2018

Expression Analyses and Identification of Key Molecular Participants in Plant Responses to Environmental Cues

Praveena Kanchupati
South Dakota State University

EXPRESSION ANALYSES AND IDENTIFICATION OF KEY MOLECULAR PARTICIPANTS IN PLANT RESPONSES TO ENVIRONMENTAL CUES

BY

PRAVEENA KANCHUPATI

A dissertation submitted in partial fulfillment of the requirements for the

Doctor of Philosophy

Major in Biological Sciences

Specialization in Biology

South Dakota State University

2018

EXPRESSION ANALYSES AND IDENTIFICATION OF KEY MOLECULAR PARTICIPANTS IN PLANT RESPONSES TO ENVIRONMENTAL CUES

PRAVEENA KANCHUPATI

This dissertation is approved as a creditable and independent investigation by a candidate for the Doctor of Philosophy in Biological Sciences degree and is acceptable for meeting the dissertation requirements for this degree. Acceptance of this does not imply that the conclusions reached by the candidates are necessarily the conclusions of the major department.

Yajun Wu, Ph.D. Date
Dissertation Advisor

Volker S. Brözel, Ph.D. Date
Head, Department of
Biology and Microbiology

Dean, Graduate School Date

Dedicated to my parents, my life-coaches Anantha Lakshmi and Venketeswar Rao: because I owe it all to you!

ACKNOWLEDGEMENTS

This work becomes a reality with the kind support and help of many individuals. I would like to extend my sincere heartfelt thanks to all of them.

I am most grateful to my dissertation advisor, Dr. Yajun Wu, for his guidance and continuous support throughout my time as a graduate student. I would like to express my sincere gratitude for his patience, motivation and enthusiasm over the years especially at times when everything seemed to be like a lost cause. I am also thankful to my committee members Dr. Yang Yen, Dr. Arvid Boe, Dr. Donald Auger, Dr. Senthil Subramanian and Dr. Robert Schmidt for their continued faith and encouragement over the years. Without the ongoing support from all these individuals, I could not have finished this work.

A big thank you to my lab mates, Rokebul Anower, Krishna Ghimire and Yafang Wang for sharing their time and expertise with me. Their close collaborative work and help will never be forgotten. I would also like to thank my colleagues for their friendship and all the fun we have had in the last six years. You guys are awesome!

Special thanks to Department of Biology and Microbiology, South Dakota State University; USDA-ARS and North Central SunGrant Initiative for research support. I would also like to extend my sincere thank you to FGCF and its staff at SDSU.

My family members have been the front and center, with their support, in my pursuit of this project. I cannot thank my parents enough, as it is because of their constant support and prayers that I was able to pursue my goal. I am ever grateful to my husband, Ghana Shyam, for his unconditional love, encouragement and positive criticism. Last but not the least, I wish to thank my son, Lohith, for providing unending inspiration and fun!

TABLE OF CONTENTS

ABSTRACT

EXPRESSION ANALYSES AND IDENTIFICATION OF KEY MOLECULAR PARTICIPANTS IN PLANT RESPONSES TO ENVIRONMENTAL CUES

PRAVEENA KANCHUPATI

2018

Plants constantly engage and interact with the environment and respond to the changes in conditions like temperature, water, and photoperiod, by regulating expression of genes of multiple regulatory and signaling pathways. Insight into these pathways and their participants has provided and will provide candidates to improve various agronomically important traits in crops through marker-assisted breeding and genetic manipulation. With this aim in mind, in the present study, I attempted to identify key candidate genes that are involved in the regulation of; i) plant response to low temperature stress, ii) plant roots' response to soil moisture content and iii) flowering time.

I identified and studied the expression of *C-repeat binding factors* (*CBFs*)-like genes and *CONSTANS* (*CO*)-, *FVE*- and *FCA*-like genes in alfalfa (*Medicago sativa* L.), a major forage crop in United States and worldwide. *C-repeat binding factors* (*CBFs*) are key transcription factors involved in plants' response to low temperatures. The results based on the gene expression and its correlation with freezing tolerance in alfalfa suggested that two *MsCBFl* genes might play important role in freezing tolerance in alfalfa. Through expression analysis of *CO-like*, *FVE-like* and *FCA*-like genes in different tissues, at differential stages, and under circadian control, I identified several genes in *CO*-, *FVE*- and *FCA*-like gene families that are potential functional homologues

involved in flowering time control. These candidate genes, once function is confirmed, can be used to delay flowering in alfalfa which will lead to higher biomass production and higher quality forage due to delayed senescence, a trait associated with flowering. An enhancement in biomass production can also pave the way for its use in cellulosic-based biofuel production.

To understand the molecular basis of plant roots' response to soil moisture gradient in major crops, a phenomenon known as hydrotropism, I attempted to identify the functional homologue of Arabidopsis *MIZ1*, one of the key regulators of hydrotropism, in maize (*Zea mays* L.). Through analysis of expression of *MIZ1*-like genes in maize, one gene, *ZmMIZ1l-K* appeared to be the candidate functioning in hydrotropic response. This study is the first attempt at understanding molecular players in hydrotropism in a crop plant and could be potentially used to enhance water acquisition of crop plants and thus their performance especially under drought conditions.

My research demonstrated that the candidate gene approach I took can be a first step to effectively identify the key players in the regulatory pathways in major crops. Through studying these genes, I also provided great insight into the complexity of molecular processes in responding to environmental cues in crop plants. Additional studies are needed to confirm the gene functionalities and their key roles in these processes. The molecular participants can then be used as resources to develop better crop varieties that could perform efficiently especially under more severe environmental stresses like drought and harsh temperatures.

CHAPTER 1: A LITERATURE REVIEW

1.1. Plant-environment interactions

Being sessile in nature, plants need to continuously modify and regulate their growth and development in accordance to the ever-changing surroundings. Such response requires complex and coordinated integration of signals from multiple regulatory pathways, that ultimately enables the plants to efficiently compete with others for the necessary resources like water, nutrients, light etc. The quick and dynamic regulation at the molecular level also allows the plants to respond to seasonal changes in day length, temperature, and other environmental stimuli important for growth and reproduction. Over the years, there has been an exponential increase in human population, industrialization and metropolitanization, that has resulted in a great change in climate globally. The availability of fertile land and water is decreasing, the soils of farmland are getting depleted of essential nutrients and the air is getting polluted. Thus, the plants are now often challenged by more severe environmental stresses like drought, harsh temperatures and limited nutrients. Plants have evolved various strategies to deal with these set of challenges. Insight into the key molecular participants of the multiple regulatory pathways has become indispensable. These studies have provided and will provide for powerful tools to improve various agronomically important traits in crops through marker-assisted breeding and genetic manipulation and allow plant biologists to reach the goal of food security through sustainable agriculture practices.

1.2. Arabidopsis and other plant species

Major biological processes in plants are under the direct impact of the environment. Photosynthesis, photorespiration, vegetative growth and flowering are but

only a few of the processes that are affected and adjusted according to those environmental conditions. The revolution in understanding the molecular mechanisms of important plant processes began with the use of thale cress, *Arabidopsis thaliana* (The Arabidopsis Genome 2000), and our knowledge advanced rapidly after the sequencing of its genome. The small genome size and easy manipulation of the genes in combination of a small plant size and a rapid life cycle (30-45 days) have established Arabidopsis as the model plant, and substantial progress has been made in understanding the molecular regulation of plant processes using the forward and reverse genetics approach. Identification of genes and determination of their function in Arabidopsis and then extrapolating that information to crop species like rice, soybean or maize is a common approach. But, with the sequencing of additional plant genomes (Eckardt 2000; Schmutz et al. 2010) and gene function analyses it is becoming more and more clear that, though there is some conservation of gene function, the divergence of gene function and emergence of novel gene function is also evident. Thus, the need to study and functionally characterize the molecular pathways in other plants is equally important and essential. In order to develop powerful tools to accommodate for growing population and decreasing cultivation land, plant biologists need to understand fine details about the molecular architecture of not just the model plants, but more importantly the crop plants.

1.3. Molecular participants regulating response to low temperature

One of the major environmental factors affecting growth, development, and biomass production in plants is temperature. Both, increase or decrease in temperature results in suboptimal plant growth and can adversely affect its production. Every year, decreased overall productivity is reported in many crops due to exposure to chilling or

subzero temperatures (Allen and Ort 2001; Thomashow 1999). Cold-induced damage is evident at both physiological and molecular levels. Formation of ice crystals in the vegetative tissues, shutting down of major metabolic pathways as the plants try to survive, are but a few of the adverse effects of low temperature stress. While most plants can't survive sudden freezing or severe freezing temperatures, it is observed that plants can show significantly improved tolerance to chilling or subfreezing temperatures if they are exposed to low nonfreezing temperatures prior to being subjected to freezing. Such adaptation mechanism is termed cold acclimation (Thomashow 1999; Chinnusamy et al. 2007). Research have now revealed that the plant cells reprogram several processes at the biochemical and physiological levels during the exposure to low nonfreezing temperatures. The plasma membranes are believed one of the first sensors. Low-temperature conditions causes rigidification of the membranes, which triggers cytoskeletal rearrangement and induces expression of cold-regulated (*COR*) genes (Viswanathan and Zhu 2002). Accumulation of solutes like proline, sugars, and similar cryprotectants is another important consequence of cold acclimation (Thomashow 1999; Guy et al. 2008). Many of the *COR* genes are involved in synthesis of the cryprotectants, change membrane fluidity, and damage repairs. Thus, altered gene expression at the molecular level plays a critical role for plant survival under low temperature (Cook et al. 2004; Hannah et al. 2005; Maruyama et al. 2009).

Among the genes whose expression is changed in cold acclimation are many transcription factors. Some of the cold-induced transcription factors include *inducer of CBF expression 1*, *ICE1* (Chinnusamy et al. 2003) and members of the calmodulin binding transcription activator family (CAMTA) (Doherty et al. 2009). But the most

studied transcription factors are the C-repeat binding factors (CBFs). They are also referred to as the dehydration-responsive element binding factors (DREBs). In Arabidopsis, three *CBFs*, namely *AtCBF1/DREB1B*, *AtCBF2/DREB1C* and *AtCBF3/DREB1A* have been demonstrated to play important roles in regulating cold stress response (Stockinger et al. 1997; Gilmour et al. 1998; Jaglo-Ottosen et al. 1998; Liu et al. 1998; Riechmann and Meyerowitz 1998; Medina et al. 1999; Gilmour et al. 2000; Zhao et al. 2016). The CBF transcription factors recognize the C-repeat (CRT)/ dehydration-responsive element (DRE) present in the promoters of the downstream cold-responsive (*COR*) genes (Stockinger et al. 1997; Liu et al. 1998; Sakuma et al. 2002; Maruyama et al. 2012) and regulate the cold response in Arabidopsis.

The *CBF* genes comprises a gene family in Arabidopsis (Riechmann and Meyerowitz 1998). *CBF1*,*2* and *3* transcripts start to accumulate rapidly and reach their peak level of expression after about 2h of exposure to low temperature treatment (Gilmour et al. 1998; Medina et al. 1999), followed by rapid upregulation of downstream *COR* genes. Among the three homologs, *CBF3* in particular has been shown to be the key regulator and integrator of multiple biochemical changes involved in the process of cold acclimation (Gilmour et al. 2000). Over-expression of *CBF3* resulted in increased freezing tolerance in the transgenic Arabidopsis plants. These plants could survive at temperatures as low as -8 °C whereas the WT plants were severely damaged at temperatures of -4°C. Given their importance in the freezing tolerance, functional homologs of Arabidopsis *CBF/DREB1* genes have been identified in many plants including crop species like barley, wheat, soybean, and rice (Choi et al. 2002; Dubouzet et al. 2003; Skinner et al. 2005; Badawi et al. 2007; Kidokoro et al. 2015; Yamasaki and

Randall 2016). These studies have established *CBFs* conserved role across diverse plant species in enhancing freezing tolerance. In addition, these studies have also shown an expansion of the *CBF* gene family in different crop plants suggesting the possibility of the involvement of the homologues in diverse functions. Soybean genome contains thirteen CBF/*DREBs* genes and *GmDREBA1* and *GmDREBA2* have been implicated in cold response where as *GmDREBE1* and *GmDREBG1* were shown to be regulated in response to heat (Kidokoro et al. 2015). Likewise, *MtCBF2* and *MtCBF3* where shown to be upregulated under cold stress, while *Medicago truncatula CBF4* has been shown to respond to salinity stress and overexpressing this gene in *Medicago* has resulted in improved salt-tolerance in the transgenic plants (Li et al. 2011).

However, few studies were aimed at identification of *CBFl* genes and determination of their role in agronomically important legume plant alfalfa (*Medicago sativa* L.). Alfalfa is a major forage crop with important agronomic and environmental traits (Castonguay et al. 2009). In 2013, approximately 18 million acres of alfalfa with a production value of $10.7 billion were harvested in United States. It has the ability to fix free-nitrogen by being in symbiotic relationship with nitrogen-fixing rhizobial bacteria. It is a very important source of protein and fiber for the livestock and is grown worldwide under varying and diverse environments. But, unpredictable drops in temperatures in early spring and late fall often result in decreased production and overall yield (Castonguay et al. 2013; Anower et al. 2016).

Conventional breeding methods have been employed with some improvement in the freezing tolerance of alfalfa. For example, Apica and Caribou are two varieties that were developed by Castonguay et al. (2009) through greenhouse screening of mature

alfalfa plants. Traditional breeding methods are extremely time-consuming and slow in generating new cultivars (Thomashow 1999; Castonguay et al. 2013). Screening for cold-tolerant germplasm is another important strategy for breeding or gene identification. Anower et al (2016) employed a freezing survival test in addition to electrolyte leakage assays and identified two freezing-tolerant alfalfa genotypes.

Given the fact the CBF3 genes appear to be structurally and functionally conserved in other plants, it is reasonable to hypothesize that improved freezing tolerance can be achieved through molecular engineering of *CBF* functional homolog in the crop. Our understanding of cold stress response in alfalfa at molecular level is limited (Castonguay et al. 2013, Castonguay et al. 2009). The importance of the *CBF* regulon and its contribution to freezing tolerance in alfalfa is still not known. The molecular dissection of freezing tolerance will provide foundation of more rapid cultivar development through marker-assisted breeding or genetic engineering.

1.4. Molecular insight into plant root response to soil moisture content

Water is one of the major essential resources and the distribution of water in the soil surrounding plant roots is constantly changing and is non-homogeneous. The situation is worsened in case of drought, one of the major abiotic stresses affecting plant growth and productivity all over the world. Drought stress can severely affect corn production, as much as 30-50% drop in yield can be seen in such conditions (http://agfax.com/2017/07/18/north-dakota-drought-75-of-state-suffering-wheat-yield-estimates-11-bu-per-acre-dtn/). Northern plains states, such as South Dakota and North Dakota, experienced another severe drought in 2017 after the historic drought in 2012

(https://www.agweb.com/article/historic-drought-hammers-dakotas-montana-naa-chris-bennett/).

Plants respond to drought through complex physiological and molecular processes (Osakabe et al. 2014; Joshi et al. 2016; Kaur and Asthir 2017). Of these processes, maintaining good water status is the key for surviving and reproducing. This is achieved by reducing water loss by closing stomates, reducing leaf expansion and production (thus reducing transpiring area) and accumulating osmotican or compatible solutes such as sugars and amino acids. Meanwhile, roots continue to explore water in the soil. Roots are less sensitive to drought stress (Sharp et al. 1988), resulting in a higher root/shoot ratio under drought. Roots continue to elongate under drought, growing deeper into the soil due to gravitropism, allowing them to reach water in deep soil. Meanwhile, roots can sense the moisture gradient to development roots toward a wetter area, a process now called hydropatterning (Bao et al. 2014) or grow directly toward water source, a process called hydrotropism. Different form gravitropism, hydropatterning and hydrtropism allows roots to explore water laterally. While gravitropism in roots is extensively studied and the molecular process is mostly understood, the sensing and signaling pathways in hydropatterning and hydrotropism is largely unknown. Part of this dissertation research is focused on identifying candidate genes involved in hydrotropism in roots.

Hydrotropism has been observed for more than 100 yrs. (Darwin and Darwin 1880; Loomis and Ewan 1936). Research on hydrotropism has been very limited. The most important reason for the limited number of studies in this area is the difficulty in separating hydrotropic response from that of gravitropic and thigmotropic responses. The pioneering research on hydrotropism was conducted by Jaffe et al. in 1985, where they

studied the pea mutant *ageotropum* that showed a positive hydrotropic response and no gravitropic response (Jaffe et al. 1985). The study highlighted the role of root caps in sensing the moisture gradient and clearly showed the presence of two separate sensing and signaling pathways in plant roots to differentiate gravity from moisture gradients.

Through physiological, biochemical, and some genetic analysis, a model of hydrotropism has been proposed (Figure 1). After the moisture gradient is sensed by the root cap, the signal is transmitted upwards to the elongation zone of the roots to initiate bending. *MIZ1* operates upstream of *MIZ2* to reduce auxin levels and this differential accumulation of auxin ultimately results in cell elongation on the dry side resulting in the bending of the roots. ABA is hypothesized to enhance the expression of *MIZ1* resulting in enhanced hydrotropic response of plant roots (Moriwaki et al. 2013). A recent study in Arabidopsis showed the root cortex cells may be a major site for sensing the moisture gradient (Dietrich et al. 2017), thus the model may need to be modified. Nakajima et al. (2017) reported that the mechanism involved in hydrotropism very likely vary depending on species.

The molecular participants of the sensing and signaling pathways regulating hydrotropism are mostly unknown. Thus far only two genes, namely *MIZ1* and *MIZ2*, have been identified based off the forward genetics approaches, i.e. mutant analysis studies, in the model plant Arabidopsis. The very first evidence of a direct link between the mechanism of hydrotropism and a molecular component in the associated pathway, came from the studies of the Arabidopsis *mizu-kussei1* mutants (Kobayashi et al. 2007). The mutants' roots displayed normal gravitropism and growth but lacked hydrotropic response. They also displayed wavy growth in roots and reduced sensitivity to light. The

modified phenotype of the mutant was linked to a recessive mutation in *MIZ1* gene. *MIZ1* encodes for a soluble protein (Yamazaki et al. 2012) with an unknown function and contains a DUF617 (domain of unknown function 617) domain and the gene homologues are only found in the genome of terrestrial plants. Expression analysis of the gene showed extensive transcript accumulation in the columella cells of the root caps. Some expression was also observed in the mature regions of the roots and in the hydathodes (water-excreting epidermal structures) of the leaves (Kobayashi et al. 2007). The overexpression lines of *MIZ1* (*MIZ1OEs*) displayed extraordinarily enhanced hydrotropic response and outnumbered the viable wild type Arabidopsis plants when grown under hydrostimulated conditions (Miyazawa et al. 2012).

The second gene that was shown to play an essential role in hydrotropism in Arabidopsis is *MIZ2* (Miyazawa et al. 2009). *miz2* mutants have mutation in *GNOM* that encodes for a guanine-nucleotide exchange factor for ADP-ribosylation factor-type G proteins. *miz2* mutants similar to *miz1* mutants, are ahydrotropic. Auxin, *MIZ1* and *MIZ2* work together to regulate lateral root development during hydrostimulated conditions. Apart from the observation that MIZ1 requires *MIZ2* activity for its function (Moriwaki et al. 2011) the role of *GNOM*/*MIZ2* in root hydrotropism is still unidentified. Altogether, various studies have suggested multiple players that regulate hydrotropism in roots, but a clear mechanism is still not understood.

Up until now, Arabidopsis is the most studied plant system in relation to the elucidation of the molecular players of hydrotropism. A recent study identified novel QTLs associated with hydrotropism in wheat (Hamada et al. 2012). Though the phenomenon of hydrotropism has been studied in some other plant species like pea

(Takahashi et al. 1992), soybean (Tsutsumi et al. 2002) and maize (Takahashi and Scott 1991) the molecular mechanism regulating the phenomenon are yet to be identified.

Maize (*Zea mays* L.) is one of the most important grain crop. Cereals including maize account for majority of the human calorie consumption across the world (Chandler and Brendel 2002), and in the United States it is a leading staple crop in term of production along with wheat and soybean. Its high yield and production depends on water availability. The water requirement for maize crop is low during the early stages of development and reaches the maximum at the reproductive stages and then diminishes again. Two weeks before and after pollination are very critical in terms of water requirement. Important traits like grain-filling and soft-dough formation are highly sensitive to water restriction. Drought at these stages can lead to severe yield loss. Thus, efficient water acquisition by roots is essential to reduce the negative impact from drought. To achieve that, one of the effective strategies is to enhance plants' hydrotropism. In other words, molecular insight into the hydrotropism mechanism will provide for clues and tools required to improve this property of roots and allow for the development of varieties that will have better chance at yield or even survival under drought stress conditions (Aslam et al. 2015).

1.5. Molecular regulation of flowering time

The life cycle of a higher plant can be divided into two major phases, vegetative and reproductive phase. For higher plants including major crops, flowering marks the beginning of reproductive phase, followed by seed formation and senescence. Numerous environmental and endogenous cues like light, temperature, the circadian clock, age of the plants and growth elicitors affect plants' transition from the vegetative phase to

reproductive phase (flowering). Flowering, an important agronomic trait, is under very tight and complex regulation. These regulatory networks monitor and coordinate subtle changes in the environment with the endogenous signals, and, then direct the plants' response accordingly. This strict monitoring ensures that the process of floral induction only occurs under most favorable conditions resulting in maximum reproductive vigor.

Most of our current understanding of the flowering process has come from molecular dissection of floral-induction pathways in the model plant, *Arabidopsis thaliana*. More than 180 genes have been identified as the regulators of these pathways (Bäurle and Dean 2006; Fornara et al. 2010). These genes act through six major pathways namely, i) photoperiod and ii) vernalization pathways that monitor the seasonal changes in day length and temperature; iii) ambient temperature pathway that responds to changes in daily temperatures; iv) gibberellin, v) autonomous and vi) age pathways that act independent of the environmental stimuli. The integration of signals from all these pathways is carried out by a set of genes named the "floral integrator". *FLOWERING LOCUS T* (*FT*), *SUPRESSOR OF CONSTANS1* (*SOC1*) and *LEAFY* (*LFY*) integrate the signals and rapidly promote floral development. These integrators then communicate with the downstream "floral meristem identity" genes like *APETALA 2* (*AP1*) to induce flowering (Parcy 2004; Simpson and Dean 2002). A model of flowering regulation has been proposed and shown in Figure 2 (Fornara et al. 2010).

Flowering is a key developmental process in a plant's life cycle and is directly linked to crop production and overall yields. Enhanced understanding of the molecular basis of this complex process in crops can be of a huge advantage to the researchers trying to develop new varieties that have improved productivity and better yield. Indeed,

plant breeders have long been manipulating the flowering regulatory mechanism as one of the key strategies to achieve optimal production of vegetative biomass or high yield of seed or grains, depending on specific crops.

As described earlier, alfalfa is a forage crop that has important agronomic and environmental traits. Along with being a legume and assisting in fixing free-nitrogen, the deep root system of alfalfa allows it to flourish under mild drought conditions. Alfalfa is also gaining ground as a potential candidate for biofuel production due to its high cellulosic biomass yield with a low input from nitrogen fertilizer. However, the high cost of biomass production in alfalfa for biofuel purpose is prohibitive. One way to reduce the price is to enhance biomass production per unit area. Flowering in alfalfa suppresses the vegetative growth and initiate senescence. Thus, a strategy to enhance biomass production is to delay flowering. A recent study reported that the genetic manipulation of a microRNA miR156 in alfalfa resulted in delayed flowering and subsequent increase in biomass. Additionally, the authors also observed reduced lignin content and enhanced cellulose content in the transgenic alfalfa overexpressing miR156 (Aung et al. 2015b, a). Another study in *M. truncatula* reported the manipulation of onset of flowering to enhance biomass and suggested genetically delaying the floral initiation as an easy tool to achieve improved biomass quality and quantity (Tadege et al. 2015). These studies support the hypothesis that genetic manipulation of flowering genes and associated signaling pathways can be used as efficient tools to delay flowering and prolong vegetative state and achieve significant increase in biomass.

Like Arabidopsis, alfalfa is considered a long day plant and may share similar molecular mechanism in flowering time regulation. Among the genes that are often

studied in Arabidopsis are *CONSTANS* (*CO*) in the photoperiod pathway and FCA and FVE genes in the autonomous pathway. *CO* genes comprises a multiple gene family. *co* mutant plants showed delay flowering in long day conditions and enhanced vegetative growth (Koornneef et al. 1991; Putterill et al. 1995) and overexpression of *CO* promoted early flowering (Onouchi et al. 2000). CO promoted flower by activating downstream genes that include floral integrator genes like *FT* and *SOC1*. *fca* and *fve* mutant showed significant delay of flowering and enhanced vegetative biomass production (Macknight et al. 1997; Morel et al. 2008). *FCA* and *FVE* are believed to suppress expression of *FLC,* that acts as a negative regulators of *FT* (Salathia et al. 2006).

Studies in other plants like rice, barley and sugar beet showed that manipulation of these gene can result in similar phenotypes as seen in Arabidopsis, suggesting the gene functions and pathways are very much conserved in different plant species (Campoli et al. 2012; Yano et al. 2000; Il-Sun et al. 2008)

1.6. Objectives, rationale and hypothesis

Objective 1: To identify the potential functional homolog of *AtCBF3*, the key regulator of cold tolerance in Arabidopsis, and to establish a better understanding of cold tolerance mechanisms at the molecular level in alfalfa.

Rationale: Anower et al reported the screening and identification of alfalfa germplasm River side (RS) and Foster ranch (FR), collected from the Grand River National Grassland in South Dakota, that had superior freezing tolerance compared to the commercial cold-tolerant germplasm Apica (AP). The transcript analyses of some of the genes involved in cold tolerance pathway showed distinct cold induction patterns in the

cold-tolerant germplasm RS, FR and AP, in contrast to the cold-susceptible germplam CUF-101(Anower et al. 2016).

Hypothesis: The superior performance of RS and FR under cold stress is the result of the differential induction of the *CBF*-regulon in contrast to the cold-susceptible alfalfa germplasm.

Objective 2: To identify the functional homologue of *AtMIZ1* and to establish a better understanding of the hydrotropic responses of maize seedling roots at the molecular level.

Rationale: *AtMIZ1* is a key gene involved in hydrotropic response in Arabidopsis roots. The overexpression lines of *MIZ1* (*MIZ1OEs*) displayed enhanced hydrotropic response and outnumbered the viable wild type Arabidopsis plants when grown under hydrostimulated conditions (Miyazawa et al. 2012).

Hypothesis: Since maize primary roots show positive hydrotropic response, the molecular mechanism of hydrotropism in maize roots may be conserved and may involve *MIZ1-like* gene. Identification of MIZ1-like gene will provide valuable insight into the hydrotropism response in maize. Manipulating the genes regulating hydrotropism will result in better performance of agriculturally important crops like maize, specifically under drought conditions.

Objective 3: To understand the conservation and divergence of two key pathways, the photoperiod and autonomous pathway in flowering control through identifying and studying *CO*-, *FCA*- and *FVE-like* genes in alfalfa.

Rationale: *CO*, *FCA*, *FVE* are key regulators of flowering time control in Arabidopsis. Delaying flowering by manipulating these genes resulted in enhanced vegetative biomass production. These strategies can be used to enhance biomass production in alfalfa.

Hypothesis: Flowering control in alfalfa may be similar to that in Arabidopsis and involve *CO*-, *FCA*- and *FVE*-like genes. Identification of these genes in alfalfa will provide valuable insight in flowering control in alfalfa.

In the following chapters, I report the expression analyses and identification of potential functional homologs of *Arabidopsis* key molecular participants, that regulate response to low temperatures and moisture gradient and control flowering, in alfalfa and maize.

1.7. References

Allen DJ, Ort DR (2001) Impacts of chilling temperatures on photosynthesis in warm-climate plants. Trends in Plant Science 6 (1):36-42. doi:https://doi.org/10.1016/S1360-1385(00)01808-2

Anower MR, Fennell A, Boe A, Mott IW, Peel MD, Wu Y (2016) Physiological and molecular characterisation of lucerne (*Medicago sativa* L.) germplasm with improved seedling freezing tolerance. Crop and Pasture Science 67 (6):655-665. doi:http://dx.doi.org/10.1071/CP15204

Aslam M, Maqbool MA, Cengiz R (2015) Effects of Drought on Maize. In: Drought Stress in Maize (Zea mays L): Effects, Resistance Mechanisms, Global Achievements and Biological Strategies for Improvement. Springer International Publishing, Cham, pp 5-17. doi:10.1007/978-3-319-25442-5_2

Aung B, Gruber MY, Amyot L, Omari K, Bertrand A, Hannoufa A (2015a) Ectopic expression of LjmiR156 delays flowering, enhances shoot branching, and improves forage quality in alfalfa. Plant Biotechnology Reports 9 (6):379-393. doi:10.1007/s11816-015-0375-2

Aung B, Gruber MY, Amyot L, Omari K, Bertrand A, Hannoufa A (2015b) MicroRNA156 as a promising tool for alfalfa improvement. Plant Biotechnol J 13 (6):779-790. doi:10.1111/pbi.12308

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 (5):533-554. doi:10.1007/s00438-006-0206-9

Bao Y, Aggarwal P, Robbins NE, Sturrock CJ, Thompson MC, Tan HQ, Tham C, Duan L, Rodriguez PL, Vernoux T, Mooney SJ, Bennett MJ, Dinneny JR (2014) Plant roots use a patterning mechanism to position lateral root branches toward available water. Proceedings of the National Academy of Sciences 111 (25):9319-9324. doi:10.1073/pnas.1400966111

Bäurle I, Dean C (2006) The Timing of Developmental Transitions in Plants. Cell 125 (4):655-664. doi:10.1016/j.cell.2006.05.005

Campoli C, Drosse B, Searle I, Coupland G, von Korff M (2012) Functional characterisation of HvCO1, the barley (Hordeum vulgare) flowering time ortholog of CONSTANS. The Plant Journal 69 (5):868-880. doi:10.1111/j.1365-313X.2011.04839.x

Castonguay Y, Dubé M-P, Cloutier J, Bertrand A, Michaud R, Laberge S (2013) Molecular physiology and breeding at the crossroads of cold hardiness improvement. Physiologia Plantarum 147 (1):64-74. doi:10.1111/j.1399-3054.2012.01624.x

Castonguay Y, Michaud R, Nadeau P, Bertrand A (2009) An indoor screening method for improvement of freezing tolerance in alfalfa. Crop Science 49 (3):809-818

Chandler VL, Brendel V (2002) The Maize Genome Sequencing Project. Plant Physiology 130 (4):1594-1597. doi:10.1104/pp.015594

Chinnusamy V, Ohta M, Kanrar S, Lee B-h, Hong X, Agarwal M, Zhu J-K (2003) ICE1: a regulator of cold-induced transcriptome and freezing tolerance in Arabidopsis. Genes & Development 17 (8):1043-1054. doi:10.1101/gad.1077503

Chinnusamy V, Zhu J, Zhu J-K (2007) Cold stress regulation of gene expression in plants. Trends in Plant Science 12 (10):444-451. doi:http://dx.doi.org/10.1016/j.tplants.2007.07.002

Choi D-W, Rodriguez EM, Close TJ (2002) Barley *Cbf3* Gene Identification, Expression Pattern, and Map Location. Plant Physiology 129 (4):1781-1787. doi:10.1104/pp.003046

Cook D, Fowler S, Fiehn O, Thomashow MF (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 (42):15243-15248. doi:10.1073/pnas.0406069101

Darwin C, Darwin F (1880) The Power of Movement of Plants. John Murray,

Dietrich D, Pang L, Kobayashi A, Fozard JA, Boudolf V, Bhosale R, Antoni R, Nguyen T, Hiratsuka S, Fujii N, Miyazawa Y, Bae T-W, Wells DM, Owen MR, Band LR, Dyson RJ, Jensen OE, King JR, Tracy SR, Sturrock CJ, Mooney SJ, Roberts JA, Bhalerao RP, Dinneny JR, Rodriguez PL, Nagatani A, Hosokawa Y, Baskin TI, Pridmore TP, De Veylder L, Takahashi H, Bennett MJ (2017) Root hydrotropism is controlled via a cortex-specific growth mechanism. Nature Plants 3:17057. doi:10.1038/nplants.2017.57

https://www.nature.com/articles/nplants201757 - supplementary-information

Doherty CJ, Van Buskirk HA, Myers SJ, Thomashow MF (2009) Roles for Arabidopsis CAMTA Transcription Factors in Cold-Regulated Gene Expression and Freezing Tolerance. The Plant Cell 21 (3):972-984. doi:10.1105/tpc.108.063958

Dubouzet JG, Sakuma Y, Ito Y, Kasuga M, Dubouzet EG, Miura S, Seki M, Shinozaki K, Yamaguchi-Shinozaki K (2003) *OsDREB* genes in rice, *Oryza sativa* L., encode transcription activators that function in drought-, high-salt- and cold-responsive gene expression. The Plant Journal 33 (4):751-763. doi:10.1046/j.1365-313X.2003.01661.x

Eckardt NA (2000) Sequencing the Rice Genome. The Plant Cell 12 (11):2011-2018

Fornara F, de Montaigu A, Coupland G (2010) SnapShot: Control of Flowering in Arabidopsis. Cell 141 (3):550-550.e552. doi:10.1016/j.cell.2010.04.024

Gilmour SJ, Sebolt AM, Salazar MP, Everard JD, Thomashow MF (2000) Overexpression of the Arabidopsis *CBF3* Transcriptional Activator Mimics Multiple Biochemical Changes Associated with Cold Acclimation. Plant Physiology 124 (4):1854-1865

Gilmour SJ, Zarka DG, Stockinger EJ, Salazar MP, Houghton JM, Thomashow MF (1998) Low temperature regulation of the*Arabidopsis* CBF family of AP2 transcriptional activators as an early step in cold-induced *COR* gene expression. The Plant Journal 16 (4):433-442. doi:10.1046/j.1365-313x.1998.00310.x

Guy C, Kaplan F, Kopka J, Selbig J, Hincha DK (2008) Metabolomics of temperature stress. Physiologia Plantarum 132 (2):220-235. doi:10.1111/j.1399-3054.2007.00999.x

Hamada A, Nitta M, Nasuda S, Kato K, Fujita M, Matsunaka H, Okumoto Y (2012) Novel QTLs for growth angle of seminal roots in wheat (Triticum aestivum L.). Plant and soil 354 (1):395-405. doi:10.1007/s11104-011-1075-5

Hannah MA, Heyer AG, Hincha DK (2005) A global survey of gene regulation during cold acclimation in *Arabidopsis thaliana*. PLoS Genetics 1 (2):e26. doi:10.1371/journal.pgen.0010026

Il-Sun B, Hyo-Young P, Min Kyoung Y, Jeong Hwan L, Jeong-Kook K (2008) Functional Conservation and Divergence of FVE Genes that Control Flowering Time and Cold Response in Rice and Arabidopsis. Mol Cells 26 (4):368-372

Jaffe MJ, Takahashi H, Biro RL (1985) A Pea Mutant for the Study of Hydrotropism in Roots. Science 230 (4724):445-447. doi:10.1126/science.230.4724.445

Jaglo-Ottosen KR, Gilmour SJ, Zarka DG, Schabenberger O, Thomashow MF (1998) *Arabidopsis CBF1* overexpression induces *COR* genes and enhances freezing tolerance. Science 280 (5360):104-106

Joshi R, Wani SH, Singh B, Bohra A, Dar ZA, Lone AA, Pareek A, Singla-Pareek SL (2016) Transcription Factors and Plants Response to Drought Stress: Current Understanding and Future Directions. Frontiers in Plant Science 7:1029. doi:10.3389/fpls.2016.01029

Kaur G, Asthir B (2017) Molecular responses to drought stress in plants. Biologia Plantarum 61 (2):201-209. doi:10.1007/s10535-016-0700-9

Kidokoro S, Watanabe K, Ohori T, Moriwaki T, Maruyama K, Mizoi J, Myint Phyu Sin Htwe N, Fujita Y, Sekita S, Shinozaki K, Yamaguchi-Shinozaki K (2015) Soybean DREB1/CBF-type transcription factors function in heat and drought as well as cold stress-responsive gene expression. The Plant Journal 81 (3):505-518. doi:10.1111/tpj.12746

Kobayashi A, Takahashi A, Kakimoto Y, Miyazawa Y, Fujii N, Higashitani A, Takahashi H (2007) A gene essential for hydrotropism in roots. Proceedings of the National Academy of Sciences 104 (11):4724-4729. doi:10.1073/pnas.0609929104

Koornneef M, Hanhart C, Van der Veen J (1991) A genetic and physiological analysis of late flowering mutants in Arabidopsis thaliana. Molecular and General Genetics MGG 229 (1):57-66

Li D, Zhang Y, Hu X, Shen X, Ma L, Su Z, Wang T, Dong J (2011) Transcriptional profiling of Medicago truncatula under salt stress identified a novel CBF transcription factor MtCBF4 that plays an important role in abiotic stress responses. BMC Plant Biology 11 (1):109. doi:10.1186/1471-2229-11-109

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. The Plant Cell 10 (8):1391-1406

Loomis WE, Ewan LM (1936) Hydrotropic Responses of Roots in Soil. Botanical Gazette 97 (4):728-743. doi:10.1086/334600

Macknight R, Bancroft I, Page T, Lister C, Schmidt R, Love K, Westphal L, Murphy G, Sherson S, Cobbett C, Dean C (1997) FCA, a Gene Controlling Flowering Time in Arabidopsis, Encodes a Protein Containing RNA-Binding Domains. Cell 89 (5):737-745. doi:10.1016/S0092-8674(00)80256-1

Maruyama K, Takeda M, Kidokoro S, Yamada K, Sakuma Y, Urano K, Fujita M, Yoshiwara K, Matsukura S, Morishita Y, Sasaki R, Suzuki H, Saito K, Shibata D, Shinozaki K, Yamaguchi-Shinozaki K (2009) Metabolic pathways involved in cold acclimation identified by integrated analysis of metabolites and transcripts regulated by DREB1A and DREB2A. Plant Physiology 150 (4):1972-1980. doi:10.1104/pp.109.135327

Maruyama K, Todaka D, Mizoi J, Yoshida T, Kidokoro S, Matsukura S, Takasaki H, Sakurai T, Yamamoto YY, 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 (1):37-49. doi:10.1093/dnares/dsr040

Medina Jn, Bargues M, Terol J, Pérez-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 (2):463-470. doi:10.1104/pp.119.2.463

Miyazawa Y, Moriwaki T, Uchida M, Kobayashi A, Fujii N, Takahashi H (2012) Overexpression of MIZU-KUSSEI1 Enhances the Root Hydrotropic Response by Retaining Cell Viability Under Hydrostimulated Conditions in Arabidopsis thaliana. Plant and Cell Physiology 53 (11):1926-1933. doi:10.1093/pcp/pcs129

Miyazawa Y, Takahashi A, Kobayashi A, Kaneyasu T, Fujii N, Takahashi H (2009) GNOM-Mediated Vesicular Trafficking Plays an Essential Role in Hydrotropism of Arabidopsis Roots. Plant Physiology 149 (2):835-840. doi:10.1104/pp.108.131003

Morel P, Tréhin C, Breuil-Broyer S, Negrutiu I (2008) Altering FVE/MSI4 results in a substantial increase of biomass in Arabidopsis—the functional analysis of an ontogenesis accelerator. Molecular Breeding 23 (2):239. doi:10.1007/s11032-008-9229-8

Moriwaki T, Miyazawa Y, Kobayashi A, Takahashi H (2013) Molecular mechanisms of hydrotropism in seedling roots of Arabidopsis thaliana (Brassicaceae). American Journal of Botany 100 (1):25-34. doi:10.3732/ajb.1200419

Moriwaki T, Miyazawa Y, Kobayashi A, Uchida M, Watanabe C, Fujii N, Takahashi H (2011) Hormonal Regulation of Lateral Root Development in Arabidopsis Modulated by MIZ1 and Requirement of GNOM Activity for MIZ1 Function. Plant Physiology 157 (3):1209-1220. doi:10.1104/pp.111.186270

Nakajima Y, Nara Y, Kobayashi A, Sugita T, Miyazawa Y, Fujii N, Takahashi H (2017) Auxin transport and response requirements for root hydrotropism differ between plant species. Journal of Experimental Botany 68 (13):3441-3456. doi:10.1093/jxb/erx193

Onouchi H, Igeño MI, Périlleux C, Graves K, Coupland G (2000) Mutagenesis of Plants Overexpressing *CONSTANS* Demonstrates Novel Interactions among Arabidopsis Flowering-Time Genes. The Plant Cell 12 (6):885-900. doi:10.1105/tpc.12.6.885

Osakabe Y, Osakabe K, Shinozaki K, Tran L-SP (2014) Response of plants to water stress. Frontiers in Plant Science 5:86. doi:10.3389/fpls.2014.00086

Parcy F (2004) Flowering: a time for integration. International Journal of Developmental Biology 49 (5-6):585-593

Putterill J, Robson F, Lee K, Simon R, Coupland G (1995) The CONSTANS gene of arabidopsis promotes flowering and encodes a protein showing similarities to zinc finger transcription factors. Cell 80 (6):847-857. doi:https://doi.org/10.1016/0092-8674(95)90288-0

Riechmann JL, Meyerowitz EM (1998) The AP2/EREBP family of plant transcription factors. Biological chemistry 379:633-646

Sakuma Y, Liu Q, Dubouzet JG, 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 (3):998-1009. doi:http://dx.doi.org/10.1006/bbrc.2001.6299

Salathia N, Davis SJ, Lynn JR, Michaels SD, Amasino RM, Millar AJ (2006) FLOWERING LOCUS C-dependent and -independent regulation of the circadian clock by the autonomous and vernalization pathways. BMC Plant Biology 6 (1):10. doi:10.1186/1471-2229-6-10

Schmutz J, Cannon SB, Schlueter J, Ma J, Mitros T, Nelson W, Hyten DL, Song Q, Thelen JJ, Cheng J, Xu D, Hellsten U, May GD, Yu Y, Sakurai T, Umezawa T, Bhattacharyya MK, Sandhu D, Valliyodan B, Lindquist E, Peto M, Grant D, Shu S, Goodstein D, Barry K, Futrell-Griggs M, Abernathy B, Du J, Tian Z, Zhu L, Gill N, Joshi T, Libault M, Sethuraman A, Zhang X-C, Shinozaki K, Nguyen HT, Wing RA, Cregan P, Specht J, Grimwood J, Rokhsar D, Stacey G, Shoemaker RC, Jackson SA (2010) Genome sequence of the palaeopolyploid soybean. Nature 463:178. doi:10.1038/nature08670 https://www.nature.com/articles/nature08670 - supplementary-information

Sharp RE, Silk WK, Hsiao TC (1988) Growth of the Maize Primary Root at Low Water Potentials : I. Spatial Distribution of Expansive Growth. Plant Physiology 87 (1):50-57

Simpson GG, Dean C (2002) Arabidopsis, the Rosetta stone of flowering time? Science 296 (5566):285-289

Skinner JS, Zitzewitz J, Szűcs P, Marquez-Cedillo L, Filichkin T, Amundsen K, Stockinger EJ, Thomashow MF, Chen THH, Hayes PM (2005) Structural, functional, and phylogenetic characterization of a large *CBF* gene family in barley. Plant Molecular Biology 59 (4):533-551. doi:10.1007/s11103-005-2498-2

Stockinger EJ, Gilmour SJ, Thomashow MF (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 (3):1035-1040

Tadege M, Chen F, Murray J, Wen J, Ratet P, Udvardi MK, Dixon RA, Mysore KS (2015) Control of Vegetative to Reproductive Phase Transition Improves Biomass Yield and Simultaneously Reduces Lignin Content in Medicago truncatula. BioEnergy Research 8 (2):857-867. doi:10.1007/s12155-014-9565-y

Takahashi H, Brown CS, Dreschel TW, Scott TK (1992) Hydrotropism in Pea Roots in a Porous-tube Water Delivery System. HortScience 27 (5):430-432

Takahashi H, Scott TK (1991) Hydrotropism and Its Interaction with Gravitropism in Maize Roots. Plant Physiology 96 (2):558-564. doi:10.1104/pp.96.2.558

The Arabidopsis Genome I (2000) Analysis of the genome sequence of the flowering plant Arabidopsis thaliana. Nature 408:796. doi:10.1038/35048692 https://www.nature.com/articles/35048692 - supplementary-information

Thomashow MF (1999) Plant cold acclimation: Freezing tolerance genes and regulatory mechanisms. Annual Review of Plant Physiology and Plant Molecular Biology 50 (1):571-599. doi:doi:10.1146/annurev.arplant.50.1.571

Tsutsumi D, Kosugi Ki, Mizuyama T (2002) Effect of Hydrotropism on Root System Development in Soybean (Glycine max): Growth Experiments and a Model Simulation. Journal of Plant Growth Regulation 21 (4):441-458. doi:10.1007/s00344-003-0006-y

