In the quest for a cold tolerant variety

– gene expression profile analysis of cold stressed oat and rice

Angelica Lindlöf

GÖTEBORGS UNIVERSITET

Department of Cell and Molecular Biology

Dissertation Göteborg 2008, Sweden

ISBN 978-91-628-7200-7

Department of Cell and Molecular Biology
Göteborg University, Sweden

Printed by Geson Hylte Tryck AB, Göteborg, Sweden

To Zlatan, Selma and Nikita

You gave me inspiration, comfort and mental relaxation
during these years of brutal brain exercises.

†

To all my near and dear ones

Especially mum and dad,
since without you I would not be in this place,
and my brother Kenneth,
who with his teasing in younger years learned me to
never give up, a valuable quality to possess when the
prospects of success were less promising.

Abstract

Cold acclimation is a process which increases the freezing tolerance of an organism, after exposure to low, non-freezing temperatures. The acclimation ensures that cold tolerant species can endure harsh winter conditions, by preparing them to sub-zero temperatures. Cold-sensitive plants such as oat and rice have limited abilities to cold acclimate and are therefore easily damaged during winter time.

The development of more tolerant varieties by using biotechnological methods is desirable, since the yields are expected to improve due to a prolonged vegetation period. However, in order to apply such methods, more knowledge about the underlying mechanisms regulating the cold tolerance and acclimation is required. One step in this direction is to analyze gene expression data generated from cold stressed oat (*Part I*) and rice plants (*Part II*).

The focus of this thesis is, consequently, analysis of expression profiling data, which was generated using the EST sequencing and cDNA microarray technologies. The results show that both oat and rice are cold responsive, with many of the previously identified cold regulated genes having a counterpart in these species. In rice, however, the response is less dynamic than in the model organism *Arabidopsis thaliana* and this may explain its inability to fully cold acclimate.

Additionally, the work in this thesis focuses on evaluating if small-scale EST sets can be used for 'digital-Northern', in order to identify genes that are involved in the regulation of the cold stress response. The results show that small-scaled EST sets are not optimal for such an analysis, since the method detected only a portion of cold responsive genes represented in the sets. This has to due with the inherent properties of EST data and limitations in the analysis steps of the sequences.

The work also concerns the identification of *cis*-elements coupled to transcription factors prominent in the regulation of the response. Since cold acclimation is a quantitative trait the response and regulation of cold stress is under combinatorial control of several transcription factors and the results show that this should be taken into account when identifying binding sites.

"Imagine for a moment that your feet are anchored to the ground and you are standing in St. Paul, Alto Rio Senguerr, Torino or Sapporo and it is summer. You are outside and can't go inside. Now imagine having to remain in that place for the entire year; all of your life.

This is the life of a tree."

Guy, Charles (1999) Molecular Responses of Plants to Cold Shock and Cold acclimation. *Journal Molecular Microbiology Biotechnology*, 1: 231-242.

List of papers

This thesis is based on the following papers, which will be referred to by Roman numerals

I. Bräutigam, M., Lindlöf, A., Zakhrabetkova, S., Gharti-Chhetri, G., Olsson, B. and Olsson, O. (2005) Generation and analysis of 9792 EST sequences from cold acclimated oat, Avena sativa. *BMC Plant Biology* 5:18.

II. Lindlöf, A., Bräutigam, M., Chawade, A., Olsson, B. and Olsson, O. (2007) Identification of Cold-Induced Genes in Cereal Crops and Arabidopsis through Comparative Analysis of Multiple EST sets. In: Hochreiter, S. and Wagner, R. (eds.), *Bioinformatics Research and Development – First International Conference, BIRD '07*, LNBI 4414: 48-65. Springer-Verlag.

III. Lindlöf A., Bräutigam M., Chawade A., Olsson O. and Olsson B. (2008) Evaluation of combining several statistical methods with a flexible cut-off for identifying differentially expressed genes in pairwise comparison of EST sets. *Biology and Bioinformatics Insights* 2: 215-237.

IV. Bräutigam, M., Lindlöf, A., Chawade, A., Gharti-Chhetri, G., Olsson, B. and Olsson, O. (2008) Transcriptional profiling of cold stress response in rice and comparative analysis to *Arabidopsis thaliana* (manuscript).

V. Lindlöf A., Bräutigam, M., Chawade, A., Gharti-Chhetri, G., Olsson, B. and Olsson, O. (2008) *In silico* analysis of promoter regions from cold-induced *CBF*s in rice (*Oryza sativa* L.) and *Arabidopsis thaliana* reveals the importance of combinatorial control (manuscript).

Table of contents

Introduction to cold stress

Tropical vs. temperate plants

The Earth can be divided into five regions with three major climates, the tropical, subtropical and temperate latitudes (Figure 1) [1][1]. The tropical region is centered as a band around the equator with the subtropics extending from it towards both north and south. The temperate regions are located even further north and south of the subtropics towards the polar circles.

The climate in the tropical region can be divided into the dry and wet season, where rain is excessive during the wet season. Otherwise the weather is commonly hot and humid, and the day temperature rarely falls below 25°C. In the subtropical and temperate regions the weather is overall cooler and the summer season gradually transfers into winter. The summers in the subtropical regions are warmer than in the temperate regions and the winters are commonly mild with the temperature only in exceptional cases falling below 0°C. In contrast, in the temperate regions the weather in winter time can be very unpredictable, with swift changes from temperatures above zero to freezing.

Plants can be broadly classified according to their main habitat, i.e. tropical, subtropical or temperate species. Tropical plants are generally unable to tolerate and survive even mildly cold weather. However, localities at high altitudes in this region can exhibit a temperate climate and plants growing at such altitudes must be able to cope with low temperatures. Subtropical plants are more tolerant to cold temperatures and can survive short periods

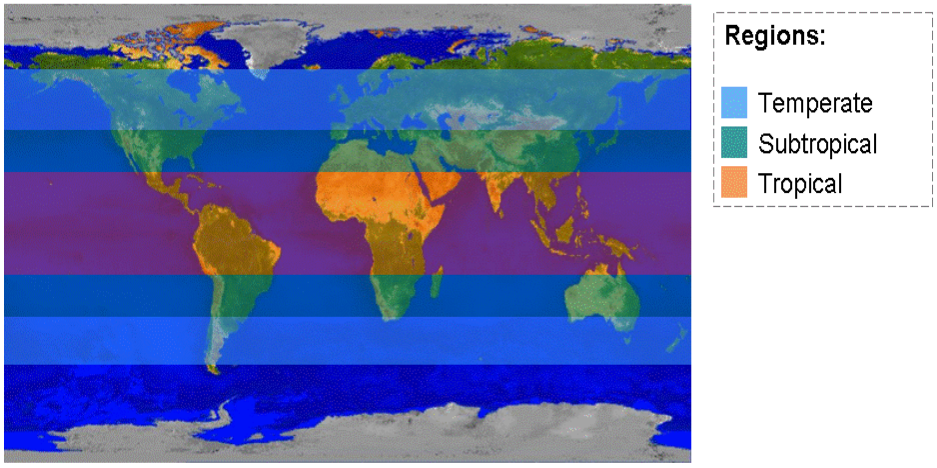

Figure 1. The Earth can be broadly divided into three major regions, the temperate, subtropical and tropical zones. Reproduced by permission from Dave Pape.

of light frost. Plants that can endure the harsh winters in the temperate regions are able to cope with freezing temperatures. The degree of tolerance, however, is highly variable among the species growing there and, additionally, can greatly vary among varieties of the same species.

Plants are forced to adapt to the prevailing environment surrounding them, since they, in contrast to many other species, are unable to migrate to more favorable localities. The survival rate in cold weather is determined by both the level as well as the duration of unfavorable temperatures and differs extensively among plant species. According to the ability to endure low temperatures, plants can be broadly classified into five major groups [2]. Rice is mainly a tropical plant and belongs to the first type. These plants are chilling sensitive and show injuries already at temperatures between 0°C and +12°C. Oat is a subtropical plant and can be found in the second group, which contains plants that are chilling-insensitive and can cope with low non-freezing temperatures. However, these plants are damaged by the slightest frost. More cold tolerant crops, such as wheat and barley, belong to the fourth group. These plants can survive temperatures as low as -30°C.

The survival ability at sub-zero temperatures that some species exhibit requires that the plant can first acclimate to unfavorable temperatures. The exposure to mild cold (temperatures slightly above 0°C) prepares the plants for more severe conditions. This process is known as cold acclimation and this topic is covered in more detail in the section 'Cold acclimation process' in this chapter.

Cold stress and crop yields

The productivity, growth and geographical distribution of important agricultural crops, such as rice, maize and oat, are severely limited by cold stress. Cold stress, which includes chilling (0-12°C) and/or freezing (<0°C) temperatures, adversely affect crop yields by causing restraints on sowing time, extensive tissue damages and stunted growth [3].

At chilling temperatures the first structural symptoms arise, such as swelling and disorganization of chloroplasts and mitochondria, reduced starch and accumulation of lipid droplets inside the cells [4]. On the metabolic level, the photosynthesis and transpiration are reduced [5-7]. When the air temperature drops below zero, ice crystals begin to form in the intracellular spaces, which can cause physical injuries. For example, ice in the membranes results in disintegration of the lipid bilayer, since the crystals do not exert the same hydrophobic forces as liquid water that is needed to maintain it [3, 8]. In addition, the formation of extracellular ice leads to a loss of water from the cells by osmosis, due to the higher water potential in

[1] Climates and Biomes: http://plantphys.info/Plant_Biology/climate.html

the intercellular spaces, which causes dehydration [3]. Consequently, the plants also suffer from drought, which will add to the physical damages.

Oat is an important commercial crop in Northern Europe, Scandinavia, Canada and the US (Figure 2). Oat yields in the Scandinavian countries are limited since only varieties planted in early spring can be used. The crop is unable to cope with freezing temperatures, in contrast to its close relatives wheat and rye, with the consequence that it can not survive the winters in the outer regions of the temperate zones [9]. However, if seeds could be planted in the autumn, the yields are expected to improve due to the longer vegetation period.

Rice is the most important staple food in the world, with more than half of the world's population depending on it as the main nutrition source (Figure 3). The species originates from tropical regions and is therefore easily damaged by low temperatures [10, 11]. The plants are especially vulnerable after sowing, during the establishment stage, where low temperature causes poor surfacing rate of seeds, but also during the reproductive stages, where low temperature can cause grain sterility. About 10% of all the localities where rice is today cultivated are subjected to low temperatures [12]. Cold stress is, together with drought and salt stress, among the major factors that constrain rice yields.

Figur 2. Oat (*Avena sativa*) field in Skåne, Sweden (June, 2007). Reproduced by permission from Olof Olsson.

Figur 3. . Rice (*Oryza sativa*) field in Kathmandu Valley, Nepal (August, 2006). Reproduced by permission from Olof Olsson.

Cold acclimation process

Plants growing in temperate regions have evolved unique traits that make it possible for them to cope and survive freezing temperatures. In response to mild cold stress, at approximately 4-6°C, a cascade of genetic reactions are triggered that greatly enhance the tolerance to later, more severe sub-zero temperatures. The tolerance to cold, as most abiotic stresses, is not a static condition but commonly varies seasonally and rapidly deteriorates at warm non-acclimating temperatures.

Cold acclimation is a set of biological processes belonging to the group of abiotic stress and stimulus associated responses. Gene Ontology [2] [13] provides the following definition and classification scheme of the phenomenon (Figure 4):

"Cold acclimation

Processes that increase freezing tolerance of an organism in response to low, nonfreezing temperatures."

[2] Gene Ontolog: http://www.geneontology.org/

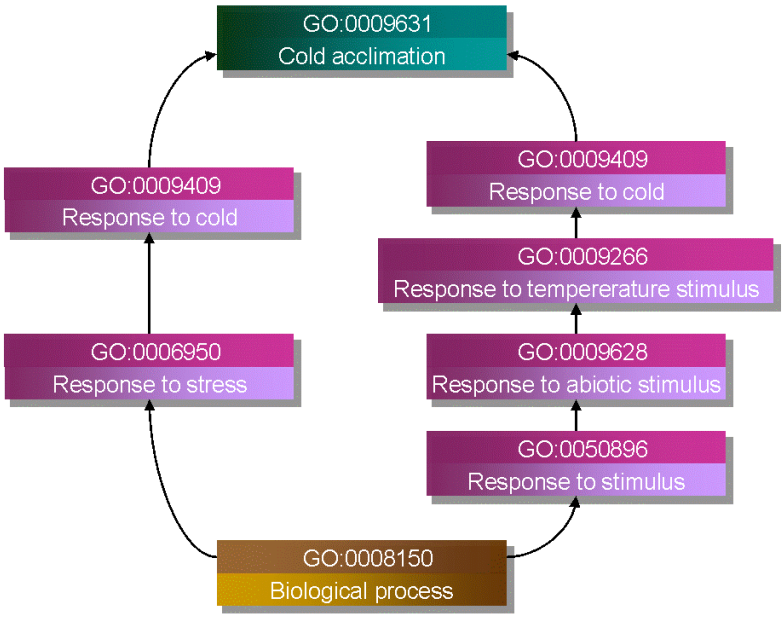

Figure 4. Gene Ontology classification scheme of cold acclimation processes.

The acclimation involves numerous physiological and molecular changes, such as alterations in the hormone balance, increase of osmolytes e.g. sugar, proline and betaine, membrane modifications and increased levels of antioxidants, as well as alterations in gene expression by a number of cold responsive genes (*COR*s) [14-17]. The main tasks of these changes are to protect the cells against freezing injuries and from the damaging effects resulting from dehydration.

Freezing temperatures result in ice crystals being formed in inter- and intracellular spaces, which may cause membrane disruption. One role of cold acclimation is to stabilize the membranes against such damages [18-23]. This is achieved by changing the membrane lipid composition, through increased levels of free sterols and glycolipids, reduction of cerebrosides as well as increased fatty acid desaturation in membrane phospholipids. Moreover, carbohydrate metabolism is an important factor involved in the protection of the tissues against freezing damage. It has been shown that the levels of sucrose, glucose and fructose increase in response to low temperatures and that these sugars have a role as cryoprotectants [24-26]. Fructose is also involved in antioxidative protection and exhibits scavenging capacities of superoxide [26]. Superoxide is a reactive oxygen species (ROS) that is toxic in high levels and there is increasing evidence that cold stress causes elevated levels of ROS [27-31]. However, plants have evolved antioxidant systems as protection against damaging effects of ROS. Cold acclimation has been shown to increase the tolerance to ROS by increasing the level of antioxidant enzymes [32-35].

The cold acclimation process in plants is primarily regulated through the signal transduction pathways that lead to the induction and enhancement of expression of *COR* and homologs of *LEA* (Late embryonic abundant) genes, commonly grouped together and referred to as *Cor*/*Lea* genes [14, 17, 19, 36]. Several of these genes are also responsive to dehydration and to the phyto-hormone abscisic acid (ABA) [37]. These genes are relatively diverse in sequence and form distinct groups regarding similarity in their amino acid sequence. However, many of them share common properties, such as being extremely hydrophilic, resistant to heat denaturation and composed largely of repeated amino acid sequence motifs. These properties are thought to enable them to protect the cells against freezing injuries by stabilizing both proteins and membranes during cold stress.

During cold acclimation, there is also a reduction in the capacity for photosynthesis, which is known as photoinhibition [38-40]. The biochemical reactions coupled to photosynthesis are inhibited by low temperatures, which cause an excess of energy that leads to an accumulation of electrons in the cells. Freezing tolerant species, such as wheat and rye, have been shown to better cope with photoinhibition than less tolerant plants and this resistance is an important factor in the acclimation [6].

Perception of low temperatures

The perception point of a decrease in temperature on the metabolic level is currently relatively uncharacterized. However, experimental studies have shown that there are three important factors involved in the initiation of the cold acclimation pathways [41-43]. Plant cells can sense cold stress through changes in the cell membrane fluidity (Figure 5). A decrease in temperature can reduce the fluidity, which causes a rigidification of the membrane. This effect is thought to activate temperature sensors located in the plasma membrane. Moreover, the acclimation is triggered by a Ca^{2+} influx into the cytosol, which is a requirement for the induction of *COR* genes [44-48]. Örvar et al. showed that the Ca^{2+} influx is dependent on the re-organization of the cytoskeleton [41]. Additionally, this re-organization was thought to serve as a link between the membrane rigidification and the Ca^{2+} influx. Consequently, the cold acclimation pathways are triggered by a change in membrane fluidity, re-organization of the cytoskeleton and the influx of calcium ions.

Calcium as a signal for cold stress

Calcium (Ca^{2+}) is the most common signal transduction element in the cells and acts as a secondary messenger [49, 50]. The elevation of the cytosolic Ca^{2+} concentration is characteristic for the response to various abiotic and

biotic stimuli. The same messenger regulates different responses by fine-tuning the frequency, duration, amplitude and/or location of the increase, which thereby creates a Ca^{2+} signature specific for each stress [51, 52]. The physical alterations that occur as a consequence of cold stress are closely followed by an influx of calcium ions. This influx is recognized by different calcium sensors, such as calmodulin, calcium-dependent protein kinases (CPKs) and calcium-sensitive phosphatases, which transduce the calcium signal into a cold acclimation signaling cascade [53].

Since prolonged elevated levels of cytosolic Ca^{2+} are toxic and may cause cell injuries, such as metabolic dysfunction and structural damage, processes that control internal Ca^{2+} homeostasis following the stimulus are needed [54]. This can be achieved by an active Ca^{2+} transport system, through, e.g., Ca^{2+} pumps located on cellular membranes. Moreover, Jian et al. made an important observation regarding the concentrations of cytosolic Ca^{2+} in a cold tolerant winter wheat and a chilling sensitive maize variety [55]. In chilled winter wheat seedlings the Ca^{2+} levels initially increased, but it was restored to base levels at normal temperature within three days. In contrast, maize was unable to achieve the same restoration under prolonged chilling and subsequently exhibited cellular damages. This result shows that cold hardy plants are able to quickly restore lower resting Ca^{2+} levels, whereas cold sensitive plants are unable to do so [55]. Further, this observation indicates that Ca^{2+} homeostasis is an important contributing factor in the ability to tolerate low temperature levels, a factor that seems to be dysfunctional in cold sensitive plants [55].

