAASAPAPA---------------------------- 49
ScCBFII-5          ---CPAAPAPSS----------SSVAAGAPPM---------------------------- 49
TaCBFII-5          ---CPAAPASSS--------AVAAVASAPAPPM--------------------------- 52
TaCBFII-5.2        ---CPAAP------------VAAEASAPAPAM---------------------------- 47
TaCBFII-5.3        ---CPAAPASSSS-------VAAAVASAPAPP--------------------------- 52
```

```
OsCBFII-1C          ---AAAAAASSSA-------AAVEASAAAAPAM-------------------------- 56
SbCBFII-5           ----SSSTPDASA-------AAAAVAVASSASVGQ------------------------ 54
ZmCBF4              ----AASQQDETA-------AVAADVVAPS---------------------------- 47
Bradi2g60331        --VIPVAVVPVA---------VVPVDHGAAGAW------------------------- 55
Bradi2g60340        --VIPVAVVPVA---------VVPVDNGAAGAW------------------------- 55
Bradi1g57970        --VPDIDDAVFS---------LATTSQPCA----------------------------- 60
Bradi1g77120        --VLDLDEAAVS---------WATTSSQLARAN-------------------------- 67
Bradi3g57360        --VPVVNEST-S---------WAT-AEPSA----------------------------- 55
Bradi4g35580        --VPVVDEVASS-------ASALPSYMDEASSW-------------------------- 66
Bradi4g35640        --VPVVDEVASS-------ASALPSYMDEASSW-------------------------- 66
Bradi4g35570        --VPVVEDIASS-------ATALPSYMDDASSW-------------------------- 65
Bradi4g35600        --VPVPVPVAASEI----TSIASMVPVNDAGSW-------------------------- 64
Bradi4g35610        --VAVAN-VGAN------VPIASMAPVDNAGPP-------------------------- 62
Bradi4g35590        --VQVVNEPPIN-------PAFAPLPPDNAVPW-------------------------- 61
Bradi4g35620        --VPVVNEAVSN-------EFASSS--DNAVSW-------------------------- 63
TmCBFIIId-17        --VPVDEVFDTSSADDAGSWSWATPQPSCAAADGMFEVPAAALASDMFDFEFDVSWVMD- 97
TaCBF17.0           --VPVDEVFDTSSADDAGSWSWATPQPSCAAADGVFEVPAAALASDMFDFEFDVSWVMD- 99
AsCBFIIId-16B       ---VPADEDTSS--------ASA----DNAGGSS------------------------- 59
LpCBF2              ---VRADEAAYS--------ASVPSSVDNAGGLS------------------------- 63
TmCBFIIId-16        ---VPADEETSG--------ASALSSADNASGSS------------------------- 63
TmCBF16             ---VPADEETSG--------ASALSSADNASGSS------------------------- 63
TaCBFIIId-A15       --TVPVDDGTCS--------QSAQSSMENTGSSW------------------------- 69
TmCBF15             --TVPVDDGSCS--------QSAQSSMENTGSSW------------------------- 69
HvCBF15B            --AALIDEGTCS--------QSAQSSMENTGSSS------------------------- 69
HvCBF15A            --AALIDEGTCS--------QSAQSSMENTGSSS------------------------- 69
TaCBF12.2           --GAVAEEAHVDG------IASAASAPDNASSS------------------------- 64
TaCBFIIId-B12       --GAVTEVASIDG------IASAASAPDNASSA------------------------- 64
HvCBFIIId-12        --GAITEVTSVDG------VASEAYAPGSASSS------------------------- 62
TaCBF12.1           --ARAAEVVPVNG------VASVAPAPGNARSS------------------------- 63
AsCBFIIId-12        --GAAAVAVPIDGV-----VASAALAPSSHADN------------------------- 63
TaCBFIIId-A19       --EAATVVPDVDE------ATSPVYLPSPVDN-------------------------- 56
TaCBF19.1           --EAATVVPDVDE------ATSPVYLPSPVDN-------------------------- 56
TaCBFIIId-D19       --EAATVVPDVDE------ATSPVYLPSPVDN-------------------------- 56
TaCBFIIId-B19       --EAATVVPDVDE------ATSPVYLPSLVDN-------------------------- 56
```

```
TaCBF19.2          --EAAAVVPIIDE------ATSPVYLPSPVDN--------------------------- 56
OsCBFIIIb-1H       --EATSGVSAPP-----PLANNAGSSETPGPSSID------------------------ 75
TmCBFIIIb-18       --EVTSSVTVPS-----AAACSVPSSET-AQTSGD------------------------ 71
HvCBFIIIc-13       --------------------DSADSSETSEPSAD------------------------- 58
TmCBFIIIc-13       --QAPSSAPAQS----SSESDSADSSETSEASAD------------------------ 73
HvCBFIIIc-3        --EASSGAPSLSP---SSGSDSAGSTGTSEPSAN------------------------ 74
TaCBFIIIc-B10      AEEASCGAPAES----SSESDDAGSSETSKPSAD------------------------ 77
TmCBFIIIc-10       -EEASCGAPAES----SSESDDAGSSETSKPSAR------------------------ 73
HvCBFIIIc-10A      -EEASCGAPAES----SSESDGAGSSETSKPSAD------------------------ 73
TaCBFIIIc-3        -EEASCDAPEES----SSESDSAGSSETSEPSAD------------------------ 73
TaCBFIIIc-D3       -EEASCGAPEES----SSESDSAGSSETSEPSAD------------------------ 73
TaCBFIIIc-3.2      -EEASCGAPEES----SSESDSVGSSETSEPSAD------------------------ 75
TmCBF3             -EEASCGAPEES----SSESDSVGSSETSEPSAD------------------------ 75
Bradi4g35630       -DEATSRVSERT-----SSSGNAGSLETSEPSADD----------------------- 73
ZmCBFIIIb-1A       -DEATSGAS-PPSSSPSLPDV-SAAGSPAAALEH------------------------ 73
ZmCBF1             -DEATSGTS-PPSSS-------SSATKPAPAIER------------------------ 67
SbCBF1a            -DEATSGVSAPPLSTSSLPGISSGSPAPAPELEQ------------------------ 82
OsCBFIII-1J        ---ATSGDA-------------ASTAPPSSSPVL------------------------ 63
FaCBFIIIa-6        ---SGTSSSTPSS-------SSLTDDEEASSQ-------------------------- 63
LpCBFIIIa-6        ---SGTESST------------LTDDEESS----------------------------- 57
TaCBFIIIa-6        ---SGTSSSAP---------SSLTDDESSSS---------------------------- 62
TaCBFIIIa-D6       ---SGTSSSAP---------SSLTDDGSSSSPL-------------------------- 64
TaCBFIIIa-6.2      ---SGTSSSAP---------SPLTDDESSSSPL-------------------------- 64
HvCBFIIIa-6        --VSGTSSSAP---------SSITDDDSSSSP--------------------------- 65
Bradi4g35650       ---SGASSSLT---------DNDTGDEMTSSR-------------------------- 66
OsCBFIIIa-1A       --ATSSSSTTPST-------PRTDDDEESAATDG------------------------ 65
SbCBFIIIa-6        ---SATSSTPPSS-------PSSSDDGSTSDGGES----------------------- 66
OsCBFIII-1I        --EDAMSATSEPS-------AASDDDAVTSSSS------------------------- 73
HvCBFIVa-2A        ----------------------CTAEES------------------------------ 48
HvCBF2B            ----------------------CTAEGS------------------------------ 48
HbCBFIVa-2         ----------------------CTAEES------------------------------ 48
TaCBF2.2           ----------------------STTEAS------------------------------ 48
TaCBFIVa-2.2       ----------------------STAEES------------------------------ 54
TaCBFIVa-2.3       ----------------------STAEES------------------------------ 54
```