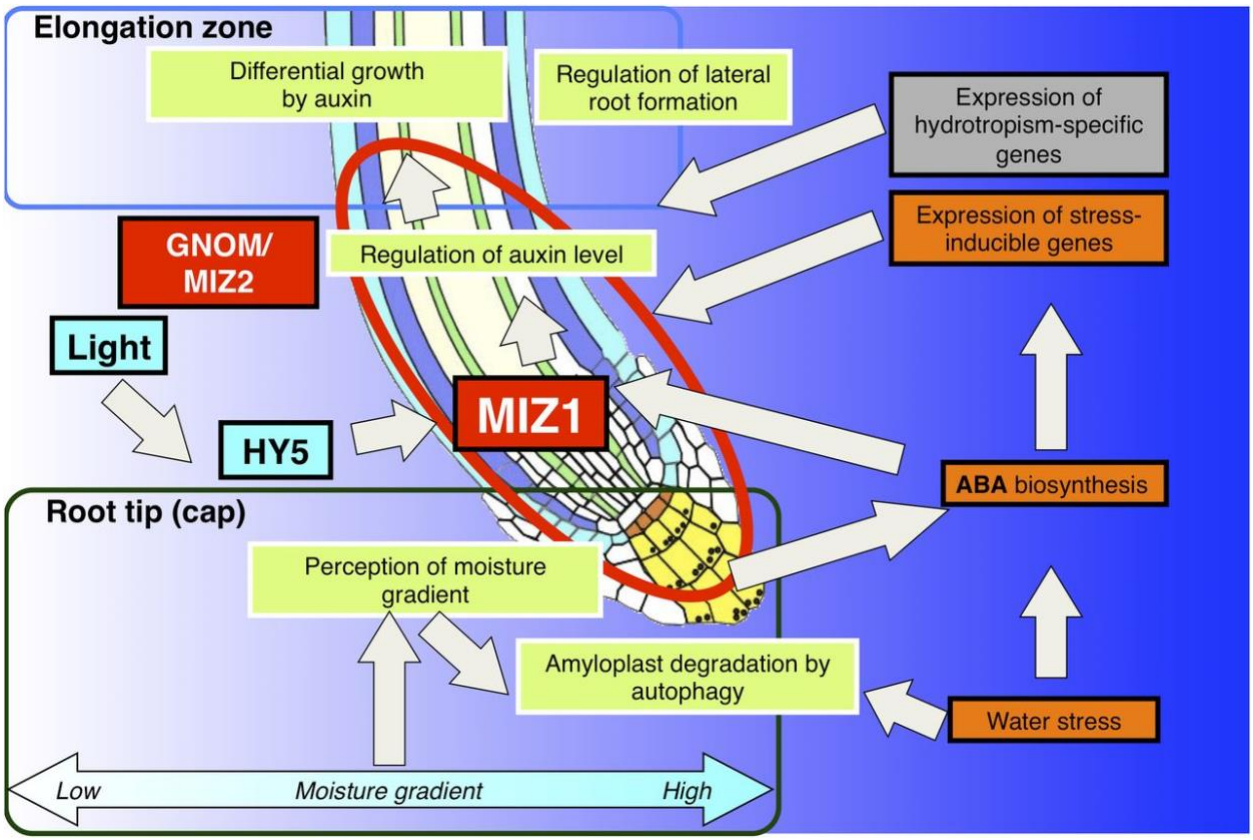
Viswanathan C, Zhu J-K (2002) Molecular genetic analysis of cold-regulated gene transcription. Philosophical Transactions of the Royal Society B: Biological Sciences 357 (1423):877-886. doi:10.1098/rstb.2002.1076

Yamasaki Y, Randall SK (2016) Functionality of soybean CBF/DREB1 transcription factors. Plant Science 246:80-90. doi:http://dx.doi.org/10.1016/j.plantsci.2016.02.007

Yamazaki T, Miyazawa Y, Kobayashi A, Moriwaki T, Fujii N, Takahashi H (2012) MIZ1, an essential protein for root hydrotropism, is associated with the cytoplasmic face of the endoplasmic reticulum membrane in Arabidopsis root cells. FEBS Letters 586 (4):398-402. doi:10.1016/j.febslet.2012.01.008

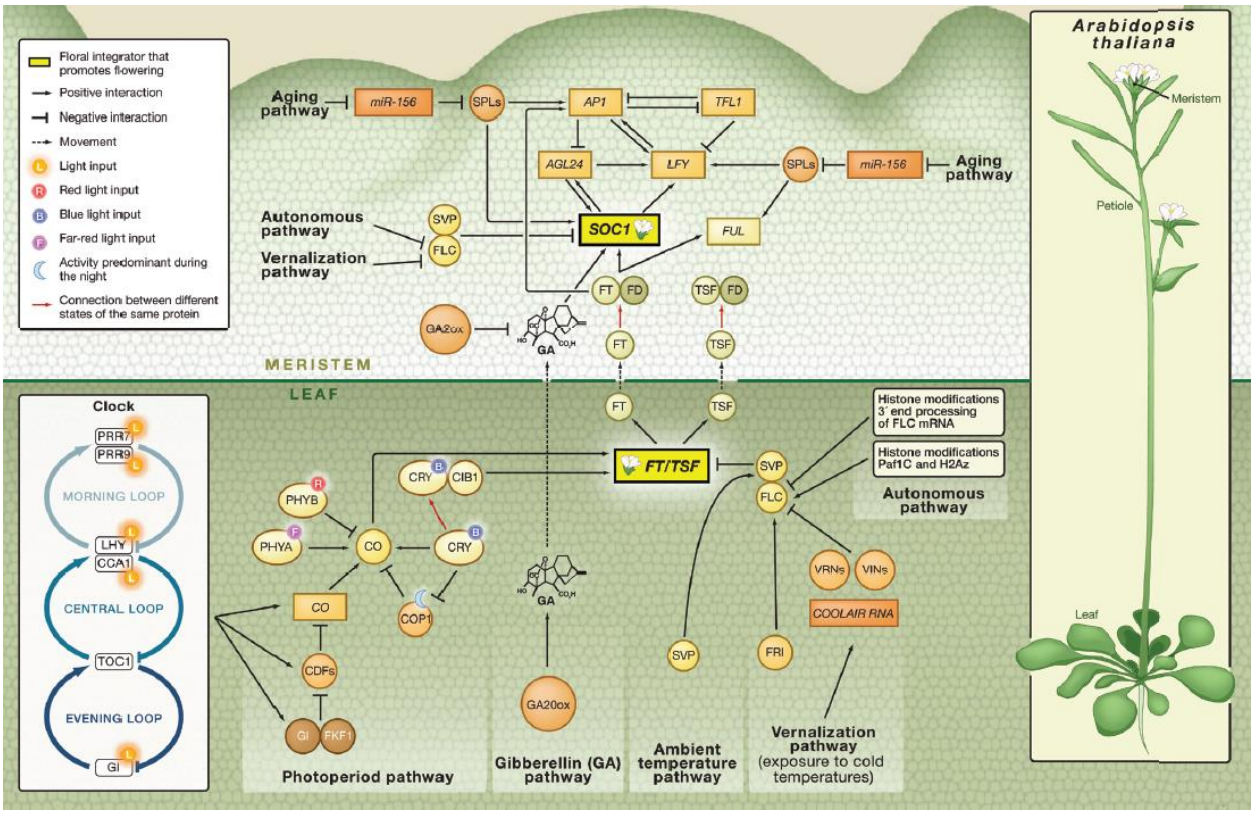
Yano M, Katayose Y, Ashikari M, Yamanouchi U, Monna L, Fuse T, Baba T, Yamamoto K, Umehara Y, Nagamura Y, Sasaki T (2000) Hd1, a Major Photoperiod Sensitivity Quantitative Trait Locus in Rice, Is Closely Related to the Arabidopsis Flowering Time Gene CONSTANS. The Plant Cell 12 (12):2473-2483. doi:10.1105/tpc.12.12.2473

Zhao C, Zhang Z, Xie S, Si T, Li Y, Zhu J-K (2016) Mutational evidence for the critical role of *CBF* genes in cold acclimation in *Arabidopsis*. Plant Physiology. doi:10.1104/pp.16.00533



Teppei Moriwaki et al. Am. J. Bot. 2012;100:25-34

Figure 1. A model of the current working hypothesis describing the mechanism underlying hydrotropism in *Arabidopsis* roots. Factors involved in the hydrotropic response are shown in relation to the root tissues specified. Red ellipsis includes main events that occur during the hydrotropic response, which connects the perception of moisture gradients in the root tip (in lower green frame) and the differential growth in the elongation zone (in upper blue frame). White arrows indicate the causal relationships among the factors. Moisture gradients are perceived in the root-cap region, and the signal is transmitted to the elongation zone where it induces bending. MIZ1 and GNOM/MIZ2 are indispensable for the induction of hydrotropism. *MIZ1* functions upstream of *MIZ2* and is hypothesized to reduce auxin level. The HY5-mediated light response and water-stress-induced biosynthesis of ABA upregulate *MIZ1* transcription, which ultimately enhances the hydrotropic response



Cell 141, April 3, 2010 ©2010 Elsevier Inc. DOI 10.1016/j.cell.2010.04.024

Figure 2. A snapshot of flowering time control in *Arabidopsis thaliana*. This snapshot presents a subset of these genes and proteins, each organized according to its spatial activity in the leaves or the shoot apical meristem of the plant. Strikingly, several genes act more than once and in several tissues during floral induction. Many of these genes occur in a network of six major pathways: the photoperiod and vernalization pathways control flowering in response to seasonal changes in day length and temperature; the ambient temperature pathway responds to daily growth temperatures; and the age, autonomous, and gibberellin pathways act more independently of environmental stimuli.

CHAPTER 2: THE *CBF-LIKE* GENE FAMILY IN ALFALFA: EXPRESSION ANALYSES AND IDENTIFICATION OF POTENTIAL FUNCTIONAL HOMOLOGS OF *ARABIDOPSIS CBF3*

Praveena Kanchupati[1], Yafang Wang[1, 3], M. Rokebul Anower[1], Arvid Boe[2], Tianming Hu[3] and Yajun Wu[1]

[1] Department of Biology and Microbiology, South Dakota State University, Brookings, SD 57007

[2]Department of Plant Science, South Dakota State University, Brookings, SD 57007

[3]Department of Grassland Science, College of Animal Science and Technology, Northwest A&F University, Yangling, China 712100

(Published in Crop Science Vol. 57 No. 4, p. 2051-2063
doi:10.2135/cropsci2016.09.0777)

2.1. Abstract

We recently identified alfalfa (*Medicago sativa* L.) germplasm River side (RS) and Foster ranch (FR), naturally adapted to the Grand River National Grassland environment in South Dakota, that showed superior freezing tolerance. To understand the molecular basis of freezing tolerance in RS and FR, we examined expression of the *C-repeat binding factor-like* (*CBF-like*) genes in alfalfa. Eighteen *CBF-like* (*CBFl*) genes were identified after examining the genome of *Medicago truncatula*, a close relative to alfalfa. Phylogenetic analysis clustered *Medicago CBFs* into 4 subgroups. Expression profiling of these genes in alfalfa seedlings revealed diverse cold-induction patterns. Four of the genes that showed an early induction as *CBF3* in Arabidopsis under cold stress were selected for detailed expression analyses. These genes varied in expression patterns, in different tissues and at different developmental stages, and exhibited different diurnal regulation without cold treatment. Two of the genes, *MsCBFl-17* and *MsCBFl-18,* showed an early and high induction under cold stress in RS and Apica, a cold-tolerant cultivar, when compared to a non-freezing tolerant germplasm; suggesting that these two genes are potentially the functional homologs of *CBF3*. On the other hand, *MsCBFl-11* was the only gene that was induced in all three cold-tolerant germplasm, including FR, but the induction was relatively late compared to *MsCBFl-17* and *MsCBFl-18.* Together, these findings suggest that the *CBFs* may play an important role in the regulation of freezing tolerance in alfalfa and additional mechanisms exist to explain the superior freezing tolerance in RS and FR.

Keywords: Alfalfa, freezing tolerance, functional homolog.

2.2. Introduction

Temperature is one of the major environmental cues regulating growth and development of plants. Chilling or subzero freezing temperature results in cold stress, which is one of the major factors limiting the production and the overall yield of plants. Cold-induced damage is evident at both the physiological and molecular levels. The formation of ice in vegetative tissues and the shutting down of important metabolic machineries are but a few of the adverse effects of freezing. However, these effects are greatly reduced when plants are exposed to low nonfreezing temperatures prior to being subjected to freezing. Such acquired tolerance to subzero temperatures by many temperate plants is termed cold acclimation (Guy, 1990; Thomashow, 1999; Chinnusamy et al., 2007). Cold acclimation is achieved through biochemical and physiological reprograming at the tissue and cellular levels. Cell plasma membranes are the first sensors and responders to low-temperature stress. Early responses include the rigidification of cellular membranes, followed by cytoskeletal rearrangement, Ca^{2+} influx, and the induction of the cold-regulated (*COR*) genes (Viswanathan and Zhu, 2002). One of the consequences of cold acclimation is the accumulation of compatible solutes like proline, sugars, mannitol, and cryoprotective compounds (Thomashow, 1999; Guy et al., 2008) that results in increased tolerance to freezing temperatures. Collectively, these responses are the direct result of altered gene expression at the molecular level (Cook et al., 2004; Hannah et al., 2005; Maruyama et al., 2009).

The process of cold acclimation involves various transcription factors and molecular switches working together in complex networks. Over the last two decades, substantial research has gone into identifying the key players regulating plants' response to chilling

or freezing stress. The most studied transcription factors in this regard are the C-repeat binding factors (CBFs), also known as dehydration-responsive element binding factors (DREBs). In *Arabidopsis*, three *CBFs*, namely *AtCBF1/DREB1B*, *AtCBF2/DREB1C*, and *AtCBF3/DREB1A*, have been demonstrated to play important roles in regulating cold-stress response (Stockinger et al., 1997; Gilmour et al., 1998; Jaglo-Ottosen et al., 1998; Liu et al., 1998; Medina et al., 1999; Gilmour et al., 2000, 2004; Zhao et al., 2016). They are the members of the Apetala 2/ethylene-responsive element-binding protein (AP2/EREBP) subfamily DREB A-1 (Riechmann and Meyerowitz, 1998) that has six members, including the genes *AtCBF4/DREB1D*, *AtDDF2/DREB1E*, and *AtDDF1/DREB1F*. Unlike *AtCBF1*, *AtCBF2*, and *AtCBF3*, the genes *AtCBF4*, *AtDDF1*, and *AtDDF2* are regulated by other abiotic stresses like salinity, dehydration, and the stress hormone abscisic aced (ABA) (Haake et al., 2002; Magome et al., 2008). The CBF transcription factors recognize the C-repeat or dehydration-responsive element present in the promoters of the downstream *COR* genes (Stockinger et al., 1997; Liu et al., 1998; Sakuma et al., 2002; Maruyama et al., 2012).

When *Arabidopsis* plants are exposed to low temperature, the transcripts of *CBFs 1*, *2*, and *3* start to accumulate rapidly and reach their peak level of expression after about 2 h of exposure to low-temperature treatment (Gilmour et al., 1998; Medina et al., 1999), followed by changes in the expression of downstream *COR* genes (Maruyama et al., 2004). Overexpression of *CBFs/DREB1s* in *Arabidopsis* induces the expression of *COR* genes and improves the freezing tolerance (Jaglo-Ottosen et al., 1998; Kasuga et al., 1999; Gilmour et al., 2004). Among the three homologs, *CBF3* in particular has been shown to be the key regulator and integrator of multiple biochemical changes involved in

the process of cold acclimation (Gilmour et al., 2000). Given their importance in the freezing tolerance of plants, functional homologs of *Arabidopsis CBF/DREB1* genes have been identified in many crop species, such as barley (*Hordeum vulgare* L.), wheat (*Triticum aestivum* L.), soybean [*Glycine max* (L.) Merr.], *Medicago truncatula* Gaertn., and rice (*Oryza sativa* L.) (Choi et al., 2002; Dubouzet et al., 2003; Skinner et al., 2005; Badawi et al., 2007; Chen et al., 2010; Jeknić et al., 2013; Ryu et al., 2014; Kidokoro et al., 2015; Yamasaki and Randall, 2016), and trees such as poplar (*Populus* spp.; Benedict et al., 2006). Many of these studies have also established their conserved role in the regulation of freezing tolerance across diverse plant species.

Alfalfa (*Medicago sativa* L.) is a major forage legume with important agronomic and environmental traits (Castonguay et al., 2009). It is a very important source of protein and fiber for livestock. It is also an agronomically important crop because of its ability to establish symbiosis with nitrogen-fixing rhizobial bacteria that decreases the need for application of nitrogen fertilizers during crop rotations. Alfalfa is grown worldwide, and in terms of production in the United States alone comes in third, only after the staple crops wheat and corn (*Zea mays* L.). However, unpredictable drops in temperature in early spring and late fall and inadequate winter hardiness adversely affect its production and overall yield in the northern climates (Castonguay et al., 2013; Anower et al., 2016). Thus, improvement in freezing tolerance of the alfalfa germplasm would be largely beneficial for farmers, as well as the livestock industry.

Previous studies in alfalfa have identified several *cold-acclimation-specific* (*CAS*) genes whose induction was correlated with the freezing tolerance of alfalfa germplasm (Mohapatra et al., 1989). One such study established a parallel relationship between the

changes in the expression levels of *CAS15B* and the varying freezing tolerance of alfalfa germplasm (Monroy et al., 1993). Another study demonstrated that alfalfa ssp. *falcata* harbored homologs of *CAS30*, a gene with a very high cold responsiveness and an expression profile similar to *Arabidopsis COR* genes (Pennycooke et al., 2008). However, few studies were aimed at the identification of the *CBFl* genes in alfalfa and their potential roles in freezing tolerance.

In an earlier study, we reported the screening and identification of alfalfa germplasm River side (RS) and Foster ranch (FR), collected from the Grand River National Grassland in South Dakota, that had superior freezing tolerance compared to the commercial cold-tolerant germplasm Apica (API) (Anower et al., 2016). Also, transcript analyses of some of the genes involved in cold tolerance pathway showed distinct cold induction patterns in the RS, FR, and API, in contrast to the cold-susceptible germplasm CUF-101 (CUF). These findings, together with the knowledge that *CBFs* are the key regulators of freezing tolerance, lead us to hypothesize that the superior performance of RS and FR under cold stress is the result of the differential induction of the *CBF* regulon, in contrast with the cold-susceptible alfalfa germplasm. The objective of this study was to identify the potential functional homolog of *AtCBF3*, the key regulator of cold tolerance in *Arabidopsis*, and to establish a better understanding of cold-tolerance mechanisms at the molecular level in alfalfa.

2.3. Results

2.3.1. Identification of *CBFl* and *CORl* genes in *M. truncatula* and phylogenetic analysis

To identify the *CBFl* genes in *M. truncatula*, we used the protein sequence of *Arabidopsis* CBF3 as a query to perform Protein Basic Local Alignment Search Tool (BLASTP) searches of the *Medicago* genome database at Phytozome (https://phytozome.jgi.doe.gov/pz/portal.html#!search?show=KEYWORD&method=Org_Mtruncatula; Goodstein et al., 2012). Eighteen CBFl proteins were identified with a default cutoff value of $E = 6.6 \times 10^{-26}$ and a score of 102.8. The protein sequences of these *CBFl* genes contain an AP2/ERF-type DNA-binding domain, which is a conserved domain in CBF/DREB-type transcription factors (Fig. 1a). An alignment of the whole-protein sequences was also conducted, showing high similarity in other regions among the proteins (Supplemental Fig. S1). Apart from the AP2-DNA binding domain and Nuclear localization signal (NLS), most of the protein sequences also had the "signature sequences" PKK/RPAGRxKFxETRHP and DSAWR (Jaglo et al., 2001). There were a few differences observed between the *Medicago* and *Arabidopsis* CBFs. Although most of the *Medicago* CBFs had the alanine-rich acidic C-terminal domains, the N-terminal and the C-terminal regions did not share extensive sequence identities.

Phylogenetic analysis of the CBFl proteins from *Medicago*, along with the CBF/DREB-type transcription factors in *Arabidopsis* and soybean, classified them into four subgroups. The majority of the *Medicago* CBFl proteins were clustered together to form an MtCBFl-only group, Subgroup 1. MtCBFl-4 was clustered with soybean DREB1As and 1Bs, forming Subgroup 2. MtCBFl-1, -16, -17, and- 18 were clustered

with AtDREB1E and -1F and soybean DREB1C and -1D in Subgroup 3. Subgroup 4 contained four CBF, all from *Arabidopsis* AtCBF1 through -4. MtCBFl-15 and soybean DREB1E-1H formed Subgroup 5 (Fig. 1b).

A single *MtCBFl* gene was found on chromosomes 1, 2, and 4, whereas a gene tandem was found on chromosomes 5 and 6. In particular, the tandem on chromosome 6 contains 12 *MtCBFl* genes presented in Subgroup 1 (Supplemental Fig. S2).

Similar procedure was followed to identify the homologs of *Arabidopsis* COR47 in *M. truncatula*. BLASTP search resulted in only one hit, an MtDehydrin-like (MtDHNl) protein, and this protein shared ~43% identity with AtCOR47. MtDHNl contained a dehydrin-specific domain characterized by the presence of a series of serines, followed by a cluster of charged amino acids. The second conserved region was marked by the presence of lysine-rich regions in Supplemental Fig. S3.

2.3.2. Thirteen *CBFl* genes are induced under cold stress in alfalfa

To identify the *CBFl* genes in alfalfa that show a similar early but transient cold-induction pattern as AtCBF3, we examined changes in transcript level of all *CBFl* genes after cold treatment in 1-wk-old SD201 seedlings, a yellow-flowered alfalfa. *MtCBFl* gene sequences were used for primer design for gene expression studies. The whole seedlings were used in this first study to identify cold-responsive *CBF* genes in different tissues. With the exception of *MsCBFl-1*, *-3*, *-8*, *-9*, and *-16*, all other *MsCBFl* genes showed response to cold treatment. Given the change in transcript level with time, at least three different patterns of induction can be seen (Fig. 2). Five genes, *MsCBFl-2*, *-4*, *-11*, *-17*, and *-18*, showed peak induction at 2 h after cold treatment, which was followed by a decrease in the transcript levels at the 12-h time point. *MsCBFl-6*, *-7*, *-10*, and *-12*

showed a significant increase at 2 h and maintained the higher level at 12 h. *MsCBFl-5*, *-13*, *-14*, and *-15* reached the highest levels at 12 h. *MsCBFl-2*, *-4*, *-11*, *-17*, and *-18* were selected for additional analyses, since their induction pattern was similar to *CBF3* in *Arabidopsis*.

2.3.3. Expression of three early cold-induced *MsCBFl* genes showed diurnal pattern

AtCBF3 expression is regulated by diurnal changes, has its expression peak at 9 h after dawn, and reaches its lowest level during the night (Grundy et al., 2015). To analyze if there was a similar regulation of *CBFl* genes in alfalfa, we examined the expression patterns of the five early cold-induced genes in young shoots of the 7-wk-old plants grown under a 16/8-h light/dark photoperiod. *MsCBFl-17* and *-18* showed peak expression at 3 h after dawn, after which the expression decreased and reached a minimum level 3 h later (Fig. 3). *MsCBFl-2*, however, showed two peaks in expression, one in daytime and the other just before dawn. *MsCBFl-11* showed a completely different expression pattern with no clear peak but reached the highest expression 9 h into daytime and remained at the higher level during the night. *MsCBFl-4* was detected at very low levels, and we were not able to obtain a reliable expression profile for this gene.

2.3.4. Expression of *MsCBFl* genes varied in different tissue types and at different developmental stages

Expression in different parts of the plant

MsCBFl-2 and *-11* had a very similar expression patterns, showing relatively similar expression levels in all three tissue types. *MsCBFl-17* and *-18* also had very similar expression patterns, with both the genes having the highest expression levels in

the root and the lowest expression levels in the leaf and stem (Fig. 4). *MsCBFl-4* was detected at very low levels, and we were not able to obtain a reliable expression profile for this gene, so the data are not presented.

Expression at different developmental stages

Expression patterns varied among the genes at different developmental stages. *MsCBFl-2* showed the highest expression during Week 2 (W_2). Its expression was significantly lower in the other stages. *MsCBFl-11* showed more than 200-fold increases in transcripts at W_2 and W_3 compared with W_1. Both *MsCBFl-17* and *-18* showed greater transcript levels during W_2 to W_4. *MsCBFl-17* exhibited the highest expression at W_3 whereas *MsCBFl-18* had the highest expression at W_2, showing more than sevenfold and 150-fold increases when compared to their respective W_1 (Fig. 5).

2.3.5. *MsCBFl-17* and *MsCBFl-18* are upregulated very early in RS and API but not in CUF under cold stress

To address the possible function of the four selected early-cold-inducible genes in cold tolerance, we examined the transcript level in three cold-tolerant germplasm, RS, FR, and API, in comparison with a cold-susceptible germplasm, CUF, when these plants were subjected to low-temperature treatment. *MsCBFl-17* and *-18* showed an early induction by cold treatment in RS and API. No induction of *MsCBFl-17* and *-18* was observed in any germplasm at 24 h. *MsCBFl-11* showed no significant change in the transcript levels at 2 h but a significant increase in RS, FR, and API (not in CUF) at 24 h. Only API showed an increase in transcript levels of *MsCBFl-2* at 2 h after cold treatment and remained at a higher induction level at 24 h (Fig. 6).

2.3.6. *MsCAS30* and *MsDHNL* genes are upregulated under cold stress in all four alfalfa germplasm

To examine whether *MsCBFl-17* and *-18* have a role in regulating downstream gene expression as *AtCBF3* does, we analyzed the expression levels of two dehydrin genes, *MsCAS30* and *MsDHNl*, in RS, FR, API, and CUF after cold treatment. *MsCAS30* has been shown to be associated with cold tolerance in alfalfa (Pennycooke et al., 2008). *COR47* is a known downstream gene regulated by *CBF3* in *Arabidopsis* (Kasuga et al., 1999; Seki et al., 2001; Fowler and Thomashow, 2002; Maruyama et al., 2004). *MsDHNl* was identified from the *M. truncatula* database using a sequence alignment analysis with AtCOR47. Both *MsCAS30* and *MsDHNl* showed great induction 24 h after cold treatment in all the four alfalfa germplasm, including CUF. More importantly, the transcript level of *MsCAS30* in CUF was similar to the level in API but was higher than in RS and FR. The transcript level of *MsDHNl* was the same for CUF, RS, and API but was lower than in FR at 24 h after cold treatment. *MsCAS30* and *MsDHNl* showed no significant change in the transcript level at 2 h after cold treatment and showed no significant difference among the four germplasm (Fig. 7).

2.4. Discussion

2.4.1. Alfalfa genome has at least 18 *CBF* homologs

The molecular dissection of complex pathways in alfalfa is restricted due to the lack of a sequenced genome. In this study, we identified a group of putative *CBF* genes by performing in silico analysis of the *M. truncatula* genome instead. *Medicago truncatula* is a close relative of alfalfa and shares a high degree of sequence similarity (Julier et al., 2003; Choi et al., 2004; Young et al., 2011). For 18 putative CBFs identified

in *M. truncatula*, we were able to amplify the gene sequences in alfalfa using the primers designed according to the *M. truncatula* genome, further supporting the fact that the genome sequences of alfalfa and *M. truncatula* are highly similar. Since most commercial varieties of alfalfa are polyploid, instead of diploid as in *M. truncatula*, alfalfa can potentially have >18 *CBF* members in its genome.

Phylogenetic analysis of CBFs from *Arabidopsis*, soybean, and *M. truncatula* suggested that *Mt*CBFs are evolved through multiple events. Genes in Subgroups 4 and 5 are potentially derived from the same ancestral sequence. Subgroups 4 and 5 are formed after the divergence of *Arabidopsis* and the legume plants, since Subgroup 4 solely contains genes from *Arabidopsis*, whereas Subgroup 5 only contains genes from soybean and *Medicago*. Subgroups 1, 2, and 3 may be derived from another ancestral sequence. However, the sequence was duplicated in legumes, possibly after the divergence of *Arabidopsis* and legume plants, forming Subgroups 1 and 2 that exclusively contain legume genes and Subgroup 3 that has both *Arabidopsis* and legume genes. Subgroup 1 contains only *M. truncatula* genes, suggesting that it resulted from recent duplication events occurring after the divergence of soybean and *M. truncatula*. This notion is further supported by the fact that all 12 homologs in Subgroup 1 are located on the long arm of chromosome 6 as tandem loci (Supplemental Fig. S2). Another possible recent duplication in *M. truncatula* is evident by two genes, *MtCBFl-17* and *-18*, in Subgroup 3 with nearly identical sequences. While it is not clear what drives these recent duplications, the consequence of the duplication is a larger *CBF* gene family in *M. truncatula* than in soybean and *Arabidopsis*. The size of the *CBF* gene family in alfalfa is unknown. However, since most of the commercial varieties of alfalfa are polyploid, there

is a strong possibility of an even larger gene family in alfalfa and presence of alfalfa-specific *CBFl* genes in its genome.

2.4.2. *MsCBFl-17* and *MsCBFl-18* are the potential functional homologs of *AtCBF3*

The biological function of *Arabidopsis CBF3* lies in its ability to be induced early under cold stress and to regulate multiple genes downstream that are responsible for the process of cold acclimation (Gilmour et al., 1998, 2000). According to the global transcriptomic analysis from multiple circadian microarray data, approximately one-third of transcripts in *Arabidopsis* oscillate under diurnal light–dark cycles (Harmer et al., 2000; Covington et al., 2008), *CBF3* being one of them. Strikingly, the diurnal regulation of the majority of genes is conserved across many crop plants and tree species, like rice, maize, and poplar (Khan et al., 2010; Filichkin et al., 2011). These key features were used to identify a functional *CBF3* homolog in alfalfa. Several lines of evidence point to *MsCBFl-17* and *-18* as potential functional homologs of *AtCBF3*. First, *MsCBFl-17* and *-18* are among the alfalfa *CBF* genes that showed early but transient cold induction; second, both showed diurnal changes in transcript level; third, they were the only ones to be induced very early in the cold-tolerant alfalfa germplasm RS and API, but not in the cold-susceptible germplasm CUF.

Interestingly, *MsCBFl-17* and *-18* are not the homologs showing the highest sequence similarity to AtCBF3. Instead, they are included in Subgroup 3 with AtDREB1F and -1E. AtDREB1F and -1E, also known as AtDDF1 and -F2, respectively, are integral to the gibberellic acid biosynthesis pathway, and overexpression of these homologs results in increased salinity-stress tolerance (Magome et al., 2004). The

soybean homologs *DREB1C* and *-1D* that are in the same subgroup as *MsCBFl-17* and *-18*, however, have been implicated as cold-inducible genes (Maruyama et al., 2012; Yamasaki and Randall, 2016). Other evidence shows that expression of *MsCBFl-17* and *-18* are regulated differently from *AtCBF3*. *MsCBFl-17* and *-18* had greater expression in the roots than in leaves and stems (Fig. 4). Information obtained using the AtGenExpress Visualization tool (www.*Arabidopsis*.org/) indicated that *AtCBF3* is expressed at relatively higher levels in mature leaves than in stems and roots. In addition, even though expression of *MsCBFl-17* and *-18* was regulated by diurnal cycle, the exact expression pattern differs slightly from that of *AtCBF3*. The transcript level of *AtCBF3* peaked at ZT9, whereas expression of *MsCBFl-17* and *-18* peaked at ZT3. The diurnal patterns of *MsCBFl-17* and *-18* were examined in the alfalfa plants grown at a 16/8-h light/dark photoperiod in this study, whereas the diurnal pattern of *AtCBF3* was examined in the plants grown at a 12/12-h light/dark photoperiod. Whether the difference in growth photoperiod affects the diurnal pattern of gene expression in alfalfa and *Arabidopsis* needs to be determined. It must be mentioned that MtCBFl-15 showed the highest sequence similarity to AtCBF3 and was not examined in detail due to its relatively late induction under cold. Further studies may be needed to determine whether it was regulated in a similar way to *AtCBF3*. The fact that the transcript levels of *MsCBFl-17* and *-18* were not induced by cold stress in FR, another cold-tolerant germplasm, and that the induction level of *MsCBFl-2* was not as great as API despite having greater cold tolerance than API suggest that *MsCBFl-17* and *-18* are not the only genes that may contribute to the freezing tolerance in alfalfa.

2.4.3. *Medicago CBFl* genes may have diverse functions

The *M. truncatula* genome contains 18 *CBFl* genes, a much larger family than the one in *Arabidopsis*. Thirteen genes in the family showed response to cold treatment, although the patterns of induction were different. The results suggested that many of these genes may be primarily involved in low-temperature tolerance. The fact that these genes are expressed in different tissues, are induced at different times after cold treatment, and have varied transcript levels at different developmental stages suggests that alfalfa has developed a complex regulatory system in cold response to minimize the damage due to stress. Alternatively, different *CBF* genes may function differently in cold response, which has been demonstrated in *Arabidopsis*. Overexpression of *AtCBF1*/*DREB1B* in *Arabidopsis* was been shown to increase the freezing tolerance of the plants and to enhance the expression of the *COR* genes that are involved in the process of cold acclimation (Gilmour et al., 1998; Jaglo-Ottosen et al., 1998). Interestingly, although *AtCBF2*/*DREB1C* has been shown to upregulate under freezing stress in a similar manner to *AtCBF1* and *-3* (Gilmour et al., 2004), a study of *cbf2* transfer DNA-insertion mutant plants elucidated that *AtCBF2* negatively regulates the activity of *AtCBF1* and *-3* (Novillo et al., 2004). The mutant plants had a higher freezing tolerance compared with wild-type cold-acclimated plants and also demonstrated increased drought and salt tolerance. The expression levels of both *AtCBF1* and *-3* were increased in the *cbf2* mutant plants. Another study showed that the expression levels of only *AtCBF3* were increased in the *cbf2* mutant plants (Kim et al., 2015).

A recent study showed the *CBF* regulon in a very different light. Single, double, and triple mutants of all the three genes, *AtCBF1*, *-2*, and *-3*, were developed using

Clustered regularly interspaced short palindromic repeats/CRISPR-associated protein-9 nuclease (CRISPR/Cas9) (Zhao et al., 2016). The *cbf2* mutant lines in this study did not show any significant increase in either *AtCBF1* or *-3*, and the plants also had lower freezing tolerance compared with the wild types, indicating *AtCBF2* to be a positive regulator of freezing tolerance and not a negative regulator of the other two genes. In the characterization of the triple mutants, *cbf1*, *-2*, and *-3* revealed additional functions of the regulon. They were shown to regulate seedling development in addition to be involved in salt-stress response.

Indeed, other *MsCBF* genes, besides *MsCBFl-17* and *-18* that are suggested in cold tolerance in this study, have also been implicated in regulation of cold stress response. For example, a previous study showed that *MtCBF2* and *MtCBF3/MtDREB1C* were induced under low-temperature treatments (Pennycooke et al., 2008). Another study reported that overexpression of *MtCBF3* in transgenic *M. truncatula* resulted in enhanced freezing tolerance, despite stunted growth (Chen et al., 2010). Additionally, the *MsCBFl* homologs belonging to the *Medicago*-specific cluster (Subgroup 1) were discovered to be part of a major freezing quantitative trait locus (Mt-FTQTL6) in *M. truncatula* (Tayeh et al., 2013). In alignment with this discovery, we saw the induction of 10 of the 13 *MsCBFl* homologs, belonging to Subgroup 1, in cold-treated seedlings. Among them, *MsCBFl-2* was induced at a significant level very early during the cold treatment in only one of the cold-tolerant germplasm, API, but not in the nonfreezing-tolerant germplasm CUF. Interestingly, cold-temperature induction of another member of Subgroup 1, *MsCBFl-11*, was observed in all three cold-tolerant alfalfa germplasm, but it was induced later in the treatment. Further studies are needed to address its potential role in freezing tolerance in

alfalfa. It is noticeable that the cold-induction patterns of *MsCBFl-2* and *-11* were different in young alfalfa plants (four different germplasm in Fig. 6) from those in seedlings of SD-201 (Fig. 2). Although the difference may be due to a variation in genotype, it is also possible that early cold induction of *MsCBFl-2* and *-11* is developmentally controlled, since the early and transient induction patterns of *MsCBFl-17* and *-18* remained relatively similar, regardless of genotype and developmental stage in cold-tolerant alfalfa.

Transcripts of *MsCBF4* were detected and induced by cold stress in young seedlings but were expressed at very low levels in young and adult plants, suggesting that its expression is developmentally regulated. Interestingly, *MtCBF4* was identified as a highly upregulated transcription factor in microarray analysis of the root samples of *M. truncatula* seedlings subjected to salinity stress. Transgenic *Arabidopsis* plants overexpressing *MtCBF4* were found to be drought- and salt-stress tolerant. *Medicago truncatula* transgenics overexpressing *MtCBF4* were also salt-stress tolerant, implicating the gene to function under drought- and salinity-stress conditions (Li et al., 2011). Thus, a few *CBFl* genes may have evolved to perform diverse functions in alfalfa, in addition to their strong potential role in cold tolerance.

2.4.4. *MsCAS30* and *MsDHNl* under cold is independent from *MsCBFl-17* and *-18* expression

During cold acclimation in *Arabidopsis*, *AtCBF3* acts as a key player and regulates expression of a set of downstream genes, including *COR* genes such as *COR47* (Kasuga et al., 1999; Seki et al., 2001; Fowler and Thomashow, 2002; Maruyama et al., 2004). *COR47* encodes a dehydrin protein that is known to be involved in cold and

dehydration stress response. Interestingly, *MsCAS30*, whose expression is closely correlated with freezing tolerance in some alfalfa germplasm, also encodes a putative dehydrin, prompting us to address the relationship between *MsCAS30* and *MsCBFl-17* and *-18* during induction. The induction of *MsCBFl-17* and *-18* (at 2 h) preceded that of *MsCAS30* (at 24 h) in RS, FR, and API. In CUF, however, *MsCAS30* was highly induced at 24 h after cold treatment, despite no induction for *MsCBFl-17* or *-18* at either 2 or 24 h, suggesting that *MsCAS30* induction may not require an induction of *MsCBFl-17* or *-18*. Through sequence analysis, we found a dehydrin that showed the highest similarity to COR47 in *Arabidopsis*. However, its high induction under cold was not correlated with the induction of *MsCBFl-17* and *-18*. These results suggested that induction of *MsCAS30* and *MsDHNl* may be regulated by genes or pathways other than *MsCBFl-17* and *-18*. In addition, *MsCAS30* as a marker for freezing tolerance in alfalfa may have its limitation.

In summary, our results suggest that the *CBFls* may play important role in the regulation of freezing tolerance in alfalfa. The three cold-tolerant alfalfa germplasm showed an induction of different *CBFl* genes, implicating that the cold-response mechanism involving CBF might vary among alfalfa germplasm. The fact that API showed induction of more *CBFl* genes than FR and RS, and that FR and RS performed better under freezing tolerance than API, suggests that additional mechanisms exist and contribute to superior freezing tolerance in RS and FR.

2.5. Materials and methods

2.5.1. Plant materials and growth conditions

Five alfalfa germplasm examined in this study include SD201, RS, FR, API, and CUF. The SD201 cultivar was developed by Dr. Arvid Boe at South Dakota State

University. Seed of RS and FR were collected from the Grand River National Grassland (45 °N, 102 °W), SD, and Thunder Butte Creek (45 °N, 101 W°), North of Faith, SD, respectively. Seed of CUF is a gift from the Desert Seed Company (Seeley, CA). Seed of API was generously provided by Dr. Yves Castonguay (Agriculture and Agri-Food Canada).

SD201 seeds were scarified with 3M 332U 150 aluminum oxide sandpaper and then sterilized in 2.5% bleach solution for 3 min. After three thorough washes with distilled water, six seeds were directly planted into each 3.79-L (one-gallon) pots (Stuewe and Sons) that were filled with potting mix (Sunshine Mix #3, Sun Gro Horticulture Canada). Another batch of seeds were grown in two 3.8-cm × 21-cm Ray Leach Cone-tainers (Stuewe and Sons) each, which were filled with 38 g of potting mix. Plants were irrigated at a 3-d interval with a Miracle-Gro (Scotts Miracle-Gro Products) nutrient solution (5 g Miracle-Gro 3.79 L^{-1} [$gallon^{-1}$] of H_2O, N–P–K = 15–30 –15). Two weeks after germination, the seedlings were thinned to three plants per pot and one per Cone-tainer. All plants were grown in a Conviron growth chamber with growth conditions, set as 22 ± 2°C day and 19 ± 2°C night thermoperiod and a 16-h photoperiod. The light intensity was set at 200 μmol m^{-2} s^{-1} (photosynthetic active radiation), and the relative humidity level was maintained at 55%.

To examine gene expression in different germplasm, plants of RS, FR, API, and CUF grown in the greenhouse were used. Seeds were scarified, sterilized, thoroughly rinsed, and soaked in distilled water overnight at 4°C. Seeds were transferred the next day to wet filter papers (Whatman No. 1, Whatman International) in a Petri plate and incubated at 25°C in the dark for 6 d. Seedlings were transplanted into two 3.8-cm × 21-

cm Ray Leach Cone-tainers (Stuewe and Sons) each, which were filled with 38 g of potting mix (Sunshine Mix #3, Sun Gro Horticulture Canada). Plants were grown in a greenhouse under the conditions described above. On the 10th day after transplanting, the seedlings were thinned to one plant per cone (Anower et al., 2016).

2.5.2. Treatment and sampling

Cold stress

For cold-stress treatment of very young seedlings, 7-d-old seedlings of SD201 grown in Petri dishes were used. The procedure for seed scarification, sterilization, and incubation is described above. The seedlings were subjected to cold temperature (4°C) by placing the plates in a cold room, and samples were harvested at 0, 2, and 12 h after the treatment. Seven seedlings were harvested per each replicate at a given time point and were immediately frozen in liquid nitrogen and stored at −80°C.

For a comparison of gene expression in different germplasm, 28- to 30-d-old plants of RS, FR, API, and CUF grown in Cone-tainers in a greenhouse were subjected to cold temperature (2°C) in a walk-in cold room under similar light intensity and the same photoperiod. Samples were harvested at 0, 2, and 24 h after the treatment. Six young shoots were harvested per replicate and stored at −80°C (Anower et al., 2016).

Diurnal samples

Young shoots from SD201 plants grown in pots at a similar developmental stage (late bud stage) were harvested every 3 h starting at dawn. Samples were harvested to represent a total of eight time points (T_0–T_7 or 0–21 h) in a complete day, with five samples collected in light and three samples in dark. Three young shoots were harvested per replicate.

Developmental stages

Samples representing different stages of development of alfalfa were harvested from SD201 plants grown in Cone-tainers every week starting from the seventh day after germination. For the first week, the whole seedlings were harvested (W_1), and from the second week onward (Weeks 2–5), the young shoots from the upper node (W_2–W_5) were harvested. The samples were harvested at the same time of the day, 3 h after lights on, every week.

Different tissues

Leaves, stems, and roots were collected separately from 28-d-old SD201 plants grown in a greenhouse before flowers were visible. The samples were harvested 3 h after lights on and were immediately frozen in liquid nitrogen and stored at −80°C.

2.5.3. Identification of *CBFl* and *CORl* genes in *M. truncatula* and phylogenetic analysis

The *Arabidopsis* CBF3 protein sequence was used to search against the *M. truncatula* genome database version Mt 4.0V1 (https://phytozome.jgi.doe.gov/pz/portal.html#!search?show=KEYWORD&method=Org_Mtruncatula). The default expected threshold value (*E*) for the *Medicago* CBF homologs that were chosen was set at −1, and the BLOSUM62 comparison matrix was used. The retrieved sequences were named according to existing nomenclature as *MtCBFl-1* to *MtCBFl-18.* Multiple sequence alignment of the protein sequences was done using Clustal Omega (Sievers et al., 2011), and the neighborhood-joining method of MEGA 6 (Tamura et al., 2013) was used to construct the phylogenetic tree. Similarly, COR47

protein sequence from *Arabidopsis* was used to identify *COR-like* genes in *M. truncatula* by following the steps described above.

2.5.4. Primer design

Gene-specific primers (Supplemental Table S1) were designed for *MtCBFls* and other genes using an online primer design tool from Integrated DNA Technologies. The efficiency and specificity for each primer pair was determined using alfalfa genomic DNA (1 ng) as the template in a 20-μL polymerase chain reaction (PCR) containing 2 μL of 10× PCR buffer, 1 μL each of 10 μM primers, 1 μL of 2 mM deoxynucleotides, and 0.1 μL of Taq polymerase (5 U μL^{-1}, BioLabs). All reactions were performed in a gradient thermocycler (Eppendorf Mastercycler) with PCR conditions set as: initial denature step at 94°C for 3 min, followed by 35 cycles of 94°C for 20 s, 20 s at annealing temperature gradient (R = 3°C s^{-1}, G = ±2.4°C), extension at 72°C for 2 min, and a final extension at 72°C for 10 min. The PCR products were then run on a 1% agarose gel stained with ethidium bromide, and images were visualized using a Bio-Rad ChemiDoc image analysis system.

2.5.5. RNA isolation and cDNA synthesis

Total RNA was isolated using Trizol reagent (Invitrogen). The RNA samples were quantified using a Nanodrop ND-1000 Spectrophotometer (ThermoFisher Scientific), and samples with a 260/280 ratio from 1.9 to 2.1 and a 260/230 ratio from 2.0 to 2.5 were used for further analysis. RNA quality was also examined by separating RNA on a 2% agarose gel stained with ethidium bromide. The samples, which showed three sharp major ribosomal RNA bands, were used for complementary DNA (cDNA) synthesis. Quantitative Real-Time PCR (qRT-PCR) reactions with the housekeeping gene

MsActin were performed directly on RNA samples without reverse transcription, in comparison with respective cDNA samples. Only the cDNA samples whose corresponding RNA samples showed no amplification or significantly greater cycle threshold (CT) values in qRT-PCR analysis were used for the gene expression analysis.

First-strand cDNA synthesis was performed using the high-capacity cDNA Reverse Transcription kit (ThermoFisher Scientific) in a 20-µL reaction according to manufacturer's instructions. Synthesized cDNA samples were validated using *MsActin* primers with 30 cycles in a regular PCR reaction. The cDNA samples were diluted 4 times for use in a real-time qRT-PCR reaction.

2.5.6. Quantification of transcripts

qRT-PCR was performed using DyNAmo Flash SYBR Green Hot Start qRT-PCR Kit (ThermoFisher Scientific) following manufacturer's instructions in a 20-µL reaction in an ABI 7900HT High-Throughput Real-Time Thermocycler (Applied Biosystems) using standard cycling conditions. Each sample from three biological experiments was assayed twice as technical replicates. The thermocycler program was set to: 15 min of activation at 95°C, followed by 40 cycles of 15 s at 94°C, 30 s at annealing temperature, 30 s of extension at 72°C, and a dissociation curve step. The dissociation curve was used to determine the primer efficiency and specificity. The normalized relative fold changes in the transcripts of *MsCBFl* or other genes were calculated using the $2^{-\Delta\Delta CT}$ or comparative CT method based on the difference between the target and reference genes, as described by Livak and Schmittgen (2001).