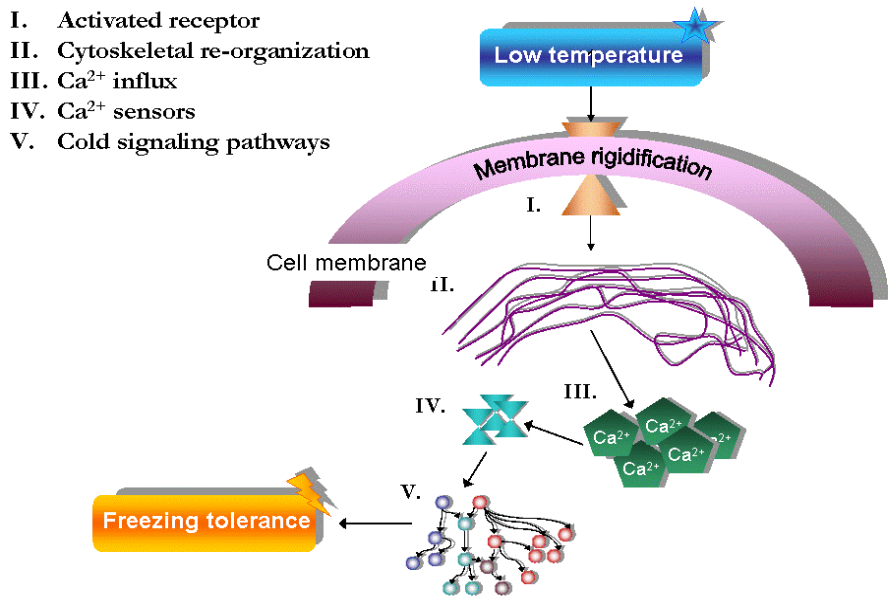

Figure 5. A drop in temperature rigidifies the cell membrane, causing a re-organization of cytoskeleton that leads to a Ca^{2+} influx. The influx activates sensors that transduce the calcium signature into a cold acclimation signal.

Secondary messengers and phosphorylation as a signal for cold stress

There is also evidence that other secondary messengers can induce Ca^{2+} signatures that impact cold stress signaling, such as reactive oxygen species (ROS), e.g. superoxide and hydrogen peroxide, and inositol (1,4,5,)-triphosphate (IP_3) [31, 52, 56]. Additionally, in a generic signaling pathway the secondary messengers modulate intracellular Ca^{2+}, which often results in the initiation of a protein phosphorylation cascade. The cascade commonly in turn activates other signal molecules, besides the secondary messengers, which can initiate another round of signaling events. For example, the plant hormones abscicic acid (ABA) and gibberillic acid (GA) act as signaling molecules and have been shown to have an impact on cold tolerance [57-61]. Moreover, the mitogen-activated protein kinases (MAPKs), which are phosphorylated and activated by MAPK kinase (MAPKK), have also been shown to be important in various stress signaling pathways, including cold stress [62, 63].

Cold acclimation and darkness

The main trigger of cold acclimation is the low non-freezing temperatures, but other factors such as the length of the daily photoperiod influence the induction of cold response. Light and temperature changes in natural environments often occur simultaneously, where the lowest temperatures are reached during night-time in winter. Moreover, since low temperatures recur annually the acclimation must begin before the incidence of the first frost event. A favorable trait for plants living in temperate regions is therefore the possibility of sensing an imminent cold period and optimizing the production of protective proteins during the night hours.

The shortening of the daily photoperiod is a strong indicator of the transition from summer to winter season, since the hours of daylight decreases continuously before the upcoming winter. Light reduction has been shown to have an impact on acquired cold tolerance and being a necessity for plants to cold acclimate, and consequently, there is an apparent coupling between light and cold acclimation signaling pathways [64-66].

The circadian rhythm/clock is a mechanism used by plants to determine the time of day [67]. Circadian signaling networks generate rhythms that maintain a period close to 24 h. This rhythm is used by plants to optimize their relationship with the environment by modulating a range of physiological and biochemical events, e.g., flowering and photosynthesis, and is known as gating. Harmer et al. showed that several hundreds of genes exhibit circadian changes in expression at warm temperatures, among those

genes that are also regulated by cold [68]. For example, the transcription factor *CBF3*, which is prominent in the regulation of cold acclimation in *A. thaliana*, was shown to peak at Zeitgeber time (ZT; hours after dawn) 4 and be expressed at its minimum at ZT16. CBF-targeted genes showed a cycle that was delayed by ~8 h from that of *CBF3*. Furthermore, Fowler et al. demonstrated that the *CBF1-3* genes, which are involved in acclimation pathways in *A. thaliana*, are gated by the circadian clock and promoter analysis of *CBF2* indicated that the gating is regulated on the transcriptional level [69].

Cao et al. showed that the *A. thaliana GIGANTEA* (*GI*) gene which regulates the circadian clock, amongst other processes, is induced by cold, but not by salt, mannitol or abscisic acid, and that mutant *gi-3* plants exhibit decreased cold tolerance and impaired acclimation ability [70]. However, there was no significant difference in transcript levels of the *CBF* genes and their target genes in wild-type and *gi-3* mutants, indicating that *GI* mediates cold response via a CBF-independent pathway.

As previously described, Ca^{2+} participates in the signaling events which lead to a development in cold acclimation. Dodd et al. demonstrated that the circadian clock can gate cold-induced Ca^{2+} signals, indicating a coupling between Ca^{2+} signaling and photoperiod [71]. In their study, cytosolic Ca^{2+} levels exhibited a 24-hour rhythm that was persistent during constant light, which was considered to be a result of circadian regulation. Low temperature induced $[Ca^{2+}]_{cyt}$ in guard cells was measured at ZT1.5, ZT6.5 and ZT11, and a peak in calcium levels was found at ZT6.5, after which there was a trough in concentration levels. There is an obvious coupling to the circadian regulation of the *CBF* genes, since they exhibit a peak at ZT4-6 and thereafter a trough to basal levels [68, 69].

During the twilight zone, there is a decrease in Red (R) and Far-Red (FR) wavelength ratio, i.e., a decrease of R and an increase of FR. Franklin and Whitelam demonstrated that plants treated with a low R/FR ratio had a much higher survival rate when transferred directly to freezing temperatures than plants treated with a high R/FR [72]. They also demonstrated that a low R/FR ratio increased *CBF* gene expression in *A. thaliana* and their target genes and that the induction of the *CBF* genes by a low R/FR ratio is gated by the circadian rhythm.

In chilling-sensitive plants, such as tomato and cucumber, several cold-induced genes display a circadian rhythm during normal long-day (16h light/8h dark) and warm temperature (>20°C) conditions, and is under the control of the circadian clock [73-76]. However, during a cold treatment the circadian expression of mRNA levels is disrupted, with the expression levels no longer oscillating in rhythms. The clock resumes upon rewarming, albeit, with an altered and out of phase clock, which causes a mistiming of the

activation of proteins and thereby a disruption in photosynthetic and cellular metabolism. This mistiming is thought to contribute to the intolerance in chilling-susceptible plants.

Cold acclimation regulatory pathways

Arabidopsis thaliana was the first plant model species with a sequenced genome and the annotation resulting from the genome project in combination with the whole-genome microarray offered by Affymetrix[3] has boosted the research on this species. Consequently, *A. thaliana* has been extensively studied during various conditions, including cold stress and acclimation, and clues to genetic regulatory pathways resulting in cold tolerance mainly emerge from studies on this species.

A general picture of the cold response in *A. thaliana* can be outlined as in Figure 6. As described previously, a drop in temperature causes a rigidification of the membrane and re-organization of the cytoskeleton [41-43]. These two physiological changes result in an influx of cytosolic Ca^{2+}, which is recognized by different calcium sensors that transduce the calcium signal into a cold acclimation signaling cascade [44-48]. The influx presumably activates multiple signaling pathways, by causing a phosphorylation of several transcription factors that are expressed in the cells during normal conditions [53, 77, 78]. These transcription factors in turn activate other transcription factors that control a number of regulons directly involved in the response [77, 79]. These regulons consist of transcription factors, other secondary messengers, signal molecules and a number of cold-responsive (*COR*) genes [14]. The transcription factors that are part of these regulons in their turn presumably control different sub-regulons, which also activate different *COR* genes [80]. Some of the activated signaling pathways are also overlapping, which adds to the complexity of the genetic regulatory network [80, 81].

Since prolonged elevated levels of cytosolic Ca^{2+} and secondary messengers are toxic, processes that control internal homeostasis following the activation are needed, i.e. negative regulation of the signaling pathways is as important as the activation in order to endure the arisen stress [54]. This regulation can be achieved at different levels, e.g., on the transcriptional level by the direct suppression of transcription factors, on the post-transcriptional level through, e.g., gene silencing by miRNAs, and on the post-translational level by, e.g., ubiquitination of genes. Consequently, the negative regulation is as intricate as the activation, which makes the regulatory network even more complex.

[3] Affymetrix: www.affymetrix.com

The following two sub-chapters summarize some of the findings regarding cold acclimation and response in *A. thaliana*. The first sub-chapter concerns the activation of cold acclimation signaling pathways, whereas the second chapter concerns negative regulation at the transcriptional level.

Activation of cold acclimation pathways

Early studies of cold acclimation in *A. thaliana* resulted in the identification of a number of genes induced by cold stress (*COR*s) in this species [14]. More thorough studies of these genes revealed that a subset of them contain the dehydration-responsive/C-repeat element (CRT/DRE motif) in their promoter regions. These two motifs are defined as 5'-*TGGCCGAC*-3'and 5'-*TACCGACAT*-3', respectively, with the shared motif of 5'-*CCGAC*-3'.

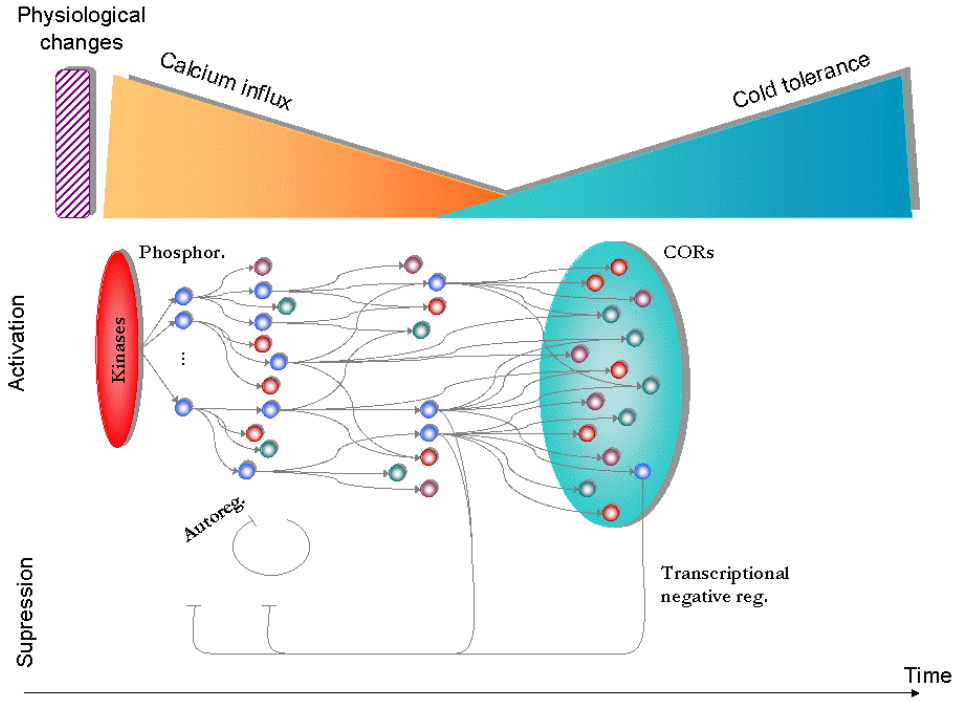

Figure 6. Schematic outline of the cold acclimation as known from studies on *A. thaliana*. The top part of the figure illustrates the physiological changes that occur in response to cold, which causes an influx of Ca^{2+}. The influx activates a number of signaling pathways, which results in an increased cold tolerance. The middle part illustrates how the signaling pathways are first activated by kinases, which phosphorylate transcription factors present in the cell. These factors thereby become activated and in their turn activate several regulons, which subsequently activate subregulons. Simultaneously many cold regulated (*COR*) genes are activated, which account for the increased cold tolerance. The bottom part illustrates the transcriptional negative control that follows the response to cold. Blue circles illustrate transcription factors, the others only that *COR*s with different function are activated.

Moreover, transcription factors that are capable of binding to and activate these genes have been identified in *A. thaliana* and are called the dehydration-responsive element binding factor1/C-repeat binding factors (DREB1/CBFs) [82].

Gene expression studies have revealed that three members of the *CBF* gene family (*CBF1-3*) are rapidly and transiently induced by cold stress in *A. thaliana* [79, 83]. They exhibit a peak in expression levels within four hours after cold treatment and thereafter a trough to basal levels at normal temperatures. Profound experimental studies have shown that the CBF regulon has a prominent role in the cold acclimation in *A. thaliana* and, moreover, homologs to the *CBF* genes have been identified in many other species, e.g., rice, wheat, barley and maize.

Through map-based cloning of the *A. thaliana ice1* mutation the transcription factor Inducer of CBF Expression 1 (ICE1) was identified and shown to regulate the expression of the *CBF3* gene, but not any of the other cold-induced *CBF* genes in *A. thaliana* [77]. ICE1 is a MYC-like bHLH protein that potentially binds to the consensus recognition site for bHLH proteins, *CANNTG*, which is present in the promoter of *CBF3*. Chinnusamy et al. also demonstrated that ICE1 is expressed in the cells during normal conditions and that phosphorylation of the TF is required for the binding to and activation of *CBF3* [77]. Furthermore, Miura et al. showed that sumoylation is also critical for the activation of ICE1 [84].

The molecular analysis of the *CBF2* gene promoter showed that the sequence *CACATG*, which is a possible match to *CANNTG*, could be a potential binding site of ICE1 [77]. However, as stated previously, ICE1 does not regulate the expression of *CBF2* and, hence, the transcription factor(s) that activate *CBF1* and *CBF2* remain to be identified.

Microarray gene expression studies have revealed that multiple regulatory pathways are activated in addition to the CBF regulon [69, 80, 81, 85]. *ZAT12* and *RAV1* are two transcription factors that are induced in parallel with the *CBF* genes, but are thought to activate distinct, although overlapping, pathways from the CBF pathway [69]. *RAV1* follows the expression pattern of the CBF genes, showing a peak at ~ZT4 after cold treatment, whereas *ZAT12* peaks at ~ZT16 after cold treatment.

Ectopic expression of *CBF* genes activates the expression of other cold-responsive transcription factors, such as RAP2.1 and RAP2.6 [81]. *RAP2.1* contains two copies of the core sequence of the CRT/DRE elements and do not increase in expression until 4-8 h after low temperature treatment, which suggests it might be a target of the *CBF* genes. These two transcription factors are thought to control subregulons of the CBF regulon.

Transcriptional negative regulation of cold acclimation pathways

The experimental studies conducted by Novillo et al. showed that in *cbf2* mutants the *CBF1* and *CBF3* genes had higher expression levels than in wild-type plants, which indicates that CBF2 is a negative regulator of *CBF1* and *CBF3* [86]. The study also indicated that CBF3 plausibly negatively regulates *CBF2* expression, since in the *ice1* mutant the transcript level was reduced for *CBF3* (ICE1 regulates *CBF3*), but enhanced for *CBF2*.

Apart from the MYC recognition site in the *CBF3* promoter, to which the ICE1 transcription factor can bind, there are also many MYB recognition sites, *(C/T)AACN(A/G)*, present in the *CBF1-3* genes. This indicates that MYB-like transcription factors can bind to and control the expression of *CBF* genes. Agarwal and colleagues identified the transcription factor MYB15, which binds to the promoters of *CBF1-3* genes and physically interacts with ICE1 [87]. MYB15 was shown to negatively regulate the expression of the *CBF1-3* genes. The expression levels of *MYB15* were accumulated at 6-12 h after cold treatment, which is slightly after the induction of the *CBF*s.

As previously described, *ZAT12* is a transcription factor that is induced in parallel with *CBF1-3* and control a separate, however overlapping, pathway from the CBF regulon [80]. In addition, ZAT12 was shown to be involved in negative regulation of the CBF cold response pathway, since constitutive expression of *ZAT12* dampened the induction of *CBF1-3* genes in response to low temperature.

There are also transcription factors that themselves are not cold responsive, but negatively regulate the cold response. For example, the two transcription factors HOS9 and HOS10, a homeodomain and a MYB transcription factor, respectively, presumably negatively regulate the CBF-regulon, without any apparent change in transcript levels in response to cold [88, 89]. There are also other proteins, which are not cold responsive but negatively regulate the response, such as the ESK1 protein that encodes a novel regulator of unknown function [90, 91]. Mutations in the *ESK1* gene result in stronger freezing tolerance, but the genes affected by the mutations differ from those of the CBF regulon.

Scientific aims

Overall aim

The overall aim of this thesis was to increase the understanding of cold response and acclimation in general, but also in particular to investigate the response in oat (*Avena sativa*) and rice (*Oryza sativa*), by analyzing experimental data from these two species when exposed to cold stress.

Specific aims

The specific aims were to:

- Survey genes that are expressed in a winter oat variety during cold stress and investigate if these genes could be coupled to the regulation of the response, by analyzing EST sequences (ESTs) collected from a cDNA library that was based on plants stressed by cold (*Paper I*).

- Identify a limited set of genes preferentially expressed during cold stress in different crop species, by utilizing EST sets as a means of gene expression profiling (also referred to as 'digital-Northern' when EST counts are used to estimate expression values) (*Paper II*).

- Investigate in more detail if 'digital-Northern' is applicable to small EST sets (~2,000-10,000 ESTs) and if a combination of statistical tests would increase the reliability of the results when deriving preferentially expressed genes from such sets (*Paper III*).

- Investigate the transcriptional dynamics of the response to cold stress in a sensitive species and compare it to the dynamics in a more tolerant one. In this case, we chose rice (*Oryza sativa*) as a cold-sensitive species and made a comparison to *A. thaliana* (*Paper IV*).