```
TaCBF2.1          ----------------------STAEES------------------------------ 54
TmCBFIVa-2        ---------------------TDSTSEES------------------------------ 57
FaCBFIVa-2        ----------------------AKDEES------------------------------ 55
TaCBF14.1         ---------------------RTTDDEKE------------------------------ 54
TaCBF14.2         ---------------------RTTDDEKE------------------------------ 54
TmCBF14           ---------------------RTTDDEKE------------------------------ 54
TaCBFIVc-B14      ---------------------RTTDDEKE------------------------------ 54
HvCBFIVc-14       ---------------------RTAVDEKE------------------------------ 54
TaCBFIVb-A20      ---------------------ETHDGEKD------------------------------ 52
TaCBF20.0         ---------------------ETHDGEKD------------------------------ 52
ScCBFIVb-20       ---------------------ETADGEKD------------------------------ 52
TaCBFIVb-B20      ---------------------ETHDGEKD------------------------------ 52
TaCBFIVb-D20      ---------------------EMHDGEKD------------------------------ 52
TaCBFIVb-21       ----------------------ADGEKD------------------------------ 49
TaCBF21.0         ----------------------ADGEKD------------------------------ 49
TaCBFIVb-D21      ----------------------ADGEKD------------------------------ 49
TaCBF4.0          ---------------------TVEAAASP------------------------------ 54
TaCBFIVd-B4       ---------------------TVEPTDSP------------------------------ 54
HvCBFIVd-4A       ---------------------SPEAPASP------------------------------ 55
TmCBFIVd-4        ---------------------AAEPAVAP------------------------------ 54
HvCBFIVd-9        ---------------------VVALQKQQ-VPVAVAVVTLQQKQQQVPVAVAVAALQQQQ 83
ScCBFIVd-9A       ---------------------VVALQKQQ-VPVAVAVVALQQRQ---------------- 67
TaCBFIVd-9        ---------------------VVALQQKQ-VPIAVAVVALQQKQ---------------- 68
TmCBF9            ---------------------VVALQQKQ-VPIAVAVVALQQKQ---------------- 68
TaCBF9.0          ---------------------VVALQQKQ-VPIAVAVVALQQKQ---------------- 68
TaCBFIVd-B9       ---------------------VVALQQQQ-VPVAVAVVALQQKQ---------------- 67
TaCBFIVd-D9       ---------------------VVALQQQQ-VPVAVAVVALQQKQ---------------- 67
ScCBFIVd-9B       ---------------------VVALQQQQ-VPVAVAVVALKQKQ---------------- 67
TaCBF22.0         ---------------------VVALQKQQIIPVAVAVVALQKQQ---------------- 68
TaCBFIVd-D22      ---------------------VVALQKQQIIPVAVAVVALQKQQ---------------- 68
TaCBFIVd-B22      ------QIIPVAV-------AVVAIQKQQ------------------------------ 68
OsCBFIV-1B        ----------------------DGDEEED------------------------------ 55
OsCBFI-1F         ---------------------PPLALPEDPQEGTS------------------------- 46
AsCBFIIId-16A     ---------------------LVGDIAALLCR-------------------------- 44
```

```
OsCBFIII-1D       ----TATATATSG-------DAASTAPPSAPVLS------------------------- 65


OsCBFIa-1G        ------------------------------------ATNSNAEQFASHPYYEV------ 65
HvCBFIa-1         ---------------------------------------TMQFAGDPYYG------- 58
HvCBFIa-11        ------------------------------------VALGNAELVADSPYYP------- 64
HvCBF11           ------------------------------------VALGNAELVADSPYYP------- 64
TaCBFIa-A11       ------------------------------------ADAGNAELVANSPYHL------- 64
ScCBFIa-11        ------------------------------------VASGNAELVESSPYCL------- 63
OsCBFIa-1E        ------------------------------------ATAASAELFADFPCYP------- 65
HvCBFII-5         ------------------------------------MVMMQEPAAVPYDSYAT-----A 70
TmCBFII-5         ------------------------------------MVITQEAAAAPYDSYA------- 65
ScCBFII-5         ------------------------------------MAMMQESAAAPYDSYA------- 65
TaCBFII-5         ------------------------------------MTMMHESAAVHYDDYP------M 69
TaCBFII-5.2       ------------------------------------MAMMQEPSAVEYDDYP------M 64
TaCBFII-5.3       ------------------------------------MTMMQ------FDDYA------M 63
OsCBFII-1C        ------------------------------------MMQYQDDMAATPSSYDY-----A 74
SbCBFII-5         ----------------------------------ATPSSSAYQAGDDATGAAMYGAEYAAAA 82
ZmCBF4            ----------------------------------------QADDVAAAAA------AAAA 61
Bradi2g60331      ----------------------PSFL--GVLFEV--PLPVPPMAMGS-GMLELELE--- 84
Bradi2g60340      ----------------------PSFL--GVLFEV--PLPVPPMAMGS-GMLELELE--- 84
Bradi1g57970      ----------------------NNGV--LLGVFD--DFEVPVAT-GS-GLFELD----- 86
Bradi1g77120      ----------------------NNGG--ML-MFD---FEVPVAAMGSDGMFELED---- 94
Bradi3g57360      ----------------------NNG-----------IFEVPVAM-GSGGMFELD----- 75
Bradi4g35580      ----------------------ASSF--QPSEIG--DFDVPVVV-GS-GMFELD----- 92
Bradi4g35640      ----------------------ASSF--QPSEIG--DFDVPVVV-GS-GMFELD----- 92
Bradi4g35570      ----------------------AASF--QPCEIG--NFDVPV------GMFELD----- 87
Bradi4g35600      ----------------------PSFQ--P-CVAG--MFDGPVVM-GS-DMFELD----- 89
Bradi4g35610      ----------------------PSFR--P-SAAG--MFEAPVAM-GS-DMFELD----- 87
Bradi4g35590      ----------------------ASSQ--PSATTG--MFGEPVAM-DS-NMFELD----- 87
Bradi4g35620      ----------------------ATSQ--ASANNG--TSEEPVVM-GS-EMFELG----- 89
TmCBFIIId-17      ------------------LGSPATSQ--PGCADK--VLEVPAAALGGGDMFEFDLE--- 131
TaCBF17.0         ------------------LGSPAASQ--PGCADK--VLEVQAAALGGGDMFEFDLE--- 133
AsCBFIIId-16B     ----------------------ATSQ---PYVDGTFDFEMPAGM-GS-DIFELD----- 86
```