2.5.7. Data analysis

Statistical analysis was performed using STATISTIX 9.0 analytical software (STATISTIX, 2011) and the Microsoft Excel 2010 data analysis tools pack. Data were subjected to ANOVA using the linear model with completely randomized design to determine significant differences among the treatments. Tukey's honestly significant difference all-pair comparison was conducted to ascertain significant differences between treatment means ($p < 0.05$).

2.6. Acknowledgements

This work was supported by the USDA National Institute of Food and Agriculture Hatch Project 3AH390, South Dakota Agricultural Experiment Station, and the Department of Biology and Microbiology at South Dakota State University. We are grateful to the Functional Genomics Core Facility (FGCF) at South Dakota State University and to Dr. Yves Castonguay (Agriculture and Agri-Food Canada) for providing Apica seeds.

2.7. Reference

Anower MR, Fennell A, Boe A, Mott IW, Peel MD, Wu Y (2016) Physiological and molecular characterisation of lucerne (*Medicago sativa* L.) germplasm with improved seedling freezing tolerance. Crop and Pasture Science 67:655-665. doi:http://dx.doi.org/10.1071/CP15204

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. doi:10.1007/s00438-006-0206-9

Benedict C, Skinner JS, Meng R, Chang Y, Bhalerao R, Huner NPA, Finn CE, Chen THH, Hurry V (2006) The CBF1-dependent low temperature signalling pathway, regulon and increase in freeze tolerance are conserved in *Populus* spp. Plant, cell & environment 29:1259-1272. doi:10.1111/j.1365-3040.2006.01505.x

Castonguay Y, Dubé M-P, Cloutier J, Bertrand A, Michaud R, Laberge S (2013) Molecular physiology and breeding at the crossroads of cold hardiness improvement. Physiologia Plantarum 147:64-74. doi:10.1111/j.1399-3054.2012.01624.x

Castonguay Y, Michaud R, Nadeau P, Bertrand A (2009) An indoor screening method for improvement of freezing tolerance in alfalfa. Crop Science 49:809-818

Chen J-R, Lü J-J, Liu R, Xiong X-Y, Wang T-x, Chen S-Y, Guo L-B, Wang H-F (2010) *DREB1C* from *Medicago truncatula* enhances freezing tolerance in transgenic *M. truncatula* and China Rose (*Rosa chinensis* Jacq.). Plant Growth Regulation 60:199-211. doi:10.1007/s10725-009-9434-4

Chinnusamy V, Zhu J, Zhu J-K (2007) Cold stress regulation of gene expression in plants. Trends in Plant Science 12:444-451. doi:http://dx.doi.org/10.1016/j.tplants.2007.07.002

Choi D-W, Rodriguez EM, Close TJ (2002) Barley *Cbf3* Gene Identification, Expression Pattern, and Map Location. Plant physiology 129:1781-1787. doi:10.1104/pp.003046

Choi HK, Kim D, Uhm T, Limpens E, Lim H, Mun JH, Kalo P, Penmetsa RV, Seres A, Kulikova O, Roe BA, Bisseling T, Kiss GB, Cook DR (2004) A sequence-based genetic map of *Medicago truncatula* and comparison of marker colinearity with *M. sativa*. Genetics 166:1463-1502

Cook D, Fowler S, Fiehn O, Thomashow MF (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. doi:10.1073/pnas.0406069101

Covington MF, Maloof JN, Straume M, Kay SA, Harmer SL (2008) Global transcriptome analysis reveals circadian regulation of key pathways in plant growth and development. Genome Biology 9:1-18. doi:10.1186/gb-2008-9-8-r130

Dubouzet JG, Sakuma Y, Ito Y, Kasuga M, Dubouzet EG, Miura S, Seki M, Shinozaki K, Yamaguchi-Shinozaki K (2003) *OsDREB* genes in rice, *Oryza sativa* L., encode transcription activators that function in drought-, high-salt- and cold-responsive gene expression. The Plant Journal 33:751-763. doi:10.1046/j.1365-313X.2003.01661.x

Filichkin SA, Breton G, Priest HD, Dharmawardhana P, Jaiswal P, Fox SE, Michael TP, Chory J, Kay SA, Mockler TC (2011) Global Profiling of Rice and Poplar Transcriptomes Highlights Key Conserved Circadian-Controlled Pathways and *cis*-Regulatory Modules. PLoS ONE 6:e16907. doi:10.1371/journal.pone.0016907

Fowler S, Thomashow MF (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. doi:10.1105/tpc.003483

Gilmour SJ, Fowler SG, Thomashow MF (2004) Arabidopsis transcriptional activators CBF1, CBF2, and CBF3 have matching functional activities. Plant Mol Biol 54:767-781. doi:10.1023/B:PLAN.0000040902.06881.d4

Gilmour SJ, Sebolt AM, Salazar MP, Everard JD, Thomashow MF (2000) Overexpression of the Arabidopsis *CBF3* Transcriptional Activator Mimics Multiple Biochemical Changes Associated with Cold Acclimation. Plant physiology 124:1854-1865

Gilmour SJ, Zarka DG, Stockinger EJ, Salazar MP, Houghton JM, Thomashow MF (1998) Low temperature regulation of the*Arabidopsis* CBF family of AP2 transcriptional activators as an early step in cold-induced *COR* gene expression. The Plant Journal 16:433-442. doi:10.1046/j.1365-313x.1998.00310.x

Goodstein DM, Shu S, Howson R, Neupane R, Hayes RD, Fazo J, Mitros T, Dirks W, Hellsten U, Putnam N, Rokhsar DS (2012) Phytozome: a comparative platform for green plant genomics. Nucleic Acids Research 40:D1178-D1186. doi:10.1093/nar/gkr944

Grundy J, Stoker C, Carré IA (2015) Circadian regulation of abiotic stress tolerance in plants. Frontiers in Plant Science 6:648. doi:10.3389/fpls.2015.00648

Guy C, Kaplan F, Kopka J, Selbig J, Hincha DK (2008) Metabolomics of temperature stress. Physiologia Plantarum 132:220-235. doi:10.1111/j.1399-3054.2007.00999.x

Guy CL (1990) Cold acclimation and freezing stress tolerance: role of protein metabolism. Annual Review of Plant Physiology and Plant Molecular Biology 41. doi:10.1146/annurev.pp.41.060190.001155

Haake V, Cook D, Riechmann JL, Pineda O, Thomashow MF, Zhang JZ (2002) Transcription factor CBF4 is a regulator of drought adaptation in Arabidopsis. Plant physiology 130:639-648. doi:10.1104/pp.006478

Hannah MA, Heyer AG, Hincha DK (2005) A global survey of gene regulation during cold acclimation in *Arabidopsis thaliana*. PLoS Genetics 1:e26. doi:10.1371/journal.pgen.0010026

Harmer SL, Hogenesch JB, Straume M, Chang H-S, Han B, Zhu T, Wang X, Kreps JA, Kay SA (2000) Orchestrated transcription of key pathways in *Arabidopsis* by the circadian clock. Science 290:2110-2113

Jaglo-Ottosen KR, Gilmour SJ, Zarka DG, Schabenberger O, Thomashow MF (1998) *Arabidopsis CBF1* overexpression induces *COR* genes and enhances freezing tolerance. Science 280:104-106

Jaglo KR, Kleff S, Amundsen KL, Zhang X, Haake V, Zhang JZ, Deits T, Thomashow MF (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

Jeknić Z, Pillman KA, Dhillon T, Skinner JS, Veisz O, Cuesta-Marcos A, Hayes PM, Jacobs AK, Chen THH, Stockinger EJ (2013) Hv-*CBF2A* overexpression in barley accelerates *COR* gene transcript accumulation and acquisition of freezing tolerance during cold acclimation. Plant Mol Biol 84:67-82. doi:10.1007/s11103-013-0119-z

Julier B, Flajoulot S, Barre P, Cardinet G, Santoni S, Huguet T, Huyghe C (2003) Construction of two genetic linkage maps in cultivated tetraploid alfalfa (*Medicago sativa*) using microsatellite and AFLP markers. BMC Plant Biology 3. doi:10.1186/1471-2229-3-9

Kasuga M, Liu Q, Miura S, Yamaguchi-Shinozaki K, Shinozaki K (1999) Improving plant drought, salt, and freezing tolerance by gene transfer of a single stress-inducible transcription factor. Nature Biotechnology 17:287-291

Khan S, Rowe SC, Harmon FG (2010) Coordination of the maize transcriptome by a conserved circadian clock. BMC Plant Biology 10. doi:10.1186/1471-2229-10-126

Kidokoro S, Watanabe K, Ohori T, Moriwaki T, Maruyama K, Mizoi J, Myint Phyu Sin Htwe N, Fujita Y, Sekita S, Shinozaki K, Yamaguchi-Shinozaki K (2015) Soybean DREB1/CBF-type transcription factors function in heat and drought as well as cold stress-responsive gene expression. The Plant Journal 81:505-518. doi:10.1111/tpj.12746

Kim YS, Lee M, Lee J-H, Lee H-J, Park C-M (2015) The unified ICE–CBF pathway provides a transcriptional feedback control of freezing tolerance during cold acclimation in *Arabidopsis*. Plant Mol Biol 89:187-201. doi:10.1007/s11103-015-0365-3

Li D, Zhang Y, Hu X, Shen X, Ma L, Su Z, Wang T, Dong J (2011) Transcriptional profiling of *Medicago truncatula* under salt stress identified a novel CBF transcription factor MtCBF4 that plays an important role in abiotic stress responses. BMC Plant Biology 11. doi:10.1186/1471-2229-11-109

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. The Plant Cell 10:1391-1406

Livak KJ, Schmittgen TD (2001) Analysis of relative gene expression data using Real-Time quantitative PCR and the 2−ΔΔCT method. Methods 25:402-408. doi:http://dx.doi.org/10.1006/meth.2001.1262

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. The Plant Journal 37:720-729. doi:10.1111/j.1365-313X.2003.01998.x

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. The Plant Journal 56. doi:10.1111/j.1365-313X.2008.03627.x

Maruyama K, Sakuma Y, Kasuga M, Ito Y, Seki M, Goda H, Shimada Y, Yoshida S, Shinozaki K, Yamaguchi-Shinozaki K (2004) Identification of cold-inducible downstream genes of the *Arabidopsis* DREB1A/CBF3 transcriptional factor using two microarray systems. The Plant Journal 38. doi:10.1111/j.1365-313X.2004.02100.x

Maruyama K, Takeda M, Kidokoro S, Yamada K, Sakuma Y, Urano K, Fujita M, Yoshiwara K, Matsukura S, Morishita Y, Sasaki R, Suzuki H, Saito K, Shibata D, Shinozaki K, Yamaguchi-Shinozaki K (2009) Metabolic pathways involved in cold acclimation identified by integrated analysis of metabolites and transcripts regulated by DREB1A and DREB2A. Plant physiology 150:1972-1980. doi:10.1104/pp.109.135327

Maruyama K, Todaka D, Mizoi J, Yoshida T, Kidokoro S, Matsukura S, Takasaki H, Sakurai T, Yamamoto YY, 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. doi:10.1093/dnares/dsr040

Medina Jn, Bargues M, Terol J, Pérez-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 119:463-470. doi:10.1104/pp.119.2.463

Mohapatra SS, Wolfraim L, Poole RJ, Dhindsa RS (1989) Molecular Cloning and Relationship to Freezing Tolerance of Cold-Acclimation-Specific Genes of Alfalfa. Plant physiology 89:375-380

Monroy AF, Castonguay Y, Laberge S, Sarhan F, Vezina LP, Dhindsa RS (1993) A New Cold-Induced Alfalfa Gene Is Associated with Enhanced Hardening at Subzero Temperature. Plant physiology 102:873-879. doi:10.1104/pp.102.3.873

Novillo F, Alonso JM, Ecker JR, Salinas J (2004) CBF2/DREB1C is a negative regulator of *CBF1/DREB1B* and *CBF3/DREB1A* expression and plays a central role in stress tolerance in Arabidopsis. Proceedings of the National Academy of Sciences of the United States of America 101:3985-3990. doi:10.1073/pnas.0303029101

Pennycooke JC, Cheng H, Stockinger EJ (2008) Comparative genomic sequence and expression analyses of *Medicago truncatula* and alfalfa subspecies *falcata COLD-ACCLIMATION-SPECIFIC* genes. Plant physiology 146. doi:10.1104/pp.107.108779

Riechmann JL, Meyerowitz EM (1998) The AP2/EREBP family of plant transcription factors. Biological Chemistry 379:633-646

Ryu JY, Hong S-Y, Jo S-H, Woo J-C, Lee S, Park C-M (2014) Molecular and functional characterization of cold-responsive C-repeat binding factors from *Brachypodium distachyon*. BMC Plant Biology 14. doi:10.1186/1471-2229-14-15

Sakuma Y, Liu Q, Dubouzet JG, 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. doi:http://dx.doi.org/10.1006/bbrc.2001.6299

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. doi:10.1105/tpc.13.1.61

Sievers F, Wilm A, Dineen D, Gibson TJ, Karplus K, Li W, Lopez R, McWilliam H, Remmert M, Söding J, Thompson JD, Higgins DG (2011) Fast, scalable generation of high-quality protein multiple sequence alignments using Clustal Omega. Mol Syst Biol 7:539. doi:10.1038/msb.2011.75

Skinner JS, Zitzewitz J, Szűcs P, Marquez-Cedillo L, Filichkin T, Amundsen K, Stockinger EJ, Thomashow MF, Chen THH, Hayes PM (2005) Structural, functional, and phylogenetic characterization of a large *CBF* gene family in barley. Plant Mol Biol 59:533-551. doi:10.1007/s11103-005-2498-2

Stockinger EJ, Gilmour SJ, Thomashow MF (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

Tamura K, Stecher G, Peterson D, Filipski A, Kumar S (2013) MEGA6: Molecular Evolutionary Genetics Analysis Version 6.0. Molecular Biology and Evolution 30:2725-2729. doi:10.1093/molbev/mst197

Tayeh N, Bahrman N, Sellier H, Bluteau A, Blassiau C, Fourment J, Bellec A, Debellé F, Lejeune-Hénaut I, Delbreil B (2013) A tandem array of *CBF/DREB1* genes is located in a major freezing tolerance QTL region on *Medicago truncatula* chromosome 6. BMC Genomics 14. doi:10.1186/1471-2164-14-814

Thomashow MF (1999) Plant cold acclimation: Freezing tolerance genes and regulatory mechanisms. Annual Review of Plant Physiology and Plant Molecular Biology 50:571-599. doi:doi:10.1146/annurev.arplant.50.1.571

Viswanathan C, Zhu J-K (2002) Molecular genetic analysis of cold-regulated gene transcription. Philosophical Transactions of the Royal Society B: Biological Sciences 357:877-886. doi:10.1098/rstb.2002.1076

Yamasaki Y, Randall SK (2016) Functionality of soybean CBF/DREB1 transcription factors. Plant Science 246:80-90. doi:http://dx.doi.org/10.1016/j.plantsci.2016.02.007

Young ND, Debelle F, Oldroyd GED, Geurts R, Cannon SB, Udvardi MK, Benedito VA, Mayer KFX, Gouzy J, Schoof H, Van de Peer Y, Proost S, Cook DR, Meyers BC, Spannagl M, Cheung F, De Mita S, Krishnakumar V, Gundlach H, Zhou S, Mudge J, Bharti AK, Murray JD, Naoumkina MA, Rosen B, Silverstein KAT, Tang H, Rombauts S, Zhao PX, Zhou P, Barbe V, Bardou P, Bechner M, Bellec A, Berger A, Berges H, Bidwell S, Bisseling T, Choisne N, Couloux A, Denny R, Deshpande S, Dai X, Doyle JJ, Dudez A-M, Farmer AD, Fouteau S, Franken C, Gibelin C, Gish J, Goldstein S, Gonzalez AJ, Green PJ, Hallab A, Hartog M, Hua A, Humphray SJ, Jeong D-H, Jing Y, Jocker A, Kenton SM, Kim D-J, Klee K, Lai H, Lang C, Lin S, Macmil SL, Magdelenat G, Matthews L, McCorrison J, Monaghan EL, Mun J-H, Najar FZ, Nicholson C, Noirot C, O/'Bleness M, Paule CR, Poulain J, Prion F, Qin B, Qu C, Retzel EF, Riddle C, Sallet E, Samain S, Samson N, Sanders I, Saurat O, Scarpelli C, Schiex T, Segurens B, Severin AJ, Sherrier DJ, Shi R, Sims S, Singer SR, Sinharoy S, Sterck L, Viollet A, Wang B-B, Wang K, Wang M, Wang X, Warfsmann J, Weissenbach J, White DD, White JD, Wiley GB, Wincker P, Xing Y, Yang L, Yao Z, Ying F, Zhai J, Zhou L, Zuber A, Denarie J, Dixon RA, May GD, Schwartz DC, Rogers J, Quetier F, Town CD, Roe BA (2011) The *Medicago* genome provides insight into the evolution of rhizobial symbioses. Nature 480:520-524. doi:http://www.nature.com/nature/journal/v480/n7378/abs/nature10625.html#supplementary-information

Zhao C, Zhang Z, Xie S, Si T, Li Y, Zhu J-K (2016) Mutational evidence for the critical role of *CBF* genes in cold acclimation in *Arabidopsis*. Plant physiology. doi:10.1104/pp.16.00533

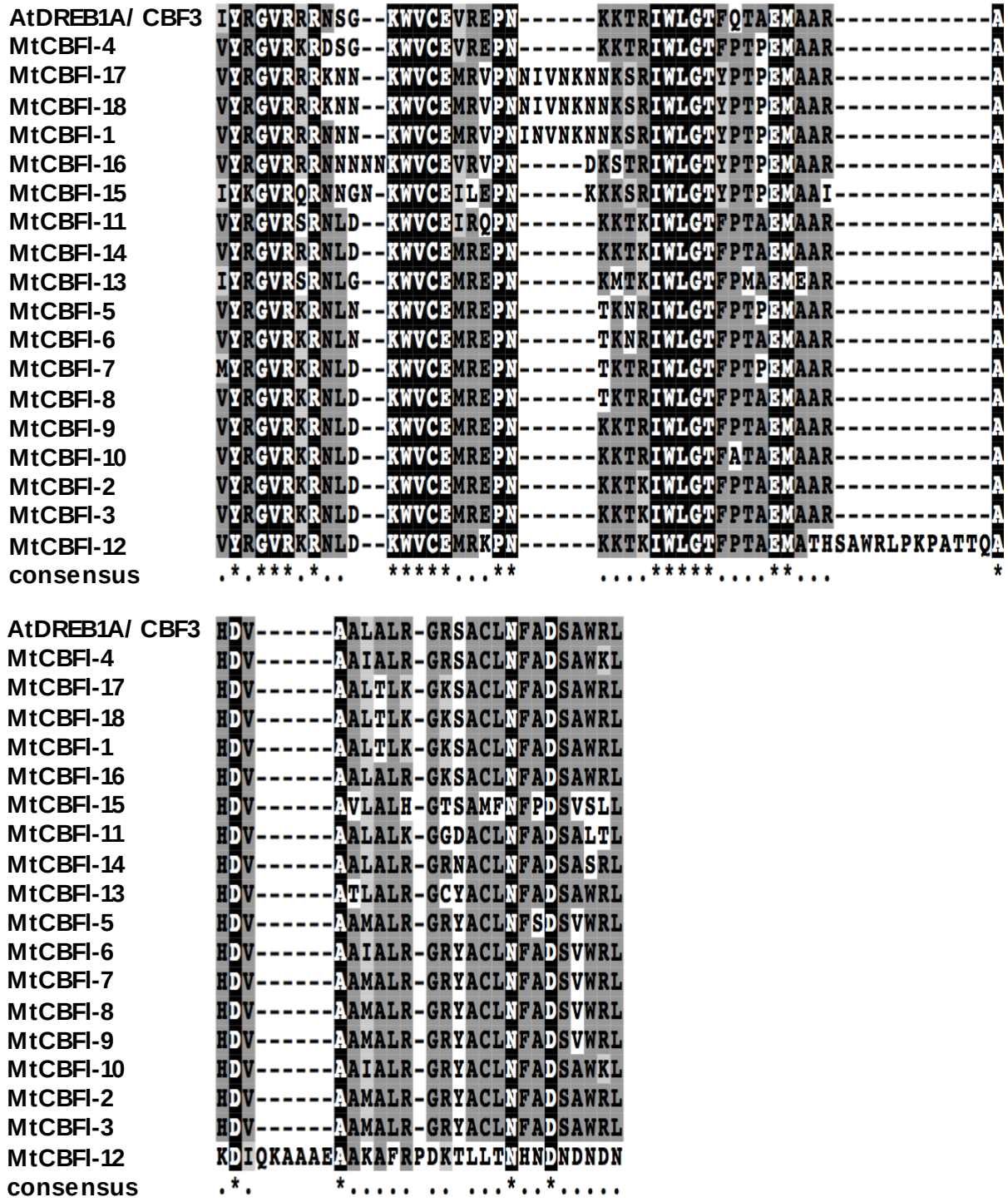


Figure 1a. Apetala 2 (AP2) domain alignment of 18 *Medicago truncatula* C-repeat binding factor-like (CBF-like) peptides with *Arabidopsis* CBF3; * represents conserved amino acid residues; · represents the identical and similar amino acid residues.

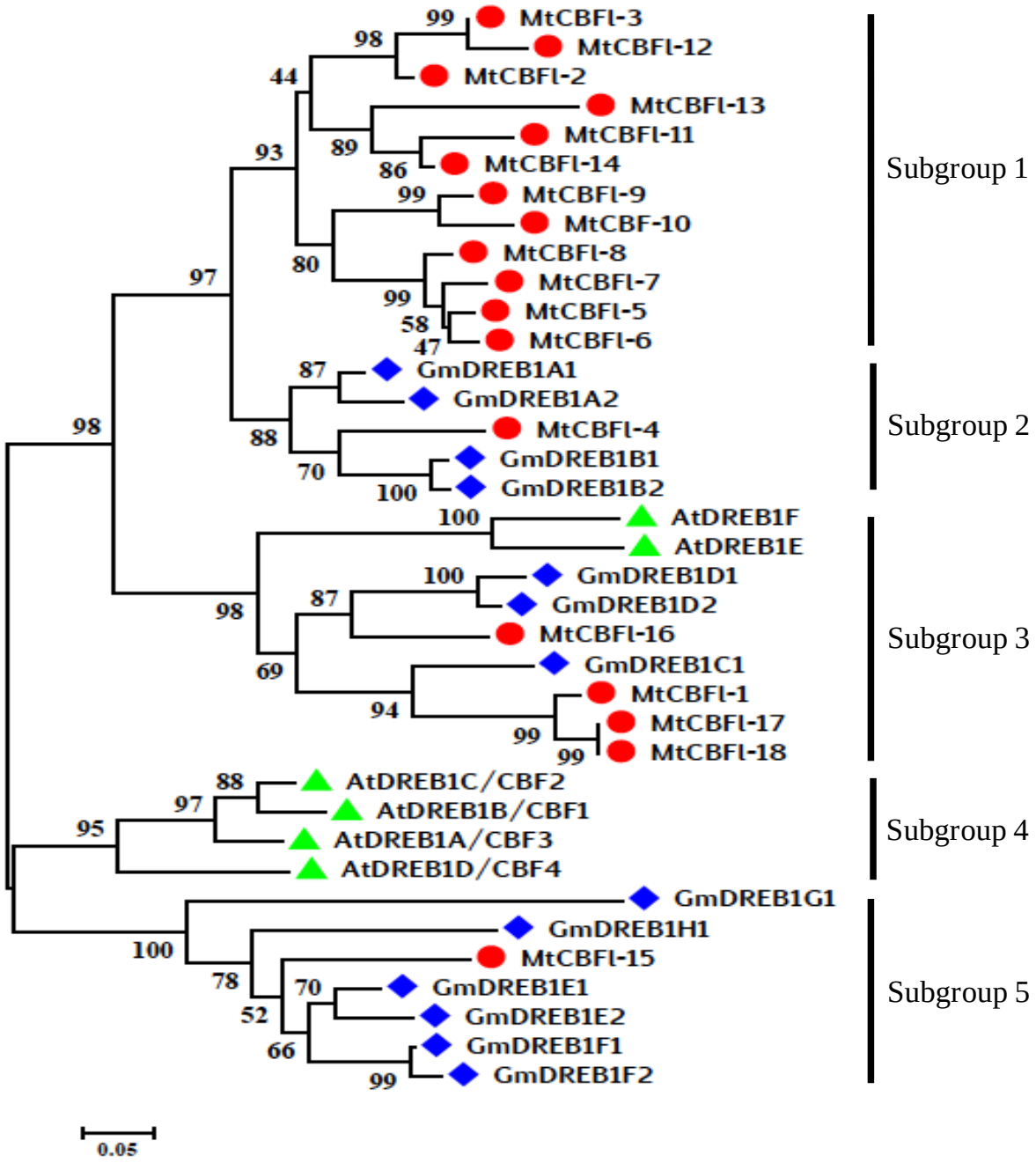


Figure 1b. Phylogenetic tree of CBF transcription factors in *M. truncatula* (circles), *Arabidopsis* (triangles), and soybean (diamonds). The numbers shown next to the branches are the bootstrap probabilities from 1000 replications. *Arabidopsis* sequences include: AtDREB1C/CBF1 (At4g25470), AtDREB1B/CBF2 (At4g25490), AtDREB1A/CBF3 (At4g25480), AtDREB1D/CBF4 (At5g51990), AtDREB1E

(At1g63030) and AtDREB1F (At1g12610). *Medicago truncatula* sequences include: MtCBF1 (Medtr5g010930), MtCBF2 (Medtr6g465690), MtCBF3 (Medtr6g466000), MtCBF4 (Medtr1g101600), MtCBF5 (Medtr6g465420), MtCBF6 (Medtr6g465430), MtCBF7 (Medtr6g465450), MtCBF8 (Medtr6g465460), MtCBF9 (Medtr6g465510), MtCBF10 (Medtr6g465530), MtCBF11 (Medtr6g465850), MtCBF12 (Medtr6g465990), MtCBF13 (Medtr6g466020), MtCBF14 (Medtr6g466130), MtCBF15 (2 g085015), MtCBF16 (Medtr4g102660), MtCBF17 (Medtr5g010910), and MtCBF18 (Medtr5g010940). Soybean sequences include: GmDREB1A1 (Glyma09g27180), GmDREB1A2 (Glyma16g32330), GmDREB1B1 (Glyma20g29410), GmDREB1B2 (Glyma10g38440), GmDREB1C1 (Glyma01g42500), GmDREB1D1 (Glyma05g03560), GmDREB1D2 (Glyma17g14110), GmDREB1E1 (Glyma12g30740), GmDREB1E2 (Glyma13g39540), GmDREB1F1 (Glyma12g09130), GmDREB1F2 (Glyma11g19340), GmDREB1G1 (Glyma13g21570), and GmDREB1H1 (Glyma12g30710)

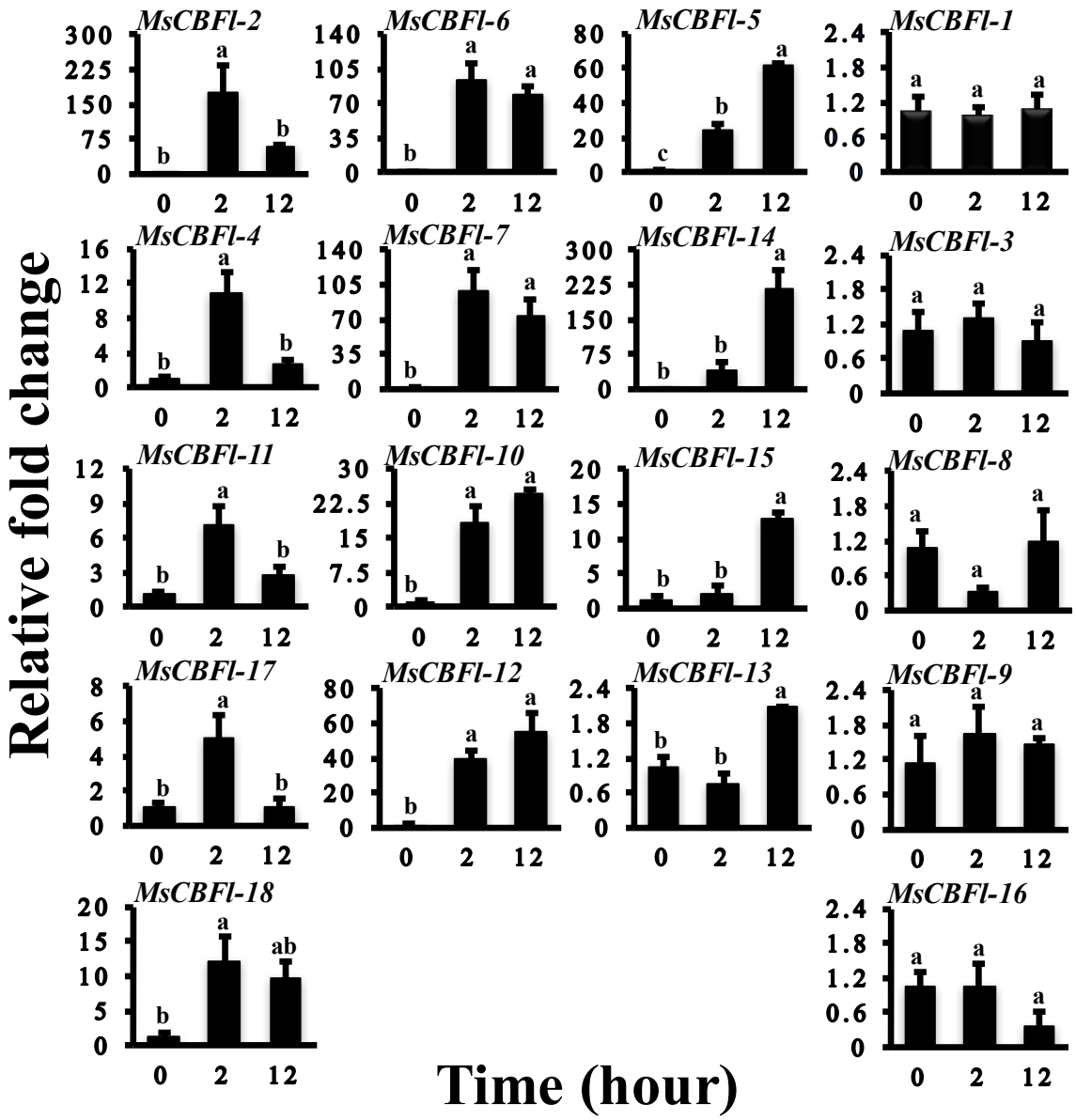


Figure 2. Cold-responsive expression of the *C-repeat binding factor-like* (*CBF-like*) genes in alfalfa. One-week-old SD201 seedlings were exposed to cold (4°C), and samples were collected at 0, 2, and 12 h after the cold treatment. The transcripts were quantified by qRT-PCR. The values represent the mean fold change ± SE (n = 3) when compared with the transcript level at 0 h. Bars with different letters are significantly different (p < 0.05).

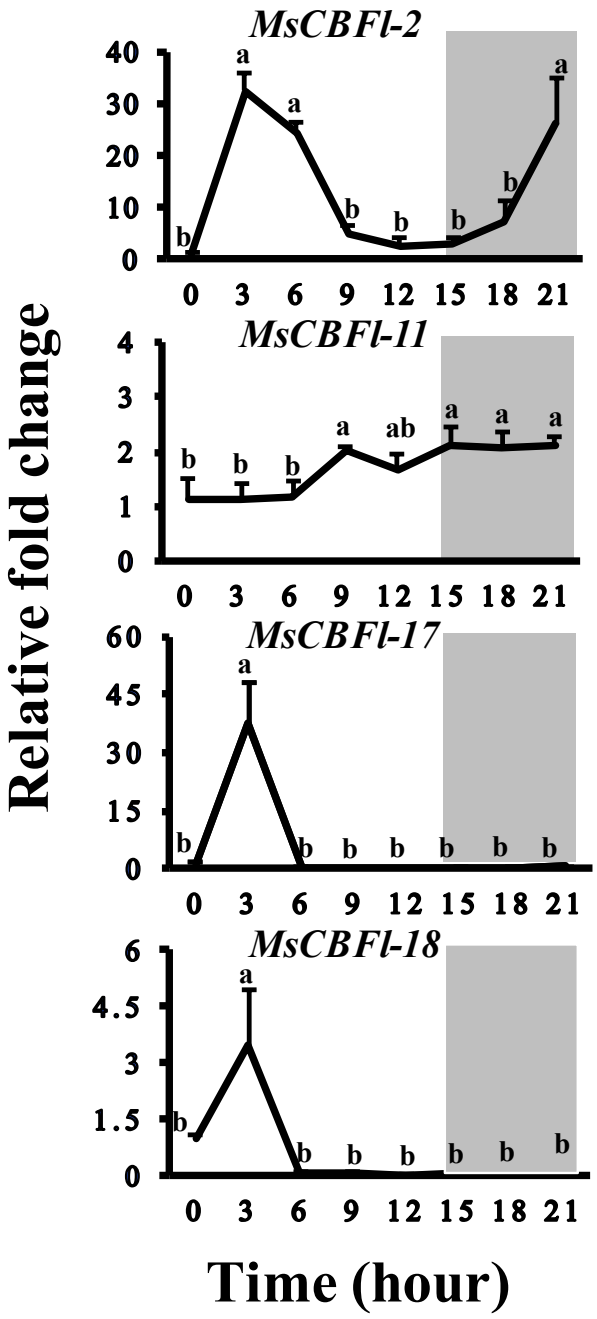


Figure 3. Diurnal regulation of four cold-induced genes in alfalfa. Young shoots were harvested from SD201 plants of the same age every 3 h after dawn. The shaded area in each graph represents sampling points during night. The values represent the mean fold change ± SE (n = 3) when compared with the transcript level at 0 h. Data points with different letters are significantly different ($p < 0.05$).

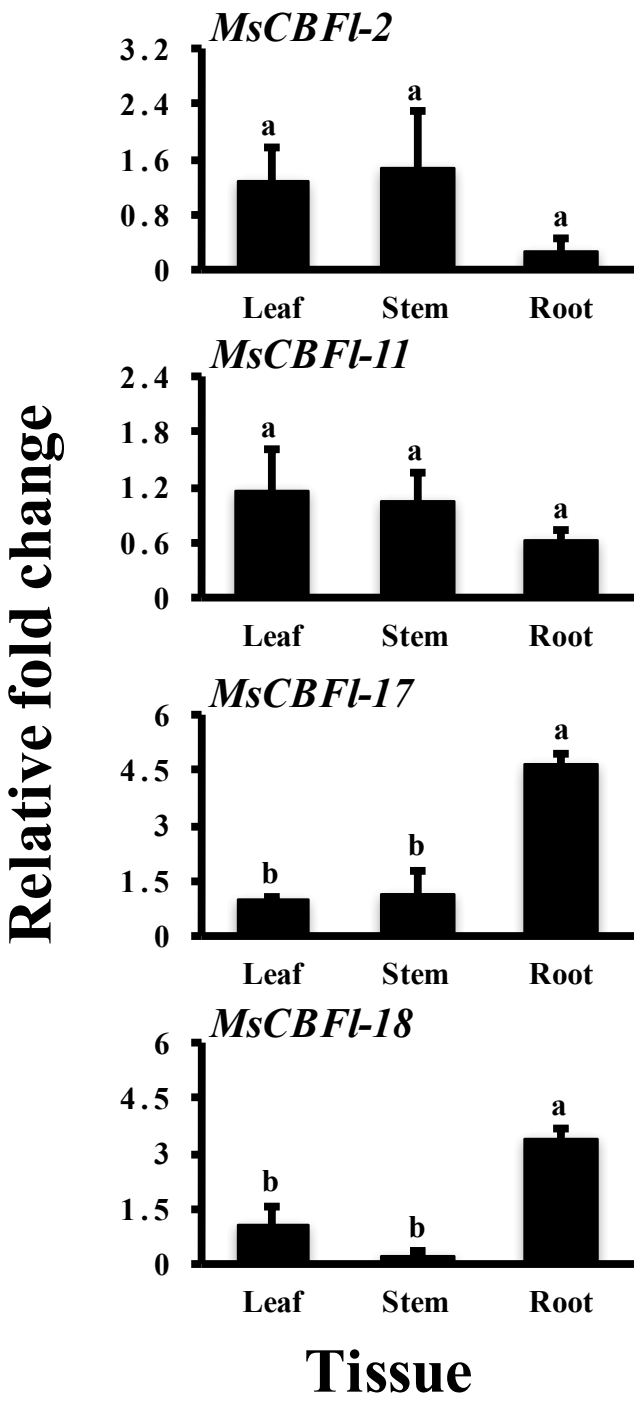


Figure 4. Expression analysis of MsCBF-like genes in different tissues; leaf, stem, and root tissues were harvested from SD201 plants for gene expression analysis. The values represent the mean fold change ± SE (n = 3) when compared with the transcript level in leaf. Bars with different letters are significantly different ($p < 0.05$).

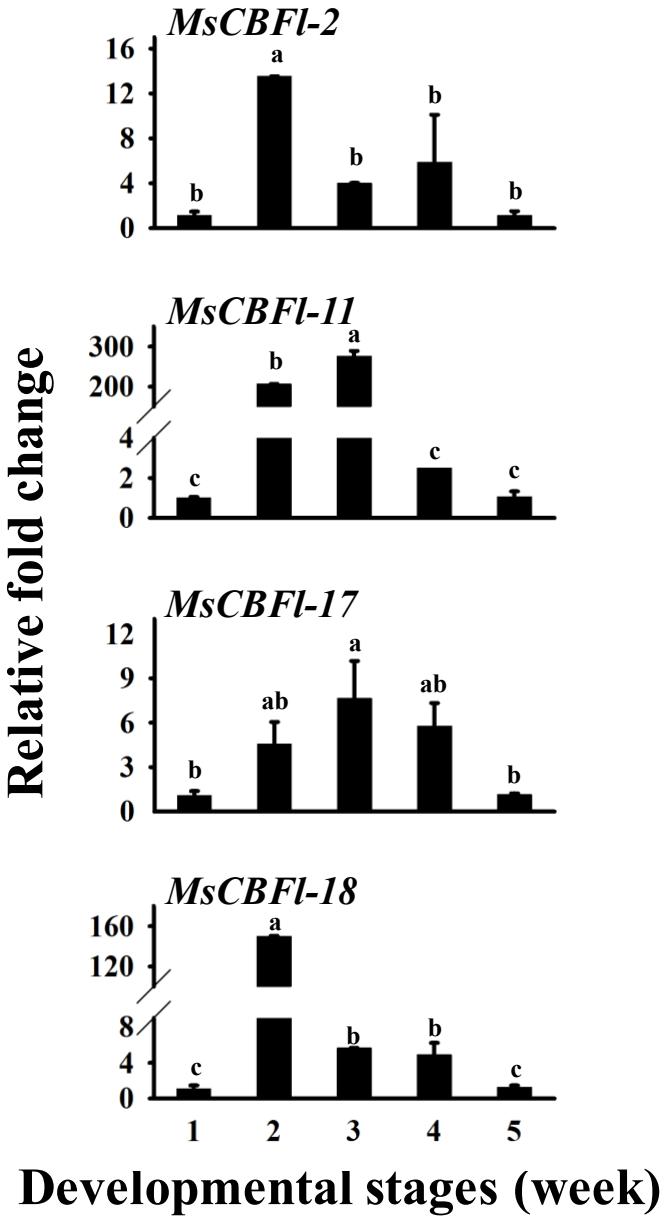


Figure 5. Expression analysis of the MsCBF-like genes at different developmental stages: samples were harvested from SD201 plants every week starting from the seventh day after germination. For the first week, the whole seedlings were harvested, and from the second week onward (Weeks 2–5), the young shoots from the upper node were harvested. The transcripts were quantified by qRT-PCR, and the values represent the mean fold change ± SE (n = 3) when compared with the transcript level at Week 1. Bars with different letters are significantly different ($p < 0.05$).

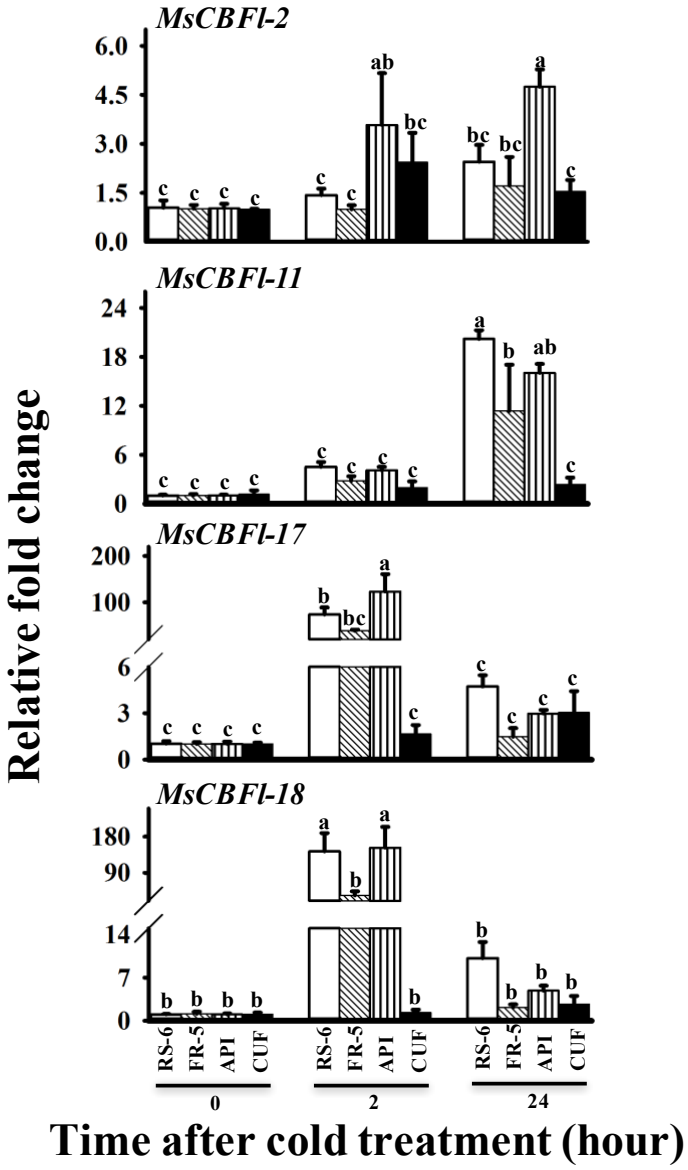


Figure 6. Cold-responsive expression of *C-repeat binding factor-like* (*CBF-like*) genes in four different alfalfa germplasm; plants (28–30 d old) of River Side (RS), Foster Ranch (FR), Apica (API), and CUF-101 (CUF) grown in Cone-tainers in a greenhouse were subjected to cold temperature (2°C) treatment, and samples were harvested at 0, 2, and 24 h after the treatment. The transcripts were quantified by qRT-PCR, and the values represent the mean fold change ± SE (n = 3) when compared with the transcript levels at 0 h of respective germplasm. Bars with different letters are significantly different (p < 0.05).