- Identify over-represented motifs in rice and *A. thaliana*, which plausibly relate to transcription factors that are prominent in cold acclimation pathways (*Paper V*).

Bioinformatic and statistical analysis

Biological data

Gene expression profiling makes it possible to conduct large-scale analysis of gene expression levels in an organism or tissue, by measuring the activity of thousands of genes simultaneously [92]. In the cells, mRNA is produced from only a fraction of the genes that are present in the genome. When the gene is required for a specific purpose, mRNA is produced from that gene and the level of expression is increased. The opposite effect of the level is observed when the gene is not needed and therefore suppressed. Multiple factors determine when the gene is activated or suppressed, such as the time of day, its local environment and chemical signals reflecting environmental stimuli both from inside and outside the cell.

In this work two different technologies have been used to identify genes that are expressed during cold stress – the EST sequencing and the DNA microarray technology. The following two sub-chapters give an outline of these two techniques.

EST sequencing

In the genomic era that we are now facing, sequencing whole genomes is becoming more common, providing us with the full repertoire of genes present in an organism. However, although complete genome sequencing has become possible, it is still not an option for many organisms. Plants commonly have a large and complex genome, containing many repeated regions and transposable elements, which make whole-genome sequencing both expensive and complicated. Consequently there are relatively few plant nuclear gene sequence entries in the public databases.

A cost efficient and rapid alternative to whole-genome sequencing, is to randomly sequence expressed genes from a cDNA library [93-95]. Applying this technology results in a collection of expressed sequence tags (ESTs), which can be utilized for the identification of characterized as well as completely novel genes. The trade-off is, however, that it does not result in the full collection of all genes present in the genome. On the other hand, the ESTs reflect genes that are expressed during the biological process under study, which is a very valuable information source. This technique was used in Paper I, for surveying genes expressed during cold stress in oat, in Paper II for comparing with publicly available EST sets extracted from related plants during cold stress and in Paper III for evaluation of ‘digital-Northern’.

DNA microarray gene expression

The microarray technology has opened up tremendous opportunities for understanding cell function by making it possible to study how the expression levels of genes are affected by environmental stimuli [96]. Physiological changes inside or outside of an organism or cell will cause alterations in the expression patterns of some genes, as the organism or cell adjusts to the arisen change by activating or suppressing genes. These changes in expression patterns can be measured with microarrays. This technique was used in Paper IV, for monitoring the expression levels of genes activated or suppressed during cold stress in rice.

The underlying technique of microarrays is that it constitutes a large array (or chip) of short immobilized target sequences attached to the surface, where each target represents a gene [96]. The array commonly represents all or the majority of genes present in the genome. Total RNA is sampled from different cells and labeled with fluorescent dye, and the sample is thereafter allowed to hybridize with the target sequences on the array. The mRNAs will bind to the complementary target sequences and thereafter the amount of fluorescence can be measured, which makes it possible to calculate the relative abundance of mRNAs for each gene on the array.

The technology has made it possible to simultaneously measure the expression level of thousands of genes in a biological sample, which is its main advantage [96]. Additionally, the production of such data can nowadays be done relatively quickly and cost efficiently. Another advantage is that the gene expression patterns in one sample can be compared to those in another sample under relatively controlled and comparable conditions, which increases the reliability of the results.

Although the technology has many advantages, it also comes with some drawbacks [96, 97]. For example, there are many parameters in the microarray data analysis which can give different results, depending on the chosen settings. The technique also have problems in distinguishing transcripts of very low abundance and are limited to only measure the expression level of known genes that are attached to the chip, in contrast to EST sequencing that does not have such requirement.

Data Analysis

Gene expression profiling generates large amount of data, which needs to be properly analyzed in order to identify genes plausibly involved in the biological process under study. The two technologies, i.e. EST sequencing and microarrays, require different approaches and methods/algorithms for this task. In this chapter, an outline is given of the methods and software that has been used in our analyses.

EST analysis

In order to identify which genes the ESTs originate from an analysis of the sequences is conducted (Figure 7Figure). The EST analysis includes 1) pre-processing, where non-nuclear DNA is removed, 2) clustering, where the ESTs are grouped by sequence similarity, and 3) assembly, where a consensus sequence is derived for each cluster of ESTs [93]. In Paper I-III, algorithms and software developed for EST analysis were used, such as the Paracel Transcript Assembler (Paracel, Pasadena, CA) and EGassembler [98].

The generated contigs and singletons from the EST analysis can be further characterized by similarity searches to previously sequenced genes, either from the species in consideration or from related species (Figure 7). The ESTs are commonly annotated by inheriting the annotation from the most

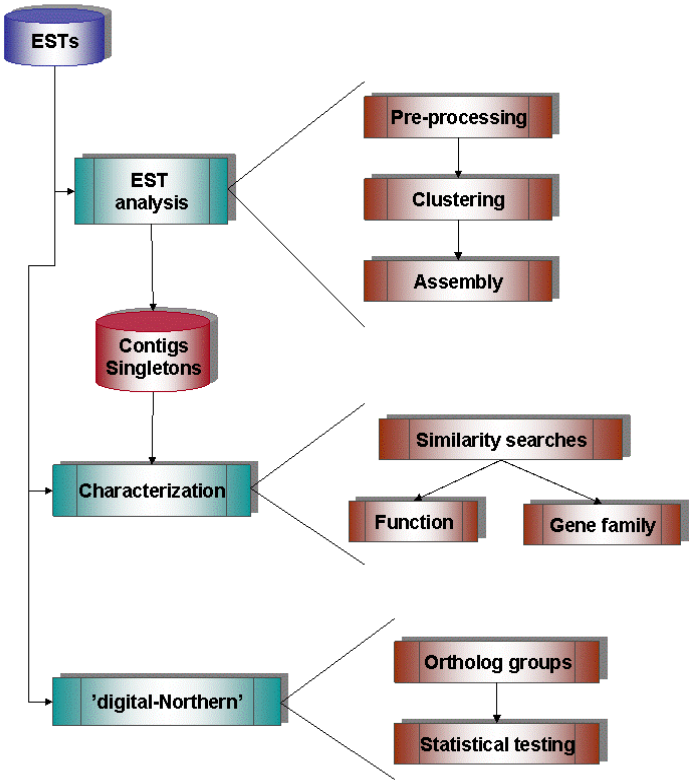

Figure 7. Outline of the analysis and characterization steps of EST sequences as well as the use of ESTs for performing 'digital-Northern'. Each step involves a number of sub-steps, such as EST analysis comprises pre-processing, clustering and assembly of the sequences, resulting in contigs and singletons, characterization comprise similarity searches against previously characterized sequences in order to elucidate the function as well gene family membership of the sequences, and 'digital Northern' , which comprise grouping of the sequences into orthologs and performing statistical tests in order to identify preferentially expressed genes in one EST set compared to another set.

similar sequence. In Paper I-III, BLAST searches [99] against characterized genes and proteins from related species were used for the functional classification of the ESTs as well as in the identification of gene origin for each EST.

In addition, methods for mapping ESTs into tentative ortholog groups (TOGs) were developed in Paper I-II, with the aim of streamlining the annotation of the ESTs as well as the identification of cold regulated genes (Figure 7). This approach made it possible to compare the expression value of each TOG in one EST set to the value in another set. In this way, preferentially expressed genes during cold stress could be identified, by comparing the expression value in the cold stress set to that in a control set.

The approach of using the cognate frequencies of gene transcripts from unbiased cDNA libraries (e.g., ESTs or SAGE tags) as an estimation of gene expression level is commonly referred to as 'digital-Northern' [100-102]. The identification of differentially expressed genes using 'digital-Northern' is commonly done by applying statistical significance tests on the data. In Paper III, several tests were applied and evaluated, e.g., Fisher's exact test and the χ^2 test, in order to optimize the identification of cold regulated genes.

Microarray gene expression analysis

There are some issues that need to be considered before making any sophisticated analysis of the results from microarray experiments. First, the expression levels have a high degree of variability from experiment to experiment, due to the random and systematic errors inherent in the microarray analysis process. Second, the number of samples is usually very small relatively to the large number of variables, which means that traditional statistical techniques cannot be used to a large extent. In order to handle these problems several pre-processing steps of the data have been developed and there is now a more or less standardized way of preparing the data for further analysis [96]. These steps include normalization to remove external influences, e.g. background intensity and relative fluorescence intensities, quality control of e.g. spots, microarray images and RNA samples, and filtering to reduce the number of genes to analyse, such as removal of uninformative genes (Figure 8). In Paper IV, the software GeneSpring version 7.3[4] was used for pre-processing the data as well as for deriving differentially expressed genes.

After deriving a set of differentially expressed genes the work begins with analyzing these in more detail. In Paper IV, the identified genes were first classified into functional groups, by utilizing functional annotation provided by MIPS[5]. The categorization gave an overview of functions required in the

[4] Agilent Technologies, www.agilent.com/
[5] MIPS, http://mips.gsf.de/projects/funcat

response to cold stress and made it possible to make an overall comparison of functions in a cold-sensitive and a tolerant species.

In Paper IV, we further classified the genes into families by utilizing the annotation in the GreenPhyl database [103]. This database contains a clustering of the full repertoire of protein sequences from rice (*Oryza sativa* cv. *Japonica*) and *A. thaliana*. The identified differentially expressed genes were mapped to either a gene in rice or *A. thaliana*, based on sequence similarity, and inherited the classification of the best match. This gave us the opportunity, in a relatively efficient way, to identify genes that are highly interesting regarding the regulation of the cold response, e.g. transcription factors and plant hormones.

Further, clustering of gene expression profiles makes it possible to identify co-expressed groups, which plausibly are also co-regulated by the same transcription factor(s) [104]. It also makes it possible to identify typical temporal or spatial expression patterns during specific conditions. In Paper

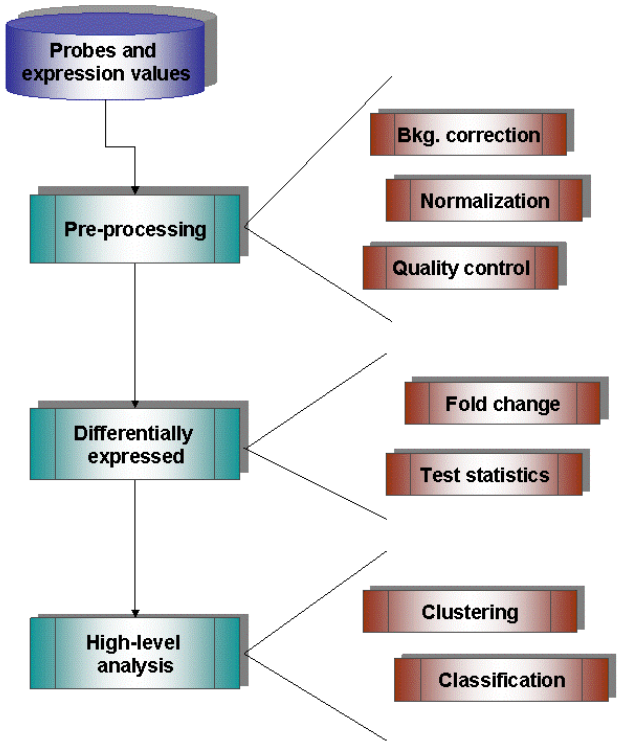

Figure 8. Outlines the steps in the analysis of microarray data. Each step involves a number of sub-steps, such that pre-processing comprises, amongst others, background correction, normalization and quality control of the expression signals. Thereafter comes the identification of differentially expressed genes, by calculating the fold change as well as applying different statistical tests on the data, and high-level analysis, by clustering the differentially expressed genes based on expression values and characterization of those genes by gene family classification. The sub-steps may include other aspects as well.

IV, clustering was used to derive an overall picture of the dynamics in the cold-sensitive species rice in response to cold stress and compare it to the dynamics in the more tolerant species *A. thaliana*. The clustering of gene expression profiles was conducted by using the QT-clustering algorithm available in the GeneSpring software[4].

Phylogenetic analysis

Phylogenetics relates to the study of the evolutionary relationships among a group of species [105]. Closely related organisms tend to be highly similar in their protein and gene sequences, whereas distantly related organisms are more dissimilar. With the aid of sequence data, it should therefore be possible to derive the relatedness between the species.

Several algorithms have been developed in order to derive the phylogeny for a group of species. The algorithms are based on slightly different assumptions, but all aim at deriving a phylogenetic tree that represents the evolutionary relationship based on a set of sequences [105]. In Paper I, IV and V, phylogenetic trees were derived using either the Neighbor-Joining or Maximum-likelihood methods implemented in the MacVector 7.2.2[6] and PHYLIP (the PHYLogeny Inference Package) software[7], respectively.

Over-represented transcription factor binding sites

The regulation of gene expression in eukaryotes is accomplished by the binding of transcription factors to short *cis*-elements, located in the promoter region of the gene. One approach to identify such elements is to derive over-represented motifs among a set of plausibly co-regulated genes when compared to a background [106]. In Paper V we derived over-represented motifs among a cluster of cold-responsive genes, which had been grouped together based on similarity in their expression profiles.

The number of genes in a cluster that a motif is occurring in as well as the number of genes in the remaining genome can easily be counted. These numbers can be put in a contingency table, to which statistical tests can be applied with the purpose of testing whether there is a significant difference in the proportions (Figure 9). In Paper V, Fisher's exact one-sided test [107] was used for deriving significantly over-represented motifs in a cluster when compared to the remaining genome. In this case, the number of occurrences T of each motif was also considered, so that a motif may not be over-represented if it occurred at least one time (T=1), but when regarding at least two times (T=2) it became over-represented.

[6] MacVector: http://www.macvector.com/
[7] PHYLIP: http://evolution.genetics.washington.edu/phylip.html

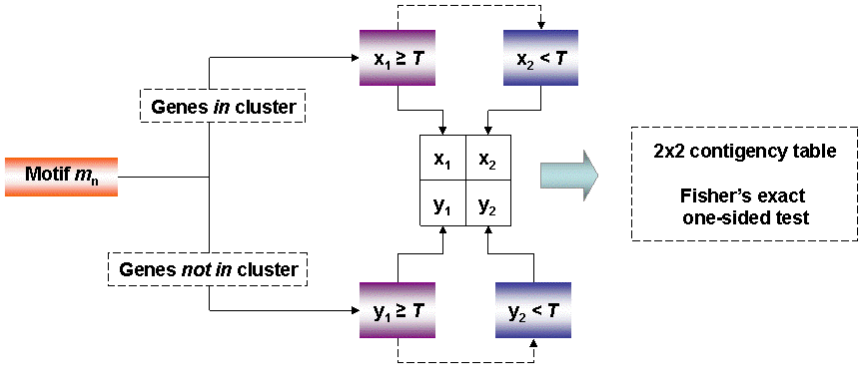

Figure 9. Representation of the number of motif occurrences in a 2x2 contingency table. m_n, the motif in consideration; x_1 and y_1 the number of genes in a cluster and the remaining genome, respectively, having at least T occurrences of the motif, x_2 and y_2 the number of genes in a cluster and the remaining genome, respectively, having less than T occurrences of the motif.

Part I

Cold response in oat

Oat as a cold hardy plant

Winter oat varieties (i.e. varieties sown in the autumn) have a better ability to withstand low temperatures than spring varieties (i.e. varieties sown in the spring), although their overall cold tolerance is not as extensive as that of more tolerant crops [108]. Regarding freezing tolerance (temperatures <0°C) rye is the most tolerant crop and oat is the least tolerant one [109]. In southern and mid-Europe winter oat varieties can be sown in the autumn, due to the rather mild winter seasons in those areas. In such locations the plants overwinter and can thereafter be harvested the next spring. Overwintering crops provide higher yields and are therefore sought after by the farmers.

Moreover, oat is a commercially interesting crop due to its high-energy grain, comparatively low demand of insecticides, fungicides and fertilizer, since it has an overall high tolerance to diseases and a low requirement for nourishment. It is also an interesting crop regarding the functional food area, since it has many qualities which positively affect the health [110-112]. However, in Northern Europe the possibility of sowing winter oat in the autumn is highly limited due to a harsher climate in this area. Consequently, the development of a winter oat that is suitable for this area is of high priority. Since cold acclimation and hardening is a complex quantitative trait with cross-talk to other abiotic stresses, traditional plant breeding programs with the aim of improving cold tolerance have so far been of limited success [14, 113, 114]. Therefore, the application of biotechnology methods appears to be a promising alternative.

In order to apply biotechnological methods in the development of a cold tolerant oat variety, more knowledge is required about the mechanisms regulating the tolerance in oat as well as the identification of candidate genes possessing a function that presumably will improve the tolerance. One step in this direction is to analyze gene expression data generated from cold stressed oat plants and compare this data with data produced from non-stressed plants as well as related crops.

EST gene expression analysis

Since oat is a cold-sensitive plant, the question was whether the species contains genes that can be coupled to the regulation of the stress response. Previous studies on cold stress have shown that cold-sensitive plants do contain cold-regulated genes in their genomes. Consequently, the first aim

was to survey genes that are expressed in a winter oat variety during cold stress and investigate if these genes could be coupled to the regulation of the stress response (Paper I).

Gene expression analysis

In paper I, we sequenced 9,792 transcripts from a cold stressed (4°C) winter variety (Gerald) incubated in the dark (for further details of the experimental protocol, see Paper I). The sequencing resulted in 8,508 high-quality ESTs, which were assembled using the Paracel Transcript Assembler (Paracel, Pasadena) into 1,100 contigs and 2,616 singletons. After removing ESTs originating from non-nuclear genes as well as redundant sequences, we arrived at a candidate gene set containing 2,800 contigs and singletons. This final set was denoted as the AsCIUniGene (*Avena sativa* Cold Induced UniGene) set.

Since very few sequences are currently available from oat, this largely leaved out the option of similarity searches against oat itself. Our annotation of the ESTs therefore relied on characterized genes and proteins from related species such as *A. thaliana*, rice, wheat and barley, using BLAST searches [99] against the sequences.