```
LpCBF2           ----------------------ATSQ---PYADG--MFELPSAL-NS-DMFELD----- 88
TmCBFIIId-16     ----------------------ATSQ---PWAEG--TFEVPSAL-GS-DMFELD----- 88
TmCBF16          ----------------------ATSQ---PWAEG--TFEVPSAL-GS-DMFELD----- 88
TaCBFIIId-A15    ----------------------TSSS--LPSGNG--MFEVPATL-GC-DMFELD----- 95
TmCBF15          ----------------------TSSSSSLPSGDG--MFAVPATL-GC-NMFELD----- 97
HvCBF15B         ----------------------TSSS--LPSADG--MLEVPATL-GS-NMFELD----- 95
HvCBF15A         ----------------------TSSS--LPSADG--MLEVPATL-GS-NMFELD----- 95
TaCBF12.2        ----------------------AAAA-HSQPPCANAGYEVP--DALCHDMFELH----- 93
TaCBFIIId-B12    ----------------------AAAS-HSQPPCANAEFEVP--DALCHDMFELH----- 93
HvCBFIIId-12     ----------------------AASSSHYQLPCANAEFVVP--DALCHDMLELH----- 92
TaCBF12.1        ----------------------ATSS---QQPCANAESEAP--DALRGGLPELH----- 90
AsCBFIIId-12     ----------------------ASPA---ATSQPSAASEVP--EALGGDMFELH----- 90
TaCBFIIId-A19    ---------------------------------ADEVFQVPTFSPLGSDMFELD----- 77
TaCBF19.1        ---------------------------------ADEVFQVPTFSPLGSDMFELD----- 77
TaCBFIIId-D19    ---------------------------------ADQVFQVPTFSPLGSDMFELD----- 77
TaCBFIIId-B19    ---------------------------------ADEVFQVPTFSPLGSDMFELD----- 77
TaCBF19.2        ---------------------------------AYEVFQVPTFSAQSSDMFELD----- 77
OsCBFIIIb-1H     ----------------------GTA-------------DTAAGA--ALDMFELD----- 92
TmCBFIIIb-18     ----------------------ANF-------------EEPGAL--SMDMFDLDC---- 89
HvCBFIIIc-13     ----------------------GDF-------------ELPVAM--DSDMFR-LD---- 75
TmCBFIIIc-13     ----------------------GEF-------------EVLATM--DIDMFR-LD---- 90
HvCBFIIIc-3      ----------------------GEF-------------EGPVVM--DSEMFR-LD---- 91
TaCBFIIIc-B10    ----------------------GDF-------------AVPGGM--DIEMFSRLD---- 95
TmCBFIIIc-10     ----------------------GDF-------------ALPGGM--DVEMFSRLD---- 91
HvCBFIIIc-10A    ----------------------GDL-------------AVPVGM--DIEMF-RLD---- 90
TaCBFIIIc-3      ----------------------REF-------------EVPVAV--DTDMFG-LD---- 90
TaCBFIIIc-D3     ----------------------AEF-------------EVPVAV--DTDMFSRLD---- 91
TaCBFIIIc-3.2    ----------------------GEF-------------EVPVAV--DTDMFR-LD---- 92
TmCBF3           ----------------------GEF-------------EVPVAV--DTDMFR-LD---- 92
Bradi4g35630     ----------------------ARFG------------EVPVAM--DSDMFSRLD---- 92
ZmCBFIIIb-1A     ----------------------VPVK----------ADEA-VALDLDGDVFGP------ 93
ZmCBF1           ----------------------VPVE----------ASET-VALD--GAVFEP------ 85
SbCBF1a          ----------------------VPVK----------ANETATALD--GDVFEPA----- 102
OsCBFIII-1J      ----------------------SPND----------DNASSASTPAVAAALDHG----- 85
```

```
FaCBFIIIa-6       ---------------------------------------------ADNSPFELE----- 72
LpCBFIIIa-6       -----------------------------------------------TPFELD----- 63
TaCBFIIIa-6       ---------------------------------------------PPEDSPFELD----- 72
TaCBFIIIa-D6      ---------------------------------------------PEEDSPFELD----- 74
TaCBFIIIa-6.2     ---------------------------------------------PEEDSPFEQD----- 74
HvCBFIIIa-6       ---------------------------------------------ADEGSPFELD----- 75
Bradi4g35650      ---------------------------------------------SEEDSPFELD----- 76
OsCBFIIIa-1A      -------------------------------------DESSSPASDLAFELD----- 80
SbCBFIIIa-6       ------------------------------------SESDSSPAATGASPFELD----- 84
OsCBFIII-1I       ----------------------------------------TTDADEEASPFELD----- 87
HvCBFIVa-2A       ------------------------------------------TTPITSSDLSGLDDEH-- 64
HvCBF2B           ------------------------------------------TTPITSSDLSGLDDEH-- 64
HbCBFIVa-2        ------------------------------------------APSITSSDLSGLDDEH-- 64
TaCBF2.2          ------------------------------------------APSITSSDLSGLDDEL-- 64
TaCBFIVa-2.2      ------------------------------------------ALSIIPSDLSGLDNEH-- 70
TaCBFIVa-2.3      ------------------------------------------ALSIIPSDLSGLDNEH-- 70
TaCBF2.1          ------------------------------------------ALSIIPSDLSGLDNEH-- 70
TmCBFIVa-2        ------------------------------------------APSITSNDLSGLDDEH-- 73
FaCBFIVa-2        ------------------------------------------ALSMS-SDLSEHDDER-- 70
TaCBF14.1         ---------------------------------IDGLPAPSALSMSS----ELLNEH-- 74
TaCBF14.2         ---------------------------------IDGLPAPSALSMSS----ELLNEH-- 74
TmCBF14           ---------------------------------IDGSPAPSALSMSS----ELLNEH-- 74
TaCBFIVc-B14      ---------------------------------IDGSPAPSALFMSS----ELLNEH-- 74
HvCBFIVc-14       ---------------------------------VDGSPAPSALFMSS----ELLNEH-- 74
TaCBFIVb-A20      ---------------------------------VQGSPTPSELSTSS----DLLDEH-- 72
TaCBF20.0         ---------------------------------VQGSPTPSELSTSC----DLLDEH-- 72
ScCBFIVb-20       ---------------------------------VQGSPRPSELSTSS----DLLDEH-- 72
TaCBFIVb-B20      ---------------------------------VQGSPTPSELSMSS----DLLDEH-- 72
TaCBFIVb-D20      ---------------------------------AQGSPTPSELSTSS----DLLDEH-- 72
TaCBFIVb-21       ---------------------------------VQGSPTPSELSTSS----DLLDEH-- 69
TaCBF21.0         ---------------------------------VQGSPTPSELSTSS----DLLDEH-- 69
TaCBFIVb-D21      ---------------------------------VQGSPTPSELSTSS----DLLDEH-- 69
TaCBF4.0          ------------------------------------SNS---LFYMSSVDLLELDEEQ-- 73
TaCBFIVd-B4       ------------------------------------SNS---LFYMSSGDLLELDEEH-- 73
```