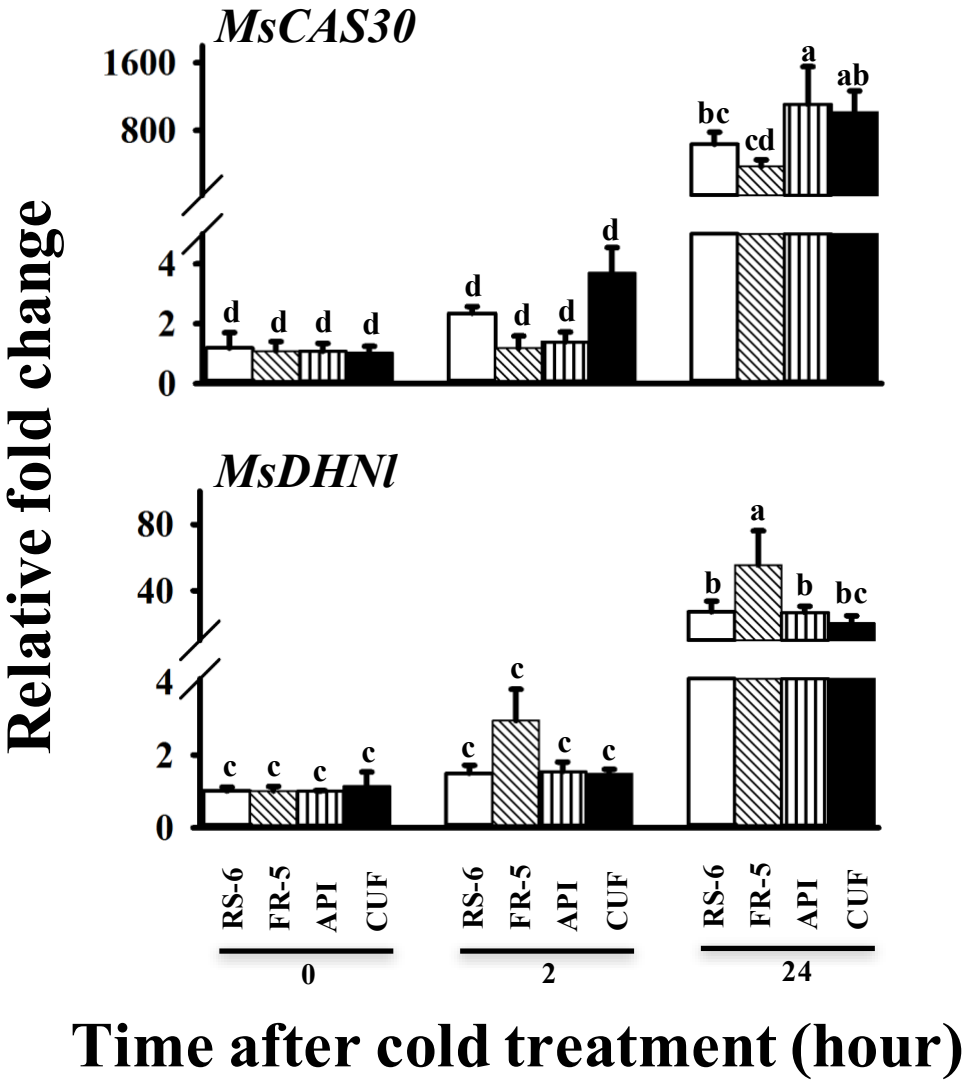
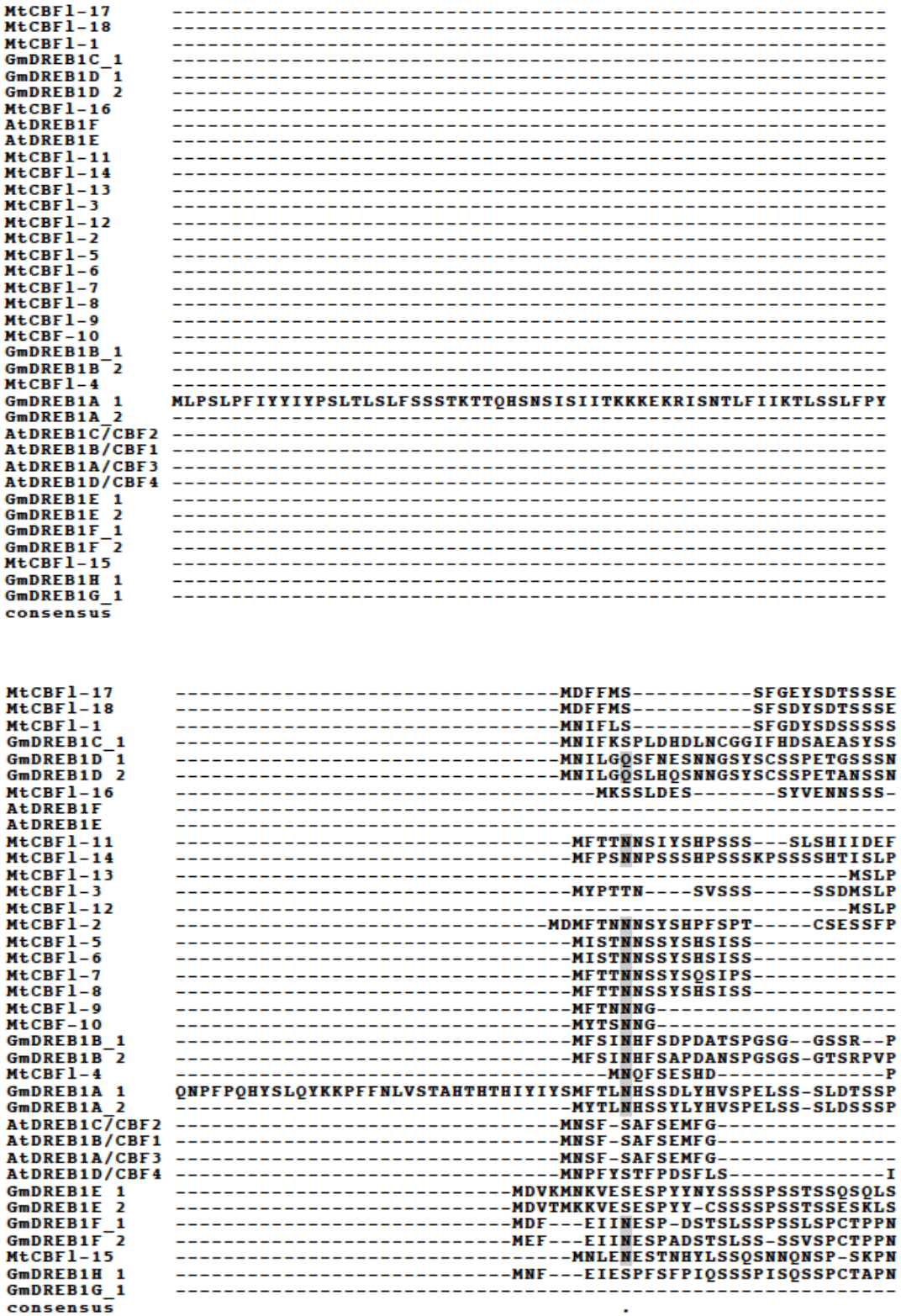


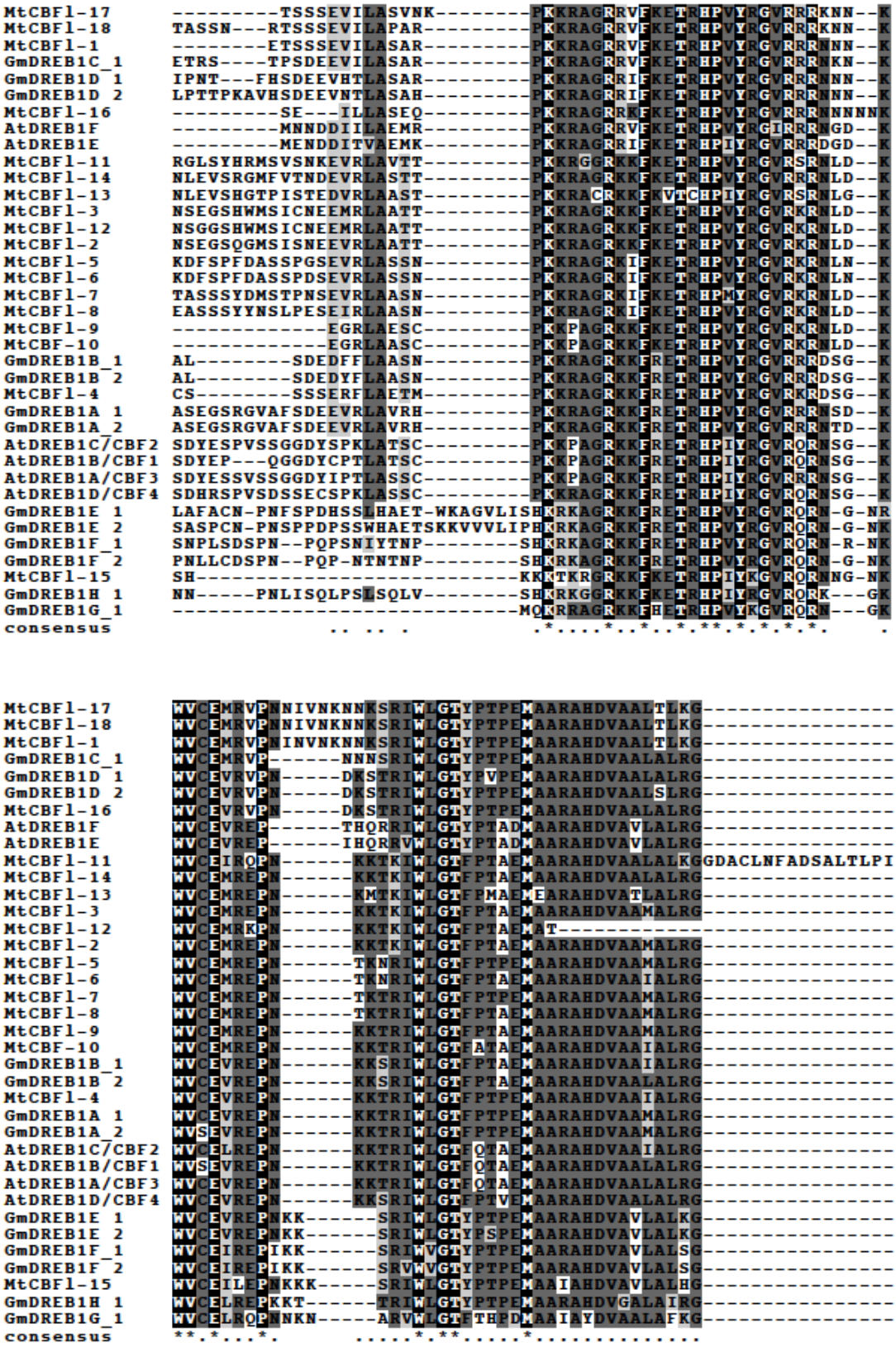
Figure 7. Cold-responsive expression of *MsCAS30* and *MsDNHl* genes in four different alfalfa germplasm; plants (28–30 d old) of River Side (RS), Foster Ranch (FR), Apica (API), and CUF-101 (CUF) grown in Cone-tainers in a greenhouse were subjected to cold temperature (2°C) treatment, and samples were harvested at 0, 2, and 24 h after the treatment. The transcripts were quantified by qRT-PCR, and the values represent the mean fold change ± SE (n = 3) when compared with the transcript levels at 0 h of respective germplasm. Bars with different letters are significantly different ($p < 0.05$).

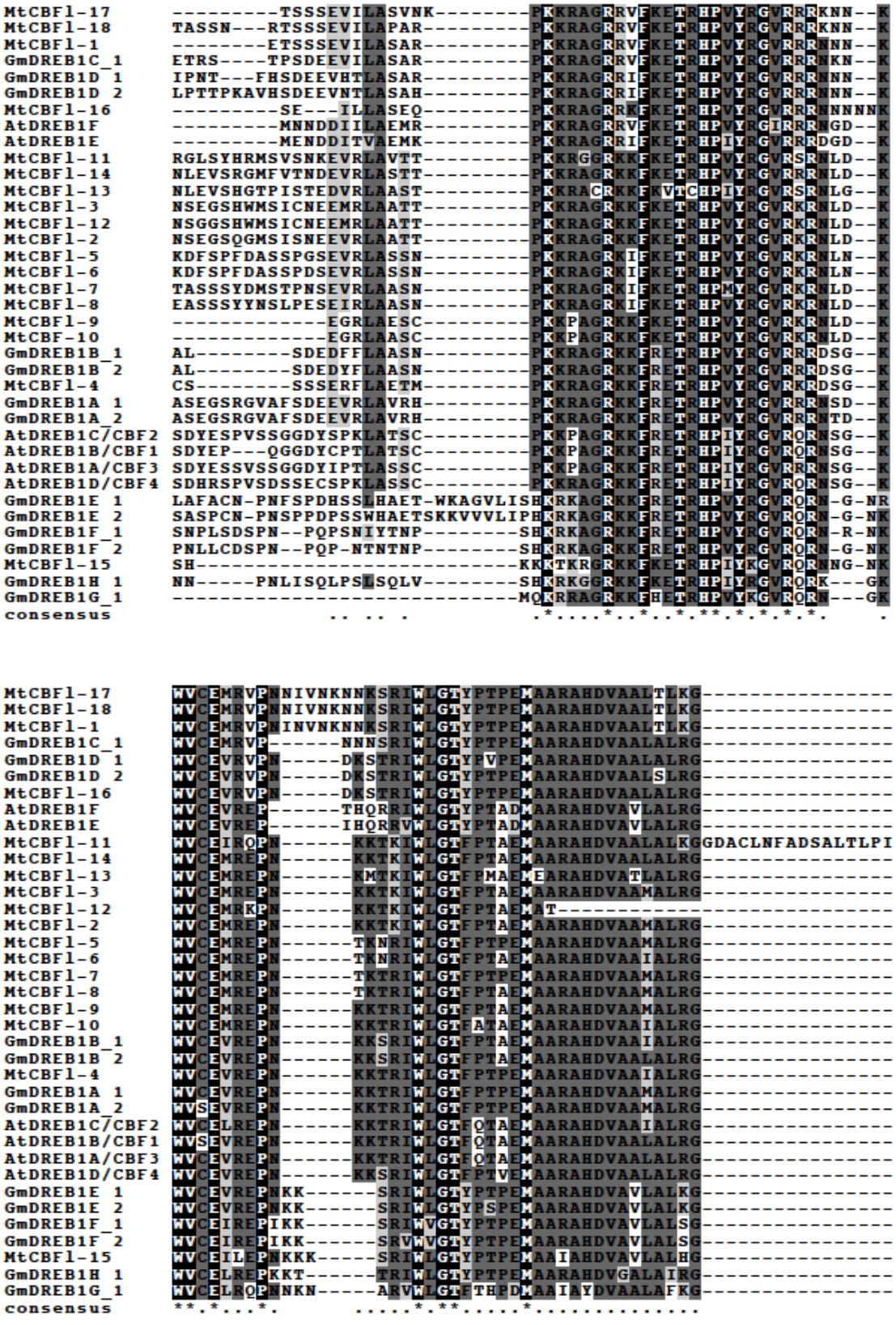
Table S1. Gene specific primer sequences used for qPCR analysis and their corresponding length, product length and melting temperature (Tm).

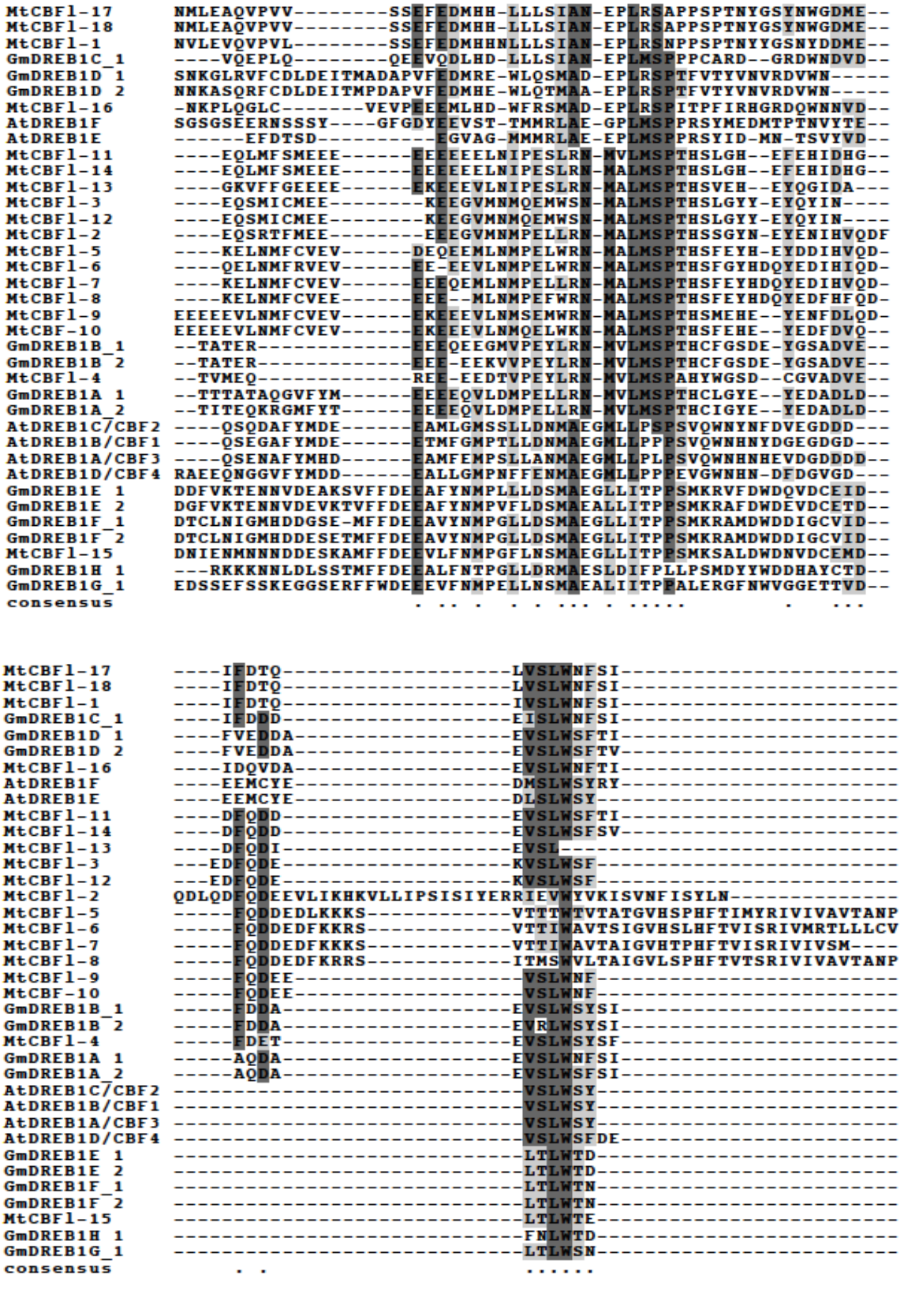
Gene-specific primer pairs for qRT-PCR

Primers	RT-Primer-Sequences	Length (bases)	Product size (bp)	Tm (°C)
MtCBF1_F	GTGATAGGGATGCTGTGGATATG	23	153	55
MtCBF1_R	GGGAAGGAGGATTAGAACGTAAAG	24		54
MtCBF2_F	ATGCCAGAGTTGTTGAGGAATA	22	156	54
MtCBF2_R	GATGAGAAGCACTTTATGCTTGAT	24		53
MtCBF3_F	GCAACGAGGAGATGCGATTA	20	107	55
MtCBF3_R	GTTCCTCTTCCTCACTCCTCTA	22		55
MtCBF4_F	GGCTTGGAACATTCCCTACA	20	110	55
MtCBF4_R	CTTCCAAGCAGAATCAGCAAAG	22		55
MtCBF5_F	GTGGAGGAATATGGCGTTGAT	21	128	55
MtCBF5_R	TCCAGGTAGTGGTAACAGACTT	22		55
MtCBF6_F	TTGTGGAGGAATATGGCACTAAT	23	131	54
MtCBF6_R	CAGATCGTCGTAACAGACCTTT	22		55
MtCBF7_F	ATGCCAGAATTGTTGAGGAATATG	24	141	53
MtCBF7_R	CCAGATCGTCGTAACAGACTTT	22		55
MtCBF8_F	TGTGTGGAAGAAGAAGAGGAAA	22	175	54
MtCBF8_R	GAACCCAACTCATGGTGATAGA	22		54
MtCBF9_F	GTCGTCGTCACAGAAGAACTTA	22	136	54
MtCBF9_R	ACGCCATATTCCTCCACATC	20		55
MtCBF10_F	CGGTGGAAGAAGAGGAAGAA	20	81	54
MtCBF10_R	ACTCTTGCATATTCAACACTTCC	23		53
MtCBF11_F	TCTACTCACACCCTTCCTCTT	21	101	55
MtCBF11_R	TAACCGCACCTCCTTGTTG	19		55
MtCBF12_F	GGAAGCCTAACAAGAAGACTAAGA	24	109	54
MtCBF12_R	TCCTTTGCCTGAGTAGTTGC	20		55
MtCBF13_F	GCATGGCGGCTCCTTATT	18	133	56
MtCBF13_R	TGTTCAACACTTCCTCCTCTTT	22		54
MtCBF14_F	CCATCTAACAACCCTTCCTCTTC	23	97	55
MtCBF14_R	TCCCACGAGAGACCTCTAAAT	21		55
MtCBF15_F	CCGGTTTCGAAATCAACTTCAC	22	111	55
MtCBF15_R	AGGCTTAGTCACAACACTCAAA	22		54
MtCBF16_F	CGTAGTTGCGGACAGTAAGG	20	98	56
MtCBF16_R	ACAGTCCCTGCAAAGGTTTAT	21		54
MtCBF17_F	CCAAAGAAGAGAGCAGGTAGAA	22	108	54
MtCBF17_R	AGGAACTCTCATCTCACAAACC	22		54
MtCBF18_F	GACCAAAGAAGAGAGCAGGTAG	22	110	55
MtCBF18_R	AGGAACTCTCATCTCACAAACC	22		54
MtActin_F	AGGCTCCACTCAATCCTAAAGCCA	24	168	59.8
MtActin_R	ACCCTTCGTATATGGGCACTGTGT	24		59.8
MtTubulin_F	ATGTTTAGGZGGGTGAGCGAGCAA	24	168	60.1
MtTubulin_R	TTCATCAGCAGTGGCATCCTGGTA	24		60
MsCAS30_F	ACAGGAACAGGAACAGGAC	20	171	61.8
MsCAS30_R	CAGTACATGATCCAGAACCAGG	22		62
MtDHNl_F	CTTCACCGATCTGATAGCTCTTC	23	119	62
MtDHNl_R	TCAACTTTCTCTACTGCCACTG	22		62


MtCBFl-17
MtCBFl-18
MtCBFl-1
GmDREB1C_1
GmDREB1D_1
GmDREB1D_2
MtCBFl-16
AtDREB1F
AtDREB1E
MtCBFl-11
MtCBFl-14
MtCBFl-13
MtCBFl-3
MtCBFl-12
MtCBFl-2
MtCBFl-5
MtCBFl-6
MtCBFl-7
MtCBFl-8
MtCBFl-9
MtCBF-10
GmDREB1B_1
GmDREB1B_2
MtCBFl-4
GmDREB1A_1
GmDREB1A_2
AtDREB1C/CBF2
AtDREB1B/CBF1
AtDREB1A/CBF3
AtDREB1D/CBF4
GmDREB1E_1
GmDREB1E_2
GmDREB1F_1
GmDREB1F_2
MtCBFl-15
GmDREB1H_1
GmDREB1G_1
consensus
MLPSLPFIYYIYPSLTLSLFSSSTKTTQHSNSISIITKKKEKRISNTLFIIKTLSSLFPY
MDFFMS
SFGEYSDTSSSE
MNIFLS
SFGDYSDSSSSS
MNIFKSPLDHDLNCGGIFHDSAEASYSS
MNILGQSFNESNNGSYSCSSPETGSSSN
MNILGQSLHQSNNGSYSCSSPETANSSN
MKSSLDES
SYVENNSSS-
MFTTNNSIYSHPSSS
SLSHIIDEF
MFPSNNPSSSHPSSSKPSSSSHTISLP
MSLP
MYPTTN
SVSSS
SSDMSLP
MDMFTNNNSYSHPFSPT
CSESSFP
MISTNNSSYSHSISS
MFTTNNSSYSQSIPS
MFTNNNG
MYTSNNG
MFSINHFSDPDATSPGSG--GSSR--P
MFSINHFSAPDANSPGSGS-GTSRPVP
MNQFSESHD
P
QNPFPQHYSLQYKKPFFNLVSTAHTHTHIYIYSMFTLNHSSDLYHVSPELSS-SLDTSSP
MYTLNHSSYLYHVSPELSS-SLDSSSP
MNSF-SAFSEMFG
MNSF-SAFSEMFG
MNPFYSTFPDSFLS
I
MDVKMNKVESESPYYNYSSSSPSSTSSQSQLS
MDVTMKKVESESPYY-CSSSSPSSTSSESKLS
MDF---EIINESP-DSTSLSSPSSLSPCTPPN
MEF---EIINESPADSTSLSS-SSVSPCTPPN
MNLENESTNHYLSSQSNNQNSP-SKPN
MNF---EIESPFSFPIQSSSPISQSSPCTAPN


MtCBFl-17
MtCBFl-18
MtCBFl-1
GmDREB1C_1
GmDREB1D 1
GmDREB1D 2
MtCBFl-16
AtDREB1F
AtDREB1E
MtCBFl-11
MtCBFl-14
MtCBFl-13
MtCBFl-3
MtCBFl-12
MtCBFl-2
MtCBFl-5
MtCBFl-6
MtCBFl-7
MtCBFl-8
MtCBFl-9
MtCBF-10
GmDREB1B_1
GmDREB1B 2
MtCBFl-4
GmDREB1A 1
GmDREB1A_2
AtDREB1C/CBF2
AtDREB1B/CBF1
AtDREB1A/CBF3
AtDREB1D/CBF4
GmDREB1E 1
GmDREB1E 2
GmDREB1F_1
GmDREB1F 2
MtCBFl-15
GmDREB1H 1
GmDREB1G_1
consensus


MtCBFl-17
MtCBFl-18
MtCBFl-1
GmDREB1C_1
GmDREB1D_1
GmDREB1D_2
MtCBFl-16
AtDREB1F
AtDREB1E
MtCBFl-11
MtCBFl-14
MtCBFl-13
MtCBFl-3
MtCBFl-12
MtCBFl-2
MtCBFl-5
MtCBFl-6
MtCBFl-7
MtCBFl-8
MtCBFl-9
MtCBF-10
GmDREB1B_1
GmDREB1B_2
MtCBFl-4
GmDREB1A_1
GmDREB1A_2
AtDREB1C/CBF2
AtDREB1B/CBF1
AtDREB1A/CBF3
AtDREB1D/CBF4
GmDREB1E_1
GmDREB1E_2
GmDREB1F_1
GmDREB1F_2
MtCBFl-15
GmDREB1H_1
GmDREB1G_1
consensus


MtCBFl-17
MtCBFl-18
MtCBFl-1
GmDREB1C_1
GmDREB1D_1
GmDREB1D_2
MtCBFl-16
AtDREB1F
AtDREB1E
MtCBFl-11
MtCBFl-14
MtCBFl-13
MtCBFl-3
MtCBFl-12
MtCBFl-2
MtCBFl-5
MtCBFl-6
MtCBFl-7
MtCBFl-8
MtCBFl-9
MtCBF-10
GmDREB1B_1
GmDREB1B_2
MtCBFl-4
GmDREB1A_1
GmDREB1A_2
AtDREB1C/CBF2
AtDREB1B/CBF1
AtDREB1A/CBF3
AtDREB1D/CBF4
GmDREB1E_1
GmDREB1E_2
GmDREB1F_1
GmDREB1F_2
MtCBFl-15
GmDREB1H_1
GmDREB1G_1
consensus

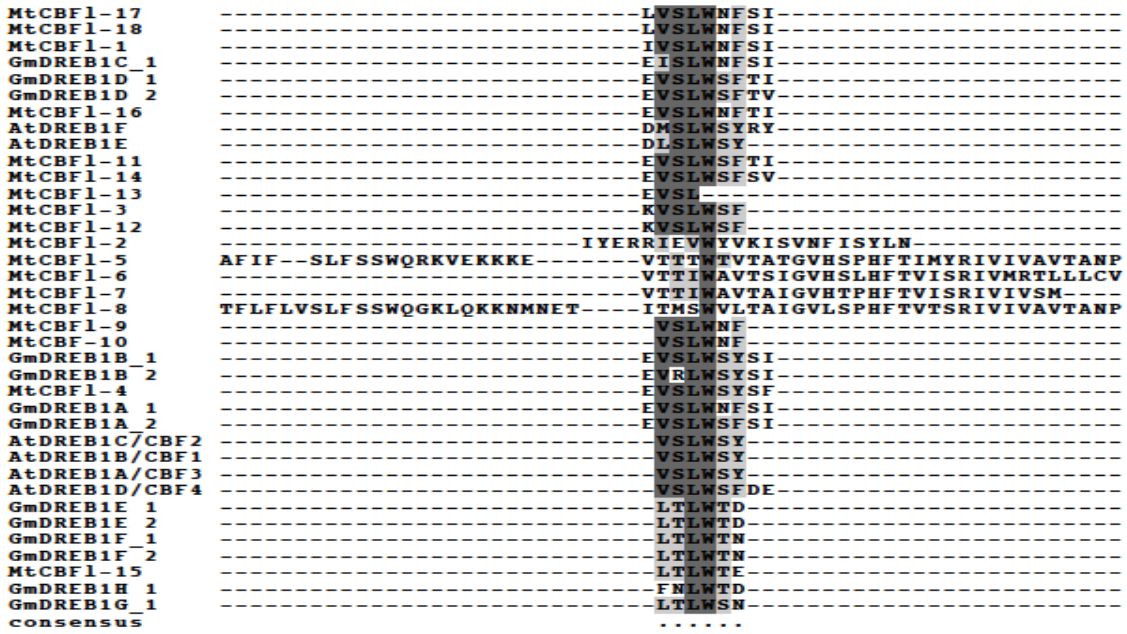


Figure S1. **Full length protein sequence alignment of CBFs from *Medicago truncatula, Glycine max* and Arabidopsis. *** represents conserved amino acid residues; • represents the identical and similar amino acid residues. The AP2 domain is highlighted with a bold black line drawn below the consensus line. *Arabidopsis* sequences include: AtDREB1C/CBF1 (At4g25470), AtDREB1B/CBF2 (At4g25490), AtDREB1A/CBF3 (At4g25480), AtDREB1D/CBF4 (At5g51990), AtDREB1E (At1g63030) and AtDREB1F (At1g12610); *Medicago truncatula* sequences include: MtCBF1 (Medtr5g010930), MtCBF2 (Medtr6g465690), MtCBF3 (Medtr6g466000), MtCBF4 (Medtr1g101600), MtCBF5 (Medtr6g465420), MtCBF6 (Medtr6g465430), MtCBF7 (Medtr6g465450), MtCBF8 (Medtr6g465460), MtCBF9 (Medtr6g465510), MtCBF10 (Medtr6g465530), MtCBF11 (Medtr6g465850), MtCBF12 (Medtr6g465990), MtCBF13 (Medtr6g466020), MtCBF14 (Medtr6g466130), MtCBF15 (2g085015), MtCBF16 (Medtr4g102660), MtCBF17 (Medtr5g010910) and MtCBF18 (Medtr5g010940); soybean sequences include: GmDREB1A1 (Glyma09g27180), GmDREB1A2 (Glyma16g32330), GmDREB1B1 (Glyma20g29410), GmDREB1B2 (Glyma10g38440), GmDREB1C1 (Glyma01g42500), GmDREB1D1 (Glyma05g03560), GmDREB1D2 (Glyma17g14110), GmDREB1E1 (Glyma12g30740), GmDREB1E2 (Glyma13g39540), GmDREB1F1 (Glyma12g09130), GmDREB1F2 (Glyma11g19340), GmDREB1G1 (Glyma13g21570) and GmDREB1H2 (Glyma12g30710)

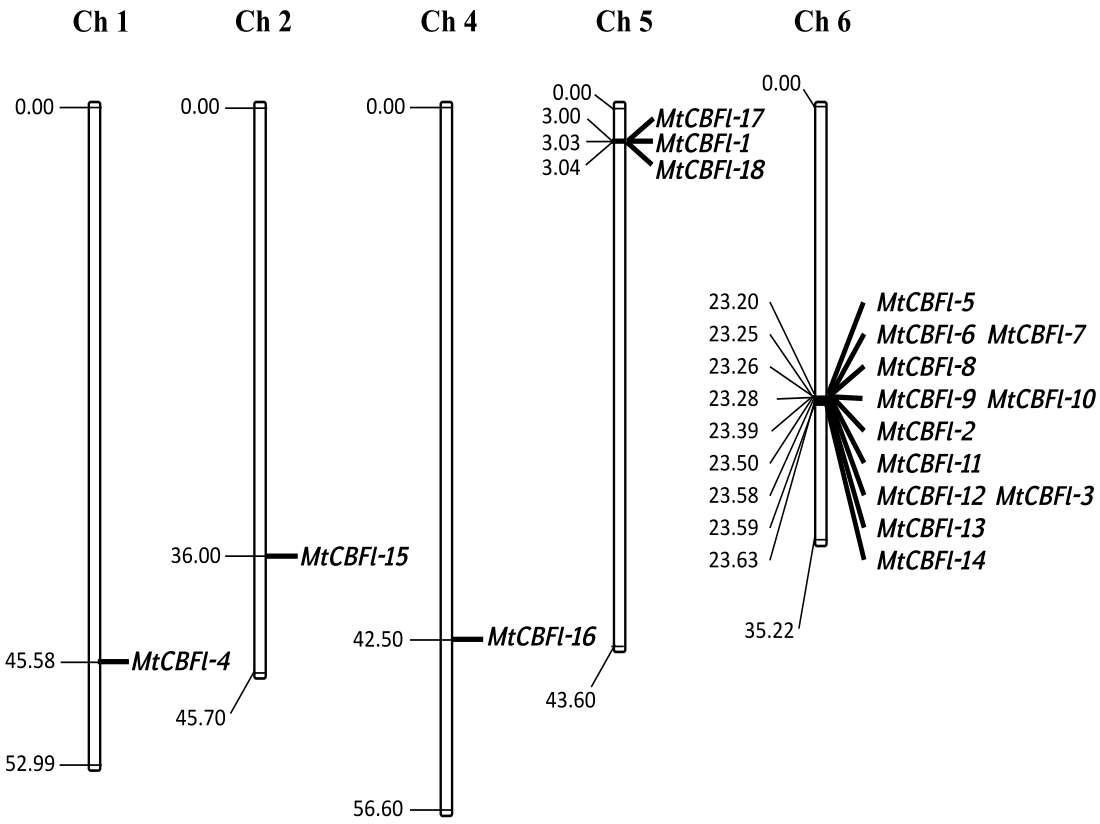


Figure S2. **Graphical representation of chromosome location of *Medicago truncatula* *CBF-like* genes.** The two groups of tandemly arranged genes, *MtCBFl-1, 17, and 18* and *MtCBFl-2, 3, 5-14* are indicated by a single position on the respective chromosomes. The length of the chromosomes is in mega base pairs (Mb).

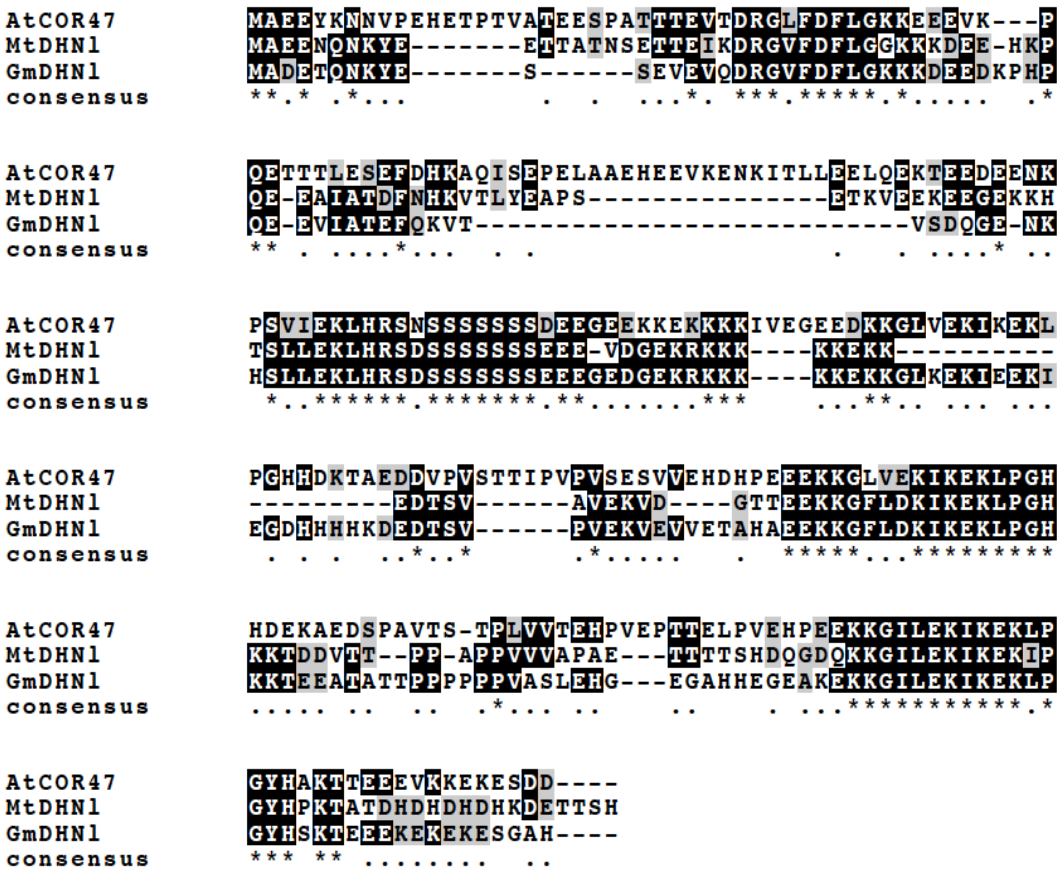


Figure S3. **Full length protein sequence alignment of Arabidopsis COR47 to DHNls from *Medicago truncatula* and *Glycine max*. *** represents conserved amino acid residues; • represents the identical and similar amino acid residues. The protein sequences include: AtCOR47 (At1g01030), MtDNH1 (Medtr3g117290), and GmDHNl (Glyma04g01130).

CHAPTER 3: THE *MIZ1-LIKE* GENE FAMILY IN MAIZE

Praveena Kanchupati[1], Yafang Wang[1, 2], Andrew Nelson[1, 3], Eric Oines[1], Viet Tran[1], Don Auger[1], Peizhi Yang[2], Tianming Hu[2] and Yajun Wu[1]

[1]*Department of Biology and Microbiology, South Dakota State University, Brookings, SD*
[2]*Department of Grassland Science, College of Animal Science and Technology, Northwest A&F University, Yangling, China*
[3]*Department of Infectious Diseases and Pathology, University of Florida, Gainesville, FL*

3.1. Abstract

The survival of land plants largely depends on the efficient functioning of the root system. Yet, very little is known about how the roots sense and grow towards a moisture gradient in soil, a phenomenon known as hydrotropism. A key advance in understanding the molecular mechanisms underlying hydrotropism was the cloning of the gene, *Mizu-kussei 1* (*MIZ1*) from the model plant Arabidopsis. A *miz1* mutant plant lacks the hydrotropic response and roots show a modified wavy growth. *MIZ1* encodes a protein that is only found in the terrestrial plants. A recent study showed that overexpression of *MIZ1* enhances the hydrotropic response of the plants leading to improved water sensing and acquisition. We hypothesize that an enhancement of hydrotropism using *MIZ1*-like genes in major crops may result in better performance under drought stress. After examining the maize (*Zea mays*) genome, we identified 15 *MIZ1-like* (*MIZl*) genes. Expression analyses of these genes in six different tissues/organs of maize seedlings reveal diverse expression profiles. Four *MIZl* genes, however, showed relatively higher levels of expression in root-tip or root-basal regions of the seedlings, a pattern very similar to that of *AtMIZ1*. Expression profiles of these four genes were studied in response to plant hormones, ABA and IAA, and to moisture gradient. *ZmMIZ1l-K* alone showed differential transcript accumulation in response to hormone treatments and appears to regulate the roots' hydrotropic response to varied moisture gradients. Based on these observations, we propose that *ZmMIZ1l-K* is the potential functional homolog of *AtMIZ1* and the gene regulates hydrotropic response in maize roots.

Keywords: MIZ1, maize, hydrotropism, functional homolog

3.2. Introduction

The distribution of water in the soil surrounding plant roots is non-homogeneous and constantly changing. The situation is worse in drought, a major abiotic stresses affecting plant growth and productivity worldwide (https://www.drought.gov/gdm/current-conditions). Plants continuously employ mechanisms to efficiently acquire and use the available water. Hydrotropism is one such mechanism, where roots sense differences in moisture gradient, and bend and grow towards higher moisture area (Darwin and Darwin 1880; Loomis and Ewan 1936).

Though root hydrotropism is recognized as an important response to avoid water deficit (Bolaños and Edmeades 1993), research on hydrotropism and its regulation at the molecular level has been very limited. The most important reason is the difficulty in separating hydrotropic response from gravitropic or thigmotropic responses. The first breakthrough came from the studies of the pea mutant *ageotropum* that shows a positive hydrotropic response without interference from gravitropic response (Jaffe et al. 1985). Other studies used different experimental setups that included clinorotation and microgravity in space to distinguish hydrotropism from other trophic responses of roots (Takahashi 1997; Cassab et al. 2013). These studies concluded that the sensing and signaling pathways in roots to gravity and moisture gradients are different. Recently two novel QTL (quantitative trait loci) were identified in wheat (*Triticum aestivum* L.), a monocot, that were associated with hydrotropism (Hamada et al. 2012). These QTL are located on different chromosomes than the QTL for gravitropism, which argue that the genetic factors controlling the hydrotropic responses of the wheat roots are separate from the genetic factors controlling gravitropism.

Early physiological studies show that the root cap cells are important in sensing a moisture gradient (Jaffe et al. 1985). A recent study with Arabidopsis indicated that root cortex cells are important for moisture sensing (Haruta et al. 2017). Ca^{2+} ions and hormones, namely auxin, cytokinin and ABA, have been shown to play important role in hydrotropic response (Takano et al. 1997; Takahashi et al. 2002; Kaneyasu et al. 2007).

The molecular participants of the sensing and signaling pathways regulating hydrotropism are poorly studied. Thus far only two genes, *MIZ1* and *MIZ2*, have been identified based on mutant analysis studies in Arabidopsis. *MIZ1* was identified in studies of *mizu-kussei1* mutants (Kobayashi et al. 2007). The mutants' roots displayed normal gravitropism and growth but lacked a hydrotropic response. They also displayed wavy growth in roots and reduced sensitivity to light. The mutant phenotype mutant was linked to a recessive mutation of the *MIZ1* gene. *MIZ1* encodes for a protein with an unknown function and contains a DUF617 (domain of unknown function 617) domain and the gene homologues are only found in terrestrial plants (Yamazaki et al. 2012). Expression analysis of the gene shows extensive transcript accumulation in the columella cells of the root caps. Some expression was also observed in the mature regions of the roots and in the hydathodes (water-excreting epidermal structures) of the leaves (Kobayashi et al. 2007). MIZ1 protein is localized to the cytoplasmic side of the endoplasmic reticulum membrane in cortical cells and lateral root cap cells (Yamazaki et al. 2012). The overexpression lines of *MIZ1* (*MIZ1OEs*) display an enhanced hydrotropic response (Miyazawa et al. 2012). Overexpression of *MIZ1* also results in fewer lateral roots. which is associated with a decreased auxin level in roots. The suppression of the lateral roots is rescued by the external application of auxin (Moriwaki et al. 2011),

suggesting a negative regulation of auxin levels by the *MIZ1* gene. This notion is supported by the fact that the *miz1* roots have increased auxin levels. Furthermore, when the overexpression lines are treated with increasing doses of auxin, reduction in the hydrotropic curvature of the roots is observed. These results indicate that *MIZ1* acts upstream of auxin synthesis and regulates the hydrotropic response.

In contrast, *MIZ1* functions downstream of cytokinin signaling to regulate the lateral root development. Exogenous application of cytokinin results in the accumulation of *MIZ1* at the lateral root primordia (Moriwaki et al. 2011). *MIZ1* was also shown to integrate signals from light and ABA signaling pathways to regulate hydrotropism in Arabidopsis (Moriwaki et al. 2012).

The second gene that was shown to play an essential role in hydrotropism in Arabidopsis is *MIZ2* (Miyazawa et al. 2009). *MIZ2* encodes for a guanine-nucleotide exchange factor for ADP-ribosylation factor-type G proteins or GNOM. Unlike other *GNOM* mutants that show altered auxin transport (Geldner et al. 2003; Steinmann et al. 1999), *MIZ2* mutants do not perturb the agravitrophism. Importantly, *miz2* mutants are similar to *miz1* mutants in that they are ahydrotropic.

Auxin, *MIZ1* and *MIZ2* together regulate lateral root development during hydrostimulated conditions. Apart from the observation that MIZ1 requires MIZ2 activity for its function in lateral root development (Moriwaki et al. 2011), the role of *GNOM*/*MIZ2* in root hydrotropism is still undetermined. Various studies have suggested that multiple players interact to regulate hydrotropism, but a clear pathway and working mechanism is still missing.

Arabidopsis is the most studied plant relative to elucidating the molecular players of hydrotropism. Hydrotropism has also been examined in wheat (Hamada et al. 2012), pea (*Pisum sativum* L.) (Takahashi et al. 1992), soybean (*Glycine max* L.) (Tsutsumi et al. 2002) and maize (*Zea mays* L.) (Takahashi and Scott 1991). These plants may have similar molecular mechanisms regulating hydrotropism based on the fact that *MIZ1* homologues are found only in the land plants, not algae (Kobayashi et al. 2007). We hypothesize that identification and manipulation of genes that are functionally homologous to *MIZ1* and *MIZ2* will result in better hydrotropic performance of agriculturally important crops like maize under drought conditions.

In the present study, we performed a genome-wide study of the *MIZ1-like* gene family in maize. The objective is to identify the candidates of functional homologue of *AtMIZ1* through gene expression analysis. This study serves as the foundation for the functional characterization of *ZmMIZ1-like* genes.

3.3. Results

3.3.1. Fifteen *MIZ1l* genes were identified in the maize genome

To identify putative *MIZ1l* genes in maize, we performed a BLASTP (basic local alignment search tool, protein) search of the maize genome using the peptide sequence of Arabidopsis MIZ1, which resulted in 25 hits. Fifteen homologues were identified based on the E values (cutoff value of 2.3E-29), score (cutoff value of 115.2) and percentage identity (between 42%-61%) with AtMIZ1. Based on chromosome location, redundant sequences were discarded. We performed the alignment of just the signature DUF617 domain (Figure 1a.) as well as the complete sequence (Figure S1). The maize and Arabidopsis MIZ1 sequences show greater homology in the DUF617 domain compared

to the N-terminal region. We named the maize *MIZ1* sequences ZmMIZ1l-A to ZmMIZ1l-O in order of the corresponding chromosome locations identified from the maize genome browser.

3.3.2. Phylogenetic analysis and chromosomal location of the *MIZ1l* gene family in maize

Based on the sequence similarity to AtMIZ1, the fifteen members of the maize *MIZ1l* gene family were divided into two subgroups, designated subgroups 1 and 2 according to the clades on the phylogenetic tree. Nine of the maize MIZ1l proteins clustered together to form the maize-only subgroup 1 and the rest six clustered with AtMIZ1 to form subgroup 2 (Figure 1b.).

A second phylogenetic tree was constructed based on the protein sequence similarity among all the members of maize and Arabidopsis *MIZ1* gene family (Figure 2a.). This tree divided the maize *MIZ1* family into four subgroups that had members from both maize and Arabidopsis.

The *ZmMIZ1l* genes are distributed randomly. There is only one *MIZ1l* homolog found on chromosomes 1, 6, 7 and 9; whereas the rest of the chromosomes, with the exception of chromosome 10, which contains no *MIZ1l* homolog, harbor more than one homolog (Figure 2b.).