Functional classification

In order to establish whether any genes in the AsCIUniGene set could be coupled to the regulation of cold stress, we first classified each gene into a functional group (see Figure 3 and Table 3 in Paper I). The functional classification was based on homology, using the inherited registered functional class in the Munich Information Centre for Protein Sequences (MIPS) *Arabidopsis thaliana* database (MAtDB)[8] [115] of the best BLASTx similarity hit against the full collection of *A. thaliana* proteins. A total of 91.3% AsCIUniGene sequences could be annotated in this way. The annotation revealed that four classes of functions were particularly prominent in the candidate gene set and related to cold acclimation: “Cell Rescue, Defence and Virulence", "Cellular Communication/Signal Transduction Mechanism", "Metabolism" and "Transcription". In total 931 (~33%) genes had been classified into one of these groups, which gave an indication of the presence of stress related genes with these types of function.

To increase the resolution of the functional classification and to improve the identification of putative cold-regulated genes we constructed a database of proteins previously characterized as related to cold stress. The database consisted of 545 entries and 398 (14.2%) of the genes in AtCIUniGene showed significant similarity (BLASTx search using an *E*-value cutoff of 10^{-10}) to a protein in the database.

[8] MIPS: http://mips.gsf.de

Gene classification

We also developed a method for classifying each EST into an orthologous group, by means of the KOG (Clusters of Eukaryotic Orthologous Group of Proteins) database [116], annotation from MAtDB [115] or the annotation of the best homolog match ($E \leq 10^{-10}$) from a BLASTx search against the non-redundant (nr) protein database at NCBI. The aim was to map each EST to a KOG, using BLASTx searches against the full collection of proteins from *A. thaliana*. However, not all proteins from *A. thaliana* were represented in the database, which meant that not all ESTs could be mapped. Therefore, the classification of the ESTs had to be complemented by using the annotation of *A. thaliana* proteins in MAtDB. Moreover, there were also some ESTs that did not receive a significant BLASTx match against an *A. thaliana* protein and these were further classified using the annotation of the best homolog match in the nr-database.

As a comparison to the cold stress set, 2,189 EST sequences from a non-induced oat leaf library were downloaded from dbEST[9] and the sequences were analyzed in the same way. We thereafter examined the 20 most frequent ESTs in the cold stress set and compared their expression values to the values in the non-induced leaf library. The results can be viewed in table 2 in Paper I. Several of the genes in the cold induced library were more abundant than in comparison to the control library. For example, the cold-induced COR410 (Wcor410) is a dehydrin [117], which is expressed during water-deficiency and cold stress, and the cold-responsive LEA/RAB-related COR protein (Wrab17), which belongs to group-3 of LEA-proteins, has previously been established as induced by cold [118]. More interestingly, these genes were not expressed in the library grown under normal conditions.

Oat CBF genes

Among the genes in AsCIUniGene, four of them could be identified as *AtCBF* homologs. As described in the section 'Cold acclimation regulatory pathways' in the introductory chapter of the thesis, the AtCBF1-3 TFs are prominent in the regulation of cold acclimation pathway(s) in *A. thaliana*, and, in addition, the genes are rapidly and transiently induced by cold stress in this species [17, 83]. Sequence analysis of the derived protein sequences of the *AsCBFs* revealed that they contain an AP2 DNA-binding domain (see Figure 5 in Paper I), which is a characteristic of CBF proteins [119].

CBF genes have also been identified in other species, such as rice, barley and wheat. A phylogenetic study of the four identified AsCBFs as well as the AtCBFs and CBFs from rice (Os), wheat (Ta), barley (Hv), tobacco (Nt) and rye (Sc) revealed that monocot and dicot CBFs are separated into two different branches (Figure 10). The AsCBFs are spread out in the monocot

[9] NCBI, Expressed Sequence Tags Database: http://www.ncbi.nlm.nih.gov/dbEST

branch, but three of the AsCBFs are most similar to CBFs in barley, which is a somewhat unexpected result since barley is among the most cold tolerant crops.

Conclusions

Methods and algorithms for pre-processing, clustering and assembling EST sequences have been developed during a long time and it therefore nowadays exists an accepted procedure for performing these steps, which is commonly termed as 'EST analysis'. However, the next step of annotating the sequences, in order to streamline the identification of key genes involved in the process under study, has been less studied and developed. In this paper we utilized functional and gene family classification in order to do so. We worked with slightly different strategies during the work, such as inheriting the classification of a previously characterized gene/protein and the mapping to an orthologous group. However, they are all based on the

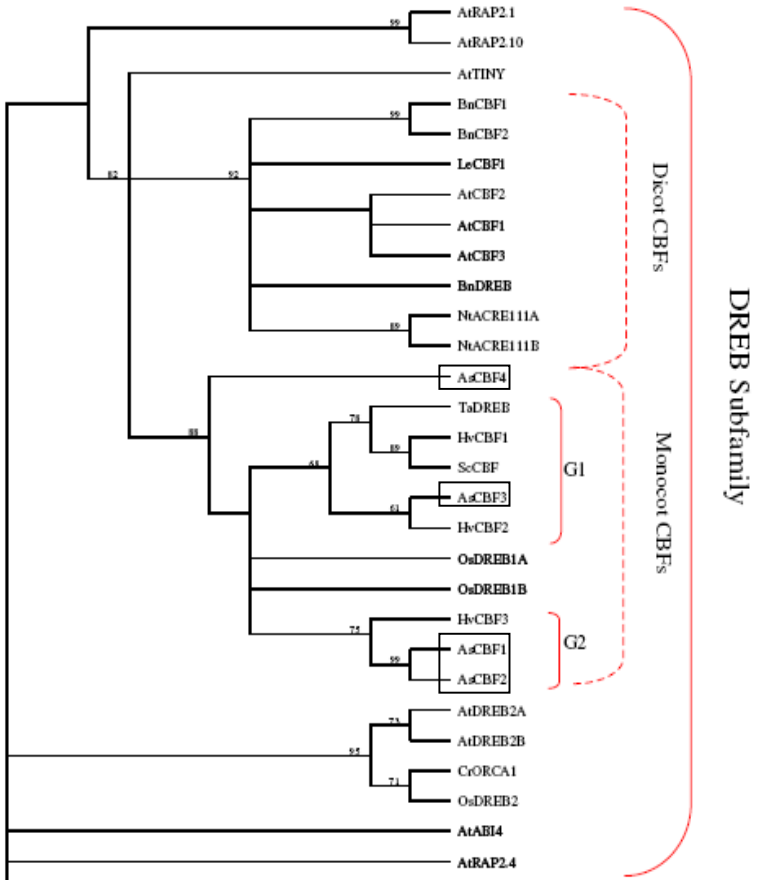

Figure 10. Phylogenetic tree of CBF factors in oat (As), barely (Hv), wheat (Ta), rye (Sc), tobacco (Nt), Tomato (Le), *A. thaliana* (At) and rice (Os). The figure shows a Neighbour-joining tree based on the AP2-domain in the CBF factors. This is an excerpt from figure 5 in Paper I.

same underlying strategy, using the annotation of the best sequence similarity hit against a gene/protein in a more characterized species. In this case, *A. thaliana* was used as a reference species, since it is one of the major model species among plants and currently the most characterized plant species. This strategy led to the identification of the four *CBF* genes in oat. However, the work also showed that this approach did not result in a classification of all sequences from oat. A development of the approach that includes other related species should therefore be sought after.

Finally, based on the presence of cold-regulated genes and *CBF* genes in particular in the EST set, we conclude that the winter oat variety Gerald contains genes in its genome that can be coupled to cold stress and the regulation of the response.

Comparative analysis of multiple EST sets

Several of the closely related species to oat are much more cold tolerant, such as wheat, barley and rye. In Paper II, we extended the comparison study made in Paper I, by comparing the ESTs derived in Paper I to other publicly available EST sets from related species as well as to sets from unstressed and etiolated plants. The comparative study aimed at identifying genes induced by cold in the related species as well.

Cold stress EST sets from rice (*Oryza sativa*), wheat (*Triticum aestivum*) and barley (*Hordeum vulgare*) and *A. thaliana*, were compared to sets from drought stressed, non-stressed and/or etiolated plants. The identification of cold-regulated genes was based on deriving preferentially expressed genes in a cold stress EST set when compared to the control sets. In this study, we wanted to investigate whether the genes identified as preferentially expressed differ among the different species or if there is a limited number of cold-regulated genes that are common to all these species.

Tentative ortholog groups

The mapping to an ortholog group was previously used for analyzing the most abundant ESTs in the oat cold stress EST set (Paper I). Besides streamlining the annotation of plausible expressed genes, using ortholog groups made it possible to compare gene expression values across multiple libraries. The ESTs in each set were mapped to a group and an expression value could subsequently be inferred by counting the number of ESTs per group. Moreover, this expression value could thereafter be compared across sets derived from different species. However, in Paper I, many of the ESTs did not receive a significant match against an *A. thaliana* gene with a KOG annotation and the classification had to be complemented with annotation from MAtDB and the best match in the nr database. The aim of this study

was therefore also to improve and develop the approach of using orthologous groups for classifying the ESTs.

The Institute of Genome Research (TIGR) have collected ESTs from various research projects and clustered sequences for each species into so called tentative consensus (TC) sequences, presumably representing expressed genes [120]. The TIGR database contains TCs from a large number of plant species, which makes it attractive for similarity searches. All sequences have also been clustered into ortholog groups, which are stored in the EGO (Eukaryotic Gene Orthologs) database [121], where each cluster contains orthologous genes (or actually TCs). The same advantages as with using KOG would be gained here, but presumably with an increased coverage of classified ESTs. However, the clusters in EGO are redundant, and consequently a TC is commonly a member of more than one ortholog group. This complicates the derivation of expression values, since the ESTs will match against several groups, thus resulting in misleading expression values.

In order to take advantage of using ortholog groups, while at the same time avoiding the redundancy in EGO, we downloaded TCs from a number of selected plant species included in the EGO database (see Table 2 in Paper II) and clustered them into a set of non-redundant tentative ortholog groups (TOGs) using the algorithm OrthoMCL [122]. The derived TOGs were thereafter used for the identification of expressed genes in the different EST sets, as well as for deriving expression values for each ortholog group. Each EST in the different sets was mapped against a TOG and the expression value for each TOG in each set was compared to the expression value in the other sets.

'Digital-Northern' for detecting preferentially expressed genes

Gene expression levels in unbiased cDNA libraries can be estimated by using the cognate frequencies of ESTs. The variation in the relative frequency of ESTs sampled from different cDNA libraries have, for example, been used for detecting genes differentially expressed in a biological sample or identifying genes with similar expression patterns in multiple cDNA libraries [100-102]. This approach is commonly referred to as 'digital-Northern'.

During the last years, various test statistics have been proposed for detecting differentially expressed genes in multiple EST sets. Such approaches have frequently been used to identify genes that differ in expression levels between different tissues [101, 123-128]. Some of the statistics have been developed more recently, taking into account the characteristics of the EST data, such as the distributions typically found in EST data [123, 127]. Four different test statistics were used in this study; the Audic and Claverie (AC)

test, the χ^2 2x2 test (Chi), the R statistic (Rstat) and the general χ^2 test over multiple data sets (MultChi) [123, 127]. The web tool IDEG6[10] was used to infer differentially expressed TOGs [129].

Identifying preferentially expressed TOGs

In total, seven cold stress EST sets and eleven control sets for the selected species were included in the comparative study (see Table 1 in Paper II). The sets were downloaded from dbEST[9], except for *A. sativa* where we used our in-house derived library (Paper I). The comparative study was performed species by species, i.e. each cold stress set was compared to control sets from the same species.

Each EST set was separately clustered and assembled into contigs and singletons (see Table 4 in Paper II). Thereafter each contig and singleton from each EST set was mapped to a TOG according to the most similar TC's TOG number, as identified by a BLASTn similarity search using an *E*-value threshold of 10^{-80}. The stringent threshold was set in an attempt to assure that each sequence would be matched to a true ortholog.

In order to derive preferentially expressed TOGs, the relative EST frequency was calculated for each detected TOG in each set and marked as up-regulated if the frequency was higher in the cold stress set than in the control set. TOGs were considered as preferentially expressed if they were: 1) detected as differentially expressed by any test statistic used and 2) up-regulated in comparison to all control sets. We also addressed uniquely expressed TOGs, i.e. those that had an expression value in the cold stress set but not in any of the control sets. In this case, we only included TOGs with $\geq n$ EST members, where n depended on the size of the EST set (for more details see Paper II), as an attempt to keep the number of false positives at a low level.

Comparison of identified cold induced genes

Analysis of the results revealed that the identified preferentially expressed TOGs differed considerably between the sets, as the percentage in the intersection of TOGs among pairs of sets was generally low (1.0-11.5%; see table 5 in Paper II). Additionally, comparing to a microarray study on cold-stressed *A. thaliana* plants showed a low overlap in identified cold-responsive genes (see Paper II for more details on the results).

At first sight, these results could indicate that the method had problems in deriving cold induced genes, but when further examining the annotation of the TOGs we found that many of them could be related to cold or other stresses (see Apendix A). For example, one of the most expressed TOGs

[10] IDEG6: http://telethon.bio.unipd.it/bioinfo/IDEG6_form/

corresponds to At1g54410, a gene that belongs to the dehydrin family. Members of this family are known to be induced by water stress [130]. Moreover, there are also different heat shock proteins, which are known to respond to stress [131] as well as the cold acclimation proteins WCOR410b and WCOR615 [19].

Many of the derived TOGs correspond to genes previously shown to be induced during a longer period of stress. This has a plausible natural explanation. For the oat, rice and *A. thaliana* sets RNA was extracted from pooled tissues collected at multiples time points, and for the barley and wheat sets the RNA was collected after two days of cold treatment. Consequently, these sets foremost represent genes that are long-term expressed or expressed after some time of stress.

Conclusions

Several of the derived TOGs could be related to cold and/or other stresses, although the results showed unexpectedly large discrepancies among the different species. This result can be explained by a number of factors, such as differences in the experimental conditions when constructing the cDNA libraries and the sequencing of ESTs. Another important factor is that the species differ both physiologically and in their level of cold tolerance. For example, oat is a monocot cold-sensitive species, whereas *A. thaliana* is a cold-tolerant dicot. It is therefore also plausible that the sets of preferentially genes differ between the species. This has also been observed in other studies, such as the comparative study by Gulick et al. between a winter and spring wheat variety in response to cold [132].

Additionally, the approach of using TOGs exhibited some limitations which can explain the differences in the results. For example, the mapping of oat contigs and singletons was still not complete, since about 34% of the sequences could not be mapped to a TOG (according to the selected BLAST *E*-value threshold). In addition, we can see that for some TOGs the OrthoMCL algorithm has clustered together too many TCs. For example, the largest TOG consists of 188 TCs with rather diverse annotations and this TOG should preferably be split up into smaller groups. The consequence is that it will generate an overestimated and misleading expression value. The method therefore needs further development in order to increase the reliability of the results.

Evaluation of ‘digital-Northern’

Since the study in Paper II produced large discrepancies in the results, we wanted to investigate in more detail what the reason could be. There were some plausible explanations available; for example, the method for mapping

ESTs to TOGs proved to have some flaws and it was found that improvements are needed in order to increase the reliability of the results. Another possibility, which other studies have also pointed at, is that EST sequencing and microarrays generate very different results, which makes a direct comparison limited and most probably misleading. A thorough comparison of the results from EST sequencing and microarrays from several experiments would give a more general picture of this issue.

There was also a possibility that the statistical tests used in Paper II were not optimal for performing 'digital-Northern', due to the characteristics of the data, e.g., some genes are highly abundant, but the majority is represented by only a few ESTs. A thorough evaluation of the test statistics used in the comparative study would give more insight into this matter. Consequently, in Paper III, we made a thorough evaluation of the applicability of the statistical tests. In addition, we wanted to investigate whether the combined outcome from several tests would improve the overall results.

Several different tests have been proposed for conducting 'digital-Northern' and some of them have also been evaluated previously [102, 123]. In this study we chose a somewhat different focus compared to previous studies. In Paper III we used relatively small EST sets (2,000-10,000 ESTs), pairwise comparisons of sets, and sets where the total number of ESTs might differ. In addition, we were interested in up-regulated (preferentially expressed) genes in a treatment set (here cold stress treated) when compared to a control set.

Statistical methods and simulated data

We tested five statistical methods, the χ^2 test (χ^2), Fisher's exact one-sided and two-sided test (Fone and Ftwo, respectively), the Audic and Claverie (AC) test and a method consisting of calculating the difference in relative frequency (Diff) [100, 107, 123, 133, 134].

The comparison of the statistical methods was first performed on simulated data, as this provides a controlled environment where the methods can be properly tested. Therefore, we created pseudo cDNA libraries following a Log normal, Gamma or Possion probability distribution, since these resembled most closely the distributions of transcript abundance in the EST data. The pseudo libraries contained about 1,000,000 clones relating to 20,000 genes. Of these genes, 4,000 (20%) were simulated as up-regulated with a fold change that was randomly selected from an underlying probability distribution (see Paper III for more details).

Thereafter, a random number of clones, representing ESTs, were picked from a pseudo library with simulated up-regulated genes, i.e., the treated condition, and another random number of clones from a control library. The

tests were subsequently applied to the data, where after genes with a test value (*p*-value or difference in relative frequency) less than a specified cutoff were considered as up-regulated.

We also introduced the use of a flexible cutoff *C*, which takes into account the size of the EST sets, and can be used as an alternative to static cutoffs. In this case, the percentage number of genes with a test value below a specified cutoff was first calculated. The level of *C* thereafter determined which test value cutoff would be used, and this would be the value that has been derived for <*C*% of the genes (for more details on how to derive the flexible cutoff, see Paper III).

The methods were evaluated by producing ROC curves, calculating the percentage derived true positives and the sensitivity (see methods section in Paper III for how this was done). The results revealed that all methods performed on a comparable level, except for χ^2 that was slightly worse than the other methods, and that the tests produced slightly different results. Based on these observations, we applied a range of rules which combined the outcome from the different tests. The results showed that a combination of the AC, Fone, Ftwo and Diff methods with a flexible cutoff should be preferred, as it increased the number of detected up-regulated genes compared to when each test was applied in isolation.

Experimentally derived data

The combination of the AC, Fone, Ftwo and Diff methods together with a flexible cutoff was also tested on experimentally generated data. In this case we used two EST sets originating from cold stressed *A. thaliana* plants and a control set from non-stressed plants. We chose to work with *A. thaliana* since this is a prominent model species for plants and most of the current knowledge about cold stress has been derived from this species.