```
HvCBFIVd-4A        ------------------------------------SAA---LFYISSGDLLELDEEQ-- 74
TmCBFIVd-4         ------------------------------------SGA---LFSMSSGDLLELDDEQ-- 73
HvCBFIVd-9         VPVAVAVVALQQLQVPVAVAVVALQEQQ-IILPVACLAP---EFYMSSGDLLELDEEQ-- 137
ScCBFIVd-9A        VPVTVAVVALQKLQVPVAVAVVALQKKQ-IILPAACLAP---EFYMSSGDLLELDEEQ-- 121
TaCBFIVd-9         VPVAVAVVALQQLPVPVAVAVVALQQQQ-IILPVACLAP---EFYMSSGDLLELDEEQ-- 122
TmCBF9             VPVAVAVVALQQLPVPVAVAVVALQQQQ-IILPVACLAP---EFYMSSGDLLELDEEQ-- 122
TaCBF9.0           VPVAVAVVALQQLPVPVAVAVVALQQQQ-IILPVACLAP---EFYMSSGDLLELDEEQ-- 122
TaCBFIVd-B9        VPLAVAVVALQQLQVPVAVAVVALQQQQQIILPVACLAP---EFYMSSGDLLELDEEQ-- 122
TaCBFIVd-D9        VPVAVAVVALQQLQVPVAVAVVALQQQQ-IILPVACLAP---EFYMSSGDLLELDEEQ-- 121
ScCBFIVd-9B        VPVAVAVVALQQLHVPVAVAVVALQQQQ-IILPVACLAP---EFYMSSGDLLELDEEH-- 121
TaCBF22.0          IPVAVALVALQEQQVPVAVAVVALHRQQVPVACPATSGPGSALFYMSSSDLLELDEEQ-- 126
TaCBFIVd-D22       IPVAVALVALQEKQIPVAVAVVALHRQQVPVDDPATSGPGSALFYMSSSDLLELDEEQ-- 126
TaCBFIVd-B22       ---IIPSPSWRSRSSRFQWPSPSWRSRNSRSQSPSPSWRFIGSRFQSRAREPPARAAL-- 123
OsCBFIV-1B         ----------------------------------NKDVLPVAAAEVFDAGAFELDDGF-- 79
OsCBFI-1F          ---------------------------------------GGGATATSGRPAAVFVD----- 63
AsCBFIIId-16A      ------------------------------------------WDIRGAGRAGQRHVR----- 59
OsCBFIII-1D        ------------------------------------AKQCEFIFLSSLDCWMLMSKLIS 88

OsCBFIa-1G         ---MDDGLDLGMQ---GYLDMAQGMLIDPPPMAGD-------PAVGSGEDD--------- 103
HvCBFIa-1          ---MDDGMDFGMQ---GYLDMAQGMLIAPPPLVG--------PSATAGDGD--------- 95
HvCBFIa-11         ----MDGLESEMQ---GYLDMAHGMLIEPPPMAW---------PSTWIEED--------- 99
HvCBF11            ----MDGLESEMQ---GYLDMAHGMLIEPPPMAW---------PSTWIEED--------- 99
TaCBFIa-A11        ----MDGLEFEMQ---GYLDMAHGMLIEPPPMAG---------PSTWIEED--------- 99
ScCBFIa-11         ----MDGLEFEMQ---GYLDMAHGMLIEPPPMAG---------PSTWIEED--------- 98
OsCBFIa-1E         ----MDGLEFEMQ---GYLDMAQGMLIEPPPLAG---------QSTWAEED--------- 100
HvCBFII-5          LYG--DLTDLDMH---SYYCYDGMS--GGGDWQSI-------SRMDGADED--------G 108
TmCBFII-5          MYG--GLADLEQH---SHCYYDGMS--GSGDWQSI-------SHMNVADED--------G 103
ScCBFII-5          MYG--GLADLDHH---SYYHYDGMSCCGGGDCQSI-------SHMNGADED--------G 105
TaCBFII-5          QYGYGGIGDLDQD---SYY-YDGMS-AAGGDWQSG-------SHMDGADDDC----NDSG 113
TaCBFII-5.2        QY--GGIGDFDQH---SYY-YDGLS-AGGGDWQSS-------SHMDGADDDS----NCGG 106
TaCBFII-5.3        QY--GGIGDLDQH---SYY-YDGLS-AAGGDWPSG-------SHMDGADDDC----NGSG 105
OsCBFII-1C         YYG---NMDFDQP---SYY-YDGMG--GGGEYQS--------WQMDGDDDG------GAG 111
SbCBFII-5          MYG--AGMDFD-H---SYYYD-GMVGGNE--WQS--------AGSSGWHSNV----DAGD 121
```