3.3.3. Gene structure is conserved between Arabidopsis and maize

The *AtMIZ1* gene has no introns and the DUF617 domain occupies more than 60% of the coding sequence. Except *ZmMIZ1l-C, ZmMIZ1l-G,* and *ZmMIZ1l-O*, the maize *MIZ1l* homologs are very similar, i.e. a larger N-terminal sequence and a very small C-terminus sequence that flank the DUF domain. This suggests a significant

conservation between monocots and dicots (Figure 1c.). *ZmMIZ1l-G* and *ZmMIZ1l-O* each contain a single intron in the DUF617 domain and form a small subclade, suggesting they are recent duplicates. *ZmMIZ1l-C* encodes the only protein that lacks the N-terminus sequence.

3.3.4. Four *ZmMIZ1l* homologs are highly expressed in roots

Using GUS (β-glucuronidase) and green fluorescent protein reporters, MIZ1 expression was detected in the root cap, the mature region of roots and the hydathodes of the leaves in Arabidopsis (Kobayashi et al. 2007; Yamazaki et al. 2012). In order to identify maize *MIZ1l* homologs that may function in root hydrotropism we determined which *MIZ1l* is expressed in roots. Six tissues were sampled from maize seedlings and examined for *MIZ1l* transcript levels. Three homologs, *ZmMIZ1l-B, ZmMIZ1l-I,* and *ZmMIZ1l-M*, showed relatively high transcript accumulation levels in the root tips including the root cap cells (Figure 3). One additional homolog, *ZmMIZ1l-K*, showed higher transcript levels in the mature region exclusive of the root tip region. Unlike *AtMIZ1*, none of the *ZmMIZ1l* homologs showed significant transcript accumulation in leaves. *ZmMIZ1l-A, ZmMIZ1l-C, ZmMIZ1l-D, ZmMIZ1l-F ZmMIZ1l-H, ZmMIZ1l-N* and *ZmMIZ1l-O* were detected in very low levels in all the tissues examined and we were not able to obtain a reliable expression profiles for these genes.

3.3.5. Exogenous application of auxin and ABA regulates the transcript accumulation of only one *MIZ1l* homolog in maize

Eight *MIZ1l* homologs that were expressed at higher levels in roots tissues were evaluated for their response to the plant growth factors, auxin and ABA. Both auxin and ABA were shown to regulate hydrotropism in roots of Arabidopsis (Miyazawa et al.

2012; Takahashi et al. 2002). Only *ZmMIZ1l-K* responded to the exogenous application of IAA (auxin) and ABA (Figure 4). *ZmMIZ1l-K* transcript levels are significantly increased by auxin and decreased by ABA. None of the other homologs show a significant response to the growth factors, and *ZmMIZ1l-M* is shown as an example (Figure 4). ABA responsive gene *ZmRAB17* and auxin response factor *ZmGH3-2* are positive controls (Feng et al. 2015; Kizis and Pagès 2002; Zheng et al. 2006) and responded as expected.

3.3.6. *ZmMIZ1l-K*, *ZmRAB17* and *ZmGH3-2* show higher accumulation on the wet-half of hydrotropic roots

We examined the expression levels of the same four *ZmMIZ1l* genes in the hydrotropic roots, together with *ZmRAB17* and *ZmGH3-2*. *ZmMIZ1l-K* shows significantly higher transcript accumulation in the root halves facing higher moisture compared to both the halves facing lower moisture and control roots (Figure 5). The control roots were also split into halves and no significant difference in transcript levels between the two halves. The expression levels of each gene in CK D and CK W samples were combined and averaged as a single CK that is compared with the transcript levels in HYD W and HYD D samples. *ZmRAB17* showed a similar expression pattern as *ZmMIZ1l-K*. In contrast, *ZmGH3-2*, an auxin response factor, shows significant transcript accumulation in both the HYD W and HYD D root halves compared to the control roots. *ZmMIZ1l-B*, *ZmMIZ1l-I* and *ZmMIZ1l-M* show no response to hydrotropic treatment.

3.4. Discussion

3.4.1. Evolutionary expansion of *MIZ1l* gene family in *Zea mays*

It is interesting to note that given the bigger size of the maize genome (2,300 Mbp) when compared to the dicot Arabidopsis (135 Mbp), maize appears to have only 15 *MIZ1l* homologues, while it has been reported that the later has a total of 11 *MIZ1l* homologues that share a varied identity ranging from 34% to 59% with the *MIZ1* gene (Kobayashi et al. 2007). Other plants like rice (13 *MIZ1l*), soybean (18 *MIZ1l*) and sorghum (11 *MIZ1l*) have similar numbers of *MIZ1l* genes irrespective of their genome size and the whole genome duplication events (Table S2.).

Phylogenetic analysis of Arabidopsis and maize *MIZ1* genes suggests that they diverged as a gene family before the divergence of Arabidopsis and maize. Nearly all subclades contain *MIZ1l* genes from both species, indicating that within each subclade *MIZ1l* orthologues evolved from a common ancestor gene (Figure 2a). The analysis indicates that the both *ZmMIZ1l* and *AtMIZ1l* genes evolved through multiple events (Figure 2a). Genes in subclade 4 that includes *AtMIZ1* likely evolved from a common ancestral gene. The ancestral gene first diverged into *AtMIZ1* and a paralogue, which further diverge to form *ZmMIZ1l*-I and other four Arabidopsis paralogues (*MIZ1l-A*, *B*, *D*, *E* and *K*). *AtMIZ1l-A* and *B*, and *AtMIZ1l-D* and *K* likely represent two duplication events. ZmMIZ1l genes also evolved through multiple duplication events. As indicated in subclade 3, *AtMIZl-F* and *I* share a common ancestor with *ZmMIZ1l-G*, *O*, *M*, *E* and *J*. While evidence of gene duplication can be seen in nearly every subclade, *ZmMIZ1l* genes with the highest identities are usually located at different chromosomes (Figure 2b), suggesting that the formation of gene family takes more than duplication.

Sequence alignment of the MIZ1l proteins reveals a high degree of conservation of the DUF 617 domain among Arabidopsis and maize. Notably, the absence of introns is conserved with the exception of *ZmMIZ1l-G* and *ZmMIZ1l-O*. The similar placement suggests that the intron formed as a single event and paralogues formed subsequently. The structural conservation of the *MIZ1l* gene family in Arabidopsis and maize suggests a functional conservation (Figure 1a & 1b).

3.4.2. Tissue types and growth hormones elicit differential expression levels of *ZmMIZ1l* transcripts

The initial breakthrough in the molecular dissection of the root hydrotropic response came from studies of the Arabidopsis mutant *miz1*. Map-based cloning and primary expression profiles of *MIZ1* suggested that the gene functions in the early phase of hydrotropic response. Transgenic Arabidopsis seedlings carrying the *pMIZ1*::*GUS* or *pMIZ1::MIZ1-GFP* fusion genes show that *MIZ1* expresses in columella cells of the root cap and in cells of the mature region of the roots. Strikingly, no expression is observed in the elongation zone of roots (Kobayashi et al. 2007). Evidence of higher transcript accumulation of the *MIZ1* gene in roots is based upon information obtained from the EST databases (Birnbaum et al. 2003). Gene expression profiles of *MIZ1* based upon microarray analysis data, available at the AtGenExpress Visualization tool (http://jsp.weigelworld.org/expviz/expviz.jsp?experiment=abiostress&normalization=absolute&probesetcsv=At2G41660&action=Run), show higher expression levels in the roots compared to the aerial parts. The expression profiles also suggest an increase in transcript levels in response to abiotic stresses like cold, osmotic and salt, predominantly in the

roots. Exogenous application of growth hormones like ABA and auxin resulted in up regulation and down regulation of *MIZ1* gene respectively.

Four of the 15 *ZmMIZ1l* genes show a higher transcript levels in roots relative to the aerial parts (Figure 3) suggesting they can be functional orthologues to *MIZ1* in Arabidopsis. When these four genes were studied for their response to the growth hormones ABA and auxin, *ZmMIZ1l-K* was the only one that showed changes in transcript level in response to the treatments (Figure 4). Unlike *AtMIZ1*, *ZmMIZ1l-K* has increased transcript levels in seedlings treated with IAA but lower in seedlings treated with ABA. While the findings suggest differential regulation of the gene in a species-specific manner, they may also suggest that *ZmMIZ1l-K* does not play a role in hydrotropism. Alternatively, maize roots may employ a different molecular process in hydrotropic response that still involves *ZmMIZ1l-K*. To distinguish these possibilities, it is necessary to determine if *ZmMIZ1l-K* is involved in hydrotropic response.

3.4.3. *ZmMIZ1l-K* expression levels are changed in hydrotropic response of maize roots

Maize roots, like Arabidopsis, when exposed to differential moisture gradients bend towards the higher moisture gradient. The involvement of *ZmMIZ1l-K* in root hydrotropism is supported by the observation that its transcript level was only increased in the root halves facing high moisture. Interestingly, the transcript levels of *ZmRAB17* and *ZmGH3-2* also increased, indicating that both ABA and IAA may contribute to the hydrotropic response (Figure 5). The greater fold increase in transcript levels of *ZmGH3-2* in hydrotropic roots may indicate a greater change in IAA content and a stronger impact on the hydrotropic response compared to ABA. This may also explain why

ZmMIZ1l-K transcript level is increased in hydrotropic roots despite the fact that ABA and IAA have an opposite effect on the transcript levels of this gene.

The fact that *ZmMIZ1l-K* transcript level is significantly increased only on the moist side suggests an important role in regulating cell elongation. Hydrotropic root bending is caused by a differential cell elongation on the sides of a root, with less cell elongation on the wet side compared to the dry side, resulting in root bending toward the water source. This differential cell elongation is proposed to be a result of differential distribution of IAA (Hirasawa et al. 1997; Takahashi et al. 2009; Takahashi and Suge 1991). Functional analysis showed that *AtMIZ1* negatively regulates IAA content in the roots. *ZmMIZ1l-K* may play a similar role as *AtMIZ1*. We propose a model to explain the role of *ZmMIZ1l-K* in hydrotropism. When a maize root is exposed to a hydrostimulant, IAA is differentially distributed to the wet side. The higher concentration of IAA in the wet side enhances expression of *ZmMIZ1l-K,* which in turn suppresses IAA synthesis. The consequence of the negative feedback leads to an overall reduction of IAA content in the wet side, resulting in a slower elongation compared to the dry side.

In summary, we propose that *ZmMIZ1l-K* is potentially the functional homolog of *AtMIZ1*. *ZmMIZ1l-K* is subject to ABA and IAA regulation and appears to play a role in the fine tuning of IAA level in hydrotropic roots, thus controlling the hydrotropic bending of maize roots.

3.5. Materials and methods

3.5.1. Plant materials and growth conditions

Seed of Dekalb hybrid DKC43-10 (Monsanto, St. Louis, MO) was used in this study. Among several varieties tested DKC43-10 forms fewer seminal roots at the early

seedling stage. Seeds were placed between two layers of wet paper towels in a glass tray. The tray was covered with plastic wrap (The Glad Products Company, Oakland, CA) to maintain high humidity. The cling wrap was perforated using a needle to ensure airflow. The tray was placed inside a dark growth chamber maintained at 25°C in the dark. The tray was placed at an angel of approximately 70° above horizontal to allow excessive water to drain and roots to grow strait down.

3.5.2. Treatment and sampling

Tissue analysis

When the primary root was about 1.5 cm, six different tissues/organs were harvested from each maize seedling . Three regions were harvested from the primary roots: root tip (first 2 mm from the tip encompassing the cell division and transition zone), root middle (next 5 mm, the elongation zone) and root base (5 mm, the mature zone). Preliminary analysis identified that the root elongation zone is about 6-7 mm long when the root was about 2 cm. Mesocotyles, young leaves and coleoptiles were also harvested. All the samples were immediately frozen in liquid nitrogen and stored at −80°C.

Hormone treatments

Maize seedlings with the primary root about 1.5 cm long were treated with 50 mL 10 μM solutions of indole-3-acetic acid (IAA, Sigma-Aldrich, St. Louis, MO) or abscisic acid (ABA, PhytoTechnology Laboratories, Lenexa, KS). Both IAA and ABA stock solutions were prepared in 1N NaOH (ThermoFisher Scientific, Waltham, MA). The control seedlings received 50 mL water containing 5 μL of 1N NaOH. The solutions were added to the seedlings covered in wet paper towels in glass trays and were carefully

drained after a few gentle swirls. The process was repeated once. The solutions were kept in the tray during the third wash. The tray was titled so that the roots were never submerged in the solution to prevent hypoxia stress. After 1 hour of treatment, 6 mm root tip sections were harvested. Eight roots tips were harvested per each replicate and were immediately frozen in liquid nitrogen and stored at -80°C.

Hydrotropic treatment

Maize seeds were germinated as described previously. The seedlings that had a straight primary root of about 1.5 cm were selected and transferred to a wet-pads cassette that had slots to hold eight seedlings. The two wet-pads of the cassette secured the kernel in place, and the primary roots were flanked by two water-saturated pads without contact. The cassettes were then placed in a chamber with near 100% humidity with the roots pointing vertically downward. The seedlings were allowed to recover for two hours. After the recovery, one of the wet-pads of the cassette was removed so that roots were exposed to dry air (65%) in the chamber to initiate hydrotropic response. The dry air in the chamber was created and maintained by placing a saturated K_2CO_3 (ThermoFisher Scientific, Waltham, MA) solution in the chamber (Takahashi et al. 2002). For the control, the roots were exposed to the water saturated air in the chamber.

After 1.5 hour of treatment hydrotropic bending becomes visible and 6 mm root tips from the control and hydrotropic seedlings were harvested under a safe green light. The roots tips were sliced longitudinally with a scalpel into halves and labeled as control-wet, control-dry, hydrotropic-wet and hydrotropic-dry. The side facing the wet-pads was designated as the wet side and the side facing away from the wet-pads was designated as

the dry side. Forty-eight half-roots were harvested per replicate and immediately frozen in liquid nitrogen and stored at -80°C.

3.5.3. Identification of *MIZ1l* homologs in *Zea mays* and phylogenetic analysis

A BLASTP (basic local alignment search tool, protein) search was conducted against the maize genome database *Zea mays* Ensembl-18 at Joint Genome Institute (https://phytozome.jgi.doe.gov/pz/portal.html#!info?alias=Org_Zmays), using the AtMIZ1 protein sequence. The default Expect (E) threshold value for the *Zea mays* MIZ1 homologs that were chosen was set at -1, and the BLOSUM62 comparison matrix was used. The retrieved sequences were named, *ZmMIZ1l-A* to *ZmMIZ1l-O*. Multiple sequence alignment of the protein sequences was done using Clustal Omega (Sievers et al. 2011), and a phylogenetic tree was constructed using the Neighborhood-joining method in MEGA 6 (Tamura et al. 2013). The information on genomic sequences, cDNA sequences, exon-intron distribution and chromosome locations of each *ZmMIZ1* gene was obtained from the genome database at the Joint Genome Institute (https://phytozome.jgi.doe.gov/pz/portal.html#).

In addition, AtMIZ1 protein homologs from other plants that include *Glycine max* (soybean, *GmMIZ1l*), *Sorghum bicolor* (sorghum, *SbMIZ1l*) and *Oryza sativa* (rice, *OsMIZ1l*) were retrieved from their respective genome databases at the Joint Genome Institute by following the steps described above. The multiple sequence alignment and phylogenetic analysis was conducted using Clustal Omega and MEGA 6 respectively.

3.5.4. Primer design

Gene specific primers (Table S1) were designed for *ZmMIZ1l* and other genes using an online primer design tool from Integrated DNA Technologies (Coralville, IA, USA). The efficiency and specificity of each primer pair was determined using B73 maize genomic DNA (1 ng) as the template in a 20 μL polymerase chain reaction (PCR) containing 2 μL of 10X PCR buffer, 1 μL each of 10 μM primers, 1 μL of 2 mM dNTPs and 0.1 μL of Taq polymerase (5 U/μL, BioLabs Inc., Ipswich, MA). All reactions were performed in a gradient thermocycler (Eppendorf Mastercycler, Eppendorf, Hauppauge, NY) with PCR conditions set as: initial denature step at 94°C for 5 min followed by 35 cycles of 94°C for 30 s, 30 s at annealing temperature gradient (R= 3°C/s, G= ± 2.4°C), extension at 72°C for 1 min and a final extension at 72°C for 10 min. The PCR products were electrophoresed on a 1% agarose gel stained with ethidium bromide, and images were visualized using a Bio-Rad ChemiDoc image analysis system (Bio-Rad Laboratories Inc., Hercules, CA).

3.5.5. RNA isolation and cDNA synthesis

Total RNA was isolated using Trizol reagent (Invitrogen, Carlsbad, CA). The RNA samples were quantified using Nanodrop ND-1000 Spectrophotometer (ThermoFisher Scientific, Waltham, MA), and samples with 260/280 ratio from 1.9 to 2.1 and 260/230 ratio from 2.0 to 2.5 were used for further analysis. RNA quality was also examined by separating RNA on a 1% agarose gel stained with ethidium bromide. The samples that showed two sharp major rRNA bands were used for cDNA synthesis. First strand cDNA synthesis was performed using the high capacity cDNA Reverse Transcription kit (ThermoFisher Scientific, Waltham, MA) in a 20 μL reaction according

to manufacturer's instructions. The cDNA samples were diluted 80 times for use in quantitative real-time PCR (qRT-PCR) reaction.

qRT-PCR reactions with the housekeeping gene *ZmUBCP* were performed directly on RNA samples without reverse transcription, in comparison with respective cDNA samples. Only the cDNA samples whose corresponding RNA samples showed no amplification or significantly greater cycle threshold (CT) values (≥35 cycles) in qRT-PCR analysis were used for the gene expression analysis.

3.5.6. Quantification of transcripts

qRT-PCR was performed using DyNAmo Flash SYBR Green Hot Start qRT-PCR Kit (ThermoFisher Scientific, Waltham, MA) following manufacturer's instructions in a 20 µL reaction in an ABI 7900HT High-Throughput Real-Time Thermocycler (Applied Biosystems, Foster City, CA). Each sample from three biological experiments was assayed twice as technical replicates. The thermocycler program was set to: 15-min activation at 95°C followed by 40 cycles of 15 s at 94°C, 30 s at annealing temperature (Supplemental Table 1), 30 s extension at 72°C, followed by a dissociation curve step. The dissociation curve was used to determine the primer efficiency and specificity. The normalized relative fold changes in the transcripts of *ZmMIZ1l* genes were calculated using the $2^{-\Delta\Delta Ct}$ or comparative Ct method based on the difference between the target and reference genes as described by Livak and Schmittgen (2001).

3.5.7. Data analysis

Statistical analysis was performed using STATISTIX 9.0 Analytical Software (Tallahassee, FL) and Microsoft Excel 2010 data analysis tools pack (Redmond, WA). Tissue-specific expression data was subjected to analysis of variance (ANOVA) using the

linear model with completely randomized design to determine significant differences among the treatments followed by Tukey's HSD all pair comparison to ascertain significant differences ($p < 0.05$). Expression data obtained from samples treated with hormones and hydrotropic conditions was analyzed using Student's T-test to ascertain significant differences between treatment means ($p < 0.05$).

3.6. Acknowledgements

This work was supported by the USDA National Institute of Food and Agriculture Hatch Project 3AH390, South Dakota Agricultural Experiment Station, and the Department of Biology and Microbiology at South Dakota State University. We are grateful to the Functional Genomics Core Facility (FGCF) at South Dakota State University.

3.7. Reference

Birnbaum K, Shasha DE, Wang JY, Jung JW, Lambert GM, Galbraith DW, Benfey PN (2003) A Gene Expression Map of the Arabidopsis Root. Science 302 (5652):1956

Bolaños J, Edmeades GO (1993) Eight cycles of selection for drought tolerance in lowland tropical maize. I. Responses in grain yield, biomass, and radiation utilization. Field Crops Research 31 (3):233-252. doi:https://doi.org/10.1016/0378-4290(93)90064-T

Cassab GI, Eapen D, Campos ME (2013) Root hydrotropism: An update. American Journal of Botany 100 (1):14-24. doi:10.3732/ajb.1200306

Darwin C, Darwin F (1880) The Power of Movement of Plants. John Murray,

Feng S, Yue R, Tao S, Yang Y, Zhang L, Xu M, Wang H, Shen C (2015) Genome-wide identification, expression analysis of auxin-responsive GH3 family genes in maize (Zea mays L.) under abiotic stresses. Journal of Integrative Plant Biology 57 (9):783-795. doi:10.1111/jipb.12327

Geldner N, Anders N, Wolters H, Keicher J, Kornberger W, Muller P, Delbarre A, Ueda T, Nakano A, Jürgens G (2003) The Arabidopsis GNOM ARF-GEF Mediates Endosomal Recycling, Auxin Transport, and Auxin-Dependent Plant Growth. Cell 112 (2):219-230. doi:10.1016/S0092-8674(03)00003-5

Hamada A, Nitta M, Nasuda S, Kato K, Fujita M, Matsunaka H, Okumoto Y (2012) Novel QTLs for growth angle of seminal roots in wheat (Triticum aestivum L.). Plant and soil 354 (1):395-405. doi:10.1007/s11104-011-1075-5

Haruta M, Tan LX, Bushey DB, Swanson SJ, Sussman MR (2017) Environmental and Genetic Factors Regulating Localization of the Plant Plasma Membrane H+-ATPase. Plant Physiology. doi:10.1104/pp.17.01126

Hirasawa T, Takahashi H, Suge H, Ishihara K (1997) Water potential, turgor and cell wall properties in elongating tissues of the hydrotropically bending roots of pea (Pisum sativum L.). Plant, Cell and Environment 20 (3):381-386. doi:10.1046/j.1365-3040.1997.d01-70.x

Jaffe MJ, Takahashi H, Biro RL (1985) A Pea Mutant for the Study of Hydrotropism in Roots. Science 230 (4724):445-447. doi:10.1126/science.230.4724.445

Kaneyasu T, Kobayashi A, Nakayama M, Fujii N, Takahashi H, Miyazawa Y (2007) Auxin response, but not its polar transport, plays a role in hydrotropism of Arabidopsis roots. Journal of Experimental Botany 58 (5):1143-1150. doi:10.1093/jxb/erl274

Kizis D, Pagès M (2002) Maize DRE-binding proteins DBF1 and DBF2 are involved in rab17 regulation through the drought-responsive element in an ABA-dependent pathway. The Plant Journal 30 (6):679-689. doi:10.1046/j.1365-313X.2002.01325.x

Kobayashi A, Takahashi A, Kakimoto Y, Miyazawa Y, Fujii N, Higashitani A, Takahashi H (2007) A gene essential for hydrotropism in roots. Proceedings of the National Academy of Sciences 104 (11):4724-4729. doi:10.1073/pnas.0609929104

Livak KJ, Schmittgen TD (2001) Analysis of relative gene expression data using Real-Time quantitative PCR and the 2−ΔΔCT method. Methods 25 (4):402-408. doi:http://dx.doi.org/10.1006/meth.2001.1262

Loomis WE, Ewan LM (1936) Hydrotropic Responses of Roots in Soil. Botanical Gazette 97 (4):728-743. doi:10.1086/334600

Miyazawa Y, Moriwaki T, Uchida M, Kobayashi A, Fujii N, Takahashi H (2012) Overexpression of MIZU-KUSSEI1 Enhances the Root Hydrotropic Response by Retaining Cell Viability Under Hydrostimulated Conditions in Arabidopsis thaliana. Plant and Cell Physiology 53 (11):1926-1933. doi:10.1093/pcp/pcs129

Miyazawa Y, Takahashi A, Kobayashi A, Kaneyasu T, Fujii N, Takahashi H (2009) GNOM-Mediated Vesicular Trafficking Plays an Essential Role in Hydrotropism of Arabidopsis Roots. Plant Physiology 149 (2):835-840. doi:10.1104/pp.108.131003

Moriwaki T, Miyazawa Y, Fujii N, Takahashi H (2012) Light and abscisic acid signalling are integrated by MIZ1 gene expression and regulate hydrotropic response in roots of Arabidopsis thaliana. Plant, Cell and Environment 35 (8):1359-1368. doi:10.1111/j.1365-3040.2012.02493.x

Moriwaki T, Miyazawa Y, Kobayashi A, Uchida M, Watanabe C, Fujii N, Takahashi H (2011) Hormonal Regulation of Lateral Root Development in Arabidopsis Modulated by MIZ1 and Requirement of GNOM Activity for MIZ1 Function. Plant Physiology 157 (3):1209-1220. doi:10.1104/pp.111.186270

Sievers F, Wilm A, Dineen D, Gibson TJ, Karplus K, Li W, Lopez R, McWilliam H, Remmert M, Söding J, Thompson JD, Higgins DG (2011) Fast, scalable generation of

high-quality protein multiple sequence alignments using Clustal Omega. Molecular Systems Biology 7:539. doi:10.1038/msb.2011.75

Steinmann T, Geldner N, Grebe M, Mangold S, Jackson CL, Paris S, Gälweiler L, Palme K, Jürgens G (1999) Coordinated Polar Localization of Auxin Efflux Carrier PIN1 by GNOM ARF GEF. Science 286 (5438):316-318. doi:10.1126/science.286.5438.316

Takahashi H (1997) Hydrotropism: The current state of our knowledge. Journal of Plant Research 110 (2):163. doi:10.1007/bf02509304

Takahashi H, Brown CS, Dreschel TW, Scott TK (1992) Hydrotropism in Pea Roots in a Porous-tube Water Delivery System. HortScience 27 (5):430-432

Takahashi H, Miyazawa Y, Fujii N (2009) Hormonal interactions during root tropic growth: hydrotropism versus gravitropism. Plant Molecular Biology 69 (4):489-502. doi:10.1007/s11103-008-9438-x

Takahashi H, Scott TK (1991) Hydrotropism and Its Interaction with Gravitropism in Maize Roots. Plant Physiology 96 (2):558-564. doi:10.1104/pp.96.2.558

Takahashi H, Suge H (1991) Root hydrotropism of an agravitropic pea mutant, ageotropum. Physiologia Plantarum 82 (1):24-31. doi:10.1111/j.1399-3054.1991.tb02898.x

Takahashi N, Goto N, Okada K, Takahashi H (2002) Hydrotropism in abscisic acid, wavy, and gravitropic mutants of Arabidopsis thaliana. Planta 216 (2):203-211. doi:10.1007/s00425-002-0840-3

Takano M, Takahashi H, Suge H (1997) Calcium Requirement for the Induction of Hydrotropism and Enhancement of Calcium-Induced Curvature by Water Stress in Primary Roots of Pea, Pisum sativum L. Plant and Cell Physiology 38 (4):385-391. doi:10.1093/oxfordjournals.pcp.a029180

Tamura K, Stecher G, Peterson D, Filipski A, Kumar S (2013) MEGA6: Molecular Evolutionary Genetics Analysis Version 6.0. Molecular Biology and Evolution 30 (12):2725-2729. doi:10.1093/molbev/mst197

Tsutsumi D, Kosugi Ki, Mizuyama T (2002) Effect of Hydrotropism on Root System Development in Soybean (Glycine max): Growth Experiments and a Model Simulation. Journal of Plant Growth Regulation 21 (4):441-458. doi:10.1007/s00344-003-0006-y

Yamazaki T, Miyazawa Y, Kobayashi A, Moriwaki T, Fujii N, Takahashi H (2012) MIZ1, an essential protein for root hydrotropism, is associated with the cytoplasmic face of the endoplasmic reticulum membrane in Arabidopsis root cells. FEBS Letters 586 (4):398-402. doi:10.1016/j.febslet.2012.01.008

Zheng J, Zhao J, Zhang J, Fu J, Gou M, Dong Z, Hou W, Huang Q, Wang G (2006) Comparative expression profiles of maize genes from a water stress-specific cDNA macroarray in response to high-salinity, cold or abscisic acid. Plant Science 170 (6):1125-1132. doi:https://doi.org/10.1016/j.plantsci.2006.01.019

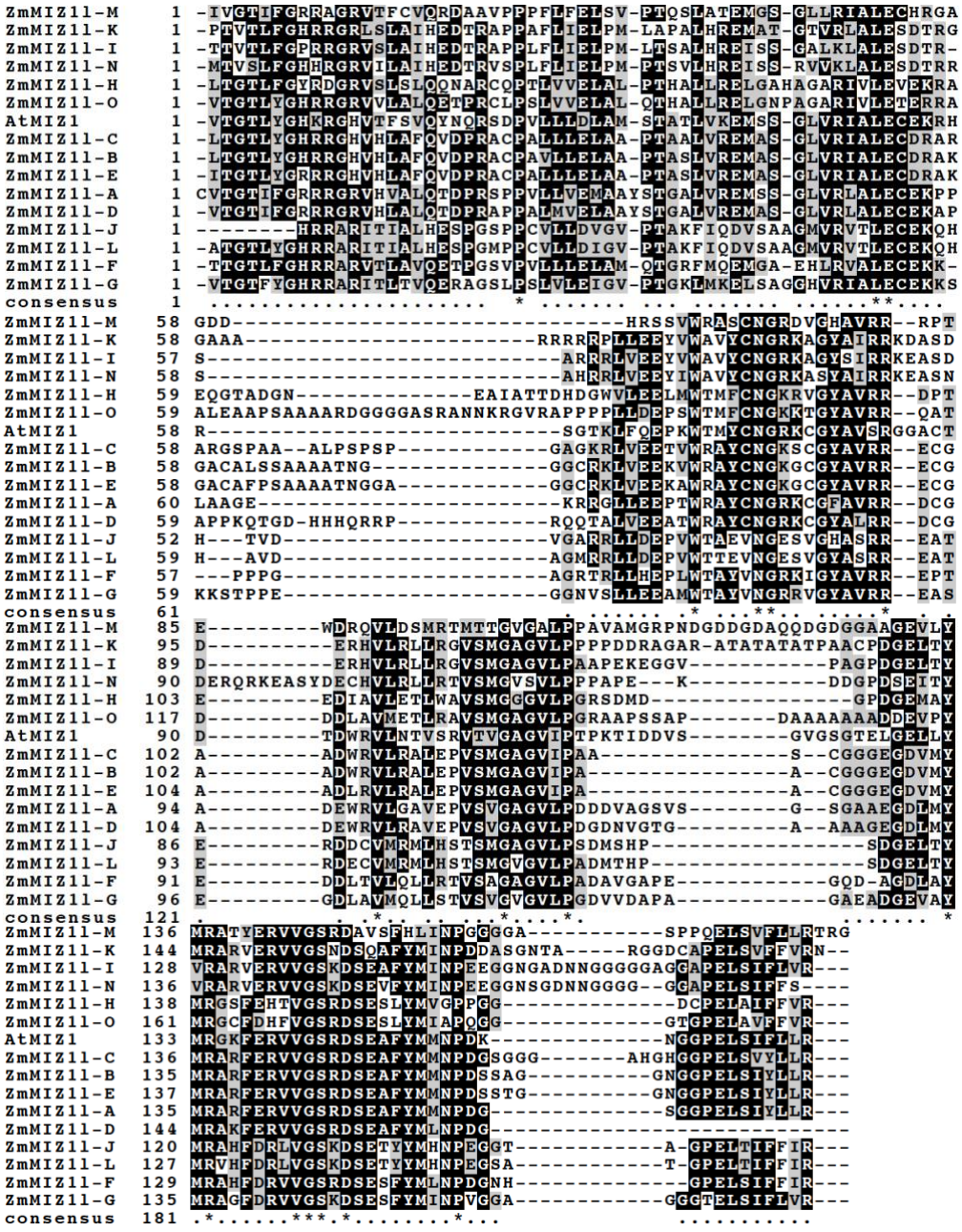


Figure 1a. **DUF617 domain alignment of 15 *Zea mays* MIZ1-like peptides with *Arabidopsis* MIZ1**. **Asterisks** (*) represent conserved amino acid residues; Dots (**•**) represent similar amino acid residues.

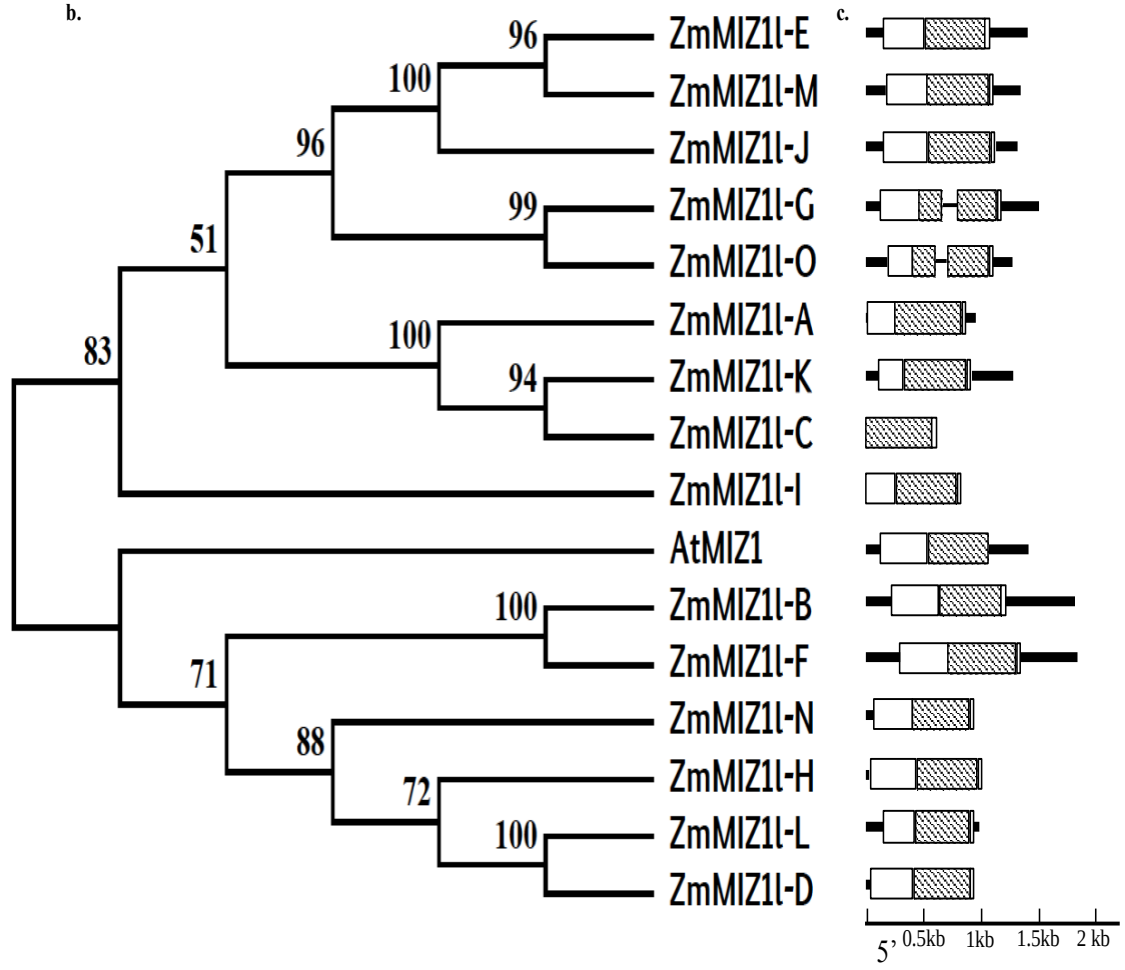


Figure 1b & c. **Phylogenetic relationship of MIZ1-like proteins and gene structure in *Zea mays*.**

b. The neighbor-joining tree includes 15 MIZ1l proteins from maize and MIZ1 protein from *Arabidopsis*. The numbers next to the branches are the bootstrap probabilities from 1,000 replications.

c. The gene structure is presented by exons (white boxes), DUF617 domain (shaded boxes), UTRs (thick black lines) and introns (thin black lines). The sizes of the structures can be estimated using the scale below.

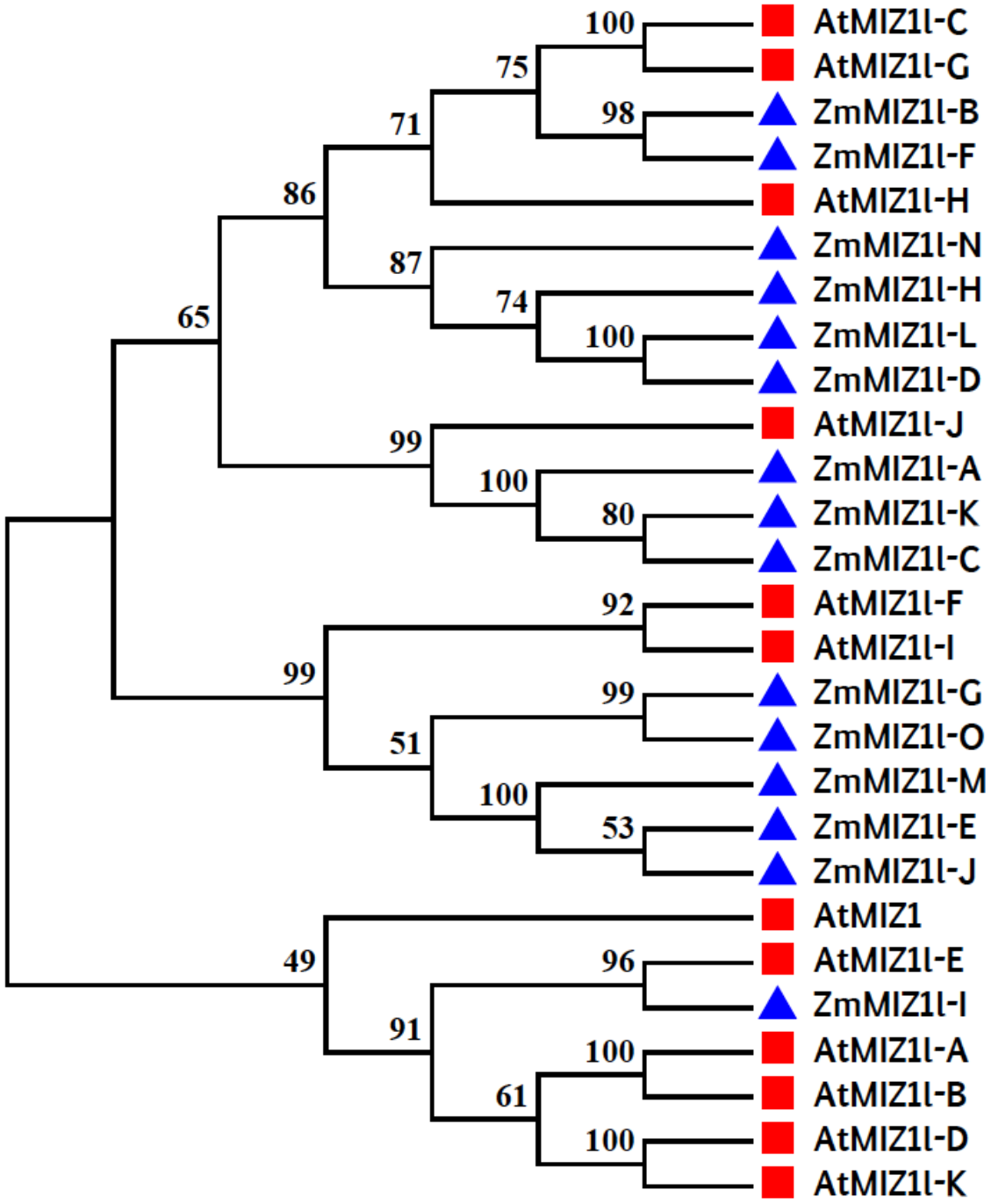


Figure 2a. **Phylogenetic relationship of MIZ1-like family in *Zea mays*** (triangles) **and *Arabidopsis*** (squares)**.** The neighbor-joining tree includes 15 MIZ1l proteins from maize and 12 MIZ1l protein from *Arabidopsis*, including AtMIZ1. The numbers shown next to the branches are the bootstrap probabilities from 1,000 replications. *Arabidopsis*

sequences include AtMIZ1 (AT2g41660), AtMIZ1l-A (At1g21050), AtMIZ1l-B (At1g76610), AtMIZ1l-C (At2g21990), AtMIZ1l-D (At2g22460), AtMIZ1l-E (At2g37880), AtMIZ1l-F (At3g25640), AtMIZ1l-G (At4g39610), AtMIZ1l-H (At5g06990), AtMIZ1l-I (At5g23100), AtMIZ1l-J (At5g42680) and AtMIZ1l-K (At5g65340). Maize sequences include ZmMIZ1l-A (Zm00001d031810), ZmMIZ1l-B (Zm00001d002136), ZmMIZ1l-C (Zm00001d005918), ZmMIZ1l-D (Zm00001d040415), ZmMIZ1l-E (Zm00001d044087), ZmMIZ1l-F (Zm00001d049489), ZmMIZ1l-G (Zm00001d051598), ZmMIZ1l-H (Zm00001d013011), ZmMIZ1l-I (Zm00001d031810), ZmMIZ1l-J (Zm00001d039108), ZmMIZ1l-K (Zm00001d020757), ZmMIZ1l-L (Zm00001d008272), ZmMIZ1l-M (Zm00001d011463), ZmMIZ1l-N (Zm00001d012555) and ZmMIZ1l-O (Zm00001d045946).

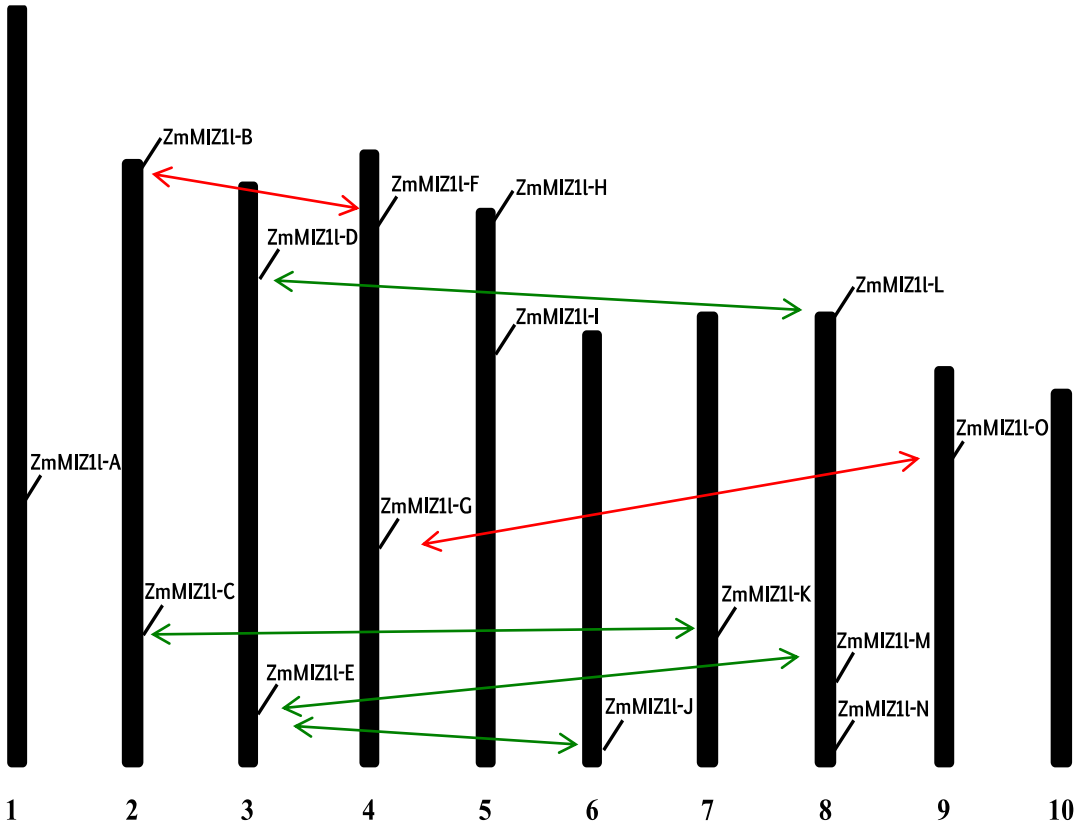


Figure 2b. **Chromosomal locations of maize *MIZ1-like* genes**. The chromosomal position of each *ZmMIZ1l* homolog are placed on the maize physical map. The chromosome number is indicated below each chromosome. The arrows connect the genes with the close homology. Green arrows indicate that these are in regions of known duplications. Red arrows indicate that the genes are not in in duplicated regions.