The concatenated EST sets were first clustered and assembled into contigs and singletons, using the EGassembler on-line tool[11] [98], and thereafter each contig and singleton was matched to a gene in *A. thaliana* by a tBLASTx search ($E \leq 10^{-5}$). For each gene with a match we calculated the number of ESTs in each matching contig, as well as all matching singletons, thereby deriving an expression value for the gene in each EST set.

In the simulation studies we recorded whether a gene was up-regulated or not and thereby we could identify which genes were true positives among the detected ones. In order to have the same template when testing on the experimental data, we used a collection of 4,037 cold responsive genes compiled by Hannah et al. as gold standard [135]. The genes in this

[11] EGassembler, http://egassembler.hgc.jp/

collection had been reported to be differentially expressed in at least two independent microarray studies.

We made some interesting observations when comparing the genes represented in the sets and with the collection of cold responsive genes. There was some overlap between the cold-stress sets, as expected, but also with the control set. Moreover, some of the genes in the control set were also represented in the collection of cold responsive genes. This means that some of the genes participating in the cold response are also expressed during normal conditions. Finally, the number of cold responsive genes was relatively low in the cold stress sets, 28% and 24%, respectively, of the total number of represented genes in the sets. In the control set 24% of the genes related to cold responsive genes.

After applying the tests, the results showed that the level of true positives was relatively low, ~25% for both sets, irrespectively of the level of *C*. On the other hand, the sensitivity increased dramatically when *C* was increased; however, this also produced a high number of false positives.

Conclusions

Based on the simulation studies, we conclude that the evaluated statistical methods are comparable regarding the percentage of derived true positives, but they produce slightly different results. The simulation studies also showed that the number of detected true positives increased when the AC, Fone, Ftwo and Diff methods were combined and that the introduction of a flexible cutoff made the inference less sensitive to the size of the EST sets.

On the other hand, based on the results from the application of the tests to experimentally derived EST sets, the tests appear to only be able to detect a portion of the cold responsive genes represented in the sets. This can relate to a number of factors, such as the inherent properties of EST data, but also to limitations in the analysis steps. For example, most of the genes were represented by a few ESTs, which will restrict the statistical inference. Further, each EST has to be matched to its true gene origin, in order to derive a correct expression value for each gene. This step is not trivial, since the ESTs commonly have poor sequence quality, are partial genes and may be polymorphic, which makes the matching complicated. If the ESTs are not correctly matched, this will lead to an over- or under-estimated expression value, which will have consequences for the detection of up-regulated genes. Moreover, when using smaller EST sets this step becomes even more important, since the mis-matching of a few ESTs will have a large impact on the results.

Consequently, based on this study, we conclude that the test statistics are of limited use when applied to small EST sets. To correctly infer preferentially

expressed genes puts large requirements on the data, i.e., the production of high-quality sequences. Furthermore, the EST sets used in Paper II and III come from cDNA libraries pooled from several time points. The ESTs should preferably come from un-pooled libraries, since the pooling of the data will result in the detection of mostly genes expressed during all or several of the time points included in the pooling (Figure 11). Therefore genes that are differentially expressed during a shorter time window might be missed.

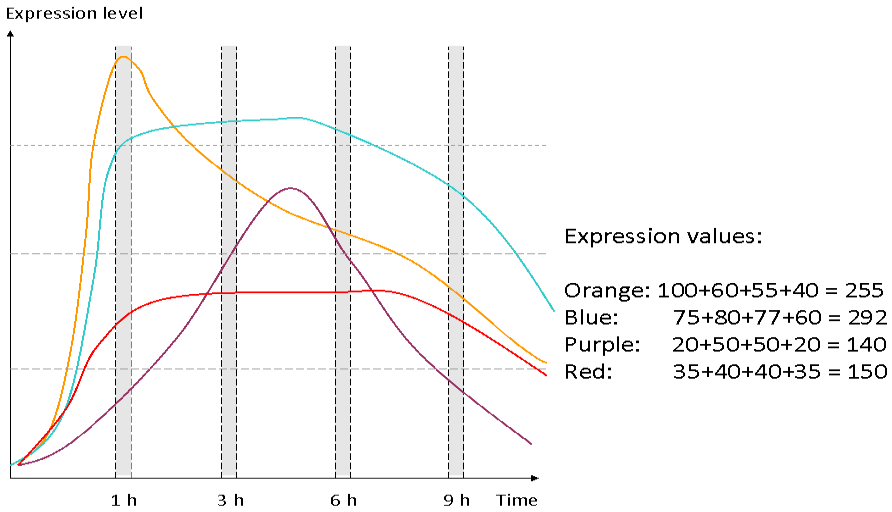

Figure 11. This figure illustrates the result in number of transcripts collected from pooled samples. Genes that are expressed at a lower level, but during a longer period will be represented approximately equally in number of ESTs to genes strongly expressed during a shorter time. The gray areas represent sampling points in the pooled data and the colored lines the expression profiles of four different genes. The calculated numbers refer to the total number of mRNAs extracted from the different samples (theoretical example).

Part II

Cold response in rice

Rice – a chilling-susceptible species

Rice is originally a tropical and subtropical species, and is therefore severely affected when the temperature drops below 10°C [136]. However, the expansion of growing rice into colder regions has led to a wide contrast in terms of sensitivity to chilling among rice cultivars. The higher sensitivity in some cultivars has been shown to be correlated to a decreased capacity of antioxidative system, a re-establishment of the water balance as well as an inability to produce cryoprotectants [137-139].

Since rice is an important crop and severely affected by cold stress, there is a demand for more tolerant varieties. In similarity to oat, due to the complexity of the cold acclimation signaling pathways, there has been limited success of improving the tolerance by traditional breeding. Consequently, in order to utilize biotechnological methods, more knowledge about the underlying regulatory mechanisms is required.

Cheng et al. made a semi-global survey of the rice transcriptome during 10°C chilling stress [136]. However, the pathways and genes involved in the response at this temperature differ to those involved in the lower 4°C responses. A whole-genome transcription profiling of the response to cold stress in rice at 4°C have, as far as we know, not previously been reported in the literature. Consequently, the first step in elucidating the response to cold stress is to conduct and analyze the data from such a study.

Microarray gene expression analysis

Previous studies have indicated that cv. *Japonica* exhibits molecular response mechanisms similar to those found in cold-tolerant plants, and, additionally, that prominent cold-responsive signaling pathways are conserved in rice [136, 140-142]. Conversely, no study has previously established which mechanisms might differ in rice, in comparison with cold-hardy species, that could explain its inability to fully acclimate.

In Paper IV, the rice cv. *Indica* v. *Jumla Marshi* was used in the microarray experiment. Regarding *Indica* genotypes, they are in general more susceptible to chilling than *Japonica* genotypes. However, *Jumla Marshi* is a high mountainous cold resistant paddy that it is still today prevailing in the harsh temperate climatic condition of around 2,600 m altitudes in the Jumla area in Nepal. The genetic base of this rice is unknown but believed to be

different from that of present-day rice varieties and it is not clear if the variety actually originates from *Indica* or *Japonica*.

Gene expression analysis

For the microarray experiments we utilized the Affymetrix GeneChip® Rice Genome Array, which represents approximately 51,000 transcripts from mostly *Japonica* genes and a few thousands originating from *Indica* genes. Rice seedlings were exposed to cold stress at 4°C during 30 min, 2 h, 4 h, 8 h and 24 h of stress treatment (for more experimental details, see Paper IV). In addition, the results were compared to a similar experiment of ~24,000 genes conducted on *A. thaliana*. This data set is publicly available and was downloaded from the AtGenExpress database[12].

The pre-processing of the data from both experiments was done in the same way, by using the GeneSpring software[13] (for more details on this part, see Paper IV). Probe sets were classified as differentially expressed if they had at least a 3-fold change in at least one time point. In total 1,617 and 1,967 probe sets were differentially expressed in rice and *A. thaliana*, respectively.

The differentially expressed probe sets in rice were matched to *Japonica* loci, through BLAST searches against the full collection of *Japonica* genes (for more details on the mapping, see Paper IV). Since *Jumla Marshi* is more tolerant to cold than *Indica* genotypes in general, we argued that it might be more similar to *Japonica* on the genetic level and therefore we utilized the rich annotation that is available for the completely sequenced genome of *Japonica*. For the probe sets in *A. thaliana*, their respective loci were downloaded from the Affymetrix[14] home site (see Paper IV for more details on this). The differentially expressed probe sets are hereafter termed the cold-responsive genes.

Most abundant gene families

In order to gain more insight on which genes that are expressed in rice during cold stress we examined the most abundant gene families among the cold-responsive genes, and in order to derive the number of cold-responsive genes per family we utilized the annotation in the GreenPhyl database [103]. Since the cold-responsive genes had previously been mapped to loci in *Japonica*, these loci were used for extracting family membership for each gene from the database. In total 1,299 (89.9% of the total) cold-responsive genes had a family classification, which is a high coverage and, hence, this approach turned out to be successful.

[12] AtGenExpress, http://www.arabidopsis.org/info/expression/ATGenExpress.jsp
[13] Agilent Technologies, www.agilent.com/
[14] Affymetrix, http://www.affymetrix.com

The most abundant gene families can be viewed in figure 5 in Paper IV. The results revealed that protein kinases and thioredoxins have a high representation. Kinases are important in a large range of abiotic stresses, including cold stress, since calcium-dependent protein kinases act as calcium sensors and consequently are important in the triggering of signaling pathways [62, 63]. Thioredoxins, on the other hand, act as antioxidants and such proteins have been shown to be prominent in the protection of the cells against ROS [143], which increases during cold stress.

There are also a number of transcription factor families represented among the most abundant gene families that are important in the response to cold, such as the AP2/EREBP, WRKY and MYB families. The CBF factors that have been shown to be prominent in cold acclimation signaling pathways in *A. thaliana* belong to the AP2/EREBP family (see section 'Cold acclimation regulatory pathways' in 'Introduction' chapter for more details). WRKY transcription factors are involved in a range of biological processes, including different abiotic stresses such as cold [144]. Moreover, they have an important role in the control of ROS that increase in response to stress [145]. MYB transcription factors are important in many hormone and stress responses [146]. For example, the MYB15 factor has been shown to negatively regulate CBF3 in *A. thaliana* and, consequently, is a suppressor of the CBF regulon (see section 'Cold acclimation regulatory pathways' for more details) [87].

The most abundant gene families clearly showed that rice is cold-responsive, since many of the affected genes could be coupled to different stress responses in general as well as to cold stress specifically.

Comparison to the cold response in A. thaliana

Although the analysis of the cold-responsive genes indicated that rice is reacting against the arisen stress, the question remained why this species is unable to fully acclimate. Therefore, we made a comparison to cold-responsive genes in *A. thaliana*.

When studying the expression profiles of the genes in the two species there is a clear difference in the overall dynamics (see figure 7 and 8 in Paper IV). In rice, the genes are mainly activated or suppressed by cold stress and thereafter remain on an elevated or reduced level throughout the entire response. In *A. thaliana*, on the other hand, the majority of the genes are transiently expressed, by making a peak or dip in expression levels and thereafter returning to base levels at normal temperatures. This transient behavior during cold stress has recently been coupled to the circadian rhythm in *A. thaliana* (see section 'Cold acclimation and darkness' in the 'Introduction' chapter). In chilling-sensitive plants, on the other hand, cold stress experiments have been shown to disrupt the circadian expression of

mRNA levels. Moreover, after rewarming the clock is altered and out of phase, which causes a mistiming of photosynthetic and cellular metabolic reactions. This mistiming is thought to contribute to the sensitivity in chilling-susceptible plants and, consequently, is a strong candidate to the susceptibility in rice.

Cold-responsive transcription factors

We also extracted the transcription factors represented among the cold-responsive genes in both rice and *A. thaliana*, as an attempt to derive more clues about differences in the transcriptional regulation (see Paper IV on how this was conducted). The results can be viewed in figure 6 in Paper IV.

The family of C2C2-CO-like factors is interesting, since the relative frequency of the number of expressed genes in this family is higher in *A. thaliana* than in rice. Members of this family control the flowering by photoperiod and, hence, are plausibly important in the timing of the acclimation processes. There are also four *PLATZ* factors expressed in *A. thaliana*, but none in rice. In peas, a factor from this family was shown to bind to A/T-rich promoter sequences and act as a repressor of transcription [147]. A potential role could be negative regulation of acclimation processes in *A. thaliana*, which seems to be lacking in rice. There are also four members of the ARF family expressed in *A. thaliana*, but none in rice. These are auxin-responsive factors, which act as either activators and or repressors of transcription [148]. Two of these factors, *At1g77850* and *At1g30330*, are located on chromosome 1 in *A. thaliana*, for which there are no orthologs in rice. These genes have been shown to be important in ovule and anther development [149].

Rice CBF genes

The mining of all transcription factors in rice made it easy to extract homologs to the *CBF*s in *A. thaliana*. Among the cold-responsive genes in rice, six up-regulated genes were found to be *AtCBF* homologs (see figure 8 in Paper IV). In *A. thaliana*, the *CBF*s are rapidly and transiently induced by cold stress [17, 83]. However, in rice the genes follow the overall dynamics previously identified, by only exhibiting an increase in expression levels. The trough that can be seen in *A. thaliana* is not apparent for the *OsCBF*s (see figure 9 in Paper IV).

In rice, two of the *CBF*s, *Os09g35010* and *Os09g35030*, exhibit a rapid increase within a few hours after the response and are therefore strong candidates to be orthologs to *AtCBF1-3*. However, according to the phylogenetic tree of the protein sequences (see figure 8 in Paper IV), these genes are not the most closely related to the *AtCBF*s. On the other hand, the

gene most closely related, *Os01g73770*, is only relatively weakly expressed and may therefore be more related to the drought-induced *AtCBF4*.

When including identified *CBF* genes from related species, i.e., wheat (*Triticum aestivum*), barley (*Hordeum vulgare*), oat (*Avena sativa*), rapeseed (*Brassica napus*) and tomato (*Lycopersicon esculentum*), in the phylogenetic analysis, we can see that *Os01g73770* is located on a separate branch from the remaining *CBF*s (Figure 12), which further indicates that this gene is not an ortholog to *AtCBF1-3*. Moreover, the monocot and dicot *CBF*s are clearly separated on different branches and the cold-responsive *OsCBF*s are spread out on the monocot branch.

Conclusions

Clustering genes into families using the full repertoire of genes/proteins in a number of related species in combination with sophisticated algorithms has streamlined the classification of the genes. Additionally, it is a valuable information source when characterizing genes responding to a specific stimulus, such as in this case where it was used for analyzing genes responding to cold stress. However, the automatic clustering is not entirely flawless and some clusters need to be manually curated. For example, the algorithms tend to cluster together genes from different families that have a

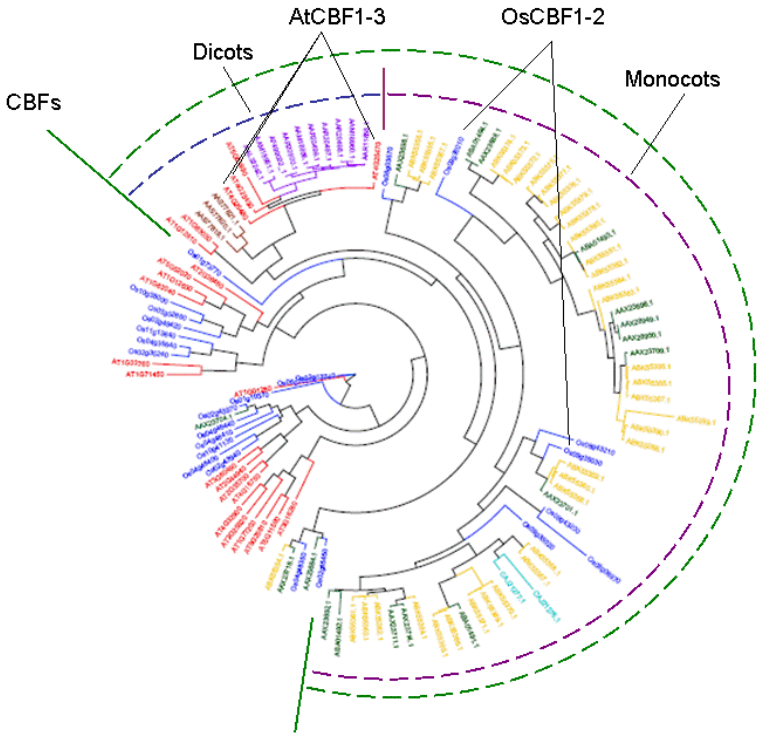

Figure 12. Phylogenetic tree of CBF genes in related crops. Colored gene IDs indicate species: rice (blue), *A. thaliana* (red), barley (yellow), wheat (green), oat (turquoise), rapeseed (purple), and tomato (brown).

conserved domain in common. Consequently, the user needs to be aware of potential problems and critically evaluate the resulting family classification, and possibly make corrections of the clusters.

There are also numerous databases that store information about a specific family or genes with a common function, e.g., transcription factors. Such databases are often more reliable, since the information commonly is manually curated; the amount of data is less and therefore easier to manage. This type of information source was also utilized in this work for extracting transcription factors among the cold-responsive genes and reduced the amount of time for identifying all factors represented in the data.

Although clustering and classification of genes based on sequence information streamlines the analysis of genes responding to some stimulus, it can still be difficult to distinguish true orthologs in different species. For example, we identified a number of CBF factors in rice, which seem to be orthologs to CBFs in *A. thaliana*. However, the fact that rice is a monocot and *A. thaliana* a dicot, the extended phylogenetic analysis, which included CBFs from a range of related species, resulted in a separation of the factors onto different branches. Additionally, the expression profiles of the *CBF*s in rice and *A. thaliana* differed considerably, and, consequently, whether the OsCBFs are true orthologs to AtCBFs is still an open question.

The importance of combinatorial control

As described previously, the activation and suppression of a gene is achieved, amongst others, by the binding of transcription factors to *cis*-elements in the promoter regions of the affected gene. These elements are commonly short (6-20 bp), tolerate some degeneracy and there are few established mechanisms on how they emerge.