```
ZmCBF4            MYG--GGMEFD-H---SYCYDDGMVSGSSDCWQSG-------AGAGGWHSIV----DGDD 104
Bradi2g60331      -LDMPSEMDLG-----MHYADLAEG-LLMEPPPD-----TAEA--PCWESG-----DYGH 125
Bradi2g60340      -LDMPSEMDLG-----MHYADLAEG-LLMEPPQD-----TNEA--SCWESG-----DFGD 125
Bradi1g57970      ---VSGDMELG-----MYYADLAEG-LLMEPPSPV----PDAG--ASLESR-----DYGH 126
Bradi1g77120      ---ICGETDLD-----MYYTELAGG-LLMEPP-------PDAG--ACWESR-----DAG- 130
Bradi3g57360      ---MSGEMDVG-----MYYADLAEG-LLMEPPQTT---PDDTG--ACWESG-----EYA- 115
Bradi4g35580      ---MSGEMDLG-----ACYADLAEG-LLMEPPQT----TPDTE--ACWGSG-----YYYD 132
Bradi4g35640      ---MSGEMDLG-----ACYADLAEG-LLMEPPQT----TPDTE--ACWGSG-----YYYD 132
Bradi4g35570      ---MAGEMDLG-----VYYADLAEG-LLMEPPQM----TPDTE--ACWEIG-----YYSH 127
Bradi4g35600      ---MPDEMDLG-----MYYADLAEG-LLMEPPLP----APDTG--ACWEIG-----EYGD 129
Bradi4g35610      ---MPDEMDLG-----MYYTDLAEG-LLMEPPPP----APDAG--ACWESG-----DCGN 127
Bradi4g35590      ---MTSEMDVG-----RYYADLAEG-LLMEPPQP----APDTG--ACWESG-----DDGD 127
Bradi4g35620      ---MPEEMDLG-----MYYKDLAEG-LLMEPPPPPP--APNTGTGVCWESG-----DYGD 133
TmCBFIIId-17      -LDMSGEMNLVG----SYYADFAEG-LLLEPPQP-----ADAT-EARWRNG-----DYCG 174
TaCBF17.0         -LDMSGEMDLVG----SYYADFAEG-LLLEPPQP-----ADAT-EARWRNG-----DYCG 176
AsCBFIIId-16B     ---MSGEMDQG-----TYYAGLAEG-LLLEPPP------QHAG--ACWDTG-----DGG- 123
LpCBF2            ---MSGEMDLG-----TYYAGLAEG-LLLDPPPP-----EHTS--ACWDTG-----DGG- 126
TmCBFIIId-16      ---LSGEMDLG-----TYYADLADG-LLLEPPPS-----LDSG--ACWDTG-----DGG- 126
TmCBF16           ---LSGEMDLG-----TYYADLADG-LLLEPPPS-----LDSG--ACWDTG-----DGG- 126
TaCBFIIId-A15     ---MSGEMDLD-----TYYAYFAEG-LLLEPPQP-----PVAG--ACWDTE-----GGG- 133
TmCBF15           ---MSGEMDLD-----TYYAYFAEG-LLLEPPQP-----PVAG--ACWDTE-----GGG- 135
HvCBF15B          ---MSGEMDLD-----TYYAYFAEG-LLLEPPQP-----PAAG--ACWDID-----GGG- 133
HvCBF15A          ---MSGEMDLD-----TYYAYFAEG-LLLEPPQP-----PAAG--ACWDTD-----GGG- 133
TaCBF12.2         ---TSGEMDAG-----TYYADLAQG-LLLEPP-------PPPS---SGA----------- 123
TaCBFIIId-B12     ---TSGEMDAG-----TYYADLAQG-LLLEPP-------PPPS---SGASS-----EQGD 129
HvCBFIIId-12      ---TSGEMDAG-----TYYADLAQG-LLLEPP-------PPPS---SGASS-----EHGD 128
TaCBF12.1         ---TSGEMDVS-----TYYADLAQG-LLLEPP-------PPAA---SDCN------DGGD 125
AsCBFIIId-12      ---TSGEMGLG-----TYYADLAEG-LLLEPP-------PAAAS--SEHGG-----DCGD 127
TaCBFIIId-A19     ---MSGEMDLD-----AYYAGFAQG-MLLEPP-------PTPAYWETGECG-----DGGA 116
TaCBF19.1         ---MSGEMDLD-----AYYAGFAQG-MLLEPP-------PTPAYWETGECG-----DGGA 116
TaCBFIIId-D19     ---MSGEMDLD-----AYYAGFAQG-MLLEPP-------PTPSYWENGECG-----DGGA 116
TaCBFIIId-B19     ---MSGEMDLD-----AYYAGFAQG-MLLEPP-------PMPSYWENGECG-----DGGA 116
TaCBF19.2         ---MSGEMDLD-----AYYAGFAQG-MLLEPP-------PTPTYWENGECG-----DGGA 116
OsCBFIIIb-1H      ---FFGEMDYDT-----YYASLAEG-LLMEPPP---------AATALWDNGD-----EGA 129
```

```
TmCBFIIIb-18       ---LFGETDSDT----YYYANLAQG-LLMEPPPS-------MATGAYWDNGDCADGGAGA 134
HvCBFIIIc-13       ---FLPEMDLGS-----YYVSLAEA-LLMDPPST------ATIIDAHRDNG-----DGAK 115
TmCBFIIIc-13       ---LFPEMDLGS-----YYVSLAEA-LLMDPPST------ATIIDAYRDNR-----DGGA 130
HvCBFIIIc-3        ---LFPEMDLGS-----YYMSLAEA-LLMDPPPT------ATIIHAYEDNG-----DGGA 131
TaCBFIIIc-B10      ---LFPEMDLGS-----YYASLAEA-LLMDPPPV------ATGTGAYWDNG-----ECGE 135
TmCBFIIIc-10       ---LFPETDLGS-----YYASLAEA-LLMDPPPV------ATGTGAYWDDG-----EFGE 131
HvCBFIIIc-10A      ---FFPEMEFGS-----YYASLAEA-LLMDPPPV------ANSTGAYWDNG-----EFGE 130
TaCBFIIIc-3        ---LFPEMDLCS-----YYASLAEA-LLVDPPAR------VTTTDTYWDNG-----DGGA 130
TaCBFIIIc-D3       ---LFPEMDLCS-----YYASLAEA-LLVDPPST------VAIIDSYWDNG-----DDGA 131
TaCBFIIIc-3.2      ---LFPELDLCS-----YYASLAEA-LLLDPPAP------VTTTYAYWDNG-----DGGA 132
TmCBF3             ---LFPELDLCS-----YYASLAEA-LLVDPPAP------VTTTYAYWDNG-----DGGA 132
Bradi4g35630       ---LFREMDLGM----YYYTSLAEA-LLVDPPPP------ETAAGAYWDNGDCV--EGGA 136
ZmCBFIIIb-1A       --DWFGDMGLELD---AYYASLAEG-LLVEPPPPP----AAWDHGDCCDSGA-------- 135
ZmCBF1             --DWFGDMDLD-----LYYASLAEG-LLVEPPPPPP--PAAWDHGDCCDSG--------- 126
SbCBF1a            --DWFGDMDMELD---VYYASLAEG-LLVEPPPAP---AAAWDHGDCCDAG--------- 144
OsCBFIII-1J        --DMFGGMR--TD---LYFASLAQG-LLIEPPPPP---TTAEGFCDDEGCGG-------- 126
FaCBFIIIa-6        ---VLSDMGWD-----LYYSSLAQGMMLMAPPFL----AASAAFGDYGEVNL-------- 112
LpCBFIIIa-6        ---VLSDMGWD-----LYYASLAQG-MLMSSPFL----AASAALGDYGEANL-------- 102
TaCBFIIIa-6        ---VLSDMGWD-----LYYASLAQA-MLMAPPS-----SMAAALGDYGEV---------- 108
TaCBFIIIa-D6       ---VLSDMGWD-----LYYASLAQA-MLMAPPS-----SMAAALGDYGEV---------- 110
TaCBFIIIa-6.2      ---VLSEMGWD-----LYYASLAQA-MLMAPP------AAAAALGDYGEAHL-------- 111
HvCBFIIIa-6        ---VLSDMGWD-----LYYASLAQG-MLMAPPA-----SLAAALGDYGEAHL-------- 113
Bradi4g35650       ---VLSDMGWD-----LYYASLAQG-MLVEPP------SAAAALGDYGEVCL-------- 113
OsCBFIIIa-1A       ---VLSDMGWD-----LYYASLAQG-MLMEPP--------SAALGDDGDAIL-------- 115
SbCBFIIIa-6        ---VFNDMSWD-----LYYASLAQG-MLVEPP---------SAVTAFMDEGF-------- 118
OsCBFIII-1I        ---VVSDMGWS-----LYYASLAEG-LLMEPP------ASGASSDDDDDAIVDS------ 126
HvCBFIVa-2A        ---WIGGMDAG-----SYYASLAQGMLMEPPAAG-----------GWREDD-----GEHD 100
HvCBF2B            ---WIGGMDAG-----SYYASLAQGMLMEPPAAG-----------GWREDD-----GEHD 100
HbCBFIVa-2         ---WISGMDAG-----SYYASLAQGMLMEPPAAG-----------AWRED------REHD 99
TaCBF2.2           ---LIDGMDAG-----SYYASLAQGMLMEPPAAG-----------AWRED------HEHD 99
TaCBFIVa-2.2       ---WIGGMEAG-----SYYASLAQGMLMEPPADGA----------SWRED------REHD 106
TaCBFIVa-2.3       ---WIGGMEAG-----SYYA------------------------------------- 82
TaCBF2.1           ---WIGGMEAG-----SYYASLAQGMLMEPPADG-----------AWQED------REHD 105
TmCBFIVa-2         ---WIGGMDAG-----SYYANLAQGMLMEPPAAG-----------AWRED------REQD 108
```