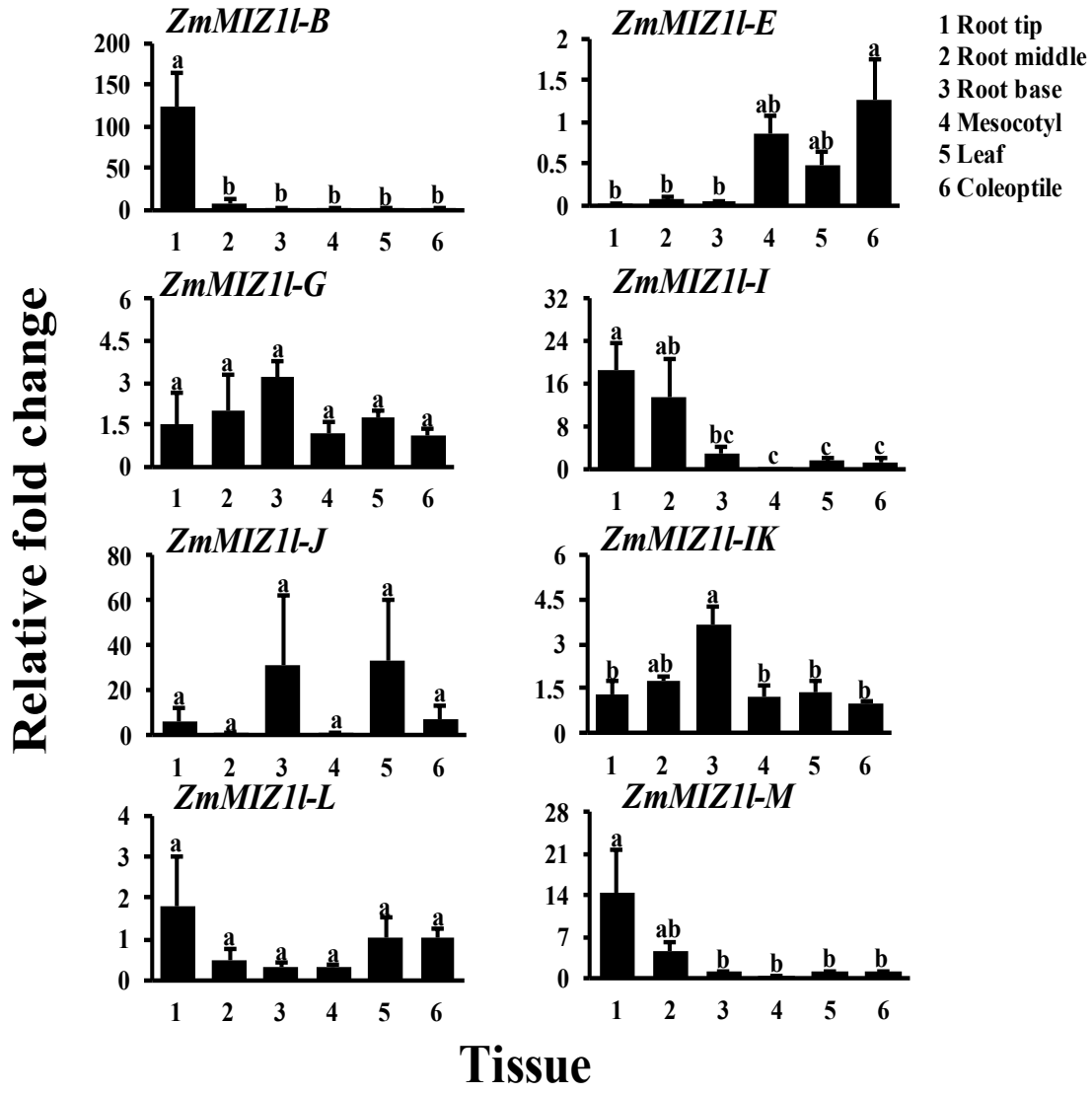


Figure 3**. Expression analysis of *ZmMIZ1-like* genes in different tissues.** Six different tissues, root tips, root middle, root base, mesocotyl, leaves and coleoptiles were harvested from maize seedlings when the roots were approximately 1.5cm long for gene expression analysis. The values represent the mean fold change ± SE (n = 3). Bars with different letters are significantly different ($p < 0.05$).

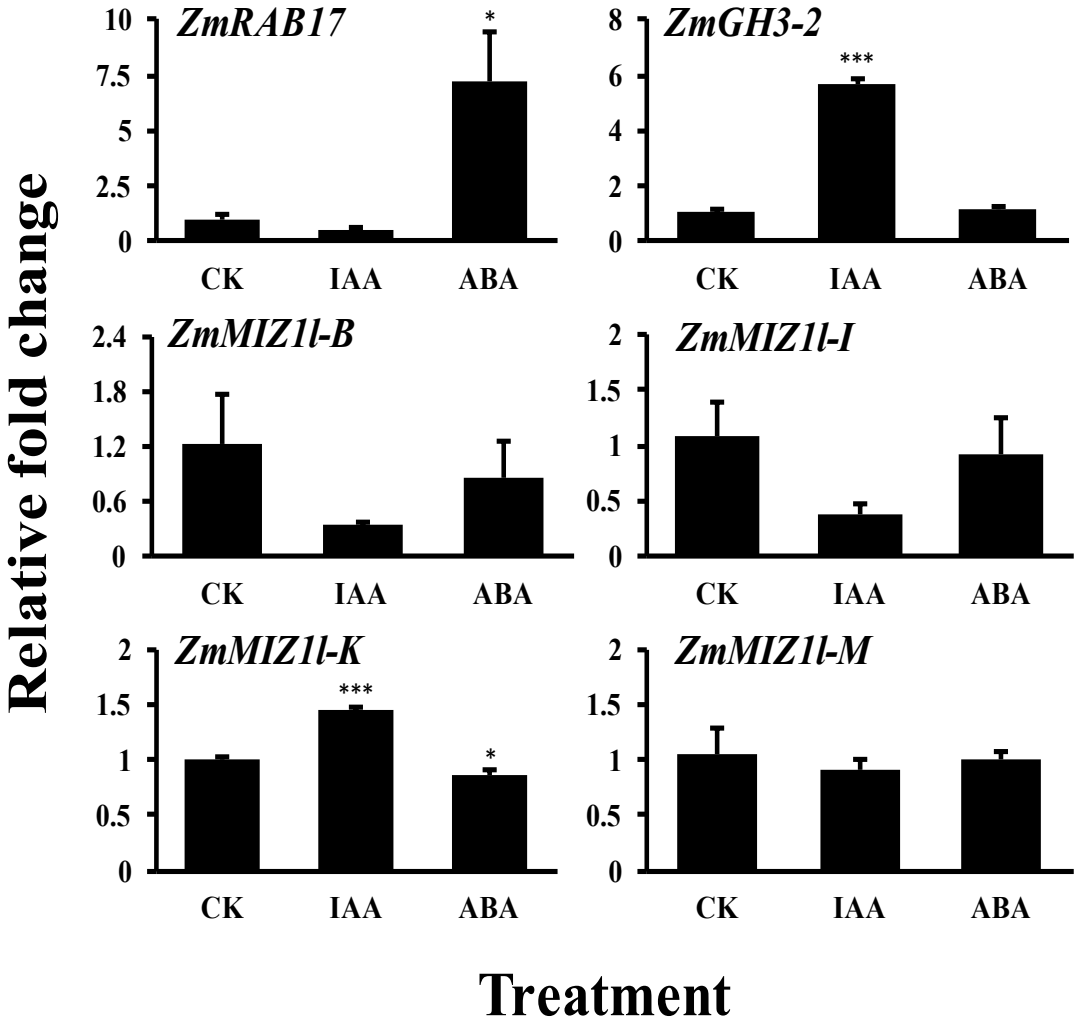


Figure 4. **Effect of plant hormones IAA and ABA on the transcript accumulation of *ZmMIZ1-like* genes**. Maize seedlings were treated with 10 μM solutions of IAA, ABA or pure water (control samples) for 1 hour and 6 mm root tips were harvested. On X-axes, CK represents the control roots, IAA and ABA represents the root samples treated with the hormones. The values on Y-axes represent the mean fold change ± SE (n = 3) relative to CK. Bars with asterisks are significantly different from control samples ($p < 0.05$).

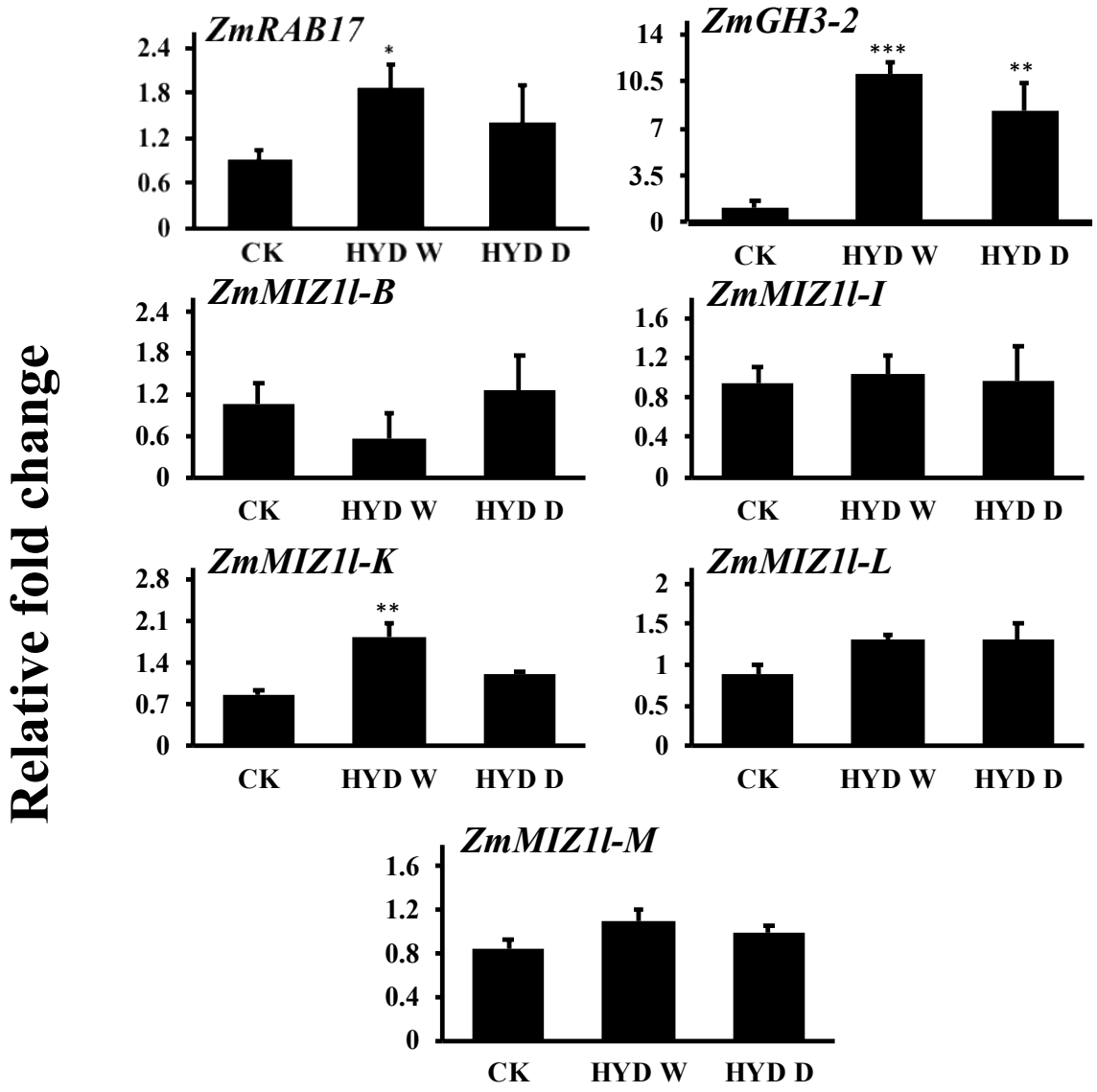
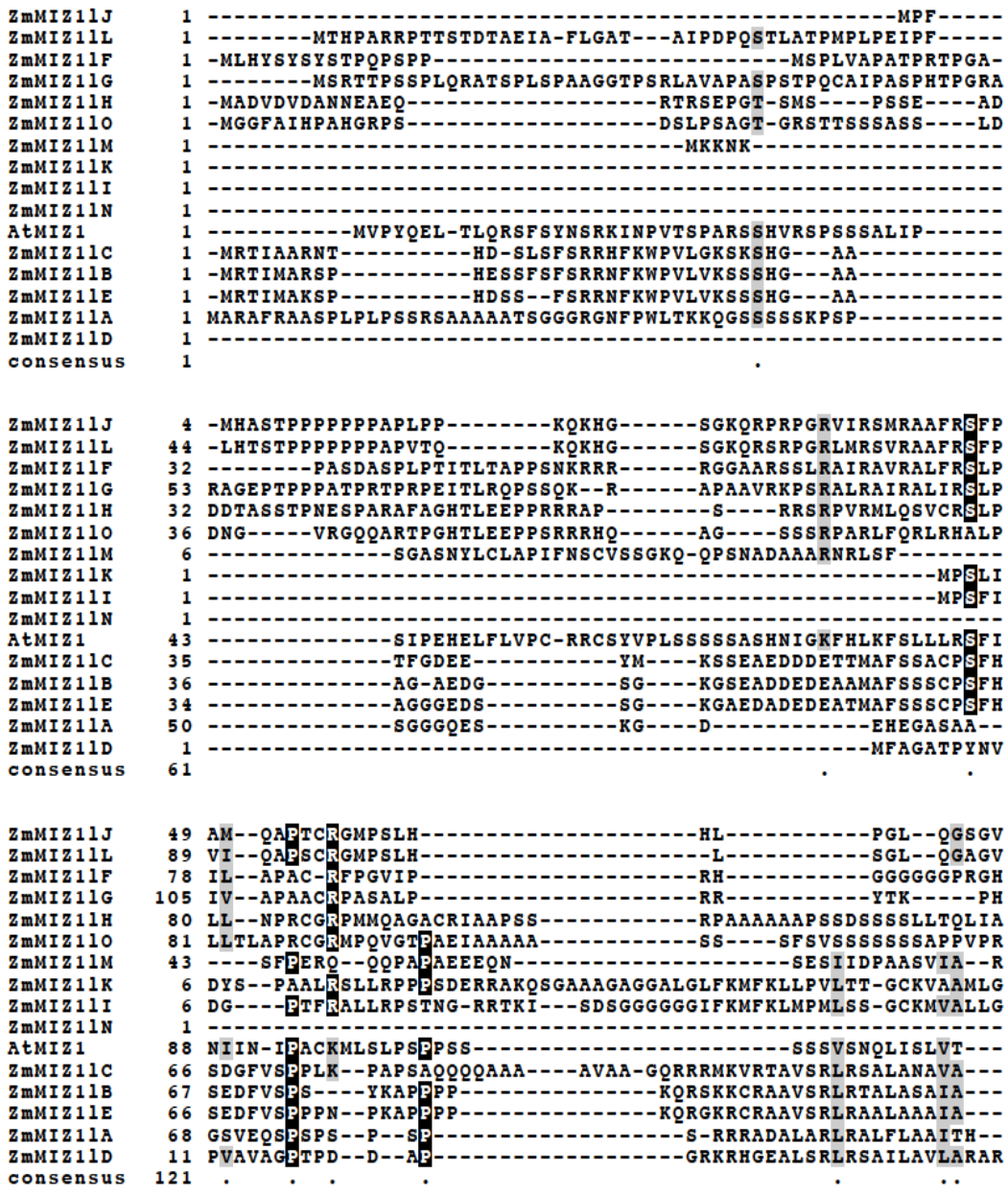
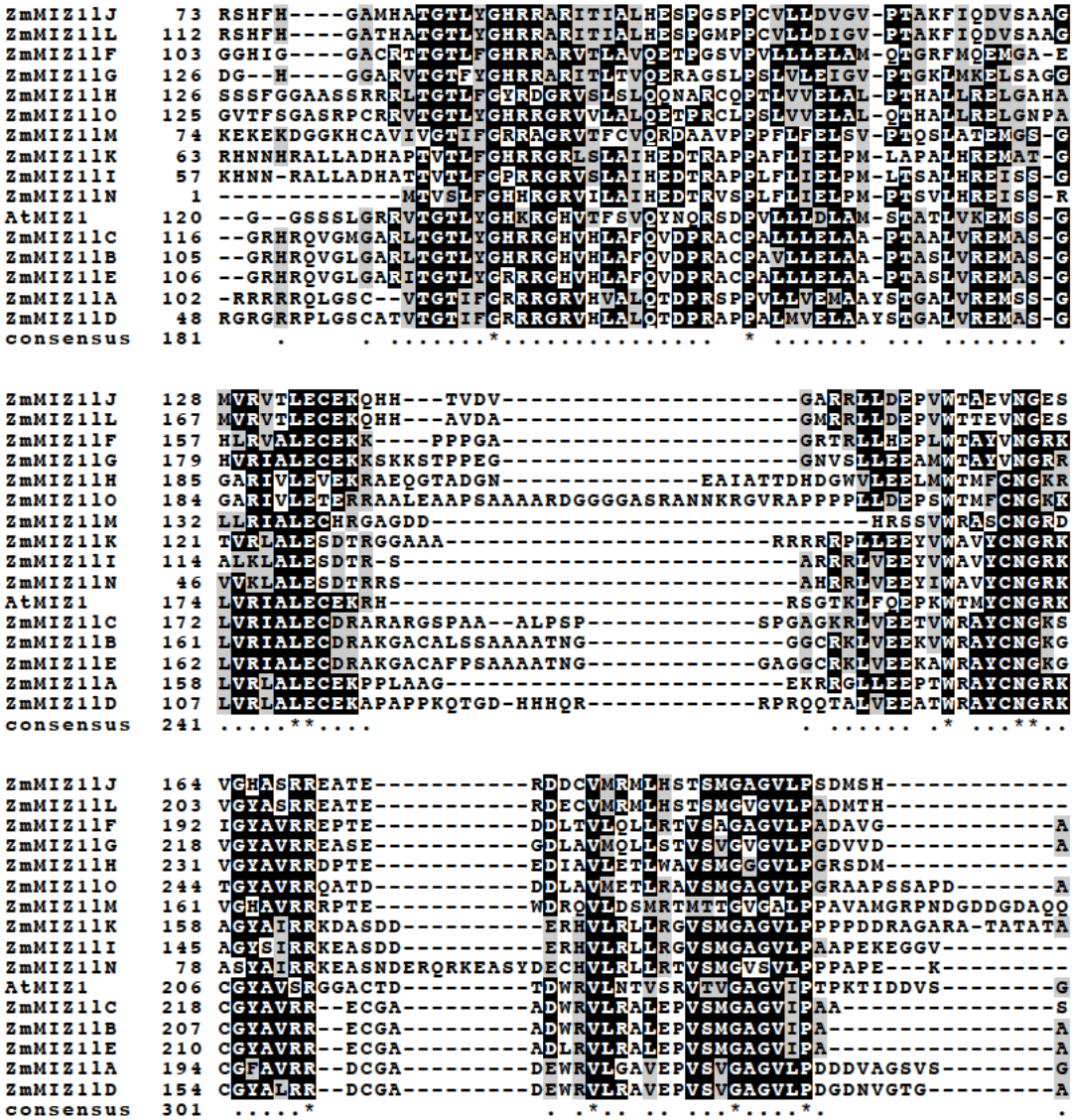


Figure 5. **Response of *ZmMIZ1-like* genes to well-watered vs. hydrotropic (selective moisture gradient) conditions.** Maize seedlings were grown in either well-watered or hydrotropic conditions for 1.5 hours and 6 mm root tips were harvested at the end of the treatment. On X-axes CK represents the control roots grown in well-watered conditions. The roots from hydrotropic conditions were sliced longitudinally; HYD W represents the half oriented toward moist conditions and HYD D represents the half exposed to dry conditions. The values on Y-axes represent the mean fold change ± SE (n = 3) relative to CK. Bars with asterisks are significantly different from the control samples ($p < 0.05$).

Table S1. Gene specific primer sequences used for qPCR analysis and their corresponding length, product length and melting temperature (Tm).

Primers	RT-Primer Sequences	Length (bases)	Product size (bp)	Tm(°C)
ZmMIZ1l-A_F	AAGCCGATGCCATCTCTTATCGAC	24	147	58.6
ZmMIZ1l-A_R	GAGCTTGAACATCTTGAACAGCCC	24		58.1
ZmMIZ1l-B_F	TGCATGAACCGATCCAAACGATGC	24	123	60
ZmMIZ1l-B_R	AGCTATGCTCTCGTGACTGACACT	24		59.3
ZmMIZ1l-C_F	CTCCTACGATGAGTGCCAC	19	199	61.4
ZmMIZ1l-C_R	GCTGTTTCCACCTTCCTCTG	20		62.5
ZmMIZ1l-D_F	CATTTTGATCGCCTCGTT	18	102	59.9
ZmMIZ1l-D_R	TCTAATGAAGAAAATTGTGAGC	22		58.8
ZmMIZ1l-E_F	CGACGGTGTTGTGGGATTAATGCA	24	112	60.5
ZmMIZ1l-E_R	TGCATTAATCCCACAACACCGTCG	24		59.5
ZmMIZ1l-F_F	CAATCCCACAACGAGACAAAC	21	164	61.6
ZmMIZ1l-F_R	CGCTAGTTATCCCCTGCAC	19		61.8
ZmMIZ1l-G_F	GTGTTTCCAAGTGTGTGCCCAAGT	24	101	60
ZmMIZ1l-G_R	AGAGGCACAGCCCAATAATCACCT	24		60.1
ZmMIZ1l-H_F	CCACGGGGAAGCTAATGAAG	24	127	53.8
ZmMIZ1l-H_R	ATCGCCTCCTCCAGCAGACTCA	21		59.8
ZmMIZ1l-I_F	CAACGACGGTGATGATGGAGACG	23	134	61.1
ZmMIZ1l-I_R	CGGGGTTGATGAGGTGGAATGACA	24		63.5
ZmMIZ1l-J_F	TCGCGACTCGGAGGCCTTCTACAT	24	195	63.9
ZmMIZ1l-J_R	TGCCTACTAATCCATTTTCACAGC	24		54.4
ZmMIZ1l-K_F	TACACACACCCTTGTGCCTTGTTG	24	138	59.5
ZmMIZ1l-K_R	CCCGGATTTGATCTCAGCACGATT	24		59
ZmMIZ1l-L_F	ATCCAACCGCCGCATGCCAA	20	129	62.8
ZmMIZ1l-L_R	TGGAAGGCGATCTCGATGGT	20		58.1
ZmMIZ1l-M_F	TTCTGTGGACTGTGGATGTG	20	162	61.8
ZmMIZ1l-M_R	CTCATGGTTCCTGGTGTGG	19		62
ZmMIZ1l-N_F	CCCCCTGCTCCGATTCACCT	20	196	59.7
ZmMIZ1l-N_R	CGCCTCCGCTTGTTG	15		48.3
ZmMIZ1l-O_F	ATATGCTGTTCGTTTGCGTGCC	22	104	58.5
ZmMIZ1l-O_R	ACTGCATGCACAATGCAGCCAA	22		60.5
ZmUBCP_F	CAGGTGGGGTATTCTTGGTG	20	97	55
ZmUBCP_R	ATGTTCGGGTGGAAAACCTT	20		54


ZmMIZ1lJ 1 ----------MPF-----
ZmMIZ1lL 1 --------MTHPARRPTTSTDTAEIA-FLGAT---AIPDPQSTLATPMPLPEIPF-----
ZmMIZ1lF 1 -MLHYSYSYSTPQPSPP----------MSPLVAPATPRTPGA-
ZmMIZ1lG 1 --------MSRTTPSSPLQRATSPLSPAAGGTPSRLAVAPASPSTPQCAIPASPHTPGRA
ZmMIZ1lH 1 -MADVDVDANNEAEQ----------RTRSEPGT-SMS----PSSE----AD
ZmMIZ1lO 1 -MGGFAIHPAHGRPS----------DSLPSAGT-GRSTTSSSASS----LD
ZmMIZ1lM 1 ----------MKKNK----------
ZmMIZ1lK 1 ----------
ZmMIZ1lI 1 ----------
ZmMIZ1lN 1 ----------
AtMIZ1 1 -----------MVPYQEL-TLQRSFSYNSRKINPVTSPARSSHVRSPSSSALIP------
ZmMIZ1lC 1 -MRTIAARNT----------HD-SLSFSRRHFKWPVLGKSKSHG---AA----------
ZmMIZ1lB 1 -MRTIMARSP----------HESSFSFSRRNFKWPVLVKSSSHG---AA----------
ZmMIZ1lE 1 -MRTIMAKSP----------HDSS--FSRRNFKWPVLVKSSSHG---AA----------
ZmMIZ1lA 1 MARAFRAASPLPLPSSRSAAAAATSGGGRGNFPWLTKKQGSSSSSKPSP----------
ZmMIZ1lD 1 ----------
consensus 1 .
ZmMIZ1lJ 4 -MHASTPPPPPPPAPLPP--------KQKHG------SGKQRPRPGRVIRSMRAAFRSFP
ZmMIZ1lL 44 -LHTSTPPPPPPPAPVTQ--------KQKHG------SGKQRSRPGRLMRSVRAAFRSFP
ZmMIZ1lF 32 --------PASDASPLPTITLTAPPSNKRRR------RGGAARSSLRAIRAVRALFRSLP
ZmMIZ1lG 53 RAGEPTPPPATPRTPRPEITLRQPSSQK--R------APAAVRKPSRALRAIRALIRSLP
ZmMIZ1lH 32 DDTASSTPNESPARAFAGHTLEEPPRRRAP--------S----RRSRPVRMLQSVCRSLP
ZmMIZ1lO 36 DNG-----VRGQQARTPGHTLEEPPSRRRHQ------AG----SSSRPARLFQRLRHALP
ZmMIZ1lM 6 --------------SGASNYLCLAPIFNSCVSSGKQ-QPSNADAAARNRLSF---------
ZmMIZ1lK 1 ----------MPSLI
ZmMIZ1lI 1 ----------MPSFI
ZmMIZ1lN 1 ----------
AtMIZ1 43 --------------SIPEHELFLVPC-RRCSYVPLSSSSSASHNIGKFHLKFSLLLRSFI
ZmMIZ1lC 35 --------------TFGDEE-----------YM----KSSEAEDDDETTMAFSSACPSFH
ZmMIZ1lB 36 --------------AG-AEDG----------SG----KGSEADDEDEAAMAFSSSCPSFH
ZmMIZ1lE 34 --------------AGGGEDS----------SG----KGAEDADEDEATMAFSSSCPSFH
ZmMIZ1lA 50 --------------SGGGQES----------KG----D----------EHEGASAA--
ZmMIZ1lD 1 ----------MFAGATPYNV
consensus 61 . .
ZmMIZ1lJ 49 AM--QAPTCRGMPSLH----------HL----------PGL--QGSGV
ZmMIZ1lL 89 VI--QAPSCRGMPSLH----------L----------SGL--QGAGV
ZmMIZ1lF 78 IL--APAC-RFPGVIP----------RH----------GGGGGGPRGH
ZmMIZ1lG 105 IV--APAACRPASALP----------RR----------YTK-----PH
ZmMIZ1lH 80 LL--NPRCGRPMMQAGACRIAAPSS----------RPAAAAAAPSSDSSSSLLTQLIA
ZmMIZ1lO 81 LLTLAPRCGRMPQVGTPAEIAAAAA----------SS----SFSVSSSSSSSAPPVPR
ZmMIZ1lM 43 ----SFPERQ--QQPAPAEEEQN----------SESIIDPAASVIA--R
ZmMIZ1lK 6 DYS--PAALRSLLRPPPSDERRAKQSGAAAGAGGALGLFKMFKLLPVLTT-GCKVAAMLG
ZmMIZ1lI 6 DG----PTFRALLRPSTNG-RRTKI---SDSGGGGGGIFKMFKLMPMLSS-GCKMVALLG
ZmMIZ1lN 1 ----------
AtMIZ1 88 NIIN-IPACKMLSLPSPPSS----------SSSVSNQLISLVT---
ZmMIZ1lC 66 SDGFVSPPLK--PAPSAQQQQAAA----AVAA-GQRRRMKVRTAVSRLRSALANAVA---
ZmMIZ1lB 67 SEDFVSPS----YKAPPPP----------KQRSKKCRAAVSRLRTALASAIA---
ZmMIZ1lE 66 SEDFVSPPPN--PKAPPPP----------KQRGKRCRAAVSRLRAALAAAIA---
ZmMIZ1lA 68 GSVEQSPSPS--P--SP----------S-RRRADALARLRALFLAAITH--
ZmMIZ1lD 11 PVAVAGPTPD--D--AP----------GRKRHGEALSRLRSAILAVLARAR
consensus 121 . . . . . .


ZmMIZ1lJ 73 RSHFH----GAMHATGTLYGHRRARITIALHESPGSPPCVLLDVGV-PTAKFIQDVSAAG
ZmMIZ1lL 112 RSHFH----GATHATGTLYGHRRARITIALHESPGMPPCVLLDIGV-PTAKFIQDVSAAG
ZmMIZ1lF 103 GGHIG----GACRTTGTLFGHRRARVTLAVQETPGSVPVLLLELAM-QTGRFMQEMGA-E
ZmMIZ1lG 126 DG--H----GGARVTGTFYGHRRARITLTVQERAGSLPSLVLEIGV-PTGKLMKELSAGG
ZmMIZ1lH 126 SSSFGGAASSRRRLTGTLFGYRDGRVSLSLQQNARCQPTLVVELAL-PTHALLRELGAHA
ZmMIZ1lO 125 GVTFSGASRPCRRVTGTLYGHRRGRVVLALQETPRCLPSLVVELAL-QTHALLRELGNPA
ZmMIZ1lM 74 KEKEKDGGKHCAVIVGTIFGRRAGRVTFCVQRDAAVPPPFLFELSV-PTQSLATEMGS-G
ZmMIZ1lK 63 RHNNHRALLADHAPTVTLFGHRRGRLSLAIHEDTRAPPAFLIELPM-LAPALHREMAT-G
ZmMIZ1lI 57 KHNN-RALLADHATTVTLFGPRRGRVSLAIHEDTRAPPLFLIELPM-LTSALHREISS-G
ZmMIZ1lN 1 ----------MTVSLFGHHRGRVILAIHEDTRVSPLFLIELPM-PTSVLHREISS-R
AtMIZ1 120 --G--GSSSLGRRVTGTLYGHKRGHVTFSVQYNQRSDPVLLLDLAM-STATLVKEMSS-G
ZmMIZ1lC 116 --GRHRQVGMGARLTGTLYGHRRGHVHLAFQVDPRACPALLLELAA-PTAALVREMAS-G
ZmMIZ1lB 105 --GRHRQVGLGARLTGTLYGHRRGHVHLAFQVDPRACPAVLLELAA-PTASLVREMAS-G
ZmMIZ1lE 106 --GRHRQVGLGARITGTLYGRRRGHVHLAFQVDPRACPALLLELAA-PTASLVREMAS-G
ZmMIZ1lA 102 -RRRRRQLGSC--VTGTIFGRRRGRVHVALQTDPRSPPVLLVEMAAYSTGALVREMSS-G
ZmMIZ1lD 48 RGRGRRPLGSCATVTGTIFGRRRGRVHLALQTDPRAPPALMVELAAYSTGALVREMAS-G
consensus 181 . . .. * ....
ZmMIZ1lJ 128 MVRVTLECEKQHH---TVDV----------GARRLLDEPVWTAEVNGES
ZmMIZ1lL 167 MVRVTLECEKQHH---AVDA----------GMRRLLDEPVWTTEVNGES
ZmMIZ1lF 157 HLRVALECEKK----PPPGA----------GRTRLLHEPLWTAYVNGRK
ZmMIZ1lG 179 HVRIALECEKKSKKSTPPEG----------GNVSLLEEAMWTAYVNGRR
ZmMIZ1lH 185 GARIVLEVEKRAEQGTADGN----------EAIATTDHDGWVLEELMWTMFCNGKR
ZmMIZ1lO 184 GARIVLETERRAALEAAPSAAAARDGGGGASRANNKRGVRAPPPPLLDEPSWTMFCNGKK
ZmMIZ1lM 132 LLRIALECHRGAGDD----------HRSSVWRASCNGRD
ZmMIZ1lK 121 TVRLALESDTRGGAAA----------RRRRRPLLEEYVWAVYCNGRK
ZmMIZ1lI 114 ALKLALESDTR-S----------ARRRLVEEYVWAVYCNGRK
ZmMIZ1lN 46 VVKLALESDTRRS----------AHRRLVEEYIWAVYCNGRK
AtMIZ1 174 LVRIALECEKRH----------RSGTKLFQEPKWTMYCNGRK
ZmMIZ1lC 172 LVRIALECDRARARGSPAA--ALPSP----------SPGAGKRLVEETVWRAYCNGKS
ZmMIZ1lB 161 LVRIALECDRAKGACALSSAAAATNG----------GGCRKLVEEKVWRAYCNGKG
ZmMIZ1lE 162 LVRIALECDRAKGACAFPSAAAATNG----------GAGGCRKLVEEKAWRAYCNGKG
ZmMIZ1lA 158 LVRLALECEKPPLAAG----------EKRRGLLEEPTWRAYCNGRK
ZmMIZ1lD 107 LVRLALECEKAPAPPKQTGD-HHHQR----------RPRQQTALVEEATWRAYCNGRK
consensus 241 .....**.... . ...* ...**..
ZmMIZ1lJ 164 VGHASRREATE----------RDDCVMRMLHSTSMGAGVLPSDMSH----------
ZmMIZ1lL 203 VGYASRREATE----------RDECVMRMLHSTSMGVGVLPADMTH----------
ZmMIZ1lF 192 IGYAVRREPTE----------DDLTVLQLLRTVSAGAGVLPADAVG----------A
ZmMIZ1lG 218 VGYAVRREASE----------GDLAVMQLLSTVSVGVGVLPGDVVD----------A
ZmMIZ1lH 231 VGYAVRRDPTE----------EDIAVLETLWAVSMGGGVLPGRSDM----------
ZmMIZ1lO 244 TGYAVRRQATD----------DDLAVMETLRAVSMGAGVLPGRAAPSSAPD-------A
ZmMIZ1lM 161 VGHAVRRRPTE----------WDRQVLDSMRTMTTGVGALPPAVAMGRPNDGDDGDAQQ
ZmMIZ1lK 158 AGYAIRRKDASDD----------ERHVLRLLRGVSMGAGVLPPPDDRAGARA-TATATA
ZmMIZ1lI 145 AGYSIRRKEASDD----------ERHVLRLLRGVSMGAGVLPAAPEKEGGV----------
ZmMIZ1lN 78 ASYAIRRKEASNDERQRKEASYDECHVLRLLRTVSMGVSVLPPPAPE---K----------
AtMIZ1 206 CGYAVSRGGACTD----------TDWRVLNTVSRVTVGAGVIPTPKTIDDVS---------G
ZmMIZ1lC 218 CGYAVRR---ECGA----------ADWRVLRALEPVSMGAGVIPAA----------S
ZmMIZ1lB 207 CGYAVRR---ECGA----------ADWRVLRALEPVSMGAGVIPA----------A
ZmMIZ1lE 210 CGYAVRR---ECGA----------ADLRVLRALEPVSMGAGVIPA----------A
ZmMIZ1lA 194 CGFAVRR---DCGA----------DEWRVLGAVEPVSVGAGVLPDDDVAGSVS---------G
ZmMIZ1lD 154 CGYALRR---DCGA----------DEWRVLRAVEPVSVGAGVLPDGDNVGTG----------A
consensus 301 .....* . .*.. .. ...*....*. .

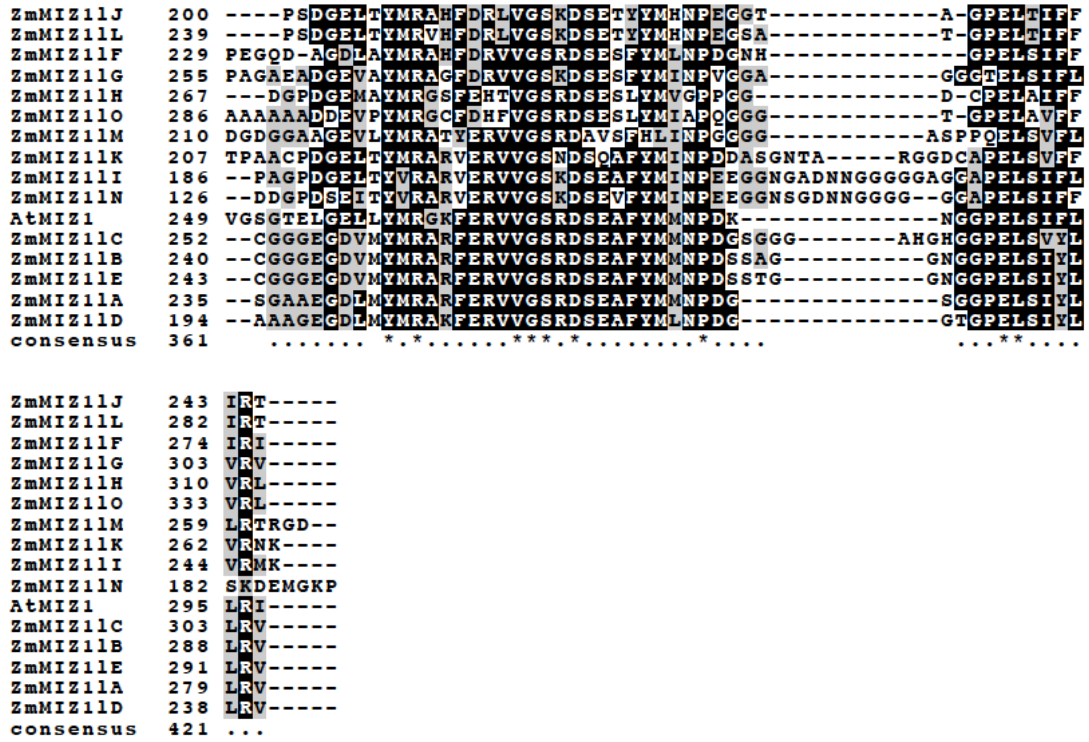


Figure S1. **Full length protein sequence alignment of MIZ1ls from *Zea mays* and *Arabidopsis* MIZ1. *** represents conserved amino acid residues; • represents the identical and similar amino acid residues. *Arabidopsis* sequences include AtMIZ1 (AT2g41660). Maize sequences include ZmMIZ1l-A (Zm00001d031810), ZmMIZ1l-B (Zm00001d002136), ZmMIZ1l-C (Zm00001d005918), ZmMIZ1l-D (Zm00001d040415), ZmMIZ1l-E (Zm00001d044087), ZmMIZ1l-F (Zm00001d049489), ZmMIZ1l-G (Zm00001d051598), ZmMIZ1l-H (Zm00001d013011), ZmMIZ1l-I (Zm00001d031810), ZmMIZ1l-J (Zm00001d039108), ZmMIZ1l-K (Zm00001d020757), ZmMIZ1l-L (Zm00001d008272), ZmMIZ1l-M (Zm00001d011463), ZmMIZ1l-N (Zm00001d012555) and ZmMIZ1l-O (Zm00001d045946).

Table S2. List of putative *MIZ1-like* homologues in soybean (*Glycine max*), rice (*Oryza sativa*) and sorghum (*Sorghum bicolor*). The score, E-value and alignment length is with respect to Arabidopsis MIZ1 protein sequence, that was used to perform BLASTP searchers of the plant genomes.

Glycine max				
Locus ID		**Score**	**E-value**	**Allignment length**
Glyma.10G113500.1 (Glyma10g13230)	GmMIZ1lA	241.5	4.90E-77	78-297
Glyma.02G185400.1 (Glyma02g33055)	GmMIZ1lB	231.5	4.60E-73	78-297
Glyma.01G090700.1 (Glyma01g22800)	GmMIZ1lC	198.7	1.50E-60	122-297
Glyma.04G145100.1	GmMIZ1lD	195.7	2.00E-59	125-297
Glyma.02G131200.1 (Glyma02g14810)	GmMIZ1lE	194.9	4.40E-59	120-297
Glyma.17G123500.1 (Glyma17g13250)	GmMIZ1lF	190.3	2.30E-57	125-297
Glyma.06G209100.1	GmMIZ1lG	189.5	4.30E-57	128-297
Glyma.05G015000.1 (Glyma05g07760)	GmMIZ1lH	185.7	1.10E-55	125-297
Glyma.06G146300.1 (Glyma06g15130)	GmMIZ1lI	179.5	1.30E-53	128-297
Glyma.04G036800.1 (Glyma04g03920)	GmMIZ1lJ	153.3	8.10E-44	130-296
Glyma.15G229300.1 (Glyma15g36740)	GmMIZ1lK	154.1	9.10E-44	126-297
Glyma.06G036900.1 (Glyma06g04020)	GmMIZ1lL	143.7	3.40E-40	130-296
Glyma.05G175200.1 (Glyma05g30890)	GmMIZ1lM	128.6	2.40E-34	129-297
Glyma.07G138800.1 (Glyma07g16680)	GmMIZ1lN	124.4	2.00E-32	129-296
Glyma.08G132500.1 (Glyma08g14090)	GmMIZ1lO	122.9	3.30E-32	129-296
Glyma.18G014300.1 (Glyma18g01770)	GmMIZ1lP	121.7	4.70E-32	129-296
Glyma.11G242700.1 (Glyma11g37870)	GmMIZ1lQ	118.2	1.50E-30	134-296
Glyma.18G189100.1 (Glyma18g41230)	GmMIZ1lR	111.7	3.50E-28	129-289
Oryza sativa				
Locus ID		**Score**	**E-value**	**Allignment length**
LOC_Os02g47980.1 DUF617 domain containing protein, expressed	OsMIZ1lA	192.2	2.60E-58	125-297
LOC_Os05g50160.1 DUF617 domain containing protein, expressed	OsMIZ1lB	182.6	1.50E-54	125-297
LOC_Os01g45510.1 DUF617 domain containing protein, expressed	OsMIZ1lC	179.1	4.70E-53	125-297
LOC_Os06g30030.1 DUF617 domain containing protein, expressed	OsMIZ1lD	172.6	1.40E-50	129-297
LOC_Os01g59200.1 DUF617 domain containing protein, expressed	OsMIZ1lE	171	3.50E-50	121-297
LOC_Os09g28880.1 DUF617 domain containing protein, expressed	OsMIZ1lF	146.7	1.80E-41	130-297
LOC_Os08g37150.1 DUF617 domain containing protein, expressed	OsMIZ1lG	146.4	3.50E-41	130-296
LOC_Os03g59690.1 DUF617 domain containing protein, expressed	OsMIZ1lH	144.1	1.00E-39	121-297
LOC_Os08g07500.1 DUF617 domain containing protein, expressed	OsMIZ1lI	137.9	2.10E-37	121-297
LOC_Os01g16320.1 plant-specific domain TIGR01570 family protein, expressed	OsMIZ1lJ	129.4	1.20E-33	128-296
LOC_Os04g54600.1 DUF617 domain containing protein, expressed	OsMIZ1lK	126.7	1.90E-33	127-297
LOC_Os05g20030.1 DUF617 domain containing protein, expressed	OsMIZ1lL	118.2	1.10E-30	129-296
LOC_Os03g52290.1 DUF617 domain containing protein, expressed	OsMIZ1lM	86.7	9.50E-20	119-297
Sorghum bicolor				
Locus ID		**Score**	**E-value**	**Allignment length**
Sobic.004G264500.1 (Sb04g030260) similar to Putative uncharacterized protein	SbMIZ1lA	197.2	2.70E-60	125-297
Sobic.010G138300.1 (Sb10g013960) similar to Putative uncharacterized protein B1089G05.18	SbMIZ1lB	183.7	1.10E-54	129-297
Sobic.009G242501.1 (Sb09g029410) similar to Os01g0642600 protein	SbMIZ1lC	178.7	8.00E-53	125-297
Sobic.003G329100.1 (Sb03g037520) similar to Putative uncharacterized protein P0702B09.7	SbMIZ1lD	177.6	1.20E-52	128-297
Sobic.001G041300.1 (Sb01g003810) similar to Putative uncharacterized protein	SbMIZ1lE	157.5	8.90E-45	120-297
Sobic.002G229800.2	SbsMIZ1lF	139	1.40E-37	130-294
Sobic.007G158400.1 (Sb07g023190) similar to Putative uncharacterized protein	SbMIZ1lG	136	3.90E-37	130-296
Sobic.006G234400.1 (Sb06g030190) similar to OSJNBb0034I13.25 protein	SbMIZ1lH	132.9	2.30E-35	120-297
Sobic.001G287700.1 (Sb01g028240) similar to Putative uncharacterized protein P0667A10.4	SbMIZ1lI	125.6	4.70E-33	129-296
Sobic.010G251800.1 (Sb10g029010) similar to Putative uncharacterized protein	SbMIZ1lJ	110.9	6.00E-28	129-296
Sobic.001G102100.1 (Sb01g009110) similar to Putative uncharacterized protein OSJNBa0079G12.6	SbMIZ1lK	61.6	5.40E-11	123-279

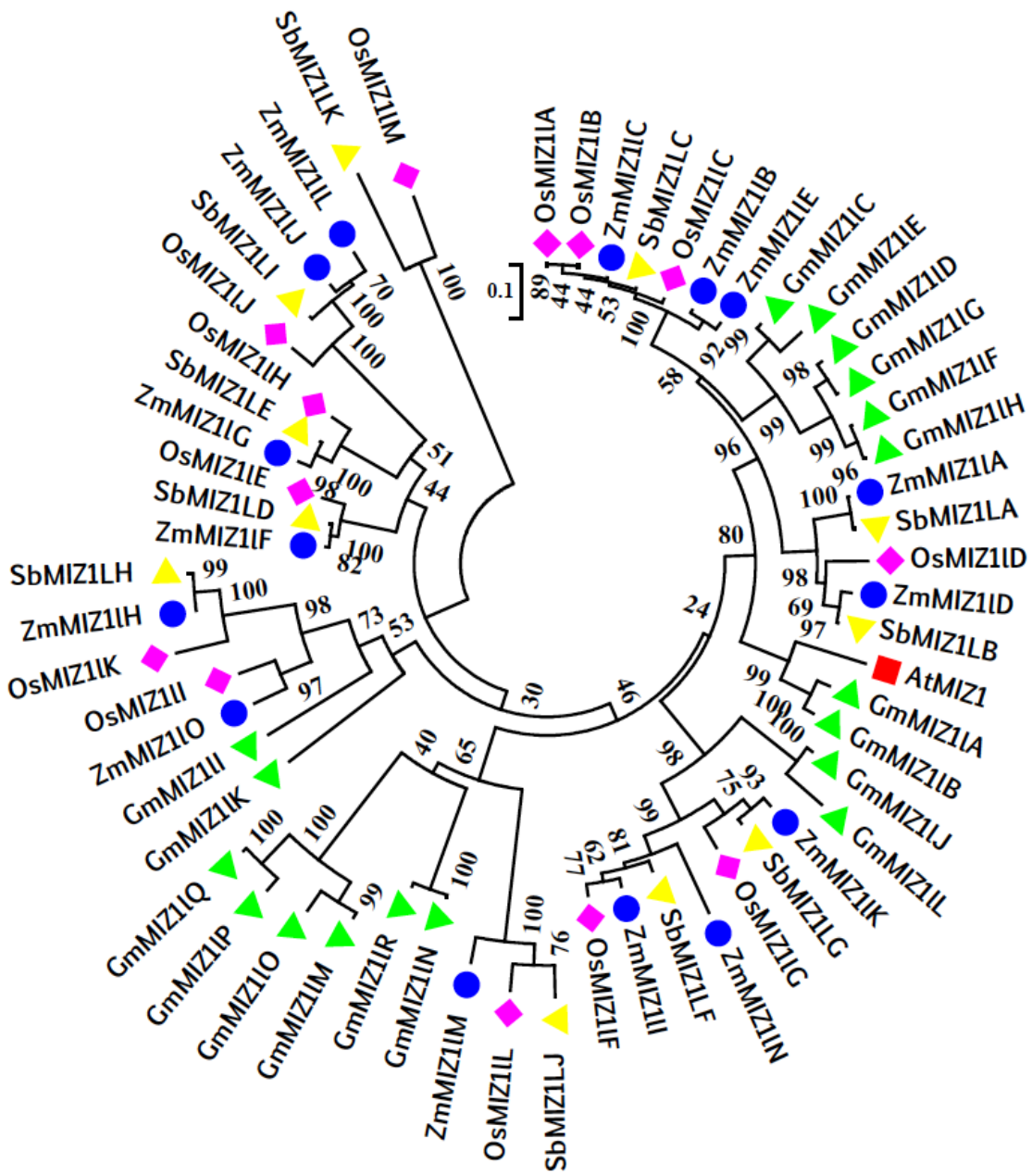


Figure S2. **Phylogenetic relationship of MIZ1-like gene family in *Zea mays*** (blue circles), ***Glycine max*** (green triangles), ***Oryza sativa*** (pink rhombuses), ***Sorghum bicolor*** (yellow triangles) **with *Arabidopsis* MIZ1** (red square)**.** The numbers shown next to the branches are the bootstrap probabilities from 1,000 replications.