Computationally, the identification of *cis*-elements coupled to transcription factors that are plausibly important in the regulation of a certain biological process is commonly performed by searching for motifs that are significantly over-represented in a set of genes sharing common properties, such as genes having highly similar expression profiles during the process or having the same biological function [106].

On the other hand, the regulation of gene expression is highly complex in eukaryotes and is commonly accomplished by the coordinated action of multiple transcription factors [150]. Since there are many more biological processes than transcription factors in the cell, this requires a gene regulation that is not one-to-one. The regulation is achieved, instead, by the use of combinatorial control, where specific combinations of transcription factors accounts for different cell responses. A single motif can,

consequently, be enriched in many processes and a single process is enriched in multiple motifs.

Since cold acclimation is a complex trait, involving thousands of genes interacting in an intricate underlying regulatory network, we wanted to investigate whether combinatorial control is important in the regulation of cold responsive genes in rice and *A. thaliana* (Paper V). Furthermore, we chose to focus on the CBF factors in these two species and genes having a similar expression profile to these genes, since the CBFs are prominent in the regulation of cold responses in *A. thaliana*.

Extracting over-represented motif combinations

In our previous analyses of gene expression data from rice and *A. thaliana* we identified a number of differentially expressed probes (Paper IV). In this study, we proceeded by extracting the promoter region of genes corresponding to these probes, by first matching each probe sequence to a *Japonica* and *A. thaliana* locus, respectively, and thereafter downloading the 1K upstream region of genes corresponding to these loci (see Paper V for more details).

In order to derive motif combinations, we first searched for over-represented single motifs in each cluster of genes. The clusters were derived by identifying genes having highly similar expression profiles to each of the *CBF*s, including only those having a Pearson correlation (PC) of $\geq$0.95 with a *CBF*. This procedure generated five clusters, one for each *CBF*. The clusters are somewhat overlapping (Figure 13) since all *CBF*s in each species have highly similar expression profiles (see figure 9 in Paper IV). One of the clusters in rice has very few members, compared to the other clusters, and this is due to that the expression level of *Os09g35030* slightly increases at time point 24 h, while for *Os09g35010* there is instead a small decrease.

We thereafter derived over-represented single motifs in a cluster by using Fisher's exact one-sided test [107]. In this case, a database of previously characterized motifs was used; where the motifs were collected from the plantCARE database [151], the PLACE database [152] and motifs extracted from the literature. The over-represented motifs were thereafter used for building motif combinations and deriving over-represented combinations in a cluster, by again using Fisher's exact one-sided test. For more details on this step, see Paper V.

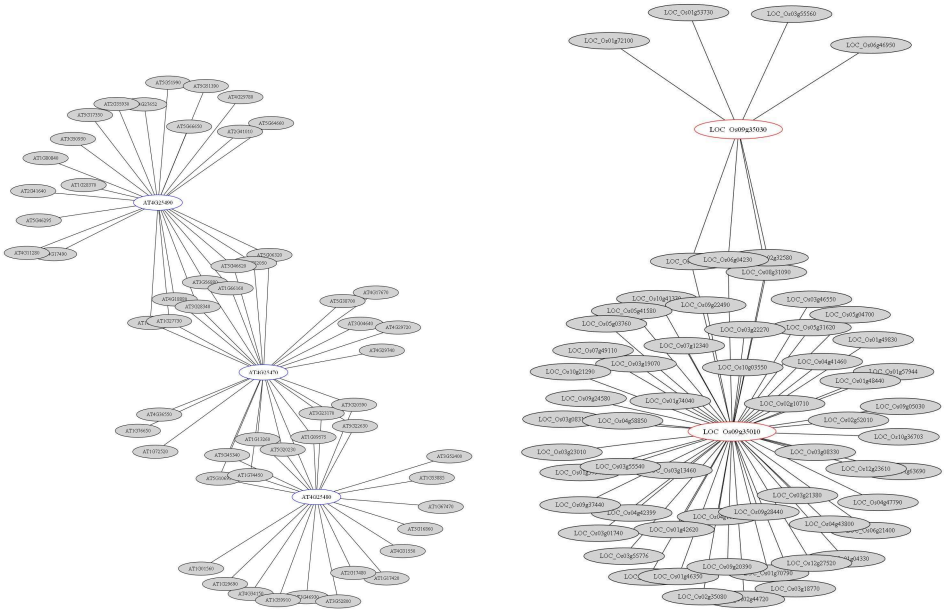

Figure 13. The figure on the left shows the gene members of the three clusters derived for *A. thaliana* and the figure on the right shows the members of the two clusters for rice.

Results and analysis

When studying single over-represented motifs, we made the observation that if a motif was common among genes in a cluster, it was also common in the remaining genome and vice versa (see figure 2 in Paper V). For example, the relative fold-change (FC) (see paper V on how FC is calculated) for the W-box motif YTGACY was ~1.3, i.e. this motif was only ~1.3 times more common among genes in a cluster than among genes in the remaining genome. On the other hand, regarding motif combinations we saw an opposite trend, i.e., a specific motif combination could be present in the majority of the genes in a cluster, but totally absent in the remaining genome.

Moreover, we could also see a decrease in *p*-values with an increased complexity in the motif combinations, i.e., the *p*-values were in general much lower among four-combinations than among two-combinations (see figure 3 in Paper V). This also meant that the number of occurring four-combinations among the genes in a cluster decreased to a much greater extent in the remaining genome.

We also examined the most significant non-overlapping non-redundant motif (MSNM) combinations in more detail, to investigate whether these could be related to cold and/or other stresses or if they were entirely novel in this context. Among the clusters in *A. thaliana*, we conclude that an ABRE-related motif, which has been identified among genes responsive to cytosolic Ca^{2+}, the binding site of the transcription factor GT-1, which has

been shown to stabilize the TFIIA-TBP-DNA complex, and the binding site of WRKY transcription factors, which are commonly induced against various abiotic stresses including cold, are important in the regulation of the *AtCBF*s. Moreover, in one of the clusters the binding site of heat shock transcription factors were in the MSNM combination. These transcription factors are commonly coupled to heat stress, but it has recently been shown that there is an overlap with non-heat signaling pathways including cold responses [153].

The ABRE-related and the binding site of GT-1 are also among the most SNM combination in the *Os09g35010* cluster, but not the WRKY motif. However, this does not mean that the motif is not present among the genes in this cluster. On the contrary, this motif is highly occurring in rice, leading to that it is not detected as over-represented in the cluster. The MSNM combinations detected in the *Os09g35030* cluster distinguishes most from the other SNM combinations, by including a low temperature responsive element (LTRE), a motif coupled to sugar starvation and the binding site of a MADS transcription factor, although the ABRE-related motif as well as the WRKY-motif are present in the *Os09g35030* gene.

Conclusions

This study clearly showed that the *CBF*s and other cold responsive genes are under combinatorial control, which should be acknowledged when detecting motifs. Some of the motifs are highly interesting, such as the motifs that correspond to WRKY and HSF binding sites, since these factors have recently been directly coupled to cold stress.

We also made some other interesting observations, such as that a motif relating to the binding site of MYC transcription factors is only over-represented in the *AtCBF2* cluster, although it is known that a MYC factor is prominent in the activation of *AtCBF3* [77]. Additionally, that, e.g., WRKY binding sites is not detected in the rice clusters, although they are commonly occurring among the genes in the clusters. This has to do with the detection of over-represented single motifs from a cluster. Which motifs that will be derived depends on the composition of genes in the cluster and this in turn highly depends on the approach/algorithm for inferring clusters. There are many options regarding this step and these will all report a somewhat different composition, which will together with the chosen statistical test have an impact on which motifs will be considered as over-represented and, consequently, which combinations are possible. This is a limitation in the method, since a motif may be over-represented in combination with some other motif, but not in isolation. Consequently, it might be better to focus on motif combinations from the beginning of the study and, hence, detect over-represented combinations in a cluster instead of single motifs.

Concluding remarks

The size of crop yields is negatively dependent on the damage of plants exposed to low temperatures. This in turn depends on a number of factors, e.g., time of planting and the time of day the cold stress occurs, but most importantly the ability to cold acclimate. The cold acclimation signaling pathways are complex and involve an intricate network of genes. Moreover, there is a coupling of cold stress with other stresses, such as wounding and drought, signifying that there is a cross-talk between these responses and their respective signaling pathways.

During the course of this work, data from gene expression profiling studies have been analyzed, with the aim of gaining more knowledge about mechanisms regulating responses to cold stress in oat and rice. Two different technologies were used, since the availability of gene sequences from the two species differed, which consequently required different types of analyses.

More specifically, based on this work, the following conclusions can be made:

(biological)

- The winter oat variety Gerald is cold responsive and contains genes homologous to previously characterized cold-regulated genes in related species.

- The oat genome contains a number of *CBF* genes, which have been shown to be prominent in acclimation pathways in related species.

- The winter rice variety Jumla Marshii is cold responsive and contain a significant amount of genes coupled to cold-regulation and acclimation pathways.

- The cold response in rice differs from the response in the more cold tolerant *Arabidopsis thaliana*. The negative repression of acclimation pathways is apparently dysfunctional in rice, which plausibly explains its inability to survive low temperatures.

- The *CBF* genes in rice and *Arabidopsis thaliana*, and other genes with similar expression profiles, are under combinatorial control of multiple transcription factors.

(data analysis)

- There is an accepted procedure for analyzing EST sequencing and microarray data, up to the point of identifying differentially expressed genes. However, much work is still needed in order to further characterize these genes and assembled procedures for doing so are less developed.

- In order to streamline the annotation of sequences, including both ESTs and probes, the use of mapping to orthologous groups or gene families should be considered, since this will facilitate the annotation to a large extent as well as the identification of biological interesting genes among vast amounts of data.

- The combined results from several test statistics when performing 'digital-Northern' should be considered, since the number of detected true positives increases.

- The interpretation of the results from the application of 'digital-Northern' to small EST sets should be made with caution, since the sets may not hold enough information for statistical significance.

- Instead of deriving over-represented single motifs among a group of related genes, one should consider over-represented motif combinations since many complex traits are under the combinatorial control of multiple transcription factors. Moreover, a specific combination of transcription factors may be more important then the presence of specific single transcription factors.

Acknowledgements

These years as a PhD student at Göteborg University as well as University of Skövde has given me valuable experience and made me grow as both a person and researcher. This experience and maturity I owe to a number of people around me.

First of all I would like to thank my supervisors Olof Olsson and Björn Olsson, both for providing me with useful advice and feedback on my work, as well as suggesting good ideas and for being my co-workers. I especially would like to thank Björn for being a good mentor. I have grown as a student and become more of an independent researcher, and I think you have grown as a supervisor, by being more patient with me and giving me the support at moments I needed them most.

Furthermore, I want to thank all the PhD student colleagues around me that have made these years more fun and intriguing. First, my colleagues at University of Skövde: Jonas, Zelmina, Jane, Simon and Kim. The small-talks at the coffee breaks have been invaluable and a true security valve for my brain. Second, I want to thank my colleagues at Göteborg University: Marcus and Aakash. The papers included in this thesis would never have been a fact without your contributions.

I also want to thank my family and relatives that are close to me, for believing in me and never giving the up hope that I some day would make it to the end, even if I have had great doubts occasionally. Especially I want to thank my partner Zlatan, my kids Selma and Nikita, mum and dad, my brother Kenneth and his family, the Hodzic family, and grandma and grandpa.

For financial support I want to thank the Swedish Research School in Genomics and Bioinformatics, the Swedish Farmers Supply and Crop Marketing Co-operative (SLF), the VL-foundation and the Swedish Research Council (VR).

References

1. Koning RE: **Climate and Biomes.** *Plant physiology information website* 1994.
2. Guy C: **Molecular responses of plants to cold shock and cold acclimation**. *J Mol Microbiol Biotechnol* 1999, **1**(2):231-242.
3. Mahajan S, Tuteja N: **Cold, salinity and drought stresses: an overview**. *Arch Biochem Biophys* 2005, **444**(2):139-158.
4. Kratsch HA, Wise RR: **The ultrastructure of chilling stress**. *Plant, Cell & Environment* 2000, **23**:337-350.
5. Allen DJ, Ort DR: **Impacts of chilling temperatures on photosynthesis in warm-climate plants**. *Trends Plant Sci* 2001, **6**(1):36-42.
6. Gesch RW, Heilman JL: **Responses of photosynthesis and phosphorylation of the light-harvesting complex of photosystem II to chilling temperature in ecologically divergent cultivars of rice.** *Environmental and Experimental Botany* 1999, **41**:257-266.
7. Hallgren J, Öquist G: **Adaptations to low temperature.** *In: Alscher, RG, Cumming, JR (Eds), Stress Responses in Plants: Adaptations and Acclimation Mechanisms Wiley-Liss, NY, USA* 1990:265-293.
8. Beck EH, Fettig S, Knake C, Hartig K, Bhattarai T: **Specific and unspecific responses to cold and drought stress.** *Journal of Biosciences* 2007, **32**:501-510.
9. Webb MS, Uemura M, Steponkus PL: **A Comparison of Freezing Injury in Oat and Rye: Two Cereals at the Extremes of Freezing Tolerance**. *Plant Physiol* 1994, **104**(2):467-478.
10. Tran DV: **World rice production main issues and technical possibilities.** *Cahiers Option Méditerranéennes* 2004, **24**:56-69.
11. Shimono H, Okada M, Kanda E, Arakawa I: **Low temperature-induced sterility in rice: Evidence for the effects of temperature before panicle initiation** *Field Crops Research* 2007, **101**:221-231.
12. Wu R, Garg A: **Engineering rice plant with trehalose-producing genes improves tolerance to drought, salt and low temperatures.** *Seedquest, ISB news report* 2003.
13. **The Gene Ontology project in 2008**. *Nucleic Acids Res* 2008, **36**(Database issue):D440-444.
14. Thomashow MF: **PLANT COLD ACCLIMATION: Freezing Tolerance Genes and Regulatory Mechanisms**. *Annu Rev Plant Physiol Plant Mol Biol* 1999, **50**:571-599.
15. Smallwood M, Bowles DJ: **Plants in a cold climate**. *Philos Trans R Soc Lond B Biol Sci* 2002, **357**(1423):831-847.
16. Stitt M, Hurry V: **A plant for all seasons: alterations in photosynthetic carbon metabolism during cold acclimation in Arabidopsis**. *Curr Opin Plant Biol* 2002, **5**(3):199-206.

17. Thomashow MF: **So what's new in the field of plant cold acclimation? Lots!** *Plant Physiol* 2001, **125**(1):89-93.
18. Bohn M, Luthje S, Sperling P, Heinz E, Dorffling K: **Plasma membrane lipid alterations induced by cold acclimation and abscisic acid treatment of winter wheat seedlings differing in frost resistance**. *J Plant Physiol* 2007, **164**(2):146-156.
19. Thomashow MF: **Role of cold-responsive genes in plant freezing tolerance**. *Plant Physiol* 1998, **118**(1):1-8.
20. Lindberg S, Banas A, Stymne S: **Effects of different cultivation temperatures on plasma membrane ATPase activity and lipid composition of sugar beet roots**. *Plant Physiol Biochem* 2005, **43**(3):261-268.
21. Martz F, Kiviniemi S, Palva TE, Sutinen ML: **Contribution of omega-3 fatty acid desaturase and 3-ketoacyl-ACP synthase II (KASII) genes in the modulation of glycerolipid fatty acid composition during cold acclimation in birch leaves**. *J Exp Bot* 2006, **57**(4):897-909.
22. Steponkus PL, Uemura M, Webb MS: **A contrast of the cryostability of the plasma membrane of winter rye and spring oat. Two species that widely differ in their freezing tolerance and plasma membrane lipid composition.** *In PL Steponkus, ed, Advances in Low-Temperature Biology, JAI Press, London* 1993, **2**:211-312.
23. Uemura M, Tominaga Y, Nakagawara C, Shigematsu S, Minami A, Kawamura Y: **Responses of the plasma membrane to low temperatures.** *Physiologia Plantarum* 2006, **126**(81-89).
24. Livingstone DP, Premakumar R, Tallury SP: **Carbohydrate partiotioning between upper and lower regions of the crown in oat and rye during cold acclimation and freezing.** *Cryobiology* 2006, **52**:200-208.
25. Bhowmik PK, Tamura K, Sanada Y, Tase K, Yamada T: **Sucrose metabolism of perennial ryegrass in relation to cold acclimation**. *Z Naturforsch [C]* 2006, **61**(1-2):99-104.
26. Bogdanovic J, Mojovic M, Milosavic N, Mitrovic A, Vucinic Z, Spasojevic I: **Role of fructose in the adaptation of plants to cold-induced oxidative stress**. *Eur Biophys J* 2008.
27. Bohnert HJ, Sheveleva E: **Plant stress adaptations--making metabolism move**. *Curr Opin Plant Biol* 1998, **1**(3):267-274.
28. Omran RG: **Peroxide Levels and the Activities of Catalase, Peroxidase, and Indoleacetic Acid Oxidase during and after Chilling Cucumber Seedlings**. *Plant Physiol* 1980, **65**(2):407-408.
29. Prasad TK, Anderson MD, Stewart CR: **Acclimation, Hydrogen Peroxide, and Abscisic Acid Protect Mitochondria against Irreversible Chilling Injury in Maize Seedlings**. *Plant Physiol* 1994, **105**(2):619-627.