```
FaCBFIVa-2          ---WIDGMDAG-----SYYASLAQGMLVEPPDAG-----------AWRED------GEHG 105
TaCBF14.1           ---WFGGMDAG-----SCYS---E--FMESPDTR-----------PWRED------FEL- 103
TaCBF14.2           ---WFGGMDAG-----SCYS---E--FMESPDTR-----------PWRED------FEL- 103
TmCBF14             ---WFGGMDAG-----SCYS---E--FMESPDTR-----------PWRED------FEL- 103
TaCBFIVc-B14        ---WFGGMDAG-----SFYS---EGLFMESPDTR-----------PWRED------LEL- 105
HvCBFIVc-14         ---WFGGMDAG-----SCYS---EGMFIESPDTR-----------PWRED------LEL- 105
TaCBFIVb-A20        ---WFGGMNAG-----SYYASLAQGMLMEPPAAR-----------ARSED------GGEY 107
TaCBF20.0           ---WFGGMNAG-----SYYASLAQGMLMEPPAAR-----------ARSED------GGEY 107
ScCBFIVb-20         ---WFSGMDAG-----SYYASLAQGMLMEPPAAR-----------AWSED------GGEY 107
TaCBFIVb-B20        ---WFGGMDAG-----SYYASLAQGMLMEPPSAR-----------TWSED------GGEY 107
TaCBFIVb-D20        ---WFGGMDAG-----SYYASLAQGMLMEPPSAR-----------TWSED------GGEY 107
TaCBFIVb-21         ---WFGGTDAG-----SYYS---PGMFMEPPERP-----------GNREL------GA-- 99
TaCBF21.0           ---WFGGTDAG-----SYYS---PGMFMEPPERP-----------ENRQL------GA-- 99
TaCBFIVb-D21        ---WFGGTNAG-----SYYS---PGMFMESPEQP-----------ENHEL------GG-- 99
TaCBF4.0            ---WFGGMDAG-----SYYESLAQGMLMAPPDDR-----------ARRED------AEQT 108
TaCBFIVd-B4         ---WFGGMDAG-----SYYDSLAQGMLVEPPDDR-----------ARRED------AEQT 108
HvCBFIVd-4A         ---WFGGMDAG-----SYYASLAQGMLVAPPDER-----------ARPEN------REHS 109
TmCBFIVd-4          ---WFGGMVAG-----SYYESLAQGMLVEPPDAG-----------AWRED------SEHS 108
HvCBFIVd-9          ---WFGGMDAG-----SYYASLAQGMLVAPPDER-----------ARPEH------GEQ- 171
ScCBFIVd-9A         ---WFGGMDAG-----SYYASLAQGMLVAPPDDR-----------ARPEN------GEQ- 155
TaCBFIVd-9          ---WFGGMEAG-----SYYASLAQGMLVAPPDER-----------ARPES------GEQ- 156
TmCBF9              ---WFGGMEAG-----SYYASLAQGMLVAPPDER-----------ARPES------GEQ- 156
TaCBF9.0            ---WFGGMEAG-----SYYASLAQGMLVAPPDER-----------ARPES------GEQ- 156
TaCBFIVd-B9         ---WFGGMEAG-----SYYASLAQGMLVAPPDER-----------AGPEH------GEQ- 156
TaCBFIVd-D9         ---WFGGMDAG-----SYYASLAQGMLVAPPDER-----------ARPEH------GEQ- 155
ScCBFIVd-9B         ---WFGGMDAG-----SYYASLAQGMLVAPPDER-----------ARPEN------GEQE 156
TaCBF22.0           ---WFGGMEAG-----SYYASLAQGMLVAPPDER-----------ARPED------GEQ- 160
TaCBFIVd-D22        ---WFGGMDAG-----SYYASLAQGMLVAPPDER-----------ARPED------GEQ- 160
TaCBFIVd-B22        ---CFTCRPAT-----CWSSTRSSGLAAWRPGRTTRAWRRGCSWHRRTKER-----GRRT 170
OsCBFIV-1B          ---RFGGMDAG-----SYYASLAQGLLVEPPAAG-----------AWWED------GEL- 113
OsCBFI-1F           -----EDAIFDMP---GLIDDMARGMMLTPPAIG-----RSLDDWAAIDDD--------D 102
AsCBFIIId-16A       ----AGRVRGNGL---RHVLRGPRGGIAAGAAAA-----------ARRRLLG-------- 93
OsCBFIII-1D         SSRAKGSLCLRKNPISFCMVTNSYTALLLEYIILQMN-----SMIVLIHELSKYQVFLLL 143
```

```
OsCBFIa-1G          --NDGEVQLWSY------------------------------ 113
HvCBFIa-1           --DDGEVSLWSY------------------------------ 105
HvCBFIa-11          --YDCEISLWNY------------------------------ 109
HvCBF11             --YDCEISLWNY------------------------------ 109
TaCBFIa-A11         --YDCEVSLWNY------------------------------ 109
ScCBFIa-11          --YDCE------------------------------------ 102
OsCBFIa-1E          --YDCEVNLWSY------------------------------ 110
HvCBFII-5           IYGAGDVALWSY------------------------------ 120
TmCBFII-5           GYGAGDVALWSY------------------------------ 115
ScCBFII-5           SYGAGD------------------------------------ 111
TaCBFII-5           GYGAGEVPLWSY------------------------------ 125
TaCBFII-5.2         GYGAGEVPLWSY------------------------------ 118
TaCBFII-5.3         GYGAGEVALWSY------------------------------ 117
OsCBFII-1C          GYGGGDVTLWSY------------------------------ 123
SbCBFII-5           DEGAGDMSLWSYY----------------------------- 134
ZmCBF4              DDGASDMTLWSY------------------------------ 116
Bradi2g60331        GD---AGDLWSY------------------------------ 134
Bradi2g60340        G-----GDLWSY------------------------------ 132
Bradi1g57970        GGS-ADADLWSCY----------------------------- 138
Bradi1g77120        ----ADADLWSCY----------------------------- 139
Bradi3g57360        -EL-WGCEICSLWL---------------------------- 127
Bradi4g35580        GG--ADAALWHYGTKHSILL---------------------- 150
Bradi4g35640        GG--ADAALWHYGTKHSILL---------------------- 150
Bradi4g35570        GG--AEATLWNY------------------------------ 137
Bradi4g35600        GG--TDATLWNY------------------------------ 139
Bradi4g35610        GG--ADASLWSY------------------------------ 137
Bradi4g35590        -----DATLWSYRNDLP------------------------- 139
Bradi4g35620        GG--ADAALWSY------------------------------ 143
TmCBFIIId-17        GDGGGDAALWSQ------------------------------ 186
TaCBF17.0           GDGGGDAAFWSQ------------------------------ 188
AsCBFIIId-16B       ----ADPALWSY------------------------------ 131
LpCBF2              ----ADSALWSY------------------------------ 134
TmCBFIIId-16        ----ADSGLWSY------------------------------ 134
```