CHAPTER 4: FLOWERING GENES IN ALFALFA: EXPRESSION ANALYSES AND IDENTIFICATION OF POTENTIAL FUNCTIONAL HOMOLOG OF *ARABIDOPSIS CONSTANS*, *FVE*, AND *FCA* GENES

Praveena Kanchupati[1], Yang Yen[1], Arvid Boe[2], and Yajun Wu[1]

[1] Department of Biology and Microbiology, South Dakota State University, Brookings, SD 57007

[2]Department of Plant Science, South Dakota State University, Brookings, SD 57007

4.1. Abstract

Alfalfa (*Medicago sativa* L.), a perennial legume, is mainly cultivated as a forage crop in US. Apart from being the third most important crop in US in terms of production, only after wheat and corn, alfalfa is emerging as a potential candidate to be used as an energy crop for cellulosic biofuel production. Classical breeding is being used to develop a variety of germplasm which have the ability to produce significantly large amounts of biomass. Though these varieties serve the purpose, they also have the limitation of getting adapted to the vast differences in the environmental conditions across the country. An alternative approach thus is to enhance biomass production of current commercial cultivars. Studies in the model plant *Arabidopsis* have shown that use of molecular tools to manipulate key flowering pathways genes can result in continued vegetative growth and subsequent biomass increase, but, most importantly in much less time compared to the conventional breeding practices. Thus, we hypothesized that a similar genetic manipulation approach in alfalfa will provide us with plants with enhanced biomass production. Based on the work in Arabidopsis and other plants, three genes *CONSTANS (CO)*; *FVE* and *FCA* controlling two independent flowering pathways namely photoperiod/light pathway and autonomous pathway respectively play important roles flowering time control. In this study, we identified potential homologues of *AtCO*, *AtFCA* and *AtFVE* genes from *M. truncatula* genome (a close relative to alfalfa) and examined their expression in different tissues and at different developmental stages and response to circadian and photoperiod in alfalfa. Our results showed alfalfa had more than one homolog for each gene. Expression analysis showed distinct patterns among the homologues. Transcript level of *FVE-like 1, 2, and FCA-like 2* in alfalfa appeared to be

associated with flowering, suggesting they may be a functional orthologue of *AtFVE* and *AtFCA*. *CO-like* genes showed different expression pattern from *AtCO*, suggesting *CO*-like genes may function differently in alfalfa. Our study provides valuable insight into the molecular control of flowering time in alfalfa. Ultimately, complementation and overexpression studies of these genes in Arabidopsis will allow us to determine their function and role in flowering time regulation. This knowledge can then be applied to manipulate the flowering genes and delay flowering to enhance biomass production in alfalfa.

Keywords: Alfalfa, *CO*, *FVE*, *FCA*, flowering, biomass

4.2. Introduction

Numerous environmental and endogenous cues like light, temperature, the circadian clock, age of the plants and growth elicitors affect plants' transition from the vegetative phase to reproductive phase (flowering). Flowering, an important agronomic trait, is under very tight and complex regulation. These regulatory networks monitor and coordinate subtle changes in the environment with the endogenous signals, and, then direct the plants' response accordingly. This strict monitoring ensures the success of reproductive growth.

Most of our current understanding of the flowering process has come from molecular dissection of floral-induction pathways in the model plant, *Arabidopsis thaliana*. More than 180 genes have been identified as the regulators of these pathways (Bäurle and Dean 2006; Fornara et al. 2010). These genes act through six major pathways namely, i) photoperiod pathway that senses day length change; ii) vernalization pathway that monitors the seasonal changes in temperature; iii) ambient temperature pathway that responds to changes in daily temperatures; iv) gibberellin that is a plant hormone, v) age that is the length of time a plant required to grow and develop and vi) autonomous pathway that act independent of the environmental stimuli and other endogenous cues (Figure 1). The integration of signals from all these pathways is carried out by a set of genes named the "floral integrators". *FLOWERING LOCUS T* (*FT*), *SUPRESSOR OF CONSTANS1* (*SOC1*) and *LEAFY* (*LFY*) integrate the signals and rapidly promote floral development. These integrators then communicate with the downstream "floral meristem identity" genes like *APETALA 2* (*AP1*) to induce flowering (Simpson and Dean 2002; Parcy 2004).

Flowering is a key developmental process in a plant's life cycle and is directly linked to crop production and overall yields. Enhanced understanding of the molecular basis of this complex process in crops can be of a huge advantage to the researchers trying to develop new varieties that have improved productivity and better yield. Alfalfa (*Medicago sativa* L.), also known as the "Queen of Forage", is a major forage crop with important agronomic traits (Castonguay et al. 2009). Alfalfa is an important source of protein and fiber for the livestock and is grown worldwide. It ranks third in terms of production in the United States trailing behind the staple crops wheat and corn. As a legume, alfalfa has the ability to establish symbiotic relationships with nitrogen-fixing bacteria reducing the need for the application of nitrogen-rich fertilizers to the soil (Wang et al. 2015). In addition, its deep root system allows the plant to flourish under mild drought conditions. Alfalfa is also gaining ground as a potential candidate for biofuel production due to its high cellulosic biomass yield with a low input from nitrogen fertilizer. Scientists have developed special alfalfa germplasm which is woody and has high cellulosic biomass yield. However, these special germplasm are less nutritious and less palatable to livestock and may be able to perform well in the diverse environmental conditions across the country.

Using current commercial forage alfalfa for cellulosic biomass production is a sound alternative. However, the high cost of biomass production in alfalfa for biofuel purpose is prohibitive. Therefore, enhancing biomass yield per unit land may reduce the price. One strategy is to delay flowering and reduce the yield loss associated with senescence in alfalfa. In Arabidopsis, many mutants with delayed flowering showed an enhanced production of vegetative tissues (Reeves and Coupland 2001; Macknight et al.

1997; Morel et al. 2008; Jung and Müller 2009). A recent study reported that the genetic manipulation of a microRNA miR156 in alfalfa resulted in delayed flowering and subsequent increase in biomass. Additionally, the authors also observed reduced lignin content and enhanced cellulose content in the transgenic alfalfa overexpressing miR156 (Aung et al. 2015b, a). Another study in *M. truncatula* reported the manipulation of onset of flowering to enhance biomass and suggested genetically delaying the floral initiation as an easy tool to achieve improved biomass quality and quantity (Tadege et al. 2015).

These studies along with the knowledge based on the mutant analysis of key flowering genes in Arabidopsis, we hypothesized that genetic manipulation of flowering genes can be used as a tool to enhance biomass production in alfalfa. By delaying flowering, we can prolong the vegetative state and thus avoid high lignin deposition and achieve significant increase in biomass, making alfalfa fit to be used as a cellulosic input in the biofuel industries. At the same time, a low-lignin alfalfa would greatly mean a better quality of forage.

Flowering control in alfalfa is not well studied. Alfalfa is generally considered a long-day plant while it was less sensitive to the day length as other plants such as Arabidopsis. Flowering regulation of the model plant *Arabidopsis thaliana* is extensively studied. The major players in each pathway have been cloned and their roles are well defined (Moon et al. 2005; Michaels et al. 2005; Putterill et al. 2004). Null mutation of some of the key flowering genes result in a significant delay of flowering and enhanced biomass production. These genes include *CO* (*COSTANS*) in the photoperiod pathway, and *FCA* and *FVE* in the autonomous pathway. Many studies have shown that these key flowering genes are conserved among long-day flowering plants. Gene *CO* in *Medicago*

truncatula, a close relative of alfalfa, has been confirmed to have the same function as in Arabidopsis (Hecht et al. 2005).

The objective of the present study was to identify and characterize homologues of three key genes, *CO*, *FVE* and *FCA* to establish a better understanding of flowering regulatory pathways in alfalfa and to provide insight into potential genes that can be manipulated to enhance biomass quantity and quality in alfalfa.

4.3. Results

4.3.1. Identification of *COL*, *FVEL* and *FCAL* genes in *Medicago truncatula* and phylogenetic analysis

Arabidopsis CO, FVE and FCA protein sequences were used as query to perform BLASTP searches of the genome database of *M. truncatula* at the Joint Genome Initiative [(https://phytozome.jgi.doe.gov/pz/portal.html#!search?show=KEYWORD&method=Org_Mtruncatula)] for the identification of putative homologues (Goodstein et al. 2012).

COL gene family in *Medicago*

Ten COL proteins were identified using the cutoff E-value 2.1E-9 and score of 59.3 with AtCO. According to the existing nomenclature the genes were named MtCOL1-MtCOL10. All the MtCOL protein sequences were checked for the presence of the signature B-BOX and CCT domains (Robson et al. 2001). Based on the group (I, II or III) to which each of the MtCOL homologues belonged, they differed in the number and sequence of the B-BOX domains they harbored (Figure S1) similar to their Arabidopsis counterparts, thus displaying a high degree of structural and functional conservation across the two species. The three homologues that are center to this study, MtCOL-1, -2 and -7 had two B-BOX domains that each contained both the conserved

CX2CX8CX7CX2C motif and the critical C and H residues (Figure 2b.). The C-terminal tail of all the three proteins showed a high degree of conservation especially within the CCT domain with intact NF-YA1/linker/NF-YA2 structure (Strayer et al. 2000). All the key residues were also present. Though all the characteristic motifs were accounted for in the MtCOLs, there was some degree of divergence observed among the amino acid sequences within the *Medicago* species as well as across species.

To get a better insight into evolution of the *Medicago CONSTANS-like* gene family, a phylogenetic tree was constructed based on the multiple protein sequence alignment that included 16 CO/COL sequences from Arabidopsis and 10 COL sequences from *Medicago*. The analysis showed the clustering of the *Medicago* COL homologues into three distinct groups I, II and III that corresponded to the number of B-BOX domains each protein harbored and their sequence similarity (Figure 2a.). Within each group, *Medicago* and Arabidopsis COL proteins grouped together indicating the independent expansion of *COL* gene family in species-specific manner. MtCOL1-3 was shown in group I with AtCO, AtCOL1-5. MtCOL-1 and -2 were chosen for this study based on their close phylogenetic relationship to AtCO, and MtCOL-7 was also chosen as it was mapped as a QTL associated with flowering in *Medicago truncatula* (Pierre et al. 2011).

FVE-like gene family in *Medicago*

The BLASTP search of the *Medicago* genome database with AtFVE sequence resulted in the identification of seven putative gene sequences encoding FVEL proteins. Redundant sequences and multiple hits were eliminated. FVE is the plant homolog of mammalian retinoblastoma-associated proteins RbAp46 and RbAp48 (Kenzior and Folk 1998), and the protein has two signature domains, the chromatin assembly factor 1

subunit C (CAF1c) domain and six WD40 repeat domains (Abou-Elwafa et al. 2011; Ausin et al. 2004). All the eight sequences were screened for the signature domains. We observed lower conservation in the N-terminus in comparison to the C-terminus of MtFVELs when compared to AtFVE (Figure 3b.). In contrast to the Arabidopsis FVE protein, the majority of the *Medicago* FVELs did not have the Nuclear Localization Signal (NLS) in the N-terminal region. Similar to MtCOLs, there was some degree of difference between the *Medicago* and Arabidopsis FVE sequences in the conserved domains as well.

When the phylogenetic analysis of the Arabidopsis and *Medicago* FVE protein sequences was performed, it was observed that both MtFVEL-1 and MtFVE-2 clustered together with AtFVE (Figure 3a.). MtFVE-1 protein showed the highest relation to the Arabidopsis counterpart indicating conservation of the structure of the protein in the *Medicago* species. MtFVEL-1 and -2 were chosen for further expression analyses, whereas MtFVEL-3 was included as a negative control.

FCA-like gene family in *Medicago*

The FCAL proteins in *Medicago* were identified by performing BLASTP search of the *Medicago* genome, in a very similar search procedure as the above gene families. Four putative FCAL proteins were identified and scanned for conserved protein domains. AtFCA have two RNA-Recognition Motifs (RRM) and a WW protein interaction domain (Macknight et al. 1997; Bork and Sudol ; Chen and Sudol 1995). All four of the identified MtFCAL proteins sequences showed significant homology within the RRM1 and 2 motifs and the WW protein binding domains (Figure 4b.). It was interesting to observe

lower degree of total protein sequence similarity across *Medicago* and Arabidopsis FCA proteins.

When the phylogenetic relationship between AtFCA and MtFCALs was studied, MtFCAL-1 was the only protein clustering with AtFCA. MtFCAL-2, -3 and -4 all together formed a separate cluster (Figure 4a.). Similar to MtCOL and MtFVEL, only MtFCAL-1 was studied in further expression analyses while MtFCAL2 was included in analysis for comparison purpose.

4.3.2. Three *MsCOL* homologues are regulated by the circadian clock

For gene expression analysis, the DNA sequences corresponding to each protein were retrieved from the *Medicago truncatula* genome and used for designing gene specific primers. These primers were tested in PCR reactions using alfalfa genomic DNA in comparison with *Medicago truncatula* DNA. Every working primer pair was able to amplify an amplicon of identical size from both *M. truncatula* and alfalfa DNA.

Arabidopsis *CO* transcript abundance oscillates with a 24h cycle and is regulated by the circadian clock (Suarez-Lopez et al. 2001). To analyze if the identified *MsCOL* homologues showed similar oscillations, we studied the expression pattern of three *MsCOL* genes in young shoots of SD201 plants grown under 16h/8h LD conditions and then transferred to continuous light (LL). Samples were harvested at an interval of 3 hours spanning a total of 72 hours. Under LD conditions, *MsCOL-1* showed an oscillation period of 24h peaking at 12h after dawn followed by downregulation at 18h time point (Figure 5). *MsCOL-1* expression continued to oscillate under LL conditions regularly but peaked at a different times from LD conditions. *MsCOL-2* and *MsCOL-7* also showed an expression pattern that continued to oscillate with a 24h period under LL

conditions. Under LD conditions *MsCOL-2* had peak transcript abundance at 3h after dawn, whereas *MsCOL-7* displayed a broad peak between 6h and 12h. In sharp contrast to *MsCOL-1, MsCOL-2* showed a decay of the expression peak on the 2nd day under LL. *MsCOL-7* maintained robust oscillations without any sharp increases in the expression amplitudes under LL conditions. However, none of the genes showed any resemblance to the expression peaks pattern of *AtCO* under LD conditions.

4.3.3. *MsFVEL1* and *MsFCAL1* genes showed diurnal changes in their transcript accumulation

Numerous genes in the autonomous flowering pathway in Arabidopsis have been implicated in the regulation of the circadian clock (Salathia et al. 2006) and in recent years, based on the mounting experimental evidence these genes are thought to become subjects to the diurnal and circadian regulation themselves (Pruneda-Paz and Kay 2010) through feedback regulation. Based on these studies we examined if *MsFVEL* and *MsFCAL* homologues displayed any response to diurnal changes.

MsFVEL1 was expressed at significantly higher levels at dawn and right before dawn. The expression levels remained at lower but relatively stable level through most part of the day. *MsFVEL-2* showed no clear peaks. *MsFVEL-3* expression showed oscillation but did not show a clear pattern overall (Figure 6). *MsFCAL-1* displayed a relative lower expression during the first 9 hour in the day, then peaked its expression at the 12h time point followed by downregulation after dusk maintaining similar expression level through the night. *MsFCAL-2* showed relatively stable expression through day and night, except for lower expression at 3 h time point after dawn, before the expression reaching significantly higher level 3 hours later during the day.

4.3.4. *MsCOL*, *MsFVEL* and *MsFCAL* genes are relatively abundant in the leaves of alfalfa

To gain a better insight into tissue-specific regulation of transcript abundance, the expression pattern of the genes was examined in four different tissues and in two different alfalfa germplasm SD201 and Alfagraze. Both *MsCOL-1* and *MsCOL-2* were extensively expressed in the leaves of SD201 and Alfagraze but expressed at significantly lower levels in roots, an expression pattern very similar to Arabidopsis *CO*. *MsCOL-2* also showed significantly higher expression in flowers compared to roots. However, there was no significant difference in the expression levels of *MsCOL-7* in the different tissue across the two germplasm examined.

The autonomous pathway genes, *MsFVEL-1* and *-2* also showed relatively higher transcript abundance in the aerial parts of SD201 and Alfagraze, when compared to the underground part; the roots. *MsFCA-2* showed relatively higher transcript abundance only in the aerial parts of SD201 and Alfagraze, when compared to the roots. As observed in case of *MsCOL-7*, *MsFVEL-3* and *MsFCAL-1* did not show any significant changes in the expression levels in the different tissues within and across the two alfalfa germplasm studied (Figure 7).

4.3.5. Expression of *MsCOL*, *MsFVEL* and *MsFCAL* genes varied at different developmental stages

Expression patterns of the eight genes examined in this study varied largely at different developmental stages. *MsCOL-1* and *-2* had the highest expression at W_0 stage in the upper node samples. W_2 developmental stage of the lower node samples had the highest transcript accumulation for these genes. In both the upper node and lower node

samples these genes were expressed at relatively constant levels after the initial higher expression, except for a few stages. *MsCOL-1* showed a higher expression at W4 in the lower node in Alfagraze. *MsCOL-7* showed a higher expression in the upper node at W3 in SD201. *MsCOL-7* showed a gradual increase in transcript level from W1-W5 and peaked at W6 in Alfagraze. A similar trend existed in the lower node but peaked at W5. *MsCOL-7* showed significantly lower expression at W6 and W7 in the lower node of SD201.

Expression of *MsFVEL1-3* displayed a gradual increase with developmental stages in the upper node of Alfagraze, reaching the highest level at W6-W7. This distinct pattern however was not observed in SD201. The expression of these genes in the lower nodes of two germplasm seemed to be relatively constant at different developmental stages with some fluctuations. An exception is *MsFVEL-2*, which showed a highest expression at W4 followed by a decrease in W5 and W6.

The transcript level of *MsFCAL-1* gradually increased with the developmental stages in the upper node of SD201, reaching the highest level at W7. This pattern was not observed in Alfagraze, with the expression remaining relatively constant. In the lower node, the gene was expressed at relatively lower levels at W6 and W7 in both SD201 and Alfagraze. The peak of expression was observed at W3 stage in lower node of SD201 and W4 and 5 in Alfagraze. Expression of *MsFCAL-2* were relatively constant in both the upper and lower nodes in both germplasm with some exceptions. For example, *MsFCAL-2* expression in the upper node showed a higher expression at W4 and W6 for SD201 and Alfagraze, respectively (Figure 8 a, b and c.).

4.4. Discussion

Alfalfa is an important forage crop and is more popularly known as the "Queen of forage". The shoots and leaves are a source of high amounts of protein and fiber for the animals. However, the forage quality decreases significantly once plants start to flower. Thus, it is believed that delaying flowering though genetic engineering or traditional breeding could enhance the forage quality as well as the biomass yield. An enhanced biomass production is also needed for using alfalfa as a potential cellulosic feedstock for biofuel production. For that, it is necessary to identify the players that are important in regulating flowering time. The transition from the vegetative to reproductive state in plants is under very tight and complex regulation. In Arabidopsis, the model plant, more than 180 genes have been identified as the regulators of the six major floral-induction pathways, namely the photoperiod, vernalization, gibberellin, autonomous, age and ambient temperature pathways; to induce and regulate the "floral integrator genes" (Fornara et al. 2010). Functional homologues of several key flowering genes, such as *CO*, *FVE* and *FCA*, have been identified in many plant species like soybean (Huang et al. 2011), sugar beet (Abou-Elwafa et al. 2011) and rice (Yano et al. 2000); implicating the functional conservation of the floral pathways across the plant kingdom. In this study, we addressed the question of whether selected genes from two major flowering regulatory pathways in Arabidopsis (the photoperiod and autonomous pathways) are conserved in alfalfa.

4.4.1. Evolutionary expansion of key flowering gene families in alfalfa

Gaining insight into the molecular participants of complex regulatory pathways in alfalfa is very restricted due to the lack of a sequenced genome. In this study, the putative

functional homologues of Arabidopsis *CONSTANS*, *FVE* and *FCA* were identified by performing the *in silico* analysis of the *Medicago truncatula* genome instead. *M. truncatula* is a close relative of alfalfa and shares a high degree of sequence similarity (Julier et al. 2003; Young et al. 2011). This statement was also supported by the fact that we were able to amplify genes from alfalfa using primers designed based on the *M. truncatula* genome. We found that alfalfa has at least ten *COL*, seven *FVEL* and four *FCAL* genes. The sequence analysis showed high homology between the *Medicago* and Arabidopsis homologues, especially in the signature domains of the proteins.

CONSTANS-like gene family in alfalfa

The phylogenetic analysis of the identified putative homologues of *CO* genes in *Medicago* showed the presence of all the three major subgroups (I, II and III) in the legume. Thus, the groups predate the divergence of Arabidopsis and the legumes (Figure 2a) and the two plant species have retained the characteristic B-BOX domains and CCT domains since (Figure 2b). It is interesting that given the greater size of the genome compared to Arabidopsis, *M. truncatula* most likely has only ten homologues of *CO*, in comparison to seventeen in Arabidopsis. Among the ten genes, MtCOL-9 and MtCOL-10 are nearly identical suggesting they arose from a recent gene duplication. However, given the ploidy level and genome size of alfalfa, it most likely has more than ten *COL* genes.

It is noted that MtCOL1-3 showed a greater similarity to AtCOL3-5 than AtCO. AtCO formed a separate subclade only with AtCOL1 and AtCOL2. The results suggested a sequence divergence occurred before forming AtCO and MtCOL1-3. One diverged ancestor sequence was only found in Arabidopsis resulting in formation of AtCO, AtCO1 and 2. The other diverged ancestor sequence were found in both Arabidopsis and *M.*

truncatula, resulting in forming AtCOL3-5 and MtCOL1-3. It can be expected that MtCOL1-3 would function similarly to AtCOL3-5.

FVE-like and *FCA-like* gene family in alfalfa

As reported earlier (Kim et al. 2013), *M. truncatula* contains 25 genes that are involved in the autonomous pathway and are homologues to 16 genes in Arabidopsis. From the *in silico* analysis we were able to find seven homologues for *FVE* and four for *FCA.* The phylogenetic analysis of the FVE homologues from Arabidopsis and *Medicago* suggested that *MtFVEL* genes evolved through multiple events (Figure 3a). *MtFVE3-7* most likely evolved after the divergence of Arabidopsis and the legume. *AtFVE* and *MtFVEL* genes are potentially derived from the same ancestral gene. However, the ancestor gene diverges before the divergence of Arabidopsis and *M. truncatula*, since MtFVEL-3 forms a parallel subclade with AtFVE and MtFVEL-1. The homologues from both plant species have retained the signature CAF1c and WD40 domains indicating the presence of these domains in the ancestral sequence from which these have evolved (Figure 3b). The size of the *FVEL* gene family in alfalfa is unknown. However, since most of the commercial varieties of alfalfa are polyploids there is high possibility of the existence of even a larger gene family in alfalfa.

The phylogenetic analysis of *FCA-Like* gene family showed that the FCA protein predates the divergence between Arabidopsis and the legume, since MtFCA-1 and AtFCA formed a cluster that is distinct from a cluster only containing FCAL from *M. truncatula*. This also made MtFCAL-1 the only possible orthologue of AtFCA (Figure 4a).

4.4.2. Diurnal and circadian changes in the photoperiod elicited conserved responses

Multiple approaches like global transcriptomic analysis and *in vivo* enhancer trapping, determined that the circadian clock controls almost all the biological functions in Arabidopsis by regulating the expression of more than one-third of genes (Harmer et al. 2000; Covington et al. 2008; Michael and McClung 2003). It was originally thought that the autonomous pathway genes contributed to the regulation of circadian clock (Salathia et al. 2006), but increasing evidence suggests that the genes themselves may also be subjected to diurnal and circadian regulation, including *FVE* and *FCA* (Pruneda-Paz and Kay 2010). Our gene expression analysis revealed that *MsFVEL-1* and *MsFCAL-1*, the two most closely related counterparts in Arabidopsis based on the phylogenetic analysis, showed clear diurnal response while others showed fluctuations of expression without a distinct diurnal pattern (Figure 6). Although *FVE* and *FCA* were shown to affect the circadian rhythm in Arabidopsis plants (Salathia et al. 2006), these genes have not been reported to be under diurnal/circadian control. The first evidence of diurnal control of *FVE* was reported in sugar beet (Abou-Elwafa et al. 2011). Our results of circadian regulation of *MsFVEL-1* and *MsFCAL-1* suggest that they are potential orthologues of FVE and FCA. Unlike *BvFVE1* that has its peak expression at 12h time point under LD conditions, *MsFVE-1* reaches its peak expression just before dawn and then remains at relatively lower levels throughout the day.

Of the one-third of circadian clock controlled genes, photoperiod pathway genes are the some of the most extensively studied genes. Robust expression oscillation under diurnal and circadian condition is one of the key features of Arabidopsis *CONSTANS*

(Suarez-Lopez et al. 2001). And this key feature has been shown to be conserved across many different plant species including legumes like soybean (Huang et al. 2011; Wu et al. 2014) and *M. truncatula* (Wong et al. 2014) and cereals like rice (Yano et al. 2000). Two *MsCOL* genes that showed reliable expression showed different circadian patterns. *MsCOL-1* did not show significant expression until the plants were under LL, where *MsCOL-1* expression clearly showed regular cycling, with higher expression in the evening. It is interesting to note that a similar expression profile was observed for *GmCOL9* under both SD and LD conditions in soybean (Huang et al. 2011). *MsCOL-2* appeared to maintain the regular expression pattern only in the first day entering, with the highest expression right after dawn. The difference in expression pattern during circadian response raised the question whether *MsCOL-1 and -2* function similarly to *AtCO*.

4.4.3. Transcript abundance of the flowering genes in different tissues is conserved within and between species

To provide further evidence conservation of these genes compared to their counterparts in Arabidopsis, we studied expression of the selected genes in different tissues.

Many of *MsCOLs, MsFVELs* and *MsFCALs* homologues showed higher expression in the leaf/stem tissues and only slightly in roots, a pattern conserved in *Medicago* and Arabidopsis (Macknight et al. 1997; Ausin et al. 2004; Suarez-Lopez et al. 2001). This includes *MsCOL-1* and *MsFVE-1*, that showed highest sequence similarity to AtCO and AtFVE. *MsFCAL-1* is however expressed at similar level in different tissues. In addition to the expression studies in different tissues we also performed experiments to

understand the transcript abundance of the selected genes at different developmental stages. When our results of the transcript abundance of the selected genes at different developmental stages were compared to profiles of *AtCO*, *AtFVE* and *AtFCA* obtained from the AtGenExpress Visualization tool (http://jsp.weigelworld.org/expviz/expviz.jsp), we noticed similarity and some divergence in transcript accumulation. *MsCOL-1* and *-2* are expressed at significantly higher levels in the seedlings and relative lower and constant level afterward. *AtCO* is expressed at the highest level in the shoots of 21 days old plants, a stage preceding flowering, a pattern that is in sharp contrast to *MsCOL-1* and -2. *MsCOL-7* however showed a peak expression at W6 and W5 in the upper and lower node respective in Alfagraze, preceding flowering, suggest *MsCO-7* may be associated flowering. Its potential role in flowering regulation is also supported by the fact that in a study by Pierre et al it was mapped as a QTL associated with flowering in *Medicago truncatula*. *MsFVEL-1* showed a gradual increase with time for in the upper node of both genotypes. It reached the highest level at W5 in SD201 and W6 for Alfagraze, at timing right before flowering. *MsFVEL-2* showed a similar pattern in the upper node of Alfagraze. Its expression in the lower node is relative constant in the lower node of both germplasm. Based on the microarray data the expression of *AtFVE* reaches peak expression late developmental stages similar to *MsFVE1-3* based on expression in the upper node of two genotypes. These data suggest a role of *MsFVEL1-3* in flowering time control in alfalfa. *AtFCA* is expressed at relatively constant level in different tissues and at different developmental stages (Macknight et al. 1997). A similar pattern was observed in *MsFCAL-1*. *MsFCAL-2* however showed peak expression in the upper node in both genotype suggesting a link with flowering in alfalfa. Altogether, the fact that these genes

have different profiles when compared to their counterparts in Arabidopsis suggests that alfalfa may have developed a different or a more complex regulatory system. Alternatively, these genes in alfalfa may play different roles from flowering timing regulation. This is supported by the results reported by other labs through functional analysis.

Functional characterization of the members of *CONSTANS-like* gene family have been conducted in soybean and *M. truncatula* (Wu et al. 2014; Wong et al. 2014) and, interestingly, resulted in two very different conclusions. The soybean *COLs* were able to complement the Arabidopsis *co* mutants, but overexpressing *MtCOLs* in Arabidopsis failed to rescue the mutants. And since alfalfa is closest to *M. truncatula* there is a possibility that the *COLs* may not play central role in the photoperiod regulation of flowering in alfalfa as well. The findings may not be surprising since alfalfa appears to be day length insensitive in flowering, although it is generally considered as a long day flowering plant.

Research on the autonomous pathway genes is very limited in legumes (Kim et al. 2013). We only know that the legumes have the homologues of the key genes like *FLC*, *FVE* and *FCA* in their genome. This study is the first evidence of structural and expression pattern conservation between Arabidopsis and *Medicago* species. Our sequence and gene expression analysis suggest that one of the *MsFVE1-3* and *MsFCAL-1* may be a functional orthologue of *AtFVE* and *AtFCA*, respectively. Functional analysis through molecular genetics such as complementation of Arabidopsis *fve and fca* mutant with alfalfa genes will provide the conclusive answer.

In summary, our study provides insight into the molecular control of flowering time in alfalfa. Our results suggested that the CO-like genes in alfalfa may be associated with flowering which is correlated with its day-length insensitive nature. *MsCOL-7* may be a functional orthologue based on its peak expression prior to flowering. Among two players in the autonomous pathway, multiple members of *MsFVEL* or both *MsFCAL-1* and *-2* are implicated a role in flowering time control based on their peak expression preceding to flowering. A systematic characterization and comparative analysis of different members of *COL*, *FVEL* and *FCAL* will provide further insight into the complexity of regulation of flowering in alfalfa and identify the candidate genes involved in flowering. Ultimately, the functional characterization of the candidate genes will allow the determination of key pathways controlling flowering specific to alfalfa. As shown in *M. truncatula* (Tadege et al. 2015), this knowledge can be applied to manipulate the flowering genes and delay flowering to enhance biomass production in alfalfa.

4.5. Materials and methods

4.5.1 Plant materials and growth conditions

Alfalfa germplasms SD201 (SD) and Alfagraze (AG) were examined in this study. SD cultivar was developed by Dr. Arvid Boe at South Dakota State University and AG seeds were obtained from The National Temperate Forage Legume Genetic Resource Unit, Prosser, WA, USA. The plants were grown in Conviron growth chamber with growth conditions set at 22±2°C, day and 19±2°C, night thermo period with a 16h photoperiod. The light intensity was set at 200 µmol m^{-2} s^{-1} (PAR), and the relative humidity level was maintained at 55%.

The SD and AG seeds were surface sterilized prior to germination. The seeds were first scarified with 3M 332U 150 aluminum oxide sandpaper, and then sterilized in 2.5% bleach solution for 3 minutes followed by 3 thorough washes with distilled water. Six seeds were then directly planted into each one-gallon pots (Stuewe and Sons, Corvallis, OR) that were filled with potting mix (Sunshine Mix #3, Sun Gro Horticulture Canada Ltd., Seba Beach, AB, Canada). Another batch of seeds were grown in 3.8 x 21-cm Ray Leach Cone-tainers (Stuewe and Sons, Corvallis, OR), two each, which were filled with 38 grams of potting mix. Plants were irrigated at a 3-day interval with a Miracle-Gro (Scotts Miracle-Gro Products, Inc., Marysville, OH) nutrient solution (5 gm Miracle-Gro/gallon of H_2O, N:P:K = 15:30:15). Two weeks after germination, the seedlings were thinned to three plants per pot and one per each Cone-tainer.

4.5.2 Treatment and sampling

Diurnal and circadian samples

To investigate the diurnal changes in gene expression of *FVE* and *FCA* homologues in alfalfa, young shoots from SD201 plants grown in pots at similar developmental stage (late bud stage) were harvested every 3h starting at dawn. Samples were harvested to represent a total of eight time points (T_0 –T_7 or 0h – 21h) in a complete day (24 hrs), with 5 samples collected in light and 3 samples in dark. Three young shoots were harvested per replicate for three replicates and immediately frozen in liquid nitrogen and stored at -80°C.

To investigate the expression of alfalfa *COL* homologues in response to circadian changes, young shoots from SD201 plants grown in pots at similar developmental stage (late bud stage) were harvested at 3h intervals over 72 h, that spanned 24h of long day

(LD) photoperiod conditions followed by 48h of continuous light. Samples were harvested to represent a total of twenty-four time points (T_0 – T_{23} or 0h – 69h). Three young shoots were harvested per replicate, and three replicates were harvested at each time point.

Developmental stages

Samples representing different stages of development of alfalfa were harvested from SD201 and AG plants grown in Cone-tainers, every week starting from the seventh day after germination. For the first two weeks the whole seedlings were harvested (W_0 and W_1), and from the third week onwards the young shoots from the upper node (UN W_2 – W_7) were harvested. The shoots from the lowest nodes were also harvested from third week onwards (LN W_2 – W_7). The samples were harvested at the same time of the day, 15h after lights on, every week.

Different tissues

Leaves, stems, flowers and roots were collected separately from mature SD201 and AG plants grown in pots. Leaves and stems were separated from young shoots of 8-week-old plants and pooled as two different samples. Flower samples included 3-5 young, unopened clusters. Each cluster contained from 10 to 20 individual flowers depending on the germplasm. The samples were harvested 15h after lights on. All the samples used for gene expression analysis were immediately frozen in liquid nitrogen after harvest and stored at -80°C until RNA extraction.

4.5.3 Identification of *COL*, *FVEL* and *FCAL* genes in *M. truncatula* and phylogenetic analysis

Arabidopsis CO, FVE and FCA protein sequences was used to search against the *M. truncatula* genome database version Mt 4.0V1 at the Joint Genome Initiative (https://phytozome.jgi.doe.gov/pz/portal.html#!search?show=KEYWORD&method=Org_Mtruncatula). The default Expect (E) threshold value for the *Medicago* homologs that were chosen was set at -1, and the BLOSUM62 comparison matrix was used. The retrieved sequences were named according to existing nomenclature as *MtCOL-1* to *MtCOL-10; MtFVEL-1* to *MtFVEL-7* and *MtFCAL-1* to *MtFCAL-4*, respectively. Multiple sequence alignment of the protein sequences was done using Clustal Omega (Sievers et al. 2011), and the Neighborhood-joining method of MEGA 6 (Tamura et al. 2013) was used to construct the phylogenetic trees.

4.5.4 Primer design

Gene specific primers (Table S1) were designed for the selected *MtCOL*, *MtFVEL* and *MtFCAL* genes using an online primer design tool from Integrated DNA Technologies (Coralville, IA, USA). The efficiency and specificity for each primer pair was determined using alfalfa genomic DNA (1 ng) as the template in a 20 µL PCR reaction containing 2 µL of 10X PCR buffer, 1 µL each of 10 µM primers, 1 µL of 2 mM dNTPs and 0.1 µL of Taq polymerase (5 U/µL, BioLabs Inc., Ipswich, MA). All reactions were performed in a gradient thermocycler (Eppendorf Mastercycler, Eppendorf, Hauppauge, NY) with PCR conditions set as; initial denature step at 94°C for 3 min followed by 35 cycles of 94°C for 20 s, 20 s at annealing temperature gradient (R= 3°C/s, G= ± 2.4°C), extension at 72°C for 2 min and a final extension at 72°C for 10 min.

The PCR products were then run on a 1% agarose gel stained with ethidium bromide, and images were visualized using a Bio-Rad ChemiDoc (Bio-Rand Laboratories Inc., Hercules, CA) image analysis system.

4.5.5 RNA isolation and cDNA synthesis

Total RNA was isolated using Trizol reagent (Invitrogen, Carlsbad, CA). The RNA samples were quantified using Nanodrop ND-1000 Spectrophotometer (ThermoFisher Scientific, Waltham, MA), and samples with 260/280 ratio from 1.9 to 2.1 and 260/230 ratio from 2.0 to 2.5 were used for further analysis. RNA quality was also examined by separating RNA on a 2% agarose gel stained with ethidium bromide. The samples, which showed three sharp major rRNA bands, were used for cDNA synthesis. First strand cDNA synthesis was performed using the high capacity cDNA Reverse Transcription kit (ThermoFisher Scientific, Waltham, MA) in a 20 µL reaction according to manufacturer's instructions. Synthesized cDNA samples were validated using *MsActin* primers with 30 cycles in a regular PCR reaction. The cDNA samples were diluted 4 times for use in real-time qRT-PCR reaction. qRT-PCR reactions with the house-keeping gene *MsActin* were then performed on cDNA samples in comparison with respective RNA samples without reverse transcription. Only the cDNA samples whose corresponding RNA samples showed no amplification or significantly greater Ct values (three cycles or more) in qRT-PCR analysis were used for the gene expression analysis.

4.5.6 Expression profiling of *M. sativa* flowering genes

qRT-PCR was performed using DyNAmo Flash SYBR Green Hot Start qRT-PCR Kit (ThermoFisher Scientific, Waltham, MA) following manufacturer's instructions in a 20 µL reaction in ABI 7900HT High-Throughput Real-Time Thermocycler (Applied

Biosystems, Foster City, CA) using standard cycling conditions. Each sample from three biological experiments was assayed twice as technical replicates. The thermocycler program was set to: 15-min activation at 95℃ followed by 40 cycles of 15 s at 94℃, 30 s at annealing temperature, 30 s extension at 72℃, followed by a dissociation curve step. The dissociation curve was used to determine the primer efficiency and specificity. The normalized relative fold changes in the transcripts of *MsCOL* or other genes were calculated using the $2^{-\Delta\Delta Ct}$ or comparative Ct method based on the difference between the target and reference genes as described by Livak and Schmittgen (Livak and Schmittgen 2001).

4.5.7 Data analysis

Statistical analysis was performed using IBM SPSS Statistics 24 (Armonk, NY) and Microsoft Excel 2016 data analysis tools pack (Redmond, WA). Data were subjected to analysis of variance (ANOVA) using the linear model with completely randomized design to determine significant differences among the treatments. Tukey's HSD all pair comparison was conducted to ascertain significant differences between treatment means ($p<0.05$).

4.6. Acknowledgements

This work was supported by the North Central Sungrant Initiative, the USDA National Institute of Food and Agriculture Hatch Project 3AH390, South Dakota Agricultural Experiment Station and the Department of Biology and Microbiology at South Dakota State University. We are grateful to the Functional Genomics Core Facility (FGCF) at South Dakota State University.