30. Wise RR, Naylor AW: **Chilling-Enhanced Photooxidation : The Peroxidative Destruction of Lipids during Chilling Injury to Photosynthesis and Ultrastructure**. *Plant Physiol* 1987, **83**(2):272-277.
31. Lee BH, Lee H, Xiong L, Zhu JK: **A mitochondrial complex I defect impairs cold-regulated nuclear gene expression**. *Plant Cell* 2002, **14**(6):1235-1251.
32. Anderson MD, Prasad TK, Stewart CR: **Changes in Isozyme Profiles of Catalase, Peroxidase, and Glutathione Reductase during Acclimation to Chilling in Mesocotyls of Maize Seedlings**. *Plant Physiol* 1995, **109**(4):1247-1257.
33. Scebba F, Sebustiani L, Vitagliano C: **Changes in activity of antioxidant enzymes in wheat (Triticum aestivum L.) seedlings under cold acclimation.** *Physiologia Plantarum* 2003, **104**:747-752.
34. Kuk YI, Shin JS, Burgos NR, Hwang TE, Han O, Cho BH, Jung S, Guh JO: **Antioxidative enzymes offer protection from chilling damage in rice plants.** *Crop Science Society of America* 2003, **43**:2109-2117.
35. Foyer C, Lelandais M, Galap C, Kunert KJ: **Effects of Elevated Cytosolic Glutathione Reductase Activity on the Cellular Glutathione Pool and Photosynthesis in Leaves under Normal and Stress Conditions**. *Plant Physiol* 1991, **97**(3):863-872.
36. Hughes MA, Dunn MA: **The molecular biology of plant acclimation to low temperature.** *Journal of Experimental Botany* 1996, **47**:291-305.
37. Bray EA: **Plants responses to water deficit.** *Trends in Plant Science* 1997(2):48-54.
38. Hurry V, Huner N, Selstam E, Garderström P, Öquist G: **Photosynthesis at low growth temperature.** *In AS Raghavendra, ed, Photosynthesis: A comprehensive treatise Cambridge University Press, Cambridge* 1998:238-249.
39. Huner N, Öquist G, Sarhan F: **Energy balance and acclimation to light and cold**. *Trends in Plant Science* 1998, **3**:224-230.
40. Hetherington SE, He J, Smillie RM: **Photoinhibition at Low Temperature in Chilling-Sensitive and -Resistant Plants**. *Plant Physiol* 1989, **90**(4):1609-1615.
41. Orvar BL, Sangwan V, Omann F, Dhindsa RS: **Early steps in cold sensing by plant cells: the role of actin cytoskeleton and membrane fluidity**. *Plant J* 2000, **23**(6):785-794.
42. Vaultier MN, Cantrel C, Vergnolle C, Justin AM, Demandre C, Benhassaine-Kesri G, Cicek D, Zachowski A, Ruelland E: **Desaturase mutants reveal that membrane rigidification acts as a cold perception mechanism upstream of the diacylglycerol kinase pathway in Arabidopsis cells**. *FEBS Lett* 2006, **580**(17):4218-4223.

43. Sangwan V, Foulds I, Singh J, Dhindsa RS: **Cold-activation of Brassica napus BN115 promoter is mediated by structural changes in membranes and cytoskeleton, and requires Ca2+ influx**. *Plant J* 2001, **27**(1):1-12.
44. Monroy AF, Dhindsa RS: **Low-temperature signal transduction: induction of cold acclimation-specific genes of alfalfa by calcium at 25 degrees C**. *Plant Cell* 1995, **7**(3):321-331.
45. Knight MR, Campbell AK, Smith SM, Trewavas AJ: **Transgenic plant aequorin reports the effects of touch and cold-shock and elicitors on cytoplasmic calcium**. *Nature* 1991, **352**(6335):524-526.
46. Knight H, Trewavas AJ, Knight MR: **Cold calcium signaling in Arabidopsis involves two cellular pools and a change in calcium signature after acclimation**. *Plant Cell* 1996, **8**(3):489-503.
47. Monroy AF, Sarhan F, Dhindsa RS: **Cold-Induced Changes in Freezing Tolerance, Protein Phosphorylation, and Gene Expression (Evidence for a Role of Calcium)**. *Plant Physiol* 1993, **102**(4):1227-1235.
48. Tahtiharju S, Sangwan V, Monroy AF, Dhindsa RS, Borg M: **The induction of kin genes in cold-acclimating Arabidopsis thaliana. Evidence of a role for calcium**. *Planta* 1997, **203**(4):442-447.
49. White PJ, Broadley MR: **Calcium in plants**. *Ann Bot (Lond)* 2003, **92**(4):487-511.
50. Knight H: **Calcium signaling during abiotic stress in plants**. *Int Rev Cytol* 2000, **195**:269-324.
51. McAinsch MR, Hetherington AM: **Encoding specificity in Ca2+ signalling systems**. *Trends in Plant Science* 1998, **3**:32-36.
52. Hetherington AM, Brownlee C: **The generation of Ca(2+) signals in plants**. *Annu Rev Plant Biol* 2004, **55**:401-427.
53. Reddy VS, Reddy AS: **Proteomics of calcium-signaling components in plants**. *Phytochemistry* 2004, **65**(12):1745-1776.
54. Berridge MJ, Lipp P, Bootman MD: **The versatility and universality of calcium signalling**. *Nat Rev Mol Cell Biol* 2000, **1**(1):11-21.
55. Jian L-C, Li J-H, Chen W-P, Li PH, Ahlstrand GG: **Cytochemical Localization of Calcium and Ca2+-ATPase Activity in Plant Cells under Chilling Stress: a Comparative Study between the Chilling-Sensitive Maize and the Chilling-Insensitive Winter Wheat** *Plant and Cell Physiology* 1999, **40**:1061-1071.
56. Xiong L, Lee B, Ishitani M, Lee H, Zhang C, Zhu JK: **FIERY1 encoding an inositol polyphosphate 1-phosphatase is a negative regulator of abscisic acid and stress signaling in Arabidopsis**. *Genes Dev* 2001, **15**(15):1971-1984.
57. Gusta LV, Trischuk R, Weiser CJ: **Plant Cold Acclimation: The role of Abscisic Acid.** *Journal of Plant Growth Regulation* 2005, **24**:308-318.

58. Farrell AD, Ougham HJ, Tomos AD: **The effect of gibberellic acid on the response of leaf extension to low temperature**. *Plant Cell Environ* 2006, **29**(7):1329-1337.
59. Pawlowski TA: **Proteomics of European beech (Fagus sylvatica L.) seed dormancy breaking: influence of abscisic and gibberellic acids**. *Proteomics* 2007, **7**(13):2246-2257.
60. Shan DP, Huang JG, Yang YT, Guo YH, Wu CA, Yang GD, Gao Z, Zheng CC: **Cotton GhDREB1 increases plant tolerance to low temperature and is negatively regulated by gibberellic acid**. *New Phytol* 2007, **176**(1):70-81.
61. Ogawa M, Hanada A, Yamauchi Y, Kuwahara A, Kamiya Y, Yamaguchi S: **Gibberellin biosynthesis and response during Arabidopsis seed germination**. *Plant Cell* 2003, **15**(7):1591-1604.
62. Mishra NS, Tuteja R, Tuteja N: **Signaling through MAP kinase networks in plants**. *Arch Biochem Biophys* 2006, **452**(1):55-68.
63. Teige M, Scheikl E, Eulgem T, Doczi R, Ichimura K, Shinozaki K, Dangl JL, Hirt H: **The MKK2 pathway mediates cold and salt stress signaling in Arabidopsis**. *Mol Cell* 2004, **15**(1):141-152.
64. Gray GR, Chauvin LP, Sarhan F, Huner N: **Cold Acclimation and Freezing Tolerance (A Complex Interaction of Light and Temperature)**. *Plant Physiol* 1997, **114**(2):467-474.
65. Wanner LA, Junttila O: **Cold-induced freezing tolerance in Arabidopsis**. *Plant Physiol* 1999, **120**(2):391-400.
66. Kim HJ, Kim YK, Park JY, Kim J: **Light signalling mediated by phytochrome plays an important role in cold-induced gene expression through the C-repeat/dehydration responsive element (C/DRE) in Arabidopsis thaliana**. *Plant J* 2002, **29**(6):693-704.
67. Hotta CT, Gardner MJ, Hubbard KE, Baek SJ, Dalchau N, Suhita D, Dodd AN, Webb AA: **Modulation of environmental responses of plants by circadian clocks**. *Plant Cell Environ* 2007, **30**(3):333-349.
68. Harmer SL, Hogenesch JB, Straume M, Chang HS, Han B, Zhu T, Wang X, Kreps JA, Kay SA: **Orchestrated transcription of key pathways in Arabidopsis by the circadian clock**. *Science* 2000, **290**(5499):2110-2113.
69. Fowler SG, Cook D, Thomashow MF: **Low temperature induction of Arabidopsis CBF1, 2, and 3 is gated by the circadian clock**. *Plant Physiol* 2005, **137**(3):961-968.
70. Cao S, Ye M, Jiang S: **Involvement of GIGANTEA gene in the regulation of the cold stress response in Arabidopsis**. *Plant Cell Rep* 2005, **24**(11):683-690.
71. Dodd AN, Jakobsen MK, Baker AJ, Telzerow A, Hou SW, Laplaze L, Barrot L, Poethig RS, Haseloff J, Webb AA: **Time of day modulates low-temperature Ca signals in Arabidopsis**. *Plant J* 2006, **48**(6):962-973.

72. Franklin KA, Whitelam GC: **Light-quality regulation of freezing tolerance in Arabidopsis thaliana**. *Nat Genet* 2007, **39**(11):1410-1413.
73. Ramos A, Perez-Solis E, Ibanez C, Casado R, Collada C, Gomez L, Aragoncillo C, Allona I: **Winter disruption of the circadian clock in chestnut**. *Proc Natl Acad Sci U S A* 2005, **102**(19):7037-7042.
74. Jones TL, Tucker DE, Ort DR: **Chilling delays circadian pattern of sucrose phosphate synthase and nitrate reductase activity in tomato**. *Plant Physiol* 1998, **118**(1):149-158.
75. Weyman PD, Pan Z, Feng Q, Gilchrist DG, Bostock RM: **DEA1, a circadian- and cold-regulated tomato gene, protects yeast cells from freezing death**. *Plant Mol Biol* 2006, **62**(4-5):547-559.
76. Martino-Catt S, Ort DR: **Low temperature interrupts circadian regulation of transcriptional activity in chilling-sensitive plants**. *Proc Natl Acad Sci U S A* 1992, **89**(9):3731-3735.
77. Chinnusamy V, Ohta M, Kanrar S, Lee BH, Hong X, Agarwal M, Zhu JK: **ICE1: a regulator of cold-induced transcriptome and freezing tolerance in Arabidopsis**. *Genes Dev* 2003, **17**(8):1043-1054.
78. Lee BH, Henderson DA, Zhu JK: **The Arabidopsis Cold-Responsive Transcriptome and Its Regulation by ICE1**. *The Plant Cell* 2005, **17**:3155-3175.
79. Medina J, Bargues M, Terol J, Perez-Alonso M, Salinas J: **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 Physiol* 1999, **119**(2):463-470.
80. Vogel JT, Zarka DG, Van Buskirk HA, Fowler SG, Thomashow MF: **Roles of the CBF2 and ZAT12 transcription factors in configuring the low temperature transcriptome of Arabidopsis**. *Plant J* 2005, **41**(2):195-211.
81. Fowler S, Thomashow MF: **Arabidopsis transcriptome profiling indicates that multiple regulatory pathways are activated during cold acclimation in addition to the CBF cold response pathway**. *Plant Cell* 2002, **14**(8):1675-1690.
82. Stockinger EJ, Gilmour SJ, Thomashow MF: **Arabidopsis thaliana CBF1 encodes and AP2 domain-containing transcriptional activator that binds to the C-repeat/DRE, a cis-acting DNA regulatory element that stimulates transcription response to low temperature and water deficit.** *Proc Natl Acad Sci U S A* 1997, **94**:1035-1040.
83. Gilmour SJ, Zarka DG, Stockinger EJ, Salazar MP, Houghton JM, Thomashow MF: **Low temperature regulation of the Arabidopsis CBF family of AP2 transcriptional activators as an early step in cold-induced COR gene expression.** *Plant Journal* 1998, **16**:433-442.

84. Miura K, Jin JB, Lee J, Yoo CY, Stirm V, Miura T, Ashworth EN, Bressan RA, Yun DJ, Hasegawa PM: **SIZ1-mediated sumoylation of ICE1 controls CBF3/DREB1A expression and freezing tolerance in Arabidopsis**. *Plant Cell* 2007, **19**(4):1403-1414.
85. Chawade A, Brautigam M, Lindlof A, Olsson O, Olsson B: **Putative cold acclimation pathways in Arabidopsis thaliana identified by a combined analysis of mRNA co-expression patterns, promoter motifs and transcription factors**. *BMC Genomics* 2007, **8**:304.
86. Novillo F, Alonso JM, Ecker JR, Salinas J: **CBF2/DREB1C is a negative regulator of CBF1/DREB1B and CBF3/DREB1A expression and plays a central role in stress tolerance in Arabidopsis**. *Proc Natl Acad Sci U S A* 2004, **101**(11):3985-3990.
87. Agarwal M, Hao Y, Kapoor A, Dong CH, Fujii H, Zheng X, Zhu JK: **A R2R3 type MYB transcription factor is involved in the cold regulation of CBF genes and in acquired freezing tolerance**. *J Biol Chem* 2006, **281**(49):37636-37645.
88. Zhu J, Shi H, Lee BH, Damsz B, Cheng S, Stirm V, Zhu JK, Hasegawa PM, Bressan RA: **An Arabidopsis homeodomain transcription factor gene, HOS9, mediates cold tolerance through a CBF-independent pathway**. *Proc Natl Acad Sci U S A* 2004, **101**(26):9873-9878.
89. Zhu J, Verslues PE, Zheng X, Lee BH, Zhan X, Manabe Y, Sokolchik I, Zhu Y, Dong CH, Zhu JK *et al*: **HOS10 encodes an R2R3-type MYB transcription factor essential for cold acclimation in plants**. *Proc Natl Acad Sci U S A* 2005, **102**(28):9966-9971.
90. Xin Z, Browse J: **Eskimo1 mutants of Arabidopsis are constitutively freezing-tolerant**. *Proc Natl Acad Sci U S A* 1998, **95**(13):7799-7804.
91. Xin Z, Mandaokar A, Chen J, Last RL, Browse J: **Arabidopsis ESK1 encodes a novel regulator of freezing tolerance**. *Plant J* 2007, **49**(5):786-799.
92. Hazen SP, Wu Y, Kreps JA: **Gene expression profiling of plant responses to abiotic stress**. *Funct Integr Genomics* 2003, **3**(3):105-111.
93. Lindlof A: **Gene identification through large-scale EST sequence processing**. *Appl Bioinformatics* 2003, **2**(3):123-129.
94. Mayer K, Mewes HW: **How can we deliver the large plant genomes? Strategies and perspectives**. *Curr Opin Plant Biol* 2002, **5**(2):173-177.
95. Rudd S: **Expressed sequence tags: alternative or complement to whole genome sequences?** *Trends Plant Sci* 2003, **8**(7):321-329.
96. White CA, Salamonsen LA: **A guide to issues in microarray analysis: application to endometrial biology**. *Reproduction* 2005, **130**(1):1-13.

97. Rensink WA, Hazen SP: **Statistical issues in microarray data analysis**. *Methods Mol Biol* 2006, **323**:359-366.
98. Masoudi-Nejad A, Tonomura K, Kawashima S, Moriya Y, Suzuki M, Itoh M, Kanehisa M, Endo T, Goto S: **EGassembler: online bioinformatics service for large-scale processing, clustering and assembling ESTs and genomic DNA fragments**. *Nucleic Acids Res* 2006, **34**(Web Server issue):W459-462.
99. Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ: **Basic local alignment search tool**. *J Mol Biol* 1990, **215**(3):403-410.
100. Ogihara Y, Mochida K, Nemoto Y, Murai K, Yamazaki Y, Shin IT, Kohara Y: **Correlated clustering and virtual display of gene expression patterns in the wheat life cycle by large-scale statistical analyses of expressed sequence tags**. *Plant J* 2003, **33**(6):1001-1011.
101. Fei Z, Tang X, Alba RM, White JA, Ronning CM, Martin GB, Tanksley SD, Giovannoni JJ: **Comprehensive EST analysis of tomato and comparative genomics of fruit ripening**. *Plant J* 2004, **40**(1):47-59.
102. Romualdi C, Bortoluzzi S, Danieli GA: **Detecting differentially expressed genes in multiple tag sampling experiments: comparative evaluation of statistical tests**. *Hum Mol Genet* 2001, **10**(19):2133-2141.
103. Conte MG, Gaillard S, Lanau N, Rouard M, Perin C: **GreenPhylDB: a database for plant comparative genomics**. *Nucleic Acids Res* 2008, **36**(Database issue):D991-998.
104. Eisen MB, Spellman PT, Brown PO, Botstein D: **Cluster analysis and display of genome-wide expression patterns**. *Proc Natl Acad Sci U S A* 1998, **95**(25):14863-14868.
105. Whelan S: **Inferring trees**. *Methods Mol Biol* 2008, **452**:287-309.
106. Pavesi G, Mauri G, Pesole G: **In silico representation and discovery of transcription factor binding sites**. *Brief Bioinform* 2004, **5**(3):217-236.
107. Blevins L, McDonald CJ: **Fisher's Exact Test: an easy-to-use statistical test for comparing outcomes**. *MD Comput* 1985, **2**(1):15-19, 68.
108. Alberdi M, Corcuera LJ, Maldona C, Barrientos M, Henriquez O: **Cold acclimation in cultivars of Avena sativa**. *Phytochemistry* 1993, **33**:57-60.
109. Petr J, Capouchová I, Stolcová M: **Physiological nature of overwintring oats forms**. *Plant Production* 2002, **48**:285-292.
110. Garsed K, Scott BB: **Can oats be taken in a gluten-free diet? A systematic review**. *Scand J Gastroenterol* 2007, **42**(2):171-178.
111. Haboubi NY, Taylor S, Jones S: **Coeliac disease and oats: a systematic review**. *Postgrad Med J* 2006, **82**(972):672-678.
112. Welch RW: **Can dietary oats promote health?** *Br J Biomed Sci* 1994, **51**(3):260-270.