```
TmCBF16           ----ADSGLWSY------------------------------ 134
TaCBFIIId-A15     ----ADAALWSY------------------------------ 141
TmCBF15           ----ADAALWSY------------------------------ 143
HvCBF15B          ----ADAALWSY------------------------------ 141
HvCBF15A          ----ADAALWSY------------------------------ 141
TaCBF12.2         ----------------------------------------
TaCBFIIId-B12     -----DAALWNH------------------------------ 136
HvCBFIIId-12      -----DAALWNH------------------------------ 135
TaCBF12.1         -----DAVLWSH------------------------------ 132
AsCBFIIId-12      -----APDMMRRYGNGATEIRLSLAAE--------------- 149
TaCBFIIId-A19     -----AAGLWSY------------------------------ 123
TaCBF19.1         -----AAGLWSY------------------------------ 123
TaCBFIIId-D19     -----AAGLWSY------------------------------ 123
TaCBFIIId-B19     -----AAGLWSY------------------------------ 123
TaCBF19.2         -----AAGLWSY------------------------------ 123
OsCBFIIIb-1H      DIA-----LWSY------------------------------ 136
TmCBFIIIb-18      DVA-----LWSY------------------------------ 141
HvCBFIIIc-13      VF-----LFWEKTLYSKDFEAVSNPSLNPEKGNIPSRLHVP- 151
TmCBFIIIc-13      DV-----ALWSY------------------------------ 137
HvCBFIIIc-3       DV-----RLWSYSVDM-------------------------- 142
TaCBFIIIc-B10     AEGATEFALWS------------------------------- 146
TmCBFIIIc-10      VA--TEFALWSL------------------------------ 141
HvCBFIIIc-10A     VA--TEFALWS------------------------------- 139
TaCBFIIIc-3       DV-----ALWS------------------------------- 136
TaCBFIIIc-D3      DV-----ALWSY------------------------------ 138
TaCBFIIIc-3.2     DV-----ALWSY------------------------------ 139
TmCBF3            DV-----ALWSY------------------------------ 139
Bradi4g35630      DV-----ALWSY------------------------------ 143
ZmCBFIIIb-1A      ----ADVALWSYY----------------------------- 144
ZmCBF1            ----ADVALWSY------------------------------ 134
SbCBF1a           ----ADVALWSY------------------------------ 152
OsCBFIII-1J       ----AEMELWS------------------------------- 133
FaCBFIIIa-6       ----ADVPLWSYQS---------------------------- 122
LpCBFIIIa-6       ----ADVPLWSYLS---------------------------- 112
```

```
TaCBFIIIa-6       -----DVPLWSYQS---------------------------- 117
TaCBFIIIa-D6      -----DVPLWSYQS---------------------------- 119
TaCBFIIIa-6.2     ----ADVPLWSYQS---------------------------- 121
HvCBFIIIa-6       ----ADVPLWSYQS---------------------------- 123
Bradi4g35650      ----ADVPLWSYQS---------------------------- 123
OsCBFIIIa-1A      ----ADVPLWSY------------------------------ 123
SbCBFIIIa-6       ----ADVPLWSY------------------------------ 126
OsCBFIII-1I       -SDIADVSLWSY----------------------------- 137
HvCBFIVa-2A       DGFNTSASLWSY----------------------------- 112
HvCBF2B           DGFSTSTSLWSY----------------------------- 112
HbCBFIVa-2        AGVDTSTSLWSY----------------------------- 111
TaCBF2.2          DGFDTPTSLWSY----------------------------- 111
TaCBFIVa-2.2      DGFDT--SLWSYQCDQLIKQCKVLESTASARLCFTKYGKNRE 146
TaCBFIVa-2.3      ----------------------------------------
TaCBF2.1          DGFDT--SLWSY----------------------------- 115
TmCBFIVa-2        DGVDT--SLWSYWLDGFG--CVKL----------------- 128
FaCBFIVa-2        -GVET--SLWSYL---------------------------- 115
TaCBF14.1         --GGVETPPWS--YLFD------------------------ 116
TaCBF14.2         --GGVETPPWS--YLFD------------------------ 116
TmCBF14           --GGVETPPWS--YLFD------------------------ 116
TaCBFIVc-B14      --CGVETPPWS--YLFD------------------------ 118
HvCBFIVc-14       --GGVQTPPWS--YLFD------------------------ 118
TaCBFIVb-A20      --SGVQTPLWN--TYPTN----------------------- 121
TaCBF20.0         --SGVQTPLWN--TYPTN----------------------- 121
ScCBFIVb-20       --SGVHTPLWN------------------------------ 116
TaCBFIVb-B20      --SGVYTPLWN------------------------------ 116
TaCBFIVb-D20      --SAVYTPLWN------------------------------ 116
TaCBFIVb-21       --GEVETLLW------------------------------- 107
TaCBF21.0         --GDV------------------------------------ 102
TaCBFIVb-D21      --GDVKTPLW------------------------------- 107
TaCBF4.0          GV-ETPTPLWS--YLFD------------------------ 122
TaCBFIVd-B4       GV-ETPTPLWS--YLFD------------------------ 122
HvCBFIVd-4A       GV-ETPIPLWS--YLFDC----------------------- 124
TmCBFIVd-4        GVAETQTPLWS------------------------------ 119
```

```
HvCBFIVd-9          --TGVQTPLWS--CLFD------------------------ 184
ScCBFIVd-9A         --SGVQTPLWS--CLFD------------------------ 168
TaCBFIVd-9          --SGVQTPLWS--CLFD------------------------ 169
TmCBF9              --SGVQTPLWS--CLFD------------------------ 169
TaCBF9.0            --SGVQTPLWS--CLFD------------------------ 169
TaCBFIVd-B9         --SGVQTPLWS--CLFD------------------------ 169
TaCBFIVd-D9         --SGVQTPLWS--CLLD------------------------ 168
ScCBFIVd-9B         RRPDAAMELFV--RLI------------------------- 170
TaCBF22.0           --SGVQTPLWSQSHLFN------------------------ 175
TaCBFIVd-D22        --SGVQTPLWSQSHLFN------------------------ 175
TaCBFIVd-B22        ASRAASRRRYGATCSPNLAV--------------------- 190
OsCBFIV-1B          --AGSDMPLWSY----------------------------- 123
OsCBFI-1F           DHYHMDYKLWMD----------------------------- 114
AsCBFIIId-16A       -----HRRRWRS----------------------------- 100
OsCBFIII-1D         TMITHHLFQWRR----------------------------- 155
```

Table 4.1. List of Primers and BAC clones used to amplify CBF genes.

Template	F / R primer	DNA sequence*	Expected size (bp)	Annealing temp. (°C)
1716E15	CBF2.1F	**GACGACGACAAG**ATGGACACCAACGCCGC	704	60
	CBF2.1R	**GAGGAGAAGCCCGGT**TAGTAGCTCCACAGCGACGTGTC		
1144N5	CBF2.2F	**GACGACGACAAG**ATGGACACCGTTGCCG	686	60
	CBF2.2R	**GAGGAGAAGCCCGGT**TAGTAGCTCCACAGCGACG		
1740J17	CBF4.0F	**GACGACGACAAG**ATGGACGTCGCCGACGCT	695	65
		GAGGAGAAGCCCGGTTAGTCAAACAAATAGCTCCATAACG		
	CBF4.0R	G		
425P7	CBF9.0F	**GACGACGACAAG**ATGGACGTCGCCGACATC	836	58
	CBF9.0R	**GAGGAGAAGCCCGGT**TAGTCGAACAAGCAGCTCCATAGC		
1144N5	CBF12.1F	**GACGACGACAAG**ATGGACACGGCCCCCG	749	60
	CBF12.1R	**GAGGAGAAGCCCGGT**TAGTGGCTCCATAGCACCG		
	CBF12.2-T2F	**GACGACGACAAG**ATGGACACGGCCCCCG	650	60
	CBF12.2-T2R	GAGGAGAAGCCCGGTTAGTACGTGCTCA		
	CBF12.1-T3F	**GACGACGACAAG**ATGGACACGGCCCCCG	560	60
	CBF12.1-T3R	**GAGGAGAAGCCCGGT**TAAGGCTGCTGG		
	CBF12.1-T4F	**GACGACGACAAG**ATGGACACGGCCCCCG	533	60
	CBF12.1-T4R	**GAGGAGAAGCCCGGT**TACCTGGCATTG		
3149L3	CBF12.2F	**GACGACGACAAG**ATGGACACGGGCCCGG	764	60
	CBF12.2R	**GAGGAGAAGCCCGGT**TAGTGGTTCCATAGCGCCG		
	CBF12.2-T2F	**GACGACGACAAG**ATGGACACGGGCCCGG		
	CBF12.2-T2R	**GAGGAGAAGCCCGGTTA**GTACGTGCCCGC	662	60
	CBF12.2-T3F	**GACGACGACAAG**ATGGACACGGGCCCGG		
	CBF12.2-T3R	**GAGGAGAAGCCCGGTCA**TGGCGGCTGTGA	575	60
	CBF12.2-T4F	**GACGACGACAAG**ATGGACACGGGCCCGG		
	CBF12.2-T4R	**GAGGAGAAGCCCGGTCA**CAATGCGTCCGGC	389	60
	CBF12.2-T5F	**GACGACGACAAG**ATGGACACGGGCCCGG		
	CBF12.2-T5R	**GAGGAGAAGCCCGGTTA**GGGGTAGTTGAGGC	332	60
	CBF12.2-T6F	**GACGACGACAAG**ATGGACACGGGCCCGG		
	CBF12.2-T6R	**GAGGAGAAGCCCGGTTA**CTCGCCGCGCTTCC	245	60
	CBF12.2-T7F	**GACGACGACAAG**ATGGACACGGGCCCGG		
	CBF12.2-T7R	**GAGGAGAAGCCCGGTTA**CGGGTGGCGCGTCTCCTT	167	60
210D22	CBF14.1F	**GACGACGACAAG**ATGGACGCCGCTGAT	665	60
	CBF14.1R	**GAGGAGAAGCCCGGT**TAGTCGAACAAGTAGCTCCA		
1179D19	CBF14.2F	**GACGACGACAAG**ATGGACGCCGTCGACG	671	58
	CBF14.2R	**GAGGAGAAGCCCGGT**TAGTCGAACAAGTAGCTCCATGG		
3149L3	CBF15.0F	**GACGACGACAAG**ATGGACATGACCGGCTCC	752	60

Table 4.1 continued

	CBF15.0R	**GAGGAGAAGCCCGGTTA**GTAGCTCGAGAGCGCGG		
	CBF15.0-T2F	**GACGACGACAAG**ATGGACATGACCGGCTC	644	60
	CBF15.0-T2R	**GAGGAGAAGCCCGGTTA**GTATGTGTCCAGGTCCA		
	CBF15.0-T3F	**GACGACGACAAG**ATGGACATGACCGGCTC	590	60
	CBF15.0-T3R	**GAGGAGAAGCCCGGTTA**TCCGTCGCCAGAAGG		
	CBF15.0-T4F	**GACGACGACAAG**ATGGACATGACCGGCTC	314	60
	CBF15.0-T4R	**GAGGAGAAGCCCGG TTA**GGCGAAGTTGAGACACG		
	CBF15.0-T7F	**GACGACGACAAG**ATGGACATGACCGGCTC	134	60
	CBF15.0-T7R	**GAGGAGAAGCCCGGTTA**CGGGTGGCGCGTCTCCTT		
425P7	CBF17.0F	**GACGACGACAAG**ATGGACATGGGCAGCGAG	899	56
	CBF17.0R	**GAGGAGAAGCCCGGT**CACTGACTCCAGAACGCGGC		
	CBF17.0-T2F	**GACGACGACAAG**ATGGACATGGGCAGCGAG	683	60
	CBF17.0-T2R	**GAGGAGAAGCCCGGTTA**CACCTCCAACACTTT		
	CBF17.0-T3F	**GACGACGACAAG**ATGGACATGGGCAGCGAG	569	60
	CBF17.0-T3R	**GAGGAGAAGCCCGGTTA**CGGCACCTCGAA		
	CBF17.0-T4F	**GACGACGACAAG**ATGGACATGGGCAGCGAG	365	60
	CBF17.0-T4R	**GAGGAGAAGCCCGGTTA**CGGAACTGCGACT		
210D22	CBF19.1F	**GACGACGACAAG**ATGGACATGGCCATCGACA	731	65
	CBF19.1R	**GAGGAGAAGCCCGGT**CAGTAGCTCCAGAGGCCG		
1144N5	CBF19.2F	**GACGACGACAAG**ATGGACACGGCCATCGAC	731	65
	CBF19.2R	**GAGGAGAAGCCCGGT**CAGTAGCTCCAGAGGCCG		
567H13	CBF20.0F	**GACGACGACAAG**ATGGACACCGCCGCCCC	680	60
	CBF20.0R	**GAGGAGAAGCCCGGTTA**ATTAGTCG		
567H13	CBF21.0F	**GACGACGACAAG**ATGGACGCCGACGCT	620	62
	CBF21.0R	**GAGGAGAAGCCCGGTTA**CACGTCGCC		
1408B5	CBF22.0F	**GACGACGACAAG**ATGGACGTCGCCG	854	60
	CBF22.0R	**GAGGAGAAGCCCGGTTA**GTTGAACAGGTG		

*Nucleotides in bold represent the overhang required for ligation independent cloning. The remaining nucleotides match 100% to CBF coding region with the ATG, start codon, boxed and stop codon underlined.