4.7. Reference

Abou-Elwafa SF, Büttner B, Chia T, Schulze-Buxloh G, Hohmann U, Mutasa-Göttgens E, Jung C, Müller AE (2011) Conservation and divergence of autonomous pathway genes in the flowering regulatory network of Beta vulgaris. Journal of Experimental Botany 62 (10):3359-3374. doi:10.1093/jxb/erq321

Aung B, Gruber MY, Amyot L, Omari K, Bertrand A, Hannoufa A (2015a) Ectopic expression of LjmiR156 delays flowering, enhances shoot branching, and improves forage quality in alfalfa. Plant Biotechnology Reports 9 (6):379-393. doi:10.1007/s11816-015-0375-2

Aung B, Gruber MY, Amyot L, Omari K, Bertrand A, Hannoufa A (2015b) MicroRNA156 as a promising tool for alfalfa improvement. Plant Biotechnol J 13 (6):779-790. doi:10.1111/pbi.12308

Ausin I, Alonso-Blanco C, Jarillo JA, Ruiz-Garcia L, Martinez-Zapater JM (2004) Regulation of flowering time by FVE, a retinoblastoma-associated protein. Nat Genet 36 (2):162-166. doi:http://www.nature.com/ng/journal/v36/n2/suppinfo/ng1295_S1.html

Bäurle I, Dean C (2006) The Timing of Developmental Transitions in Plants. Cell 125 (4):655-664. doi:10.1016/j.cell.2006.05.005

Bork P, Sudol M The WW domain: a signalling site in dystrophin? Trends in Biochemical Sciences 19 (12):531-533. doi:10.1016/0968-0004(94)90053-1

Castonguay Y, Michaud R, Nadeau P, Bertrand A (2009) An indoor screening method for improvement of freezing tolerance in alfalfa. Crop Science 49 (3):809-818

Chen HI, Sudol M (1995) The WW domain of Yes-associated protein binds a proline-rich ligand that differs from the consensus established for Src homology 3-binding modules. Proceedings of the National Academy of Sciences of the United States of America 92 (17):7819-7823

Covington MF, Maloof JN, Straume M, Kay SA, Harmer SL (2008) Global transcriptome analysis reveals circadian regulation of key pathways in plant growth and development. Genome Biology 9 (8):1-18. doi:10.1186/gb-2008-9-8-r130

Fornara F, de Montaigu A, Coupland G (2010) SnapShot: Control of Flowering in Arabidopsis. Cell 141 (3):550-550.e552. doi:10.1016/j.cell.2010.04.024

Goodstein DM, Shu S, Howson R, Neupane R, Hayes RD, Fazo J, Mitros T, Dirks W, Hellsten U, Putnam N, Rokhsar DS (2012) Phytozome: a comparative platform for green plant genomics. Nucleic Acids Research 40 (Database issue):D1178-D1186. doi:10.1093/nar/gkr944

Harmer SL, Hogenesch JB, Straume M, Chang H-S, Han B, Zhu T, Wang X, Kreps JA, Kay SA (2000) Orchestrated transcription of key pathways in *Arabidopsis* by the circadian clock. Science 290 (5499):2110-2113

Hecht V, Foucher F, Ferrandiz C, Macknight R, Navarro C, Morin J, Vardy ME, Ellis N, Beltran JP, Rameau C, Weller JL (2005) Conservation of Arabidopsis flowering genes in model legumes. Plant Physiology 137 (4):1420-1434. doi:10.1104/pp.104.057018

Huang G, Ma J, Han Y, Chen X, Fu Y-F (2011) Cloning and Expression Analysis of the Soybean CO-Like Gene GmCOL9. Plant Molecular Biology Reporter 29 (2):352-359. doi:10.1007/s11105-010-0240-y

Julier B, Flajoulot S, Barre P, Cardinet G, Santoni S, Huguet T, Huyghe C (2003) Construction of two genetic linkage maps in cultivated tetraploid alfalfa (*Medicago sativa*) using microsatellite and AFLP markers. BMC Plant Biology 3. doi:10.1186/1471-2229-3-9

Jung C, Müller AE (2009) Flowering time control and applications in plant breeding. Trends in Plant Science 14 (10):563-573. doi:10.1016/j.tplants.2009.07.005

Kenzior AL, Folk WR (1998) AtMSI4 and RbAp48 WD-40 repeat proteins bind metal ions 1. FEBS Letters 440 (3):425-429. doi:10.1016/S0014-5793(98)01500-2

Kim MY, Kang YJ, Lee T, Lee S-H (2013) Divergence of Flowering-Related Genes in Three Legume Species. The Plant Genome 6 (3). doi:10.3835/plantgenome2013.03.0008

Livak KJ, Schmittgen TD (2001) Analysis of relative gene expression data using Real-Time quantitative PCR and the 2−ΔΔCT method. Methods 25 (4):402-408. doi:http://dx.doi.org/10.1006/meth.2001.1262

Macknight R, Bancroft I, Page T, Lister C, Schmidt R, Love K, Westphal L, Murphy G, Sherson S, Cobbett C, Dean C (1997) FCA, a Gene Controlling Flowering Time in Arabidopsis, Encodes a Protein Containing RNA-Binding Domains. Cell 89 (5):737-745. doi:10.1016/S0092-8674(00)80256-1

Michael TP, McClung CR (2003) Enhancer Trapping Reveals Widespread Circadian Clock Transcriptional Control in Arabidopsis. Plant Physiology 132 (2):629-639. doi:10.1104/pp.021006

Michaels SD, Himelblau E, Kim SY, Schomburg FM, Amasino RM (2005) Integration of flowering signals in winter-annual Arabidopsis. Plant Physiology 137 (1):149-156

Moon J, Lee H, Kim M, Lee I (2005) Analysis of flowering pathway integrators in Arabidopsis. Plant and Cell Physiology 46 (2):292-299

Morel P, Tréhin C, Breuil-Broyer S, Negrutiu I (2008) Altering FVE/MSI4 results in a substantial increase of biomass in Arabidopsis—the functional analysis of an ontogenesis accelerator. Molecular Breeding 23 (2):239. doi:10.1007/s11032-008-9229-8

Parcy F (2004) Flowering: a time for integration. International Journal of Developmental Biology 49 (5-6):585-593

Pierre J-B, Bogard M, Herrmann D, Huyghe C, Julier B (2011) A CONSTANS-like gene candidate that could explain most of the genetic variation for flowering date in Medicago truncatula. Molecular Breeding 28 (1):25-35. doi:10.1007/s11032-010-9457-6

Pruneda-Paz JL, Kay SA (2010) An expanding universe of circadian networks in higher plants. Trends in Plant Science 15 (5):259-265. doi:10.1016/j.tplants.2010.03.003

Putterill J, Laurie R, Macknight R (2004) It's time to flower: the genetic control of flowering time. Bioessays 26 (4):363-373

Reeves PH, Coupland G (2001) Analysis of Flowering Time Control in Arabidopsis by Comparison of Double and Triple Mutants. Plant Physiology 126 (3):1085-1091

Robson F, Costa MMR, Hepworth SR, Vizir I, Pin˜eiro M, Reeves PH, Putterill J, Coupland G (2001) Functional importance of conserved domains in the flowering-time gene CONSTANS demonstrated by analysis of mutant alleles and transgenic plants. The Plant Journal 28 (6):619-631. doi:10.1046/j.1365-313x.2001.01163.x

Salathia N, Davis SJ, Lynn JR, Michaels SD, Amasino RM, Millar AJ (2006) FLOWERING LOCUS C-dependent and -independent regulation of the circadian clock by the autonomous and vernalization pathways. BMC Plant Biology 6 (1):10. doi:10.1186/1471-2229-6-10

Sievers F, Wilm A, Dineen D, Gibson TJ, Karplus K, Li W, Lopez R, McWilliam H, Remmert M, Söding J, Thompson JD, Higgins DG (2011) Fast, scalable generation of high-quality protein multiple sequence alignments using Clustal Omega. Molecular Systems Biology 7:539. doi:10.1038/msb.2011.75

Simpson GG, Dean C (2002) Arabidopsis, the Rosetta stone of flowering time? Science 296 (5566):285-289

Strayer C, Oyama T, Schultz TF, Raman R, Somers DE, Más P, Panda S, Kreps JA, Kay SA (2000) Cloning of the Arabidopsis Clock Gene TOC1, an Autoregulatory Response Regulator Homolog. Science 289 (5480):768-771. doi:10.1126/science.289.5480.768

Suarez-Lopez P, Wheatley K, Robson F, Onouchi H, Valverde F, Coupland G (2001) CONSTANS mediates between the circadian clock and the control of flowering in Arabidopsis. Nature 410 (6832):1116-1120

Tadege M, Chen F, Murray J, Wen J, Ratet P, Udvardi MK, Dixon RA, Mysore KS (2015) Control of Vegetative to Reproductive Phase Transition Improves Biomass Yield and Simultaneously Reduces Lignin Content in Medicago truncatula. BioEnergy Research 8 (2):857-867. doi:10.1007/s12155-014-9565-y

Tamura K, Stecher G, Peterson D, Filipski A, Kumar S (2013) MEGA6: Molecular Evolutionary Genetics Analysis Version 6.0. Molecular Biology and Evolution 30 (12):2725-2729. doi:10.1093/molbev/mst197

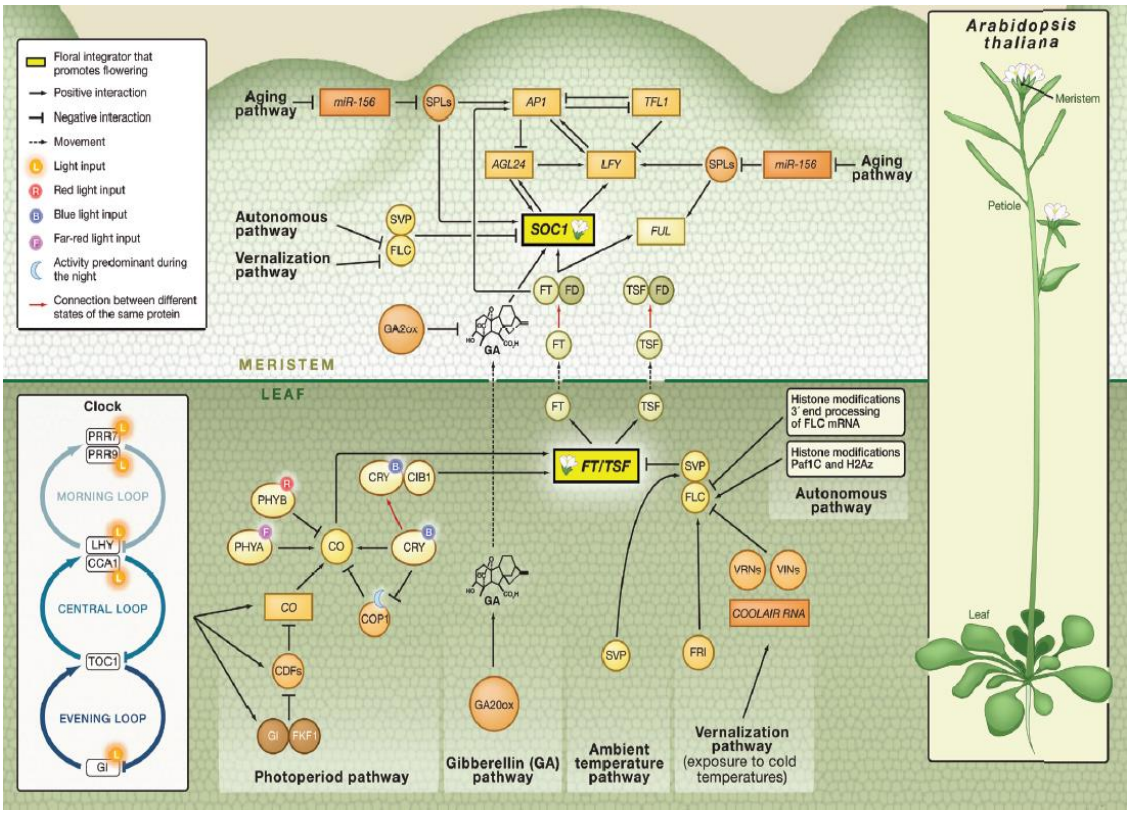
Wang Z, Ke Q, Kim MD, Kim SH, Ji CY, Jeong JC, Lee H-S, Park WS, Ahn M-J, Li H, Xu B, Deng X, Lee S-H, Lim YP, Kwak S-S (2015) Transgenic Alfalfa Plants Expressing the Sweetpotato Orange Gene Exhibit Enhanced Abiotic Stress Tolerance. PLoS ONE 10 (5):e0126050. doi:10.1371/journal.pone.0126050

Wong ACS, Hecht VFG, Picard K, Diwadkar P, Laurie RE, Wen J, Mysore K, Macknight RC, Weller JL (2014) Isolation and functional analysis of CONSTANS-LIKE genes suggests that a central role for CONSTANS in flowering time control is not evolutionarily conserved in Medicago truncatula. Frontiers in Plant Science 5:486. doi:10.3389/fpls.2014.00486

Wu F, Price BW, Haider W, Seufferheld G, Nelson R, Hanzawa Y (2014) Functional and Evolutionary Characterization of the CONSTANS Gene Family in Short-Day Photoperiodic Flowering in Soybean. PLoS ONE 9 (1):e85754. doi:10.1371/journal.pone.0085754

Yano M, Katayose Y, Ashikari M, Yamanouchi U, Monna L, Fuse T, Baba T, Yamamoto K, Umehara Y, Nagamura Y, Sasaki T (2000) Hd1, a Major Photoperiod Sensitivity Quantitative Trait Locus in Rice, Is Closely Related to the Arabidopsis Flowering Time Gene CONSTANS. The Plant Cell 12 (12):2473-2483. doi:10.1105/tpc.12.12.2473

Young ND, Debelle F, Oldroyd GED, Geurts R, Cannon SB, Udvardi MK, Benedito VA, Mayer KFX, Gouzy J, Schoof H, Van de Peer Y, Proost S, Cook DR, Meyers BC, Spannagl M, Cheung F, De Mita S, Krishnakumar V, Gundlach H, Zhou S, Mudge J, Bharti AK, Murray JD, Naoumkina MA, Rosen B, Silverstein KAT, Tang H, Rombauts S, Zhao PX, Zhou P, Barbe V, Bardou P, Bechner M, Bellec A, Berger A, Berges H, Bidwell S, Bisseling T, Choisne N, Couloux A, Denny R, Deshpande S, Dai X, Doyle JJ, Dudez A-M, Farmer AD, Fouteau S, Franken C, Gibelin C, Gish J, Goldstein S, Gonzalez AJ, Green PJ, Hallab A, Hartog M, Hua A, Humphray SJ, Jeong D-H, Jing Y, Jocker A, Kenton SM, Kim D-J, Klee K, Lai H, Lang C, Lin S, Macmil SL, Magdelenat G, Matthews L, McCorrison J, Monaghan EL, Mun J-H, Najar FZ, Nicholson C, Noirot C, O/'Bleness M, Paule CR, Poulain J, Prion F, Qin B, Qu C, Retzel EF, Riddle C, Sallet E, Samain S, Samson N, Sanders I, Saurat O, Scarpelli C, Schiex T, Segurens B, Severin AJ, Sherrier DJ, Shi R, Sims S, Singer SR, Sinharoy S, Sterck L, Viollet A, Wang B-B, Wang K, Wang M, Wang X, Warfsmann J, Weissenbach J, White DD, White JD, Wiley GB, Wincker P, Xing Y, Yang L, Yao Z, Ying F, Zhai J, Zhou L, Zuber A, Denarie J, Dixon RA, May GD, Schwartz DC, Rogers J, Quetier F, Town CD, Roe BA (2011) The *Medicago* genome provides insight into the evolution of rhizobial symbioses. Nature 480 (7378):520-524. doi:http://www.nature.com/nature/journal/v480/n7378/abs/nature10625.html - supplementary-information



Cell 141, April 3, 2010 ©2010 Elsevier Inc. DOI 10.1016/j.cell.2010.04.024

Figure 1. A snapshot of flowering time control in *Arabidopsis thaliana*. This snapshot presents a subset of these genes and proteins, each organized according to its spatial activity in the leaves or the shoot apical meristem of the plant. Strikingly, several genes act more than once and in several tissues during floral induction. Many of these genes occur in a network of six major pathways: the photoperiod and vernalization pathways control flowering in response to seasonal changes in day length and temperature; the ambient temperature pathway responds to daily growth temperatures; and the age, autonomous, and gibberellin pathways act more independently of environmental stimuli.

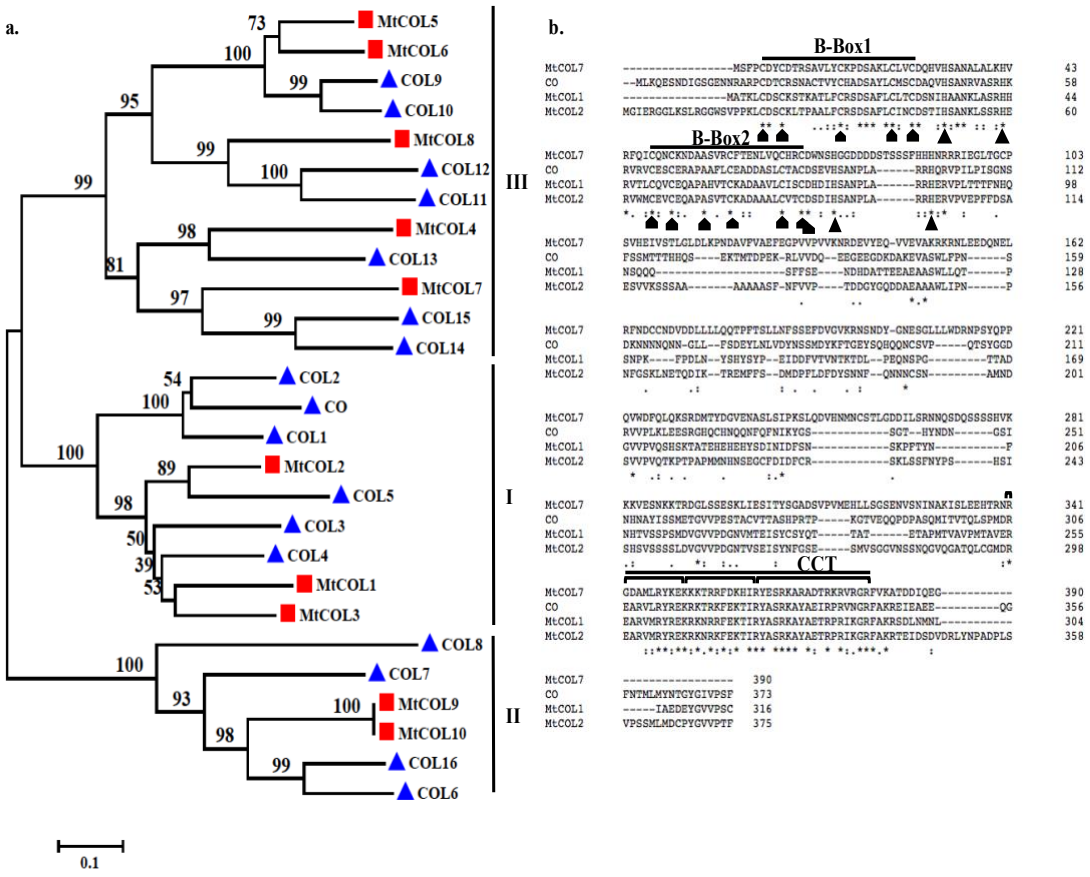


Figure 2a & 2b. **Phylogenetic relationship and protein sequence alignment of CO-like proteins in *Medicago truncatula.***

b. The neighbor-joining (NJ) tree includes 10 COL proteins from *Medicago truncatula* and 16 COL proteins from *Arabidopsis*. The numbers shown next to the branches are the bootstrap probabilities from 1,000 replications.

c. Whole protein alignment of MtCOL1,2 and 7 with AtCO. * represents conserved amino acid residues; • represents the identical and similar amino acid residues. The conserved B-BOX and CCT domains are highlighted. Also highlighted are the characteristic essential amino acids for the functioning of CO protein.

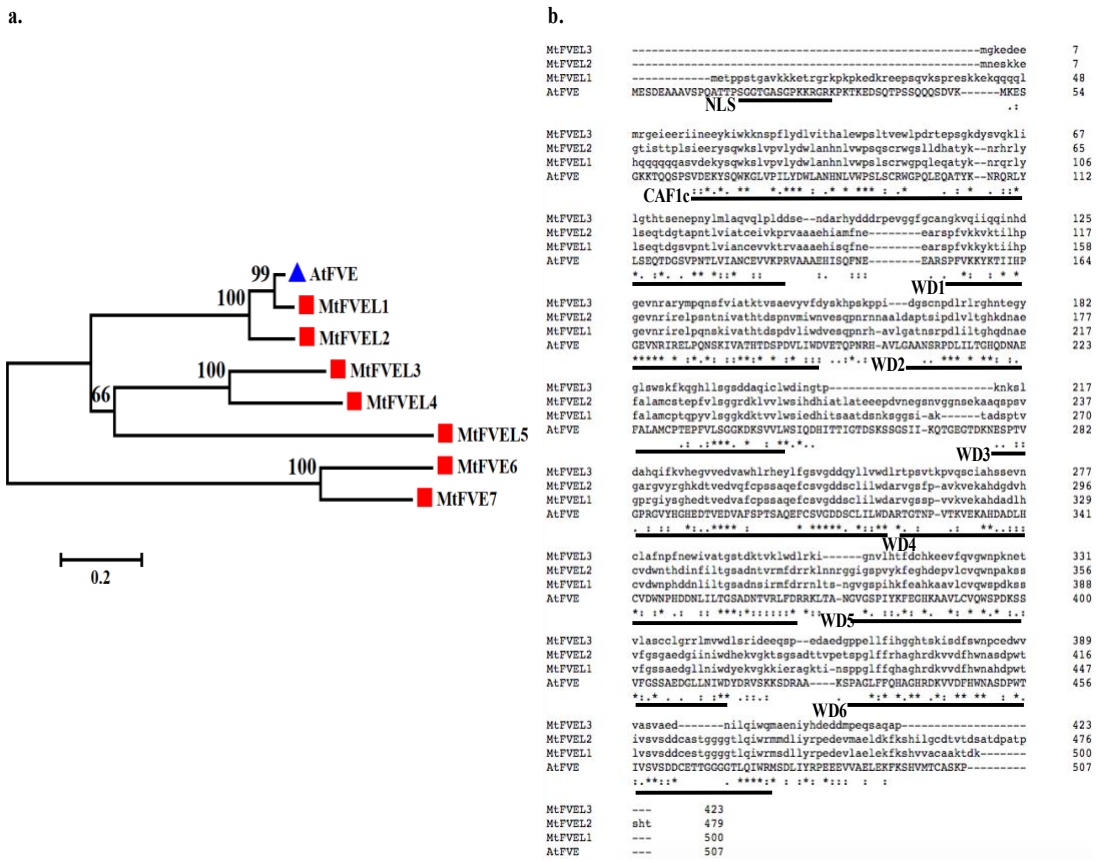


Figure 3a & 3b. **Phylogenetic relationship and protein sequence alignment of FVE-like proteins in *Medicago truncatula.***

b. The neighbor-joining (NJ) tree includes 7 FVEL proteins from *Medicago truncatula* and *Arabidopsis* FVE protein. The numbers shown next to the branches are the bootstrap probabilities from 1,000 replications.

c. Whole protein alignment of MtFVE1,2 and 3 with AtFVE. * represents conserved amino acid residues; • represents the identical and similar amino acid residues. The conserved NLS, CAF1c and WD40 domains are highlighted.

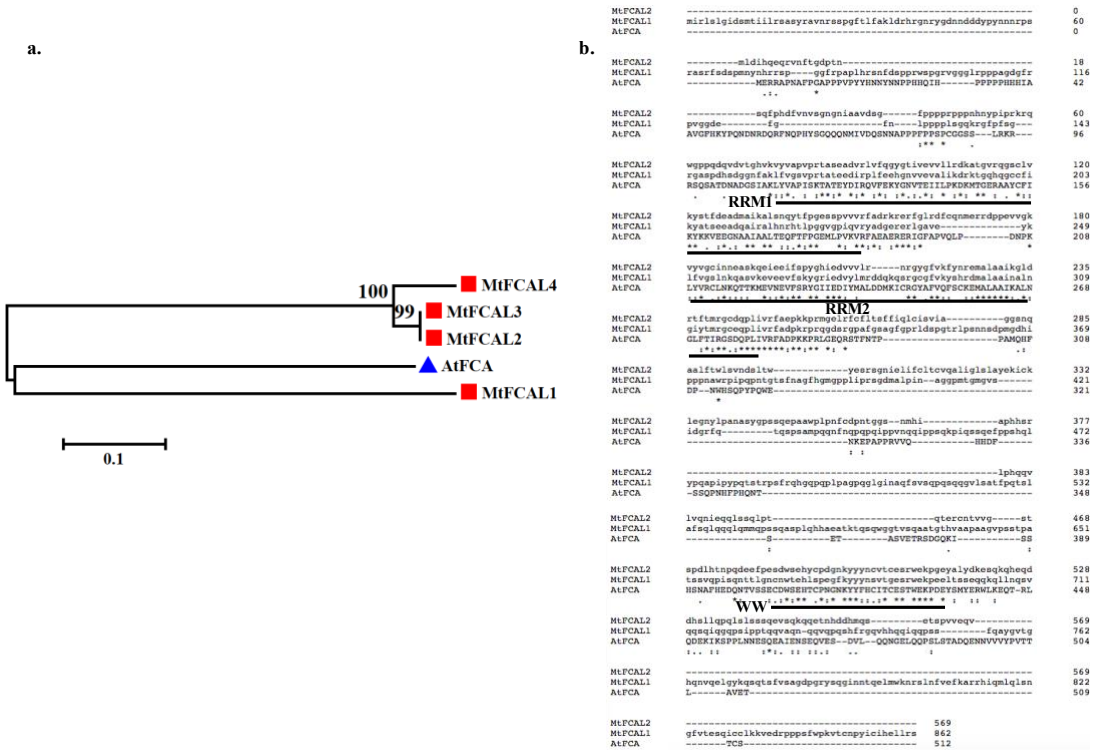


Figure 4a & 4b. **Phylogenetic relationship and protein sequence alignment of CO-like proteins in *Medicago truncatula.***

b. The neighbor-joining (NJ) tree includes 4 FCAL proteins from *Medicago truncatula* and *Arabidopsis* FCA protein. The numbers shown next to the branches are the bootstrap probabilities from 1,000 replications.

c. Whole protein alignment of MtFCA1 and 2 with AtFCA. ***** represents conserved amino acid residues; **•** represents the identical and similar amino acid residues. The conserved RRM1&2 and WW domains are highlighted.

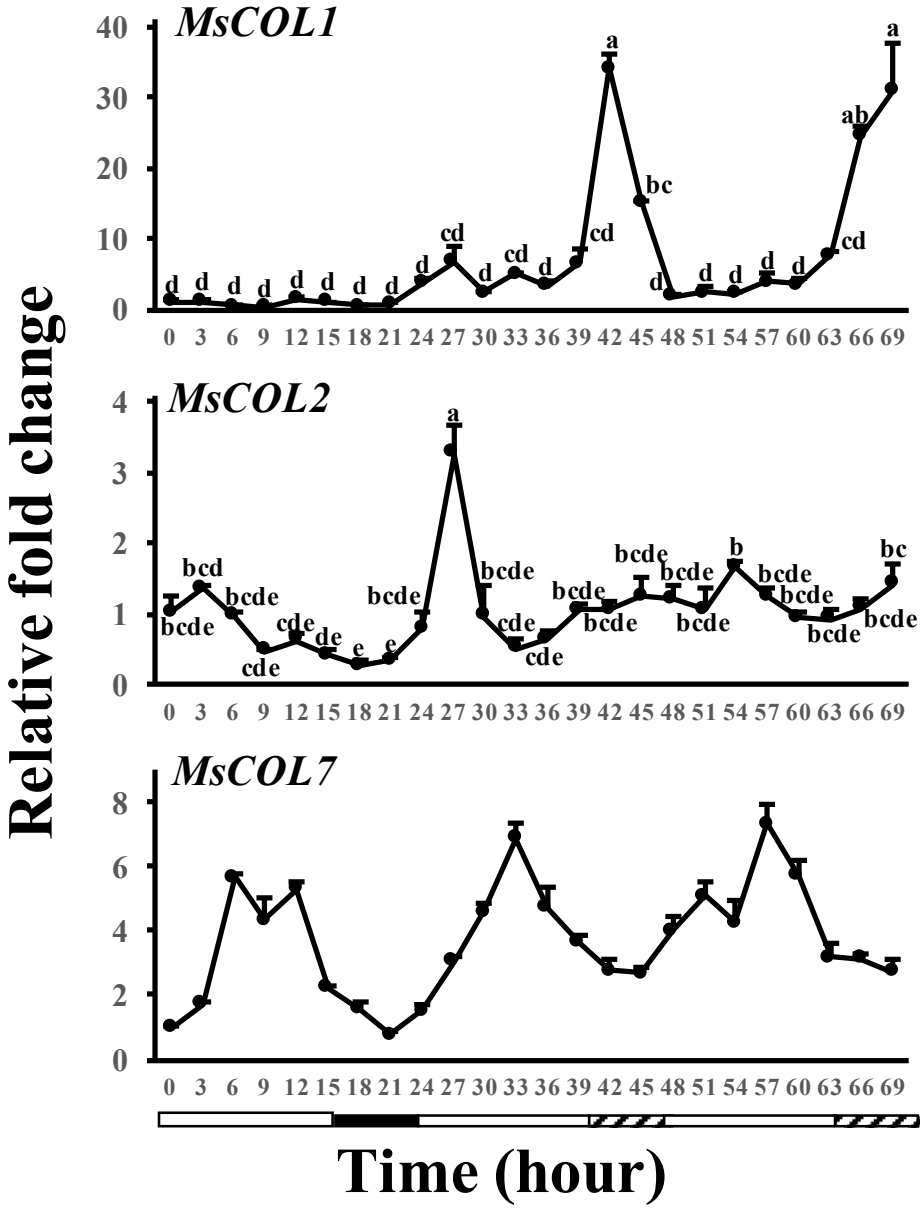


Figure 5. Circadian regulation of *COL* genes in alfalfa. Young shoots were harvested from SD201 plants of the same age every 3 h after dawn. Relative expression levels are shown for a 24h period under long-day (LD) conditions, followed by 48h under continuous light (LL) conditions. The values represent the mean fold change ± SE (n = 3) when compared with the transcript level at 0 h. Data points with different letters are significantly different ($p < 0.05$).

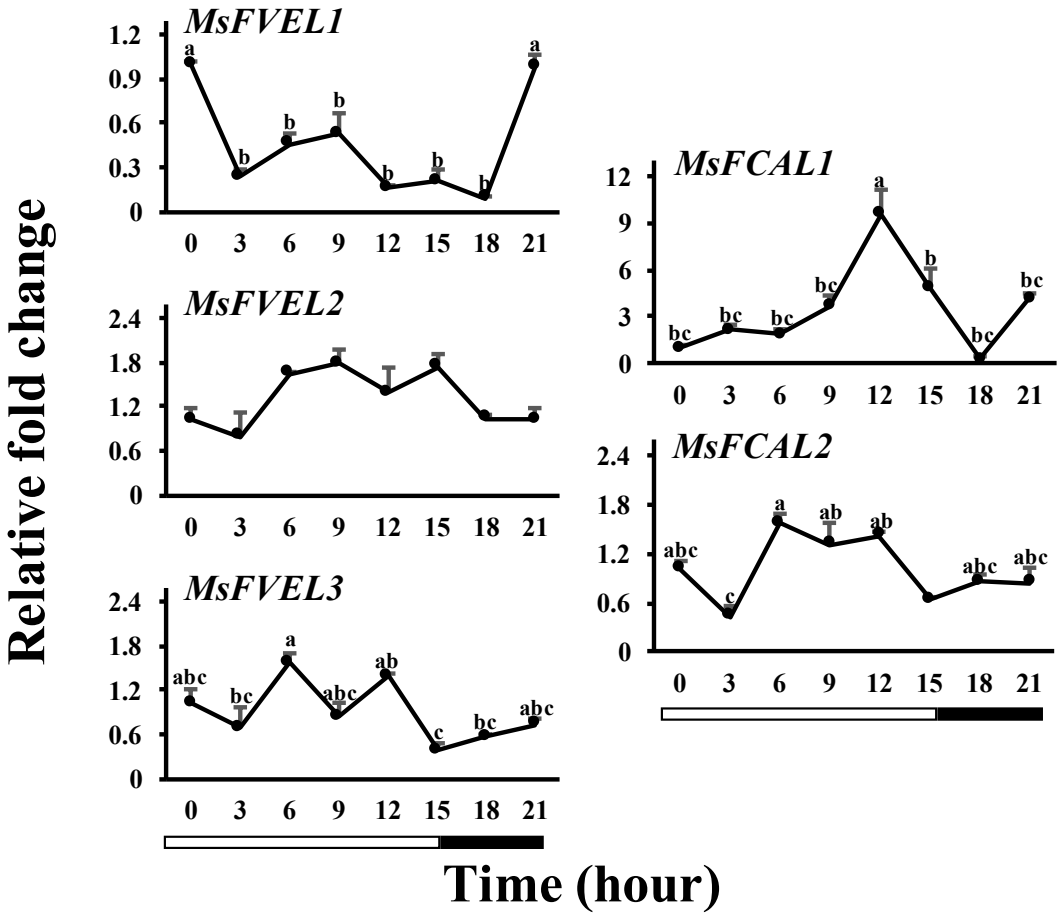


Figure 6. Diurnal regulation of *FVEL* and *FCAL* genes in alfalfa. Young shoots were harvested from SD201 plants of the same age every 3 h after dawn. Relative expression levels are shown for a 24h period under long-day (LD) conditions. The values represent the mean fold change ± SE (n = 3) when compared with the transcript level at 0 h. Data points with different letters are significantly different ($p < 0.05$).

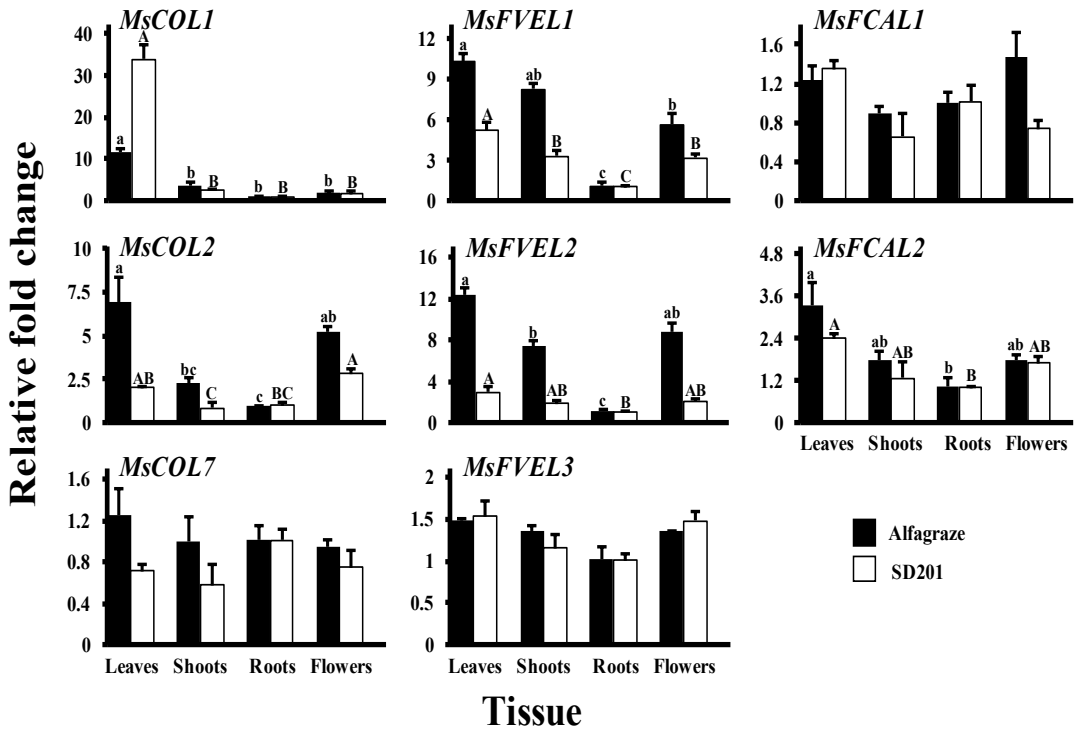


Figure 7. Expression analysis of *MsCOLs*, *MsFVELs* and MsFCAL*s* in different tissues; leaf, stem, root and flower. Tissues were harvested from SD201 and Alfagraze plants for gene expression analysis. The values represent the mean fold change ± SE (n = 3) when compared with the transcript level in leaf. Bars with different letters are significantly different ($p < 0.05$).

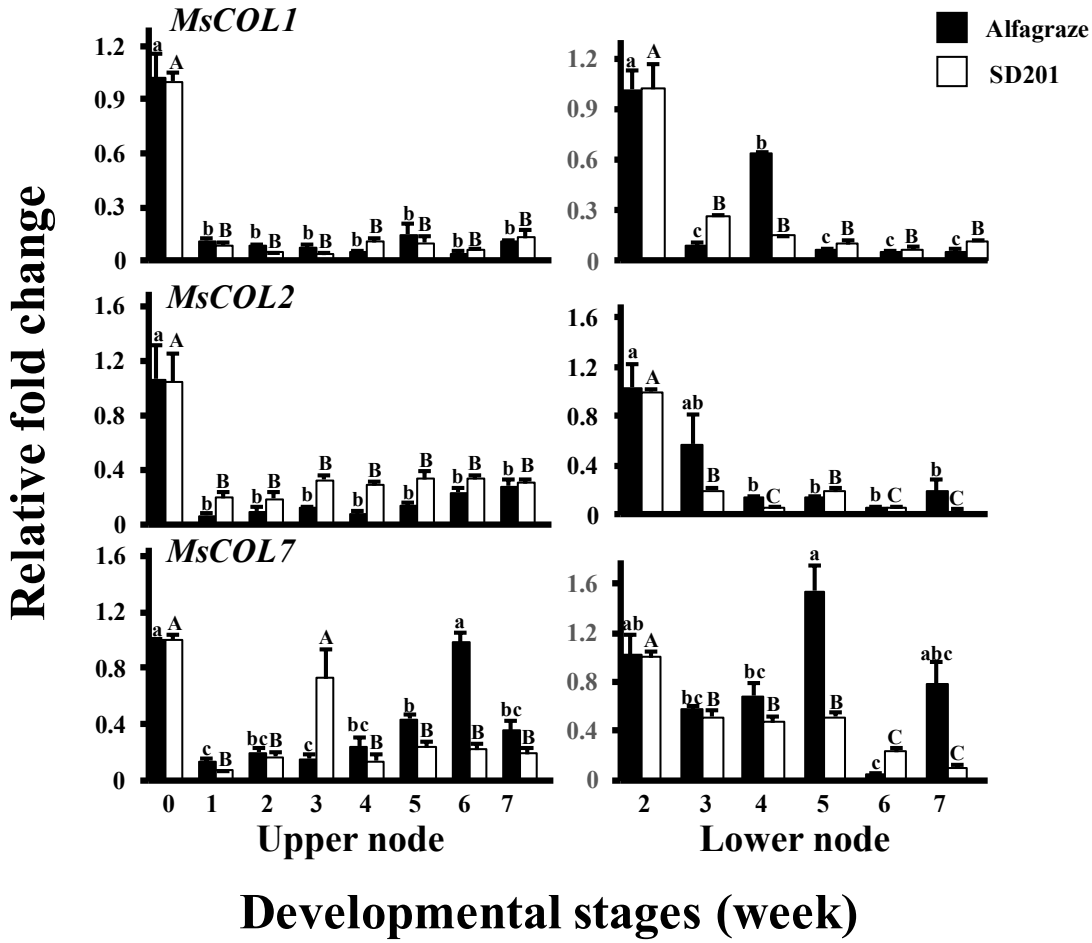


Figure 8a. Expression analysis of the *MsCO-like* genes at different developmental stages: samples were harvested from SD201 and Alfagraze plants every week starting from the seventh day after germination. For the first two weeks (Weeks 0&1), the whole seedlings were harvested, and from the second week onward (Weeks 2–7), the young shoots from the upper node and shoots from the lower nodes were harvested. The transcripts were quantified by qRT-PCR, and the values represent the mean fold change ± SE (n = 3) when compared with the transcript level at Week 0. Bars with different letters are significantly different ($p < 0.05$).

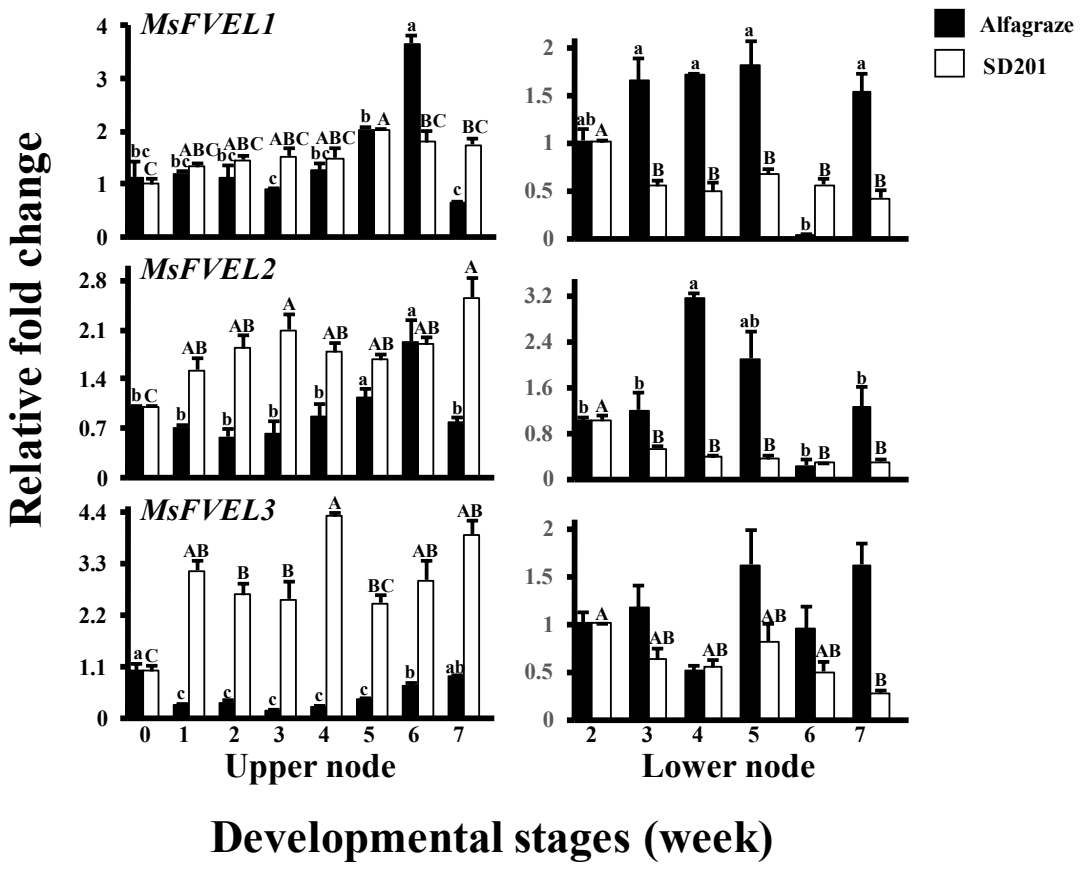


Figure 8b. Expression analysis of the *MsFVE-like* genes at different developmental stages: samples were harvested from SD201 and Alfagraze plants every week starting from the seventh day after germination. For the first two weeks (Weeks 0&1), the whole seedlings were harvested, and from the second week onward (Weeks 2–7), the young shoots from the upper node and shoots from the lower nodes were harvested. The transcripts were quantified by qRT-PCR, and the values represent the mean fold change ± SE (n = 3) when compared with the transcript level at Week 0. Bars with different letters are significantly different ($p < 0.05$).

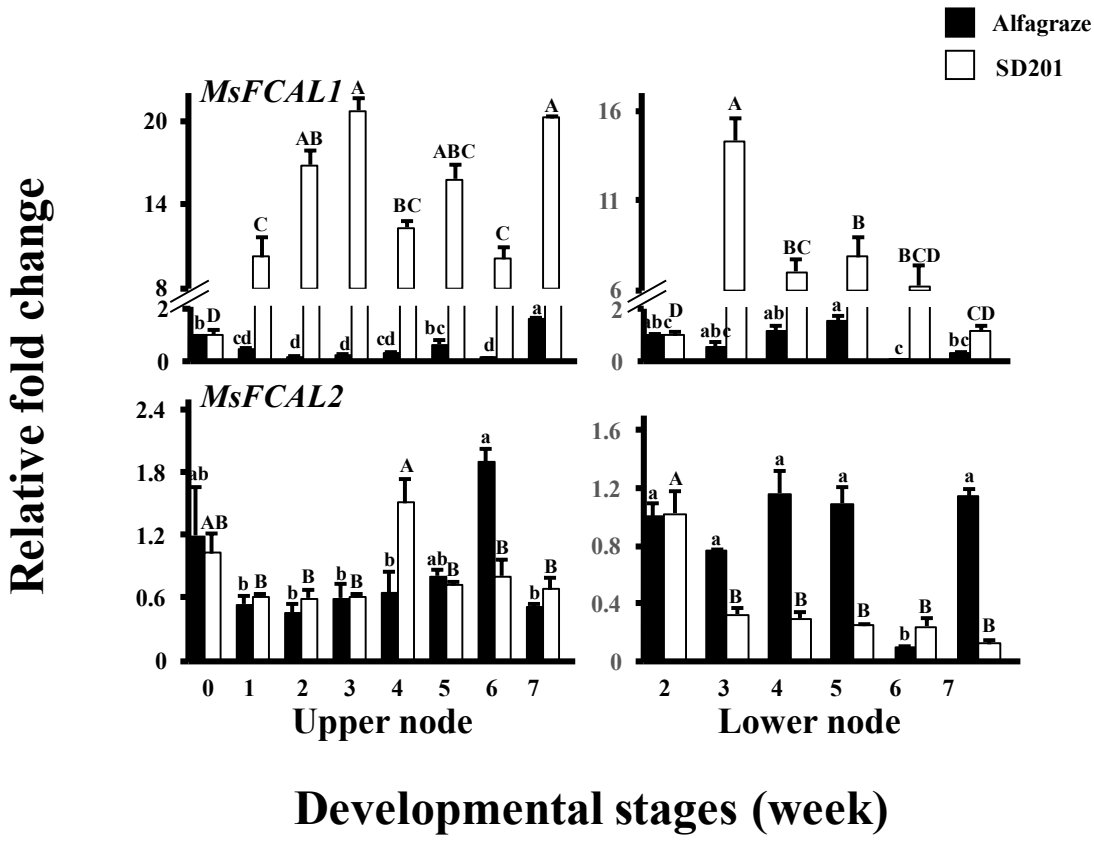


Figure 8c. Expression analysis of the *MsFCA-like* genes at different developmental stages: samples were harvested from SD201 and Alfagraze plants every week starting from the seventh day after germination. For the first two weeks (Weeks 0&1), the whole seedlings were harvested, and from the second week onward (Weeks 2–7), the young shoots from the upper node and shoots from the lower nodes were harvested. The transcripts were quantified by qRT-PCR, and the values represent the mean fold change ± SE (n = 3) when compared with the transcript level at Week 0. Bars with different letters are significantly different ($p < 0.05$).

Table S1. Gene specific primer sequences used for qPCR analysis and their corresponding length, product length and melting temperature (Tm).

Primers	RT- primer Sequences	Length (bases)	Product size (bp)	Tm(°C)
MtFVE1_F	TCTAATGGGGTTGGGTCTCCTAT	23	172	54.7
MtFVE1_R	TTTTTCCAGCTCGCTCTATCTTC	23		53.9
MtFVE2_F	AAGGCCATGATGAACCAGTCCTCT	24	191	59.9
MtFVE2_R	ATCCCTATGCCCTGCATGACGAAA	24		60.3
MtFVE3_F	TTTGCTTAGTGGTTCTGATGATG	23	124	51.4
MtFVE3_R	AGCCACGTCTTCCACAACACC	21		56.2
MtFCA1_F	CCACAATCCCCTTCTCAGTTAGC	22	199	55.8
MtFCA1_R	GCGATACAGTCCCACCCCATTG	23		59.9
MtFCA2_F	GGAGCGAACATTACTGCCCTGAT	23	127	58.2
MtFCA2_R	TCTTGCTCGTGTTGCTTCTGTGA	23		57.4
MtCO 1_F	AAACCGCCACCGAACACGAACAT	23	129	63.1
MtCO 1_R	CCGGCACAACTCCAACATCCATC	23		62
MtCO 2_F	TTCCTTTAACTATCCATCACATT	23	196	47.2
MtCO 2-R	CCCTTGGTTACTACTATTCACAC	23		47.1
MtCO 3_F	GGCGCCGACAGTGTACCAGTTAT	23	190	59
MtCO 3_R	CAGCCCTAGCCTTCCTTGATTC	22		56
Mt Actin_F	AGGCTCCACTCAATCCTAAAGCCA	24	168	59.8
Mt Actin_R	ACCCTTCGTATATGGGCACTGTGT	24		59.8

CHAPTER 5: CONCLUSIONS AND FUTURE DIRECTIONS

5.1. Conclusions and Future Directions

The expression analyses and identification of genes involved in plants' response to low temperature and soil moisture content, and regulation of flowering time emphasized the need to study these genes in agronomically important crops like maize and alfalfa. Extrapolating the knowledge, gained through the studies done in the model plant *Arabidopsis thaliana*, to the crop plants is essential. The present study suggested that the key pathways, and their molecular participants, that regulate plants responses to various environmental cues are potentially conserved across different plant species, but only to a certain degree. There is an indication of divergence and greater complexity in the molecular mechanisms in response to the same cues in the crop plants. The observations made in this study placed high uncertainty on the candidate genes identified in maize and alfalfa and demands additional studies in the future:

a. Functional characterization of the potential homologues: In this study using expression analyses, the potential functional homologues were narrowed down. Now, complementation, overexpression and knockdown/knockout studies are needed to confirm the function and the role of the homologues in regulating flowering time or in the plants' response to low temperature or soil moisture content.

b. Systemic understanding of signaling and response at the molecular level: The current studies concentrated on a single component of very complex regulatory networks. Further studies are required to elucidate the complete signaling and response pathway, when plants are exposed to various environmental conditions. Experiments like Yeast 2 hybrid assays, or ChIP assays could be employed to gain further understanding of the

interacting partners and downstream genes and thus the molecular regulation of plants' responses.

c. Extensive transcriptomic and proteomic studies for new pathway discovery: Gaining in-depth knowledge into molecular biology of multiple signaling pathways in alfalfa is limited as the genome has not been sequenced yet. So, spatial and temporal transcriptomic and proteomic studies using RNA-seq analyses and shot-gun followed by MS respectively, can be used to compare the germplasms varying in stress tolerance, or RNA-seq analysis can be used to study maize roots from germplasms varying hydrotropic response. These open-end approaches allow discovery of novel molecular components or pathways in these crop plants.

d. Molecular breeding: Once the key genes are identified and their functions are confirmed, plant biologists will be able to achieve their ultimate goal of developing better performing crops, through conventional and molecular breeding, that could provide better food security and a dependable source of energy to the ever-growing population in the world.