113. Chinnusamy V, Zhu J, Zhu JK: **Cold stress regulation of gene expression in plants**. *Trends Plant Sci* 2007, **12**(10):444-451.
114. Sarhan F, Danyluk J: **Engineering cold-tolerant crops -throwing the master switch.** *Trends Plant Sci* 1998, **3**:289-291.
115. Schoof H, Ernst R, Nazarov V, Pfeifer L, Mewes HW, Mayer KF: **MIPS Arabidopsis thaliana Database (MAtDB): an integrated biological knowledge resource for plant genomics**. *Nucleic Acids Res* 2004, **32**(Database issue):D373-376.
116. Tatusov RL, Fedorova ND, Jackson JD, Jacobs AR, Kiryutin B, Koonin EV, Krylov DM, Mazumder R, Mekhedov SL, Nikolskaya AN *et al*: **The COG database: an updated version includes eukaryotes.** *BMC Bioinformatics* 2003, **4**:41.
117. Danyluk J, Houde M, Rassart E, Sarhan F: **Differential expression of a gene encoding an acidic dehydrin in chilling sensitive and freezing tolerant gramineae species.** *FEBS Lett* 1994, **344**:20-24.
118. Tsuda K, Tsvetanov S, Takumi S, Mori N, Atanassov A, Nakamura C: **New members of a cold-responsive group-3 Lea/Rab-related Cor gene family from common wheat (Triticum aestivumL.).** . *Genes Genet Syst* 2000, **75**:179-188.
119. Sakuma Y, Liu Q, Dubouzet JG, Abe H, Shinozaki K, Yamaguchi-Shinozak K: **DNA-binding specificity of the ERF/AP2 domain of Arabidopsis DREBs, transcription factors involved in dehydration- and cold-inducible gene expression.** *Biochem Biophys Res Commun* 2002, **290**:998-1009.
120. Quackenbush J, Cho J, Lee D, Liang F, Holt I, Karamycheva S, Parvizi B, Pertea G, Sultana R, White J: **The TIGR Gene Indices: analysis of gene transcript sequences in highly sampled eukaryotic species**. *Nucleic Acids Res* 2001, **29**(1):159-164.
121. Lee Y, Sultana R, Pertea G, Cho J, Karamycheva S, Tsai J, Parvizi B, Cheung F, Antonescu V, White J *et al*: **Cross-referencing eukaryotic genomes: TIGR Orthologous Gene Alignments (TOGA)**. *Genome Res* 2002, **12**(3):493-502.
122. Li L, Stoeckert CJ, Jr., Roos DS: **OrthoMCL: identification of ortholog groups for eukaryotic genomes**. *Genome Res* 2003, **13**(9):2178-2189.
123. Audic S, Claverie JM: **The significance of digital gene expression profiles**. *Genome Res* 1997, **7**(10):986-995.
124. Claverie JM: **Computational methods for the identification of differential and coordinated gene expression**. *Hum Mol Genet* 1999, **8**(10):1821-1832.
125. Jung SH, Lee JY, Lee DH: **Use of SAGE technology to reveal changes in gene expression in Arabidopsis leaves undergoing cold stress**. *Plant Mol Biol* 2003, **52**(3):553-567.
126. Schmitt AO, Specht T, Beckmann G, Dahl E, Pilarsky CP, Hinzmann B, Rosenthal A: **Exhaustive mining of EST libraries for**

genes differentially expressed in normal and tumour tissues. *Nucleic Acids Res* 1999, **27**(21):4251-4260.
127. Stekel DJ, Git Y, Falciani F: **The comparison of gene expression from multiple cDNA libraries**. *Genome Res* 2000, **10**(12):2055-2061.
128. Strausberg RL, Greenhut SF, Grouse LH, Schaefer CF, Buetow KH: **In silico analysis of cancer through the Cancer Genome Anatomy Project**. *Trends Cell Biol* 2001, **11**(11):S66-71.
129. Romualdi C, Bortoluzzi S, D'Alessi F, Danieli GA: **IDEG6: a web tool for detection of differentially expressed genes in multiple tag sampling experiments**. *Physiol Genomics* 2003, **12**(2):159-162.
130. Close TJ: **Dehydrins: emergence of a biochemical role of a family of plant dyhadration proteins.** *Physiol Plant* 1996, **97**:795-803.
131. Li QB, Haskell DW, Guy CL: **Coordinate and non-coordinate expression of the stress 70 family and other molecular chaperones at high and low temperature in spinach and tomato**. *Plant Mol Biol* 1999, **39**(1):21-34.
132. Gulick PJ, Drouin S, Yu Z, Danyluk J, Poisson G, Monroy AF, Sarhan F: **Transcriptome comparison of winter and spring wheat responding to low temperature**. *Genome* 2005, **48**(5):913-923.
133. Ugoni A, Walker BF: **THE CHI SQUARE TEST: An Introduction**. *COMSIG Rev* 1995, **4**(3):61-64.
134. Mochida K, Kawaura K, Shimosaka E, Kawakami N, Shin IT, Kohara Y, Yamazaki Y, Ogihara Y: **Tissue expression map of a large number of expressed sequence tags and its application to in silico screening of stress response genes in common wheat**. *Mol Genet Genomics* 2006, **276**(3):304-312.
135. Hannah MA, Heyer AG, Hincha DK: **A global survey of gene regulation during cold acclimation in Arabidopsis thaliana**. *PLoS Genet* 2005, **1**(2):e26.
136. Cheng C, Yun KY, Ressom HW, Mohanty B, Bajic VB, Jia Y, Yun SJ, de los Reyes BG: **An early response regulatory cluster induced by low temperature and hydrogen peroxide in seedlings of chilling-tolerant japonica rice**. *BMC Genomics* 2007, **8**:175.
137. Guo Z, Ou W, Lu S, Zhong Q: **Differential responses of antioxidative system to chilling and drought in four rice cultivars differing in sensitivity**. *Plant Physiol Biochem* 2006, **44**(11-12):828-836.
138. Yu X, Peng YH, Zhang MH, Shao YJ, Su WA, Tang ZC: **Water relations and an expression analysis of plasma membrane intrinsic proteins in sensitive and tolerant rice during chilling and recovery**. *Cell Res* 2006, **16**(6):599-608.
139. Kawakami A, Sato Y, Yoshida M: **Genetic engineering of rice capable of synthesizing fructans and enhancing chilling tolerance**. *J Exp Bot* 2008, **59**(4):793-802.

140. Rabbani MA, Maruyama K, Abe H, Khan MA, Katsura K, Ito Y, Yoshiwara K, Seki M, Shinozaki K, Yamaguchi-Shinozaki K: **Monitoring expression profiles of rice genes under cold, drought, and high-salinity stresses and abscisic acid application using cDNA microarray and RNA gel-blot analyses**. *Plant Physiol* 2003, **133**(4):1755-1767.
141. Dubouzet JG, Sakuma Y, Ito Y, Kasuga M, Dubouzet EG, Miura S, Seki M, Shinozaki K, Yamaguchi-Shinozaki K: **OsDREB genes in rice, Oryza sativa L., encode transcription activators that function in drought-, high-salt- and cold-responsive gene expression**. *Plant J* 2003, **33**(4):751-763.
142. Ito Y, Katsura K, Maruyama K, Taji T, Kobayashi M, Seki M, Shinozaki K, Yamaguchi-Shinozaki K: **Functional analysis of rice DREB1/CBF-type transcription factors involved in cold-responsive gene expression in transgenic rice**. *Plant Cell Physiol* 2006, **47**(1):141-153.
143. Meyer Y, Siala W, Bashandy T, Riondet C, Vignols F, Reichheld JP: **Glutaredoxins and thioredoxins in plants**. *Biochim Biophys Acta* 2008, **1783**(4):589-600.
144. Eulgem T, Rushton PJ, Robatzek S, Somssich IE: **The WRKY superfamily of plant transcription factors**. *Trends Plant Sci* 2000, **5**(5):199-206.
145. Miller G, Shulaev V, Mittler R: **Reactive oxygen signaling and abiotic stress**. *Physiol Plant* 2008.
146. Yanhui C, Xiaoyuan Y, Kun H, Meihua L, Jigang L, Zhaofeng G, Zhiqiang L, Yunfei Z, Xiaoxiao W, Xiaoming Q *et al*: **The MYB transcription factor superfamily of Arabidopsis: expression analysis and phylogenetic comparison with the rice MYB family**. *Plant Mol Biol* 2006, **60**(1):107-124.
147. Nagano Y, Furuhashi H, Inaba T, Sasaki Y: **A novel class of plant-specific zinc-dependent DNA-binding protein that binds to A/T-rich DNA sequences**. *Nucleic Acids Res* 2001, **29**(20):4097-4105.
148. Ulmasov T, Hagen G, Guilfoyle TJ: **Activation and repression of transcription by auxin-response factors**. *Proc Natl Acad Sci U S A* 1999, **96**(10):5844-5849.
149. Wu MF, Tian Q, Reed JW: **Arabidopsis microRNA167 controls patterns of ARF6 and ARF8 expression, and regulates both female and male reproduction**. *Development* 2006, **133**(21):4211-4218.
150. Remenyi A, Scholer HR, Wilmanns M: **Combinatorial control of gene expression**. *Nat Struct Mol Biol* 2004, **11**(9):812-815.
151. Lescot M, Dehais P, Thijs G, Marchal K, Moreau Y, Van de Peer Y, Rouze P, Rombauts S: **PlantCARE, a database of plant cis-acting regulatory elements and a portal to tools for in silico analysis of promoter sequences**. *Nucleic Acids Res* 2002, **30**(1):325-327.

152. Higo K, Ugawa Y, Iwamoto M, Higo H: **PLACE: a database of plant cis-acting regulatory DNA elements**. *Nucleic Acids Res* 1998, **26**(1):358-359.
153. Swindell WR, Huebner M, Weber AP: **Transcriptional profiling of Arabidopsis heat shock proteins and transcription factors reveals extensive overlap between heat and non-heat stress response pathways**. *BMC Genomics* 2007, **8**:125.

Appendix

Appendix A.

Most expressed TOGs among the preferentially derived.

Set	Putative function	Stress	Ref
AtCI I	Ribulose-1,5-bisphosphate carboxylase/oxygenase small subunit	Cold, light	(Zhang et al., 2001, Gulick et al., 2005)
	Translation elongation factor 1	Cold	(Berberich et al., 1995)
	Cyclophilin	Cold	(Savitch et al., 2005)
	Translationally controlled tumor protein homolog	-	
	Glyceraldehyde-3-phosphate dehydrogenase	Cold	(Jeong et al., 2000, Hurry et al., 2000)
	Glycine-rich RNA-binding protein	Cold	(Kwak et al., 2005, Shinozuka et al., 2006)
	Wali3 protein	Alu, wound	(Snowden and Gardner, 1993, Snowden et al., 1995)
	Chlorophyll a/b-binding protein CP29	Light	(Yang et al., 2000)
	Rieske Fe-S protein	-	
	Carbonic anhydrase	Cold, salt	(Yamada et al., 1998, Gulick et al., 2005)
	Phosphoglycerate kinase	Cold, heat	(Piper et al., 1986, Hurry et al., 2000)
	60S ribosomal protein L35	-	
	60S ribosomal protein L10	-	
	At1g54410, Dehydrin	Water	TAIR locus
	60S ribosomal protein L10	-	
AtCI II	Heat shock protein cognate 70	Cold	(Li et al., 1999)
	Translation elongation factor 1	Cold	(Berberich et al., 1995)
	Ribulose-1,5-bisphosphate carboxylase/oxygenase small subunit	Cold, light	(Zhang et al., 2001, Gulick et al., 2005)
	Thiazole biosynthetic enzyme, mitochondrial precursor (Stress-inducible protein sti35)	Heat, copper, etc.	(Choi et al., 1990)

	Adenosylhomocysteinase (S-adenosyl-L-homocysteine hydrolase)	Cold	(Jung et al., 2003)
	Glyceraldehyde-3-phosphate dehydrogenase A	Cold	(Jeong et al., 2000, Hurry et al., 2000)
	Transketolase	Cold	(Hurry et al., 2000)
	Glycine hydroxymethyltransferase	-	
	Glyceraldehyde-3-phosphate dehydrogenase B	Cold	(Jeong et al., 2000, Hurry et al., 2000)
	Glyceraldehyde-3-phosphate dehydrogenase, cytosolic 1	Cold	(Jeong et al., 2000, Hurry et al., 2000)
	Ribulose bisphosphate carboxylase/oxygenase activase	Heat	(Law and Crafts-Brandner, 2001)
	Phosphoglycerate kinase	Cold	(Hurry et al., 2000)
	Carbonic anhydrase	Cold, salt	(Yamada et al., 1998, Gulick et al., 2005)
	Fructose 1,6-bisphosphate aldolase	Cold	(Hurry et al., 2000)
	Beta-glucosidase	-	
AtCI III	Ribulose-1,5-bisphosphate carboxylase/oxygenase small subunit	Cold, light	(Zhang et al., 2001, Gulick et al., 2005)
	Glycine-rich RNA-binding protein	Cold	(Nomata et al., 2004, Stephen et al., 2003, Zchut et al., 2003)
	GSK3/Shaggy-related protein kinase	Osmotic	(Jonak and Hirt, 2002, Richard et al., 2005)
	Cinnamyl-alcohol dehydrogenase	-	
	Mitochondrial 60S ribosomal protein L6	-	
	Glutaredoxin protein	Cold	(Gidekel et al., 2003)
	At1g60710/F8A5_23	Hypoosmolarity, light	(Rook et al., 1998, Satoh et al., 2004)
	GSK3/Shaggy-related protein kinase	Osmotic	(Jonak and Hirt, 2002, Richard et al., 2005)
	Ubiquitin-conjugating enzyme OsUBC5a	-	
	Tonneau 1	-	
	At5g12470	Unknown	TAIR locus detail
	60S acidic ribosomal protein P0	-	
	Voltage dependent anion channel	-	
	Heat shock protein 80/82/90	Cold	(Rinehart and Denlinger, 2000, Freitag et al., 1997)
	40S ribosomal protein SA (p40)	-	

AsCI	Ribulose-1,5-bisphosphate carboxylase/oxygenase small subunit	Cold, light	(Zhang et al., 2001, Gulick et al., 2005)
	Chlorophyll a/b-binding protein WCAB precursor	Light	(Yang et al., 2000)
	Ribulose bisphosphate carboxylase/oxygenase activase	Heat	(Law and Crafts-Brandner, 2001)
	Fructose 1,6-bisphosphate aldolase	Cold	(Hurry et al., 2000)
	Cold acclimation protein WCOR410b	Cold	(Gulick et al., 2005)
	Chlorophyll a/b-binding protein CP29	-	
	Glyceraldehyde-3-phosphate dehydrogenase	Cold	(Jeong et al., 2000, Hurry et al., 2000)
	Chlorophyll a/b-binding protein	Light	(Yang et al., 2000)
	Oxygen-evolving enhancer protein 2	Salt	(Abbasi and Komatsu, 2004, Sugihara et al., 2000)
	Oxygen-evolving enhancer protein 1	Salt	(Abbasi and Komatsu, 2004, Sugihara et al., 2000)
	Photosystem II 10 kDa polypeptide/40S ribosomal protein S10-1	-	
	Ferredoxin-NADP(H) oxidoreductase	-	
	Ferredoxin	-	
	Chlorophyll A-B binding protein of LHCII type III	Light	(Yang et al., 2000)
	Chlorophyll a-b binding protein 4	Light	(Yang et al., 2000)
	Photosystem I reaction center subunit XI, chloroplast precursor (PSI-L) (PSI subunit V)	Light	(Toyama et al., 1996)
HvCI	Chlorophyll a/b-binding protein WCAB precursor	Light	(Yang et al., 2000)
	Nonspecific lipid-transfer protein 4.1 precursor (LTP 4.1) (CW21)	-	
	Chlorophyll a-b binding protein of LHCII type III, chloroplast precursor (CAB)	Light	(Yang et al., 2000)
	Unknown	-	
	Ribulose-1,5-bisphosphate carboxylase/oxygenase small subunit	Light	(Zhang et al., 2001)
	Glutathione-S-transferase	Cold	(Seppanen et al., 2000, Jung et al., 2003)

	Alcohol dehydrogenase	Cold	(Christie et al., 1991, Jarillo et al., 1993)
	Chlorophyll a-b binding protein CP26	Light	(Yang et al., 2000)
	Glutathione-S-transferase	Cold	(Seppanen et al., 2000, Jung et al., 2003)
	Photosystem II subunit PsbS	Light	(Toyama et al., 1996)
	Glyceraldehyde-3-phosphate dehydrogenase	Cold	(Jeong et al., 2000, Hurry et al., 2000)
	Photosystem I reaction center subunit XI, chloroplast precursor (PSI-L) (PSI subunit V)	Light	(Toyama et al., 1996)
	Selenium binding protein	Cold	(Machuka et al., 1999)
	Cold acclimation protein WCOR615	Cold	-
	ES2A protein	GA3	(Speulman and Salamini, 1995)
TaCI	Low temperature-responsive RNA-binding protein	Cold	
	Translation elongation factor EF-G	-	
	Photosystem II protein W-like protein	-	
	Weakly similar to UP|Q8VYV1 (Q8VYV1) AT5g08050	-	
	Glycine rich protein, RNA binding protein	-	
	(S)-2-hydroxy-acid oxidase	-	
	50S ribosomal protein L6	Cold	(Bosl and Bock, 1981)
	G5bf	-	
	Mitochondrial aspartate-glutamate carrier	-	
	Cold acclimation protein WCOR615	Cold	
	Xyloglucan endo-1,4-beta-D-glucanase	-	
	Weakly similar to UP|Q6YZD3 (Q6YZD3) Lipid transfer protein	-	
	Pyruvate/2-oxoglutarate dehydrogenase/ Fibrillarin	-	
	Coproporphyrinogen III oxidase	-	
	Embryonic abundant protein-like (At5g10830)	-	

OsCI	Translation elongation factor 1	Cold	(Berberich et al., 1995)
	Glyceraldehyde-3-phosphate dehydrogenase	Cold	(Jeong et al., 2000)
	26S ribosomal RNA	-	
	Unknown	-	
	Metallothionein-like protein type 3	Abiotic	
	Heat shock protein 80/82/90	Cold	(Rinehart and Denlinger, 2000, Freitag et al., 1997)
	Progesterone-binding	-	
	60S ribosomal protein L10	-	
	Alcohol dehydrogenase	Cold	(Christie et al., 1991, Jarillo et al., 1993)
	Fructose 1,6-bisphosphate aldolase	Cold	(Hurry et al., 2000)
	Enolase 1	Salt	(Yan et al., 2005)
	NADP-specific isocitrate dehydrogenase	Salt	(Popova et al., 2002)
	Glucose-6-phosphate isomerase	-	
	14-3-3-like protein GF14	-	
	Triosephosphate isomerase	Cold	(Graumann et al., 1